microorganisms

Phototrophic Bacteria

Edited by
Robert Blankenship and Matthew Sattley
Printed Edition of the Special Issue Published in *Microorganisms*

MDPI

Phototrophic Bacteria

Phototrophic Bacteria

Editors

Robert Blankenship
Matthew Sattley

MDPI

Editors
Robert Blankenship
Biology and Chemistry
Washington University in St. Louis
St. Louis
United States

Matthew Sattley
Division of Natural Sciences
Indiana Wesleyan University
Marion, IN
United States

Contents

About the Editors

Robert Blankenship

Robert Blankenship is the Lucille P. Markey Distinguished Professor of Arts and Sciences, Emeritus, at Washington University in St. Louis. He has spent his entire 50-year scientific career studying the essential biological process of photosynthesis. He has researched nearly every type of phototrophic bacteria, including both anoxygenic and oxygenic forms. His particular research interests include energy transfer in antenna complexes, as well as electron transfer in reaction centers and electron transfer proteins. Additionally, he has also researched the origin and evolution of the photosynthetic process.

Matthew Sattley

Matthew Sattley is a Professor of Biology at Indiana Wesleyan University (Marion, IN), where he teaches laboratory courses in general microbiology, environmental microbiology, and environmental science. He has mentored dozens of undergraduate research students in his laboratory and has served as the director of the Hodson Research Institute (IWU), a faculty-led summer research program for undergraduate students in the natural sciences. Matt's research group investigates the ecology and diversity of bacteria that inhabit extreme environments, with a special focus on the physiological and genomic characterization of anoxygenic phototrophs.

Preface to "Phototrophic Bacteria"

Microorganisms is pleased to publish this book, which reprints papers that appeared in a Special Issue on "Phototrophic Bacteria", with Guest Editors Robert Blankenship and Matthew Sattley. This Special Issue included research on all types of phototrophic bacteria, including both anoxygenic and oxygenic forms. Research on these bacterial organisms has greatly advanced our understanding of the basic principles that underlie the energy storage that takes place in all types of photosynthetic organisms, including both bacterial and eukaryotic forms. Topics of interest include: microbial physiology, microbial ecology, microbial genetics, evolutionary microbiology, systems microbiology, agricultural microbiology, microbial biotechnology, and environmental microbiology, as all are related to phototrophic bacteria.

Robert Blankenship and Matthew Sattley
Editors

Article

Elioraea tepida, sp. nov., a Moderately Thermophilic Aerobic Anoxygenic Phototrophic Bacterium Isolated from the Mat Community of an Alkaline Siliceous Hot Spring in Yellowstone National Park, WY, USA

Mohit Kumar Saini [1,†] **, Shohei Yoshida** [1,†] **, Aswathy Sebastian** [2] **, Eri Hara** [3] **, Hideyuki Tamaki** [3] **, Nathan T. Soulier** [4] **, Istvan Albert** [2,4] **, Satoshi Hanada** [1] **, Marcus Tank** [1,4,5,*] **and Donald A. Bryant** [4,*]

1   Department of Biological Sciences, Tokyo Metropolitan University, Tokyo 192-0397, Japan;
    mohitvtm@gmail.com (M.K.S.); rhcp.mariabeetle@gmail.com (S.Y.); shanada2@icloud.com (S.H.)
2   The Huck Institutes for the Life Sciences, The Pennsylvania State University, University Park, PA 16802, USA;
    azs13@psu.edu (A.S.); iua1@psu.edu (I.A.)
3   Bioproduction Research Institute—National Institute of Advanced Industrial Science and Technology (AIST),
    1-1-1, Higashi, Tsukuba 305-8566, Japan; hara-eri@aist.go.jp (E.H.); tamaki-hideyuki@aist.go.jp (H.T.)
4   Department of Biochemistry and Molecular Biology, The Pennsylvania State University,
    University Park, PA 16802, USA; riftreiluos@gmail.com
5   DSMZ-German Culture Collection of Microorganisms and Cell Cultures, GmbH Inhoffenstraße 7B,
    38124 Braunschweig, Germany
*   Correspondence: mat19@dsmz.de (M.T.); dab14@psu.edu (D.A.B.); Tel.: +49-531-2616-395 (M.T.);
    +1-814-777-9699 (D.A.B.)
†   These authors contributed equally to this work.

Abstract: Strain MS-P2[T] was isolated from microbial mats associated with Mushroom Spring, an alkaline siliceous hot spring in Yellowstone National Park, WY, USA. The isolate grows chemoheterotrophically by oxygen-dependent respiration, and light stimulates photoheterotrophic growth under strictly oxic conditions. Strain MS-P2[T] synthesizes bacteriochlorophyll *a* and the carotenoid spirilloxanthin. However, photoautotrophic growth did not occur under oxic or anoxic conditions, suggesting that this strain should be classified as an aerobic anoxygenic phototrophic bacterium. Strain MS-P2[T] cells are motile, curved rods about 0.5 to 1.0 µm wide and 1.0 to 1.5 µm long. The optimum growth temperature is 45–50 °C, and the optimum pH for growth is circum-neutral (pH 7.0–7.5). Sequence analysis of the 16S rRNA gene revealed that strain MS-P2[T] is closely related to *Elioraea* species, members of the class *Alphaproteobacteria*, with a sequence identity of 96.58 to 98%. The genome of strain MS-P2[T] is a single circular DNA molecule of 3,367,643 bp with a mol% guanine-plus-cytosine content of 70.6%. Based on phylogenetic, physiological, biochemical, and genomic characteristics, we propose this bacteriochlorophyll *a*-containing isolate is a new species belonging to the genus *Elioraea*, with the suggested name *Elioraea tepida*. The type-strain is strain MS-P2[T] (= JCM33060[T] = ATCC TSD-174[T]).

Keywords: aerobic anoxygenic phototrophic bacteria; chlorophototroph; thermophile; hot spring; bacteriochlorophyll; *Alphaproteobacteria*

1. Introduction

Aerobic anoxygenic phototrophic bacteria (AAPB) are bacteria that grow chemoheterotrophically by oxygen-dependent respiration, but they also synthesize photosynthetic reaction centers and light-harvesting complexes containing bacteriochlorophyll (BChl) *a* under oxic conditions [1,2]. Although AAPB can produce a photosynthetic apparatus, AAPB are unable to reduce carbon dioxide; light usually stimulates growth slightly under chemoorganoheterotrophic conditions. Many species of AAPB have been found in a variety of habitats, including oceans, freshwater lakes and rivers, and soils [3]. Although most

AAPB are mesophiles, some species are moderate thermophiles isolated from hot springs. Examples include *Porphyrobacter tepidarius* [4], *Porphyrobacter cryptus* [5], and *Rubritepida flocculans* [6].

AAPB do not form a monophyletic clade but are widely distributed within the classes *Alphaproteobacteria*, *Betaproteobacteria,* and *Gammaproteobacteria* [7]. Phylogenetically, AAPB are often more closely related to non-phototrophic bacteria than to phototrophic bacteria [3]. The family *Acetobacteraceae* in the *Alphaproteobacteria* contains several examples of AAPB, e.g., *Roseococcus thiosulfatophilus* [8], *Humitalea rosea* [9], *Craurococcus roseus* [10], *Roseomonas aestuarii* [11], and *Acidiphilium rubrum* [12]; these AAPBs are interspersed among non-phototrophic bacteria. Of the AAPB belonging to the *Acetobacteraceae*, *Rubritepida flocculans* was once thought to be the sole thermophilic species [6]. Interestingly, *Rubritepida flocculans* cells grown at high temperatures do not produce BChl *a*, but BChl *a* and carotenoids are produced when cells are grown at 30 °C.

The type species of the genus *Elioraea*, i.e., *Elioraea tepidiphila*, is a slightly thermophilic bacterium that was isolated from a hot spring in the Furnas area of the island of São Miguel in the Azores [13]. The genus *Elioraea* belongs to the class *Alphaproteobacteria,* which was proposed to be the type genus of a separate family, *Elioraeaceae,* which was suggested to be a closely related sister clade to *Acetobacteraceae* in the order *Rhodospirillales* [13,14]. However, more recently the genus *Elioraea* was moved back into the family *Acetobacteraceae* [15,16]. The optimal growth temperature range for *E. tepidiphila* is 45 to 50 °C and the optimal pH is between pH 8.0 and 8.5. Initially, bacteriochlorophyll (BChl) *a* and carotenoids were not detected [13,14]. However, subsequent sequencing of the *E. tepidi-phila* genome (GenBank Project Accession NZ_ARKI00000000) revealed these capabilities, which were then verified experimentally by others (see Habib et al. [14]).

Previous studies reported the occurrence of an organism phylogenetically related to *E. tepidiphila* in microbial mats associated with Mushroom Spring in Yellowstone National Park (YNP), WY, USA [17,18]. Spectroscopic measurements revealed that this new isolate, strain MS-P2[T], produces substantial amounts of BChl *a* even under aerobic conditions in the light at high temperature [18], which suggested that this new isolate should be considered to be a thermophilic member of the AAPB. Based on its growth at elevated temperatures and its relationship to *E. tepidiphila*, this new organism was named "*Candidatus* Elioraea thermophila" strain MS-P2[T] by Tank et al. [18].

More recently still, another organism closely related to *E. tepidiphila* was isolated from a hot spring in Yunnan province, China [14]. These authors were apparently unaware of the previous use of the species epithet "thermophila" to describe the organism from Mushroom Spring, and they also named their isolate *E. thermophila*. The authors reported that both *E. thermophila* and *E. tepidiphila* could synthesize BChl *a* and carotenoids and that both possessed *pufLM* genes. Importantly, Habib et al. [14] also reported that their *E. thermophila* isolate could grow photoautotrophically when thiosulfate or hydrogen served as electron donors. This is possibly because this bacterium is thus far the only *Elioraea* species that has genes encoding phosphoribulokinase and a type-1 ribulose 1,5-bisphosphate carboxylase-oxygenase (RuBisCO), and thus it should have the capacity to fix CO_2 by the Calvin–Benson–Bassham cycle.

A mesophilic *Elioraea* species that promotes plant growth, *E. rosea*, was also recently described [19]. This bacterium, which was isolated from the floodwater of a paddy field in South Korea, is strictly aerobic, motile by swimming, and pink-pigmented with a growth temperature optimum of 28 °C. Although it has not yet been validly described, yet another thermophilic *Elioraea* sp., strain "Yellowstone", was isolated from the runoff channel of Octopus Spring in YNP at 50 °C, and an incomplete, draft genome sequence is available [20].

In this manuscript, we describe the major characteristics of the AAPB isolate MS-P2[T] from Mushroom Spring and compare its properties with those of *E. tepidiphila*, *E. thermophila*, *E. rosea*, and *Elioraea* sp. "Yellowstone". Additionally, phylogenomic analyses are presented that are based upon the complete genomic sequence data for the Mushroom Spring isolate.

Based on its phenotypic and phylogenetic characteristics, we propose that this isolate represents a new species within the genus *Elioraea* with the suggested name, *Elioraea tepida*.

2. Materials and Methods

2.1. Strain Isolation and Cultivation Conditions

'*Chloracidobacterium thermophilum* Midnight medium' (CTM medium) [21,22] supplemented with 0.8% (*w/v*) agar (pH 7.0) was used as the basal medium for the isolation. A microbial mat sample collected from Mushroom Spring in the Lower Geyser Basin of Yellowstone National Park, WY, USA (GPS coordinates: Lat.: 44.5387, Long.: –110.798) was inoculated into sterile liquid CTM medium, and the culture was incubated in a beaker covered with aluminum-foil under continuous illumination from an incandescent light (approx. 20–50 μmol photons m^{-2} s^{-1}) at 50 °C for >10 days. Diluted samples from the enrichment culture were mixed with liquified agar prepared with CTM medium at 45 °C (tempered agar), and the agar mixtures were then poured into Petri dishes. After cooling to solidify the agar, the plates were incubated at 50 °C. Pink colonies that formed in the solidified agar medium within a week were picked up with a sterilized Pasteur pipet, resuspended into 1.0 mL of sterile distilled water, and mixed with tempered agar in CTM medium, cooled to solidify the agar, and incubated at 50 °C. This procedure was repeated until axenic cultures were obtained. Liquid cultivation was also successful in basal CTM medium containing succinate, acetate, and yeast extract (1.0 g L^{-1} of each). The axenic cultures could be maintained as 60-mL cultures in 100-mL Erlenmeyer flasks covered with aluminum foil, with or without shaking (150 rpm) at 50 °C (see Supplementary Figure S1A).

2.2. Microscopy and Analytical Procedures

The size and shape of the cells of strain MS-P2^T were determined by phase-contrast microscopy. Autofluorescence from BChl *a* of cells of strain MS-P2^T was visualized under an epifluorescence microscope (Nikon Eclipse E600 (NIS-Elements software D), Nikon Xenon power supply XPS-100, monochromatic CMOS camera (Orcaflash 4.0, Hamamatsu), Filterset Excitation: 350–510 nm dichroic mirror:665 nm/Emission: 830 nm LP (SEMROCK, Rochester, NY, USA)).

Absorbance spectra were measured with a Shimadzu model UV-1800 spectrophotometer (Shimadzu Corporation, Kyoto, Japan). Pigments were extracted with acetone–methanol (7:2, *v/v*), and the absorbance spectra of extracts were measured to assess in a preliminary manner the pigmentation of cells. Extracted pigments were also analyzed by reversed-phase, high-performance liquid chromatography (HPLC) as described [23]. Pigments were extracted with acetone–methanol (7:2, *v/v*) and were filtered with a 0.2-μm polytetrafluoroethylene, single-use filter device (Cytiva, Marlborough, MA, USA) prior to injection into the column (25 cm × 4.6 mm MilliporeSigma™ Supelco™ Discovery™ 5-μm C-18 column; Fisher Scientific, Hampton, NH, USA). The filtered pigment solution was analyzed with an Agilent 1100 HPLC system equipped with a diode-array detector (Santa Clara, CA, USA), as previously described [24]. The identities of pigments were determined by comparison of elution times and in-line absorbance spectra to those of authenticated standards that are maintained in the laboratory.

Gram-staining was performed by using the Favor-G kit (Nissui Pharmaceutical, Tokyo, Japan). Production of catalase was assessed by measuring bubble production in 3% (*v/v*) H$_2$O$_2$, and cytochrome *c* oxidase activity was assessed by using the oxidase reagent (bioMérieux, Marcy l'Etoile, France). The presence of other selected enzymes was tested using the APIZYM system (bioMérieux, Marcy l'Etoile, France). The ability to oxidize a variety of substrates was tested using the Biolog GEN III plate system (Biolog, Hayward, CA, USA). The Biolog Gen III system tests 71 carbon substrates and 23 potentially inhibitory growth conditions in a 96-well microtiter plate format. Cells were resuspended at a recommended concentration in a proprietary "inoculation solution", that was amended to contain 0.1 g/L yeast extract. The cell suspension was then pipetted into each of the 96 wells, which included positive and negative control wells. Each well contained a carbon substrate and

other nutrients and buffer or contained a potential growth-inhibiting condition (e.g., salt, pH, antibiotic, etc.). The plates were incubated at 45 °C for 24 to 48 h, and the OD_{590} of each well was recorded at 8-h intervals. Positive wells turned purple in color due to the reduction of a tetrazolium redox dye, indicating growth and/or oxidation of the included substrate. The small amount of yeast extract added to the inoculation medium was insufficient to produce a positive color reaction.

Finally, fatty acid methyl esters, respiratory quinones, and the GC content of the DNA were analyzed according to previously described procedures [25]. Polar lipids were extracted using the Bligh–Dyer method and were analyzed by two-dimensional thin-layer chromatography as previously described [26,27].

2.3. Genome Sequencing and Bioinformatic Analyses

Total genomic DNA of strain MS-P2[T] was extracted, sequenced, and assembled as previously described [28]. The genome of strain MS-P2[T] was sequenced using the PacBio Sequel platform. Sequence assembly was performed using a Canu 1.8 assembler, yielding a high-quality, closed circular genome. The completeness and contamination of the genome were checked using the online version of CheckM implemented in the Kbase software and data platform [29]. The assembled genome was annotated by the RAST annotation system (Rapid Annotation using SEED Technology) [30–32] as well as by using the NCBI Prokaryotic Genome Annotation Pipeline (www.ncbi.nlm.nih.gov/genome/annotation_prok/ (accessed on 15 July 2021)) [33–35] to predict the number of coding genes, total RNA genes, and functional genes. The gene locus designations and gene names used in the manuscript are those from the NCBI Prokaryotic Genome Annotation Pipeline. Average nucleotide identity (ANI) values were calculated by the OrthoANI algorithm method as implemented in the OAT software package [36]. Digital DNA–DNA hybridization (dDDH) values were determined with the in-silico genome-to-genome distance calculator (GGDC2.1; http://ggdc.dsmz.de/distcalc2.php (accessed on 26 May 2021)) using the alignment method blast+ [37–39]. Genomes of type species belonging to the genus *Elioraea* (*E. tepidiphila* TU-7[T] [13], *E. thermophila* YIM 72297[T] [14], and *E. rosea* PF-30[T] [19]) were compared with the genome of strain MS-P2[T]. The MS-P2[T] genome was also compared to the genome of *Elioraea* sp. strain "Yellowstone" [20].

A phylogeny based on concatenated proteins was constructed using a set of 49 universally conserved proteins/gene as defined by COG (Clusters of Orthologous Groups) gene families on the Kbase platform using the "SpeciesTree builder version 2.2.0" [40]. This analysis included the genomes of five *Elioraea* spp. with a set of closely related genomes selected from the public KBase genomes import from RefSeq. Relatedness was determined by alignment similarity to a select subset of 49 COG domains. FastTree2 was used to prepare the phylogenetic tree [40]. A list of the 49 proteins used, as well as other information, can be found online at https://narrative.kbase.us/#catalog/apps/SpeciesTreeBuilder/insert_set_of_genomes_into_species_tree/release (accessed on 17 December 2021).

3. Results and Discussion

3.1. Isolation and Initial Cultivation

Strain MS-P2[T] was isolated from the microbial mats associated with the main runoff channel at Mushroom Spring in the Lower Geyser Basin of Yellowstone National Park, WY, USA (GPS coordinates: Lat.: 44.5387, Long.: −110.798). The temperature at the sampling site was 52 °C, and the pH was ~8.0. In this hot spring, the microbial mats mainly consist of thermophilic cyanobacteria of the genus *Synechococcus* [41–43] and phototrophic *Chloroflexota* from the genera *Roseiflexus*, *Chloroflexus*, "*Candidatus* Roseilinea", and "*Candidatus* Chloranaerofilum" [17,18,42–44]. Several other chlorophototrophic organisms, including two *Chloracidobacterium* species, *C. thermophilum* and *C. aggregatum* [21–23], "*Candidatus* Thermochlorobacter aerophilum" [45], and at least four chlorophototrophic members of the *Proteobacteria* also occur in these mats [18]. Strain MS-P2[T] was isolated as described in Section 2.1 of the Materials and Methods using CTM medium solidified with 0.8% (*w/v*)

agar at pH 7.0. Pink colonies appeared within a week and were purified by restreaking. Liquid cultures were pinkish-orange in color, and concentrated cells were bright pink (Supplementary Figure S1A,B).

3.2. Phenotypic, Biochemical, and Chemotaxonomic Characterization

Cells of strain MS-P2[T] grown in liquid CTM medium were motile, curved rods with dimensions of 0.5 to 1.0 μm (width) by 1.0 to 1.5 μm (length) (Table 1). Strain MS-P2[T] cells stained Gram-negative and were positive for both catalase and cytochrome *c* oxidase. Division occurred by binary fission (Figure 1A,B). When viewed by epifluorescence microscopy with a filter set specific for BChl *a*, cells exhibited intense autofluorescence, which suggested that the cells contain a substantial amount of BChl *a* (Figure 1B).

Table 1. Differential characteristics of species in the genus *Elioraea*. Positive [+], negative [−], not determined [ND]. All species can synthesize bacteriochlorophyll *a* and carotenoids, are pink in color, are positive for cytochrome *c* oxidase, and employ ubiquinone-10 as the major respiratory quinone.

Property	[1] *E. tepidiphila* DSM17972[T]	[2] *E. thermophila* YIM 72297[T]	[3] *E. rosea* PF-30A[T]	*E. tepida* MS-P2[T]
Cell morphology	Rods	Curved rods	Rods	Curved rods
Cell size (width × length (μm))	0.5–1.0 × 1.0–1.5	0.7–0.9 × 2.2–3.2	1.0–1.2 × 2.2–2.6	0.5–1.0 × 1.0–1.5
Motility (flagella genes present)	+	−	+	+
Temperature optimum (range) °C	45–50 (30–50)	55 (45–60)	28 (20–40)	45–50 (>35–<55)
pH optimum (range)	8.0–8.5 (6.0–9.8)	7.0–7.5 (5.0–9.0)	7.0 (5.0–9.0)	7.0–7.5 (6.0–10.0)
Catalase	+	+	−(but *katG* is present) [4]	+
Photoautotrophic growth	−	+	−	−
Substrate Utilization [5]				
D-Fructose	−	+	−	+
D-Fucose	−	−	−	+
D-Galactose	−	−	−	+
D-Mannose	−	−	−	+
D-Melibiose	−	−	−	+
D-Psicose	−	−	ND	+
Glutamate	+	ND	ND	−
Glutamine	+	ND	ND	−
Proline	+	ND	ND	+
Compounds Hydrolyzed				
Starch	+	−	−	−
Gelatin	−	−	−	+
Aesculin	−	−	+	+
Pectin	−	ND	ND	+
Casein	−	−	−	ND

[1] Data are from references [13,14]. [2] Data are from reference [14]. [3] Data are from reference [19]. [4] *katG*, catalase. [5] Positive reactions detected in Biolog Gen III plates for strain MS-P2[T] reflect a chemical transformation and can be due to enhanced respiration and/or growth and do not necessarily mean that a substrate supports growth. The complete list of compounds tested is available online at https://www.biolog.com/wp-content/uploads/2020/04/00P_185_GEN_III_MicroPlate_IFU.pdf (accessed on 21 December 2021).
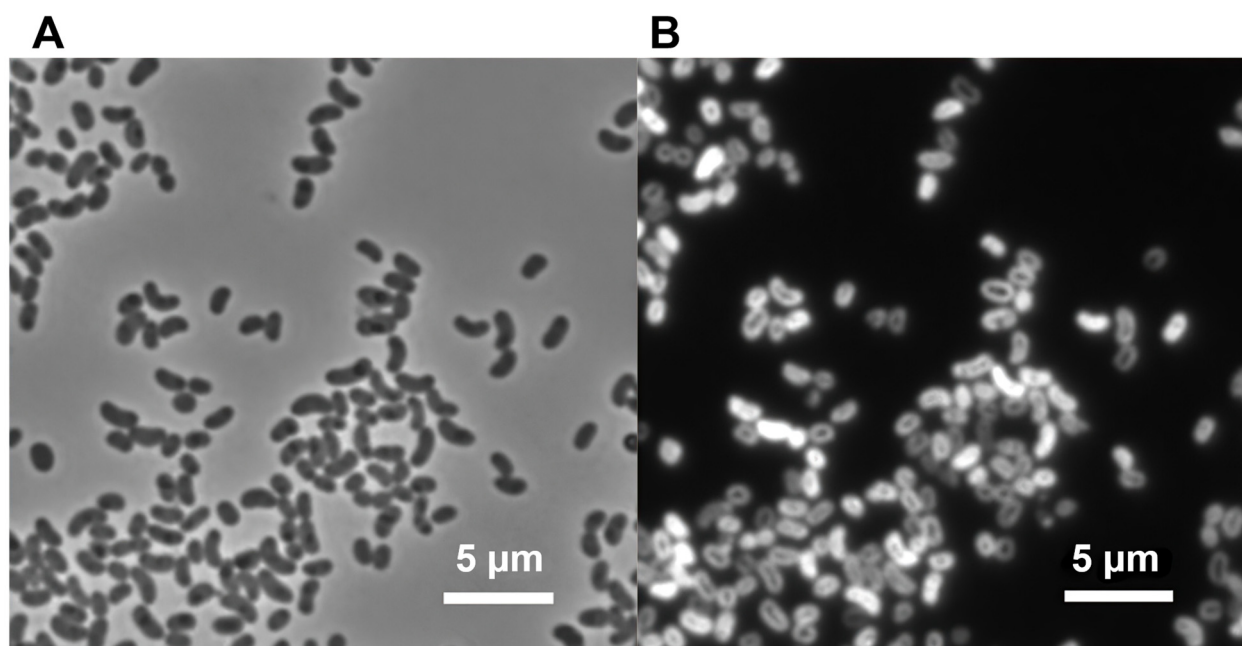
Figure 1. Phase-contrast fluorescence and microscopic images of *E. tepida* strain MS-P2[T]. (**A**) Phase-contrast micrograph of strain MS-P2[T] showing curved, vibrio-shaped cells that multiply by binary fission. (**B**) Fluorescence micrograph of the same field as panel (**A**). Scale bars = 5 μm.

The in vivo absorbance spectrum of strain MS-P2[T] cells that had been disrupted by ultrasonication in phosphate-buffered saline buffer (0.137 M NaCl, 0.0081 M Na_2HPO_4, 0.00268 M KCl, 0.00147 M KH_2PO_4 pH 7.0) is shown in Figure 2. The in vivo spectrum had distinctive absorbance peaks at 800 and 865 nm in the infrared region and a smaller maximum at 590 nm in the visible region, consistent with the presence of BChl *a* (Figure 2). The spectrum also suggested that strain MS-P2[T] has light-harvesting complex 1 (LH1) but that it might lack light-harvesting complex 2 (LH2; however, see below). The three absorbance maxima between 450 and 550 nm suggest that the isolate also produces carotenoids. Pigments were extracted with acetone-methanol (7:2, *v/v*) and an absorbance spectrum was also recorded. The presence of BChl *a* (Q_y band absorbance maximum at 770 nm) and carotenoids (absorbance maxima at 467, 494, and 530 nm) (Figure 2) were also detected in the spectrum of the pigment extract. To confirm this preliminary assessment, extracted pigments were analyzed by reversed-phase HPLC as described [24]. BChl *a* esterified with phytol, a small amount of bacteriopheophytin *a*, and the carotenoids spirilloxanthin (major) and 3,4-dehydrorhodopin (minor) were identified by comparison to authentic standards (Supplementary Figure S2).

Strain MS-P2[T] was able to grow chemoheterotrophically under aerobic conditions in the dark but was unable to grow photoautotrophically under oxic or anoxic conditions in the light. Cells grown in light grew slightly faster than cells grown under the same aerobic chemoheterotrophic conditions in the dark.

The growth temperature range of strain MS-P2[T] was examined by measuring growth at 35, 45, 50, 55, and 60 °C by measuring the OD_{660} of cultures incubated in 60 mL of liquid CTM medium containing succinate, acetate, and yeast extract (contained 1.0 g L^{-1}, respectively) in 100-mL aluminum-foil-capped flasks, with shaking at 150 rpm. The optimal growth of strain MS-P2[T] occurred between 45 and 50 °C (Figure 3A). The strain also grew at 40 °C but did not grow above 55 °C or below 35 °C. Thus, the growth temperature range can be described as ~40 °C and higher but less than 55 °C (Table 1).
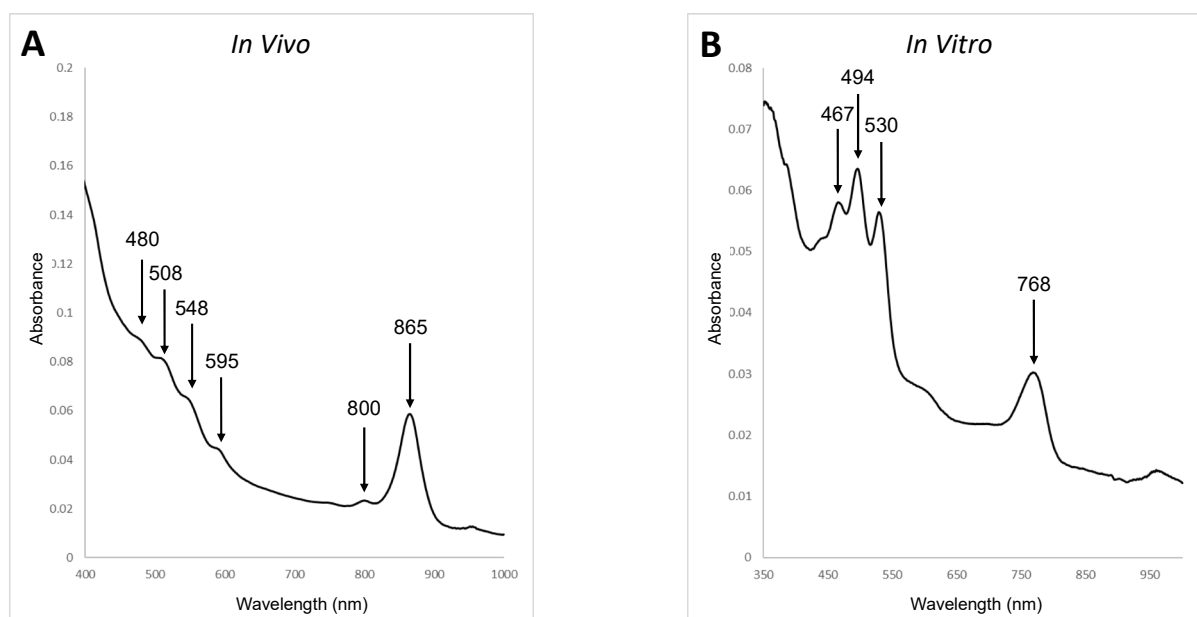
Figure 2. Absorbance spectra of *E. tepida* strain MS-P2[T]. (**A**). An in vivo absorbance spectrum of strain MS-P2[T]. The cells were grown in CTM medium under aerobic conditions in the dark and then were disrupted by ultrasonication. Absorbance maxima are indicated by the arrows. (**B**). Absorbance spectrum of pigments extracted from cells with acetone–methanol (7:2, *v/v*). The absorbance maxima indicate the presence of BChl *a* and carotenoids. For additional details, see text. Similar results were obtained with cells grown in the light.

The pH range for growth was tested from pH 4.0 to 10.5 at 50 °C in the CTM medium by using appropriate buffering agents: 10 mM 2-(N-morpholino)ethanesulfonic acid (MES), 10 mM (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), 10 mM Bicine, and 50 mM N-cyclohexyl-3-aminopropanesulfonic acid (CAPS). Strain MS-P2[T] grew at pH values between 6.0 and 10.0 but was unable to grow at pH 5.5 or at pH 10.5. The optimum pH for growth was pH 7.0 to 7.5 (Figure 3B; Table 1).

Succinate and acetate (1.0 g L^{-1}) clearly promoted growth of strain MS-P2[T] when added to CTM media containing yeast extract (1.0 g L^{-1}). Based on results from Biolog GN2 and GEN3 testing, strain MS-P2[T] was able to oxidize a wide variety of compounds, including 3-methyl-D-glucose, citric acid, D-arabitol, D-fructose, D-fucose, D-galactose, D-mannose, D-melibiose, D-psicose, D-sorbitol, glucuronamide, hydroxybutyric acid, L-arabinose, L-proline, methyl pyruvate, *p*-hydroxy-phenylacetic acid, Tween 40, α-D-glucose, α-D-glucose-1-phosphate, β-hydroxy-D, L-butyric acid, and β-methyl-D-glucoside. Gelatin, aesculin, and pectin were hydrolyzed. Anaerobic growth with nitrate (0.1% (*w/v*) KNO₃) as an electron acceptor was not observed.

Testing with the APIZYM system revealed that alkaline phosphatase, esterase (C4), esterase lipase (C8), leucine arylamidase, valine arylamidase, acid phosphatase, and naphthol-AS-BI-phosphohydrolase were produced.

Strain MS-P2[T] contains the following fatty acids: 15:0 *iso* (7.8%), 16:0 (12.3%), 18:0 (30.0%), 18:1 *ω*7c (9.9%), 19:0 cyclo *ω*8c (13.0%), 18:1 *ω*7c 11-methyl (22.0%), and 18:0 3-OH (4.7%) (Table 2). The major respiratory quinone was ubiquinone-10. The DNA base composition of the isolate as determined by HPLC was 69.1 mol % ± 0.35 mol % G + C; this is in reasonably good agreement with the actual value of 70.6% calculated from the genome sequence (see below).
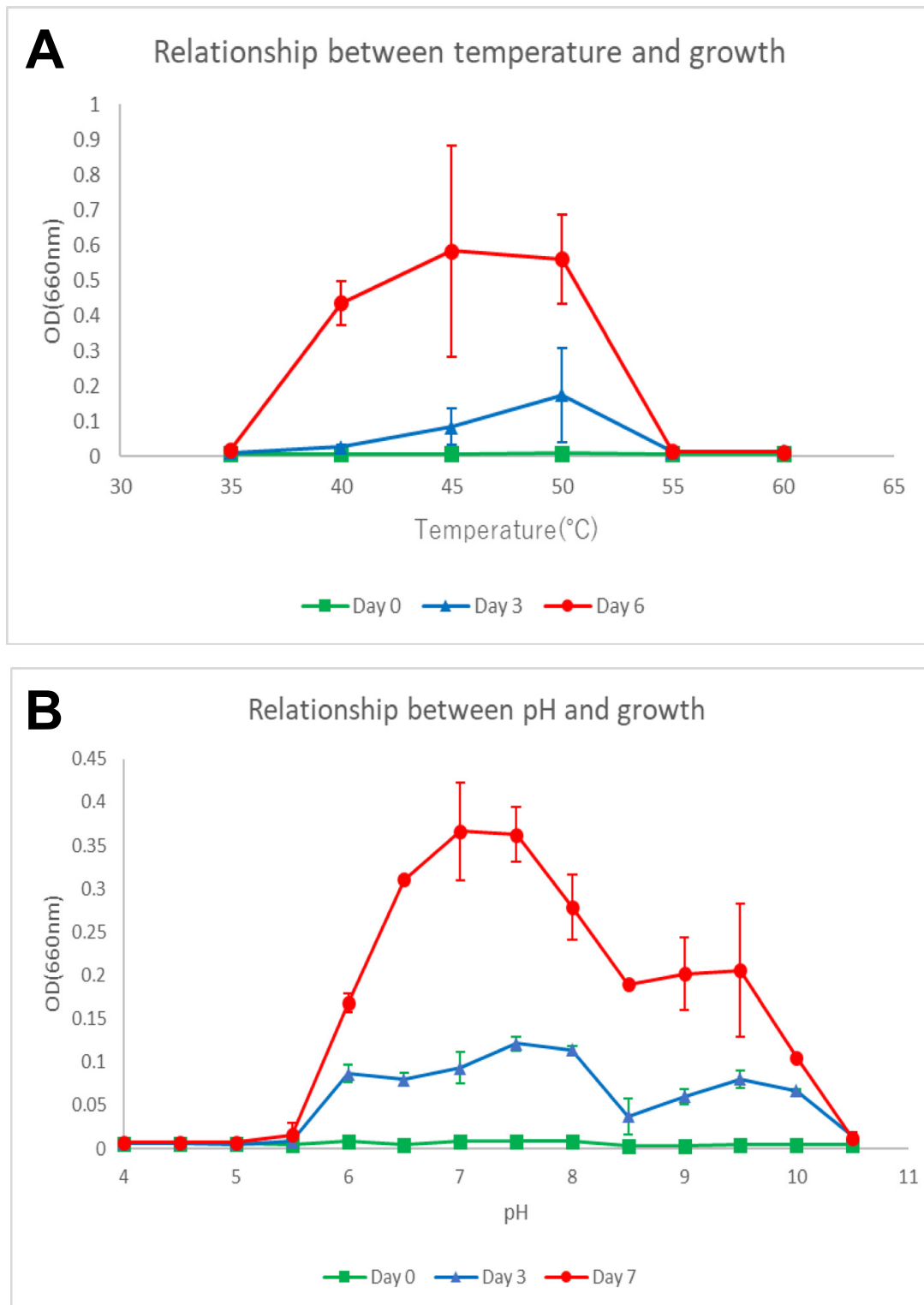
Figure 3. Growth behavior of *Elioraea tepida* strain MS-P2[T] as a function of temperature (**A**) and pH (**B**). Growth of this strain occurred at temperatures greater than 35 °C and lower than 55 °C with an optimum between 45 and 50 °C. Growth occurred over a wide range of pH values above pH 5.5 and below pH 10.5, with an optimum value at pH 7.0 to 7.5. The OD_{660} values at time zero (green line), after three days of growth (blue line), and 6 (temperature) or 7 days (pH) of growth (red line) are plotted. The plotted values are the mean and standard deviation values from triplicate cultures.

Cells in late-exponential growth phase were harvested for polar lipid analysis. Four major and four minor polar lipids were found in strain MS-P2[T]. Phosphatidylcholine (PC), an unidentified aminophospholipid (AP; possibly phosphatidylethanolamine), and two unidentified aminolipids (AL2 and AL3) comprise the major polar lipids. Diphosphatidylglycerol (DPG), phosphatidylglycerol (PG), an unidentified aminolipid (AL1), and an unidentified phospholipid (PL) were the minor polar lipids (Supplementary Figure S3). The polar lipid composition of strain MS-P2[T] is similar to the polar lipid composition of *E. tepidiphila* TU-7[T] [13] and *E. thermophila* YIM 72297[T] [14].

Table 2. Fatty acid composition of *Elioraea* species type strains.

Fatty Acid	[1] *E. tepidiphila* DSM17972[T]	[2] *E. thermophila* YIM 72297[T]	[3] *E. rosea* PF-30A[T]	*E. tepida* MS-P2[T]
14:0	0.3%	–	–	–
15:0 *iso*	–	–	–	7.8%
16:0	5.4%	12.6%	3.6%	12.3%
17:0	0.4%	–	1.6%	–
16:0 2-OH	1.9%	5.4%	2.7%	–
18:1 ω7c	19.0% [4]	30.1%	35.7%	9.9%
18:0	24.8%	35.8%	22.7%	30.0%
18:1 ω7c 11-methyl	8.0%	–	3.4%	22.0%
19:0 cyclo ω8c	12.4%	4.2%	5.7%	13.0%
18:0 2-OH	0.6%	1.5%	2.7%	–
18:0 3-OH	3.8%	2.5%	2.9%	4.7%
19:0 2-OH cyclo ω8 (Δ11:12)	18.6%	–	7.9%	–

[1] Data are from references [13,14]. [2] Data are from reference [14]. [3] Data are from reference [19]. [4] The 18:1 ω7c value for *C. tepidiphila* may also include 18:1 ω6c, which could not be resolved.

3.3. Genomic Features

The assembled genome of MS-P2[T] comprises a single circular contig with a total length of 3,367,643 bp exhibiting 100% completeness with no contamination. No plasmids are present. The mol % G + C content calculated from the genome sequence is 70.6%. The genome encoded 3083 protein-coding sequences (CDS), a single rRNA operon, and 46 tRNAs (Table 3). The genome contains a complete set of genes for the synthesis of bacteriochlorophyll *a*, carotenoids of the spirilloxanthin series, and the photosynthetic apparatus (e.g., *pufABLM*, *puhA*). The presence of *acsF* and *bchE* allows BChl synthesis to occur under both oxic and anoxic conditions, respectively [46]. Surprisingly, although the absorbance spectrum of cells suggests the absence of LH2 antenna complexes, the genome encodes a *pucBAC* operon that could potentially produce such LH2 complexes. Three terminal oxidases (cytochrome aa_3, bb_3, and *bd* oxidases) are encoded in the genome. Genes (*soxBCDYZ*) for oxidation of thiosulfate are present, but consistent with the absence of photoautotrophic growth, genes for phosphoribulokinase, and ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO), and other enzymes for CO_2 fixation are absent. Genes encoding nitrogenase and nitrate and nitrite reductase are missing, but genes for urease (*ureABC*) and its assembly (*ureDEFGJ*) are present (Table 3). Finally, *E. tepida* lacks genes for the synthesis of vitamin B_{12} and methionine synthase, but the genome encodes a methionine transporter.

Figure 4 shows the pairwise relationships based on the calculated average nucleotide identity (ANI) for all five *Elioraea* spp. strains for which genome sequence information was available. These data show that *E. tepidiphila* and *E. thermophila*, with a pairwise ANI

value of 85.08%, are slightly more similar than any other pair of strains, but that all strains are otherwise similarly and distantly related with pairwise ANI values ranging from 73 to 78%. Because ANI values above 95–96% are expected when two strains belong to the same species, strain MS-P2[T] belongs to the genus *Elioraea* but is obviously different from the other three type-species of this genus. Strain MS-P2[T] is also distinct from another recently sequenced isolate, *Elioraea* sp. strain "Yellowstone", from Octopus Spring [20]. Genomic relatedness by digital DNA-DNA hybridization (dDDH) indicates that genomic similarities between strain MS-P2[T] and *E. rosea* PF-30[T], *Elioraea* sp. strain "Yellowstone", *E. tepidiphila* TU-7[T], and *E. thermophila* YIM 72297[T] were only 21.90%, 20.50%, 19.80%, and 18.70%, respectively (Table 3). The low pairwise dDDH values (19 to 22%) are far below the threshold values for species-level relatedness (70% dDDH) [47], and thus these genome-wide comparisons strongly support the proposal that strain MS-P2[T] represents a new species within the genus *Elioraea*.

Table 3. Selected genomic characteristics of strains and species in the genus *Elioraea*.

Genomic Properties	[1] *E. tepidiphila* DSM17972[T]	[2] *E. thermophila* YIM 72297[T]	[3] *E. rosea* PF-30A[T]	[4] *Elioraea* sp. "Yellowstone"	*E. tepida* MS-P2[T]
Size (bp)	>4,304,240	3,029,970	>4,487,660	>3,824,070	3,367,643
Proteins (CDS)	>4014	2845	>4189	>3647	3083
rRNA genes	3	3	3	3	3
tRNAs	48	46	46	46	46
DNA Mol % G + C	71.3	70.9	69.9	72.4	70.6
Average nucleotide identity (%) [5]	75.06	73.34	78.59	75.84	—
Digital DNA-DNA hybridization (%) [5]	19.8	18.7	21.9	20.5	—
16S rRNA % identity [6]	97.26	96.58	97.99	97.52	—
Nitrate reductase (NarG) /growth with nitrate	+/−	−/−	+/−	+/ND	−/−
RubisCO (RbcLS) and Phosphoribulokinase (PRK)	−	+	−	−	−
Urease (UreABCDEFGJ)	−	+	+	−	+
Thiosulfate oxidation (SoxBCDYZ)	+	+	−	+	+

[1] Data are from references [13,14], and GenBank entry NZ_ARKI00000000. [2] Data are from reference [14]. [3] Data are from reference [19]. [4] Data are from reference [20]. [5] Values are the results from comparisons with the *E. tepida* MS-P2[T] genome. [6] Values are the results from comparisons with the *E. tepida* MS-P2[T] 16S rRNA gene sequence.

For phylogenetic assessment of strain MS-P2[T], the complete 16S rRNA gene sequence (1494 bp, NCBI acc. no. MZ358392) was retrieved from the sequenced whole genome. This sequence was 100% identical to the partial 16S rRNA gene sequence derived from an amplified PCR product (NCBI acc. no. MN600983) for this isolate, and the complete sequence was used in phylogenetic calculations. Pairwise nucleotide sequence similarity values for the 16S rRNA gene were calculated with the robust global sequence alignment algorithms in the EzTaxon server (https://www.ezbiocloud.net/ (accessed on 26 May 2021)) [48]. Phylogenetic trees were constructed with the MEGA 7 program [49] using the neighbor-joining, maximum parsimony, and maximum likelihood methods with Kimura's two-parameter model [50] (Figure 5 and Supplementary Figure S4A,B). The maximum likelihood tree based on 16S rRNA gene sequences shows that strain MS-P2[T] is a member of the genus *Elioraea* near the root of the *Acetobacteraceae* together with *E. tepidiphila*, *E. thermophila*, *E. rosea*, and *Elioraea* sp. strain "Yellowstone" (Figure 5). These five isolates form a mono-

phyletic group that is an early-diverging clade within the family *Acetobacteraceae* (identity values of 90–92% for other members of this family). Similar results were obtained for phylogenetic trees calculated by the neighbor-joining and maximum parsimony methods (Supplementary Figure S4A,B). Surprisingly, strain MS-P2[T] is closest in all cases to *E. rosea* with a pairwise sequence identity value of 97.99%, which is below the threshold value for novel species demarcation (proposed 16S rRNA gene sequence similarity threshold value of <98.6% [51]). Sequence identities to *E. tepidiphila* TU-7[T] (=DSM 17972[T]), *E. thermophila* YIM 72297[T], and *Elioraea* sp. strain "Yellowstone" were 97.26%, 96.58%, and 97.52%, respectively (Table 3). These findings support the proposal that strain MS-P2[T] is a novel species within the genus *Elioraea.* Moreover, our findings support the proposal by Hördt et al. that the genus *Elioraea* should be placed in the family *Acetobacteraceae* [15,16].
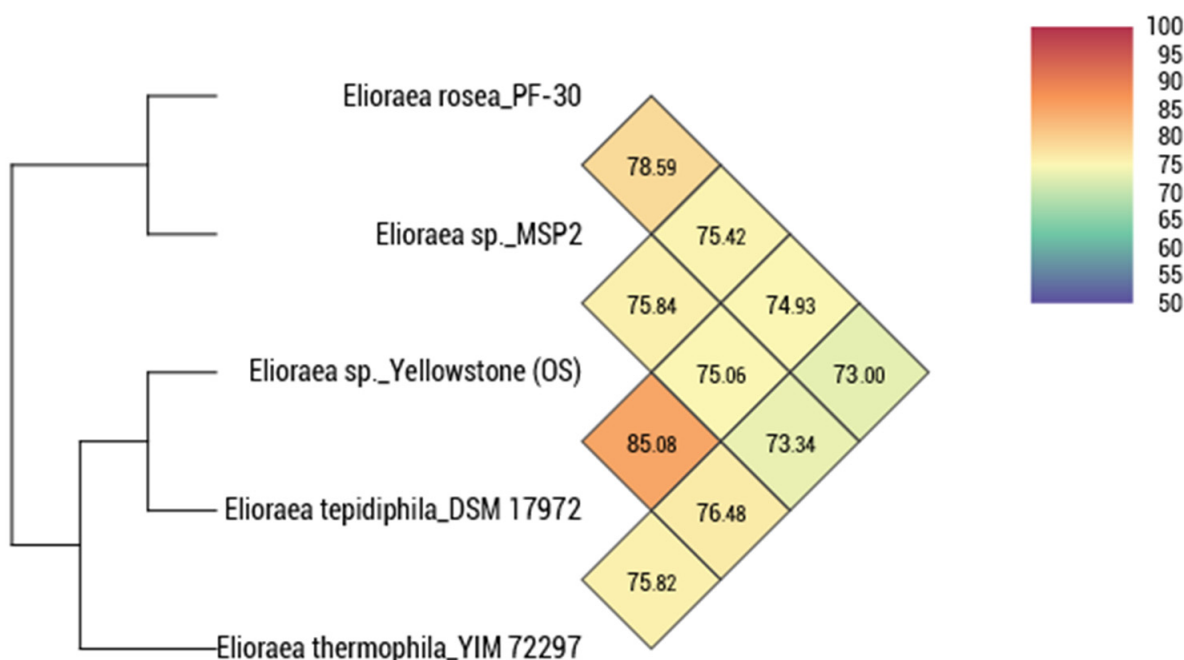
Figure 4. Heat-map and cladogram showing the average nucleotide identity (ANI) values for pairwise comparisons of the genome sequences of five *Elioraea* spp. strains. *Elioraea* sp. strain MS-P2[T] is very slightly more closely related to *E. rosea* (78.59%) than to the other strains. Otherwise, all pairs are roughly equally dissimilar except for *E. tepidiphila* and *Elioraea* sp. "Yellowstone". However, the values for all comparisons are well below the threshold of 95–96 ANI expected for members of the same species.

To assess the relationships among the five *Elioraea* strains currently available further and to assess the relationship of the genus *Elioraea* to other members of the *Alphaproteobacteria*, a phylogenetic analysis was performed that is based upon an alignment of a concatenation of 49 universally conserved proteins (Figure 6). As found in the ANI analysis described above, strain MS-P2[T] was most closely related to the mesophilic strain *E. rosea*; and *E. tepidiphila* and *Elioraea* sp. "Yellowstone" were the second pair of more closely related strains. The five *Elioraea* spp. strains collectively form a monophyletic clade near the base of the strains forming the family *Acetobacteraceae* (Figure 6).
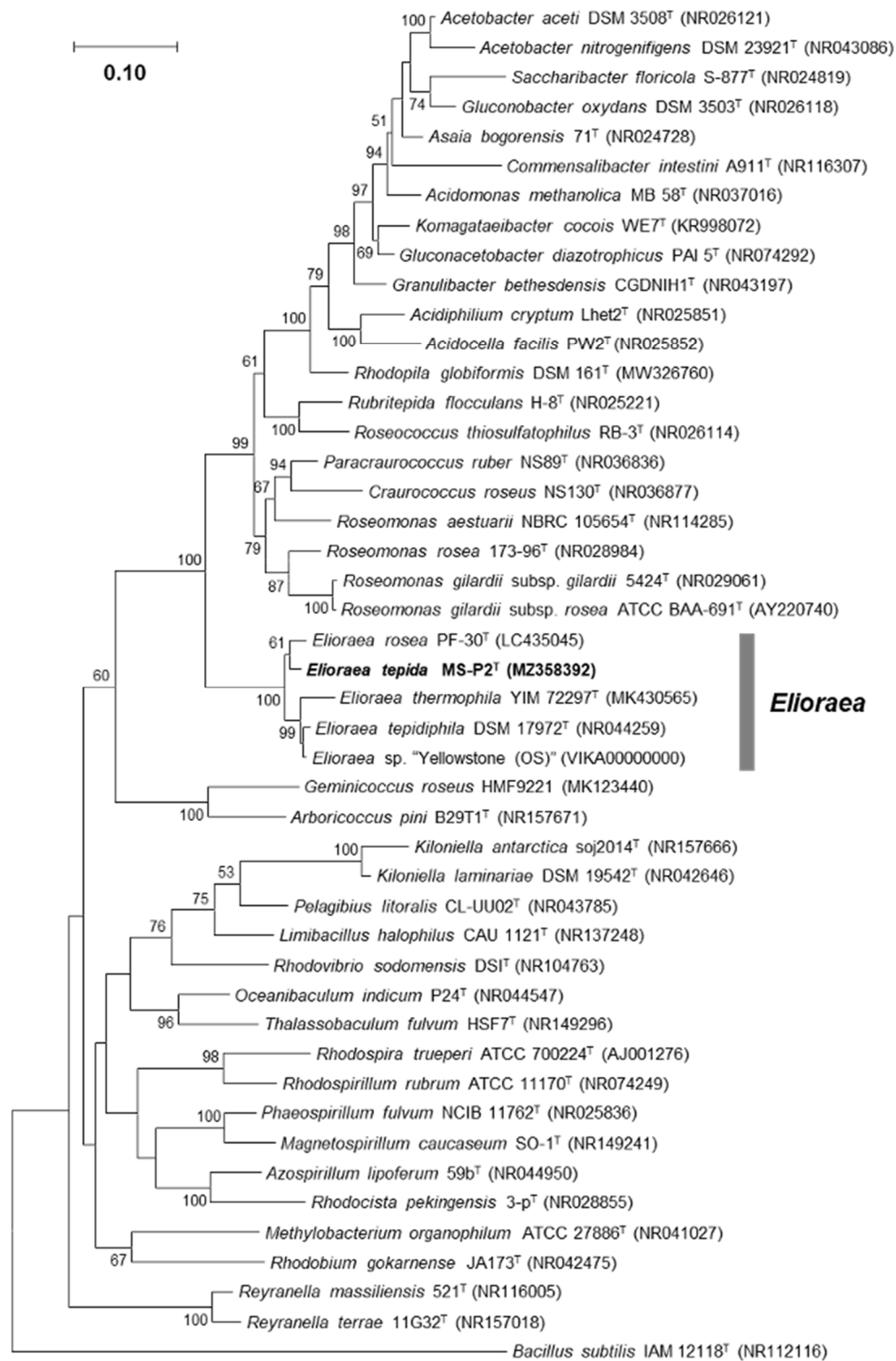
Figure 5. Phylogenetic tree based on 16S rRNA gene sequences constructed by the maximum-likelihood method based on the Kimura 2-parameter model [50] showing the phylogenetic position of strain MS-P2[T] within the order *Rhodospirillales*. Robustness of the maximum-likelihood tree was tested by bootstrapping (100 resamplings, values > 50 are given at the nodes). *Bacillus subtilis* IAM 12118[T] was used as an outgroup. The scale bar represents 0.10 substitution per site. Phylogenetic analyses were conducted in MEGA7 [49]. Trees made with neighbor-joining and maximum parsimony methods are very similar and can be seen in Supplementary Figure S4A,B.
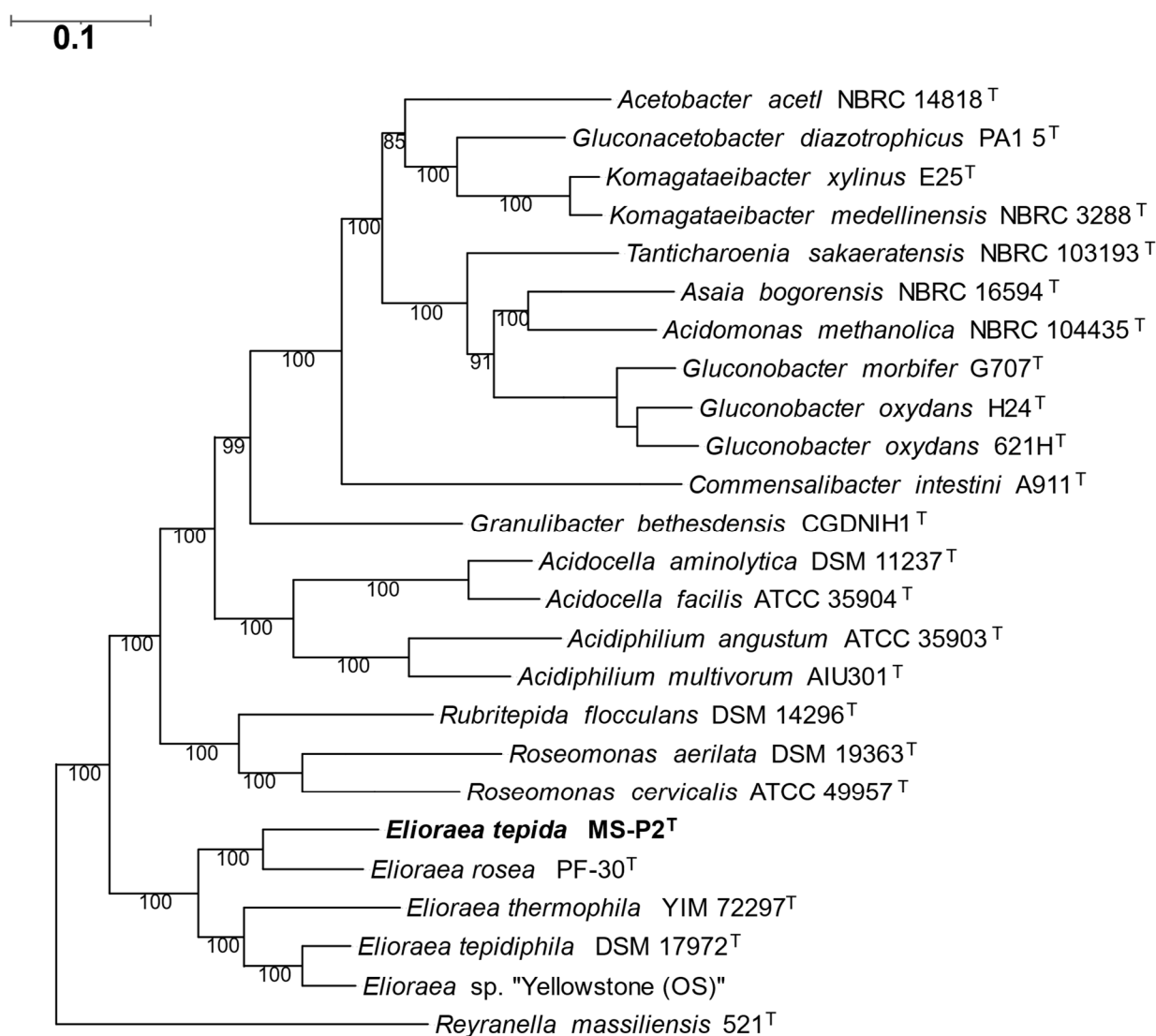
Figure 6. Inferred phylogenetic tree based on concatenated proteins using a set of 49 core universally conserved proteins/genes that shows the phylogenetic position of *Elioraea tepida* MS-P2[T] within the genus *Elioraea* in comparison to selected members of the family *Acetobacteraceae*. The outgroup was *Rayranella massiliensis*, a member of the order *Hyphomicrobiales*. The other strains belonged to the order *Rhodospirillales* and the family *Acetobacteraceae*. The scale bar represents 0.10 substitution per site.

It is interesting and surprising that strain MS-P2[T] isolated from Mushroom Spring is distinctly different from *Elioraea* sp. "Yellowstone" isolated from nearby Octopus Spring. These two hot springs are separated by only about a quarter of a mile in the Lower Geyser Basin of Yellowstone National Park, and they are chemically very similar and have associated mat communities that are also quite similar [42]. However, *Elioraea* sp. "Yellowstone" from Octopus Spring is more similar to *E. tepidiphila* and not to *E. tepida* or *E. rosea*.

Phenotypic, fatty acid, and genotypic characteristics of strain MS-P2[T] and the three validly described species (*E. tepidiphila*, *E. thermophila*, *E. rosea*) [13,14,19] are summarized in Tables 1–3, respectively. These organisms generally share several common phenotypic traits, such as cell shape, swimming motility (except *E. thermophila*), moderate thermophily (except *E. rosea*), use of ubiquinone-10 as the respiratory quinone, production of catalase (except *E. rosea*, but *katG* is present in its genome), and the presence of cytochrome *c* oxidase. Similarities in cellular fatty acid and polar lipid composition also suggest that these organisms are closely related (Table 2). However, there are also important differences

among these four isolates as follows: (1) The moderately thermophilic species were isolated from circum-neutral to slightly alkaline hot spring microbial mats, while the mesophilic *E. rosea* strain was isolated from floodwaters from a paddy field in South Korea. (2) *E. thermophila* is unique among the strains in having both phosphoribulokinase and type-1 ribulose 1,5-bisphosphate carboxylase-oxygenase, suggesting that this strain fixes CO_2 by the Calvin–Benson–Bassham cycle. This provides an explanation for the unique ability among *Elioraea* spp. of *E. thermophila* to grow photoautotrophically [14]. (3) All strains including MS-P2[T] produced BChl *a*, carotenoids, and a similar photosynthetic apparatus, although both *E. tepidiphila* and *Elioraea* sp. strain "Yellowstone" were initially reported to be non-pigmented [13,14,19]. (4) Like *E. rosea* and *E. tepidiphilia*, strain MS-P2[T] showed optimal growth between pH 7.0–7.5, but *E. tepidiphila* prefers slightly more alkaline conditions (pH 8.0–8.5). (5) Strain MS-P2[T] lacked C17:0, C16:0 2-OH, C18:0 2-OH, and 19:0 cyclo ω8 (Δ11:12) cellular fatty acids that were detected in other *Elioraea* spp. and uniquely possesses a significant amount of C15:0 *iso*, a fatty acid not found in any other *Elioraea* spp. (Table 2) (6) Strain MS-P2[T] oxidizes a wide variety of sugars (e.g., fructose, fucose, galactose, mannose, melibiose, and psicose), while in general other *Elioraea* spp. cannot. (7) Strain MS-P2[T] is the only *Elioraea* spp. strain that can hydrolyze gelatin. (8) Strain MS-P2[T] could only oxidize proline, but *E. tepidiphila* can oxidize glutamate and glutamine as well as proline. Based on these phenotypic and genotypic differences and others (Tables 1–3), in addition to the phylogenetic and genomic differences among the strains described above, we propose strain MS-P2[T] to be a new species belonging to the genus *Elioraea* with the suggested name *Elioraea tepida* (See description in Section 3.5).

3.4. Distribution and Ecological Considerations

Ward et al. [52] reported a 16S rRNA sequence, designated "Type O", in the microbial mats associated with Octopus Spring, Yellowstone National Park, WY, USA, in 1992. However, because the type species of the genus *Elioraea* was not described until 2008 [13], the Type O sequence (*Elioraea* spp.) initially could not be attributed to any specific organism. Later, using 16S rRNA amplicon sequencing and metagenomic sequencing of the upper green euphotic layer, a survey study of the mats of nearby Mushroom Spring provided the first evidence that members of the genus *Elioraea* were present in these hot-spring mat communities [17,53]. The 16S rRNA amplicon, which at that time was most similar to the 16S rRNA sequence of *E. tepidiphila*, could also be associated with a bin of sequences derived from the metagenome, which represented about 50% of the genome of the organism [17,53]. From the partial genomic information included in the *Elioraea* bin (Bin 22, OTU-46), it was apparent that the organisms associated with these sequences would likely have the capacity to synthesize BChl *a* and to produce bacterial reaction centers [17,18,53]. Strain MS-P2[T], described herein, and strain "Yellowstone" [20] were subsequently isolated. Interestingly, the latter strain was reported to be non-pigmented [20], as was initially the case for *E. tepidiphila* [13], although the genomes of both suggested that they should produce BChl *a* and carotenoids like other *Elioraea* sp. strains [14,19]; this study. It is possible that these two closely related strains may only synthesize BChl *a* under specific growth conditions.

The physiology of *E. tepida* strain MS-P2[T] described here and the distribution data from the 16S rRNA amplicon sequencing and metagenomic analyses do not necessarily agree [17,53]. The isolate clearly requires oxygen for growth and does not exhibit sensitivity to oxygen that might suggest it to be a microaerophile. However, 16S rRNA amplicon sequencing suggests that members of the genus *Elioraea* are not present or are rare in the uppermost green layer of the mat, where members of *Synechococcus* (*Cyanobacteria*) and *Roseiflexus* (*Chloroflexota*) spp. are the predominant organisms [7,17,18]. Instead, amplicon sequencing showed that sequences from *Elioraea* spp. are found in the undermat [17,53]. This observation indicates that strain MS-P2[T] probably occurs naturally near but below the surface of the mat, i.e., in the upper portion of the undermat that also includes microaerophiles like *Chloracidobacterium thermophilum*. As previously noted, the genome encodes cytochrome *bd*- and *bb*$_3$-type terminal oxidases, which characteristically have a

higher affinity for oxygen than aa_3-type oxidases. The ability of *E. tepida* to oxidize thiosulfate might also help to explain its location below the surface of the mats, away from the highest concentrations of oxygen. Sulfate reduction occurs away from the surface of the mat deeper in the anoxic zone and mostly at night [54]. Because *Elioraea* spp. are found in similar mat communities of both Mushroom and Octopus Springs, and because the two strains are similar but not very closely related, it is likely that these two strains play a specific role in the mat ecophysiology. However, at this time it is unclear what that role might be, and it is likewise uncertain whether *Elioraea* spp. can form specific associations with other microbes in these mat communities. Given the overall complexity of the chlorophototrophs in the Mushroom Spring mats, with at least 18 types of phototrophs identified to date [7,18], as well as dozens of ecotypes of the major mat inhabitants [55], deciphering the interactions among this panoply of phototrophs and the major chemoheterotrophs in the mat community will likely keep microbiologists busy for many years.

3.5. Description of Elioraea tepida sp. nov.

Elioraea, type genus of the family *Elioraeaceae*; te.pi'.da; L. fem. adj. *tepida*, warm.

Cells are motile curved rods, 0.5 to 1.0 μm wide and 1.0 to 1.5 μm long, containing bacteriochlorophyll *a* and carotenoids of the spirilloxanthin series with spirilloxanthin predominant. Chemoheterotrophic growth occurred under strictly aerobic conditions, and light stimulated growth yield slightly; photoautotrophic growth with light, thiosulfate, and bicarbonate did not occur under oxic or anoxic conditions. Colonies in CTM medium containing succinate, acetate, and yeast extract are pinkish in color. Moderately thermophilic; the optimum growth temperature is about 45–50 °C and growth does not occur below 35 or above 55 °C. The optimum pH is between 7.0 and 7.5; growth does not occur at pH 5.5 or below nor above pH 10.5. Growth with nitrate by anaerobic respiration is not observed. Major fatty acids are 15:0 *iso* (7.8%), 16:0 (12.3%), 18:0 (30.0%), 18:1 ω7c (9.9%), 19:0 cyclo ω8c (13.0%), 18:1 ω7c 11-methyl (22.0%), and 18:0 3-OH (4.7%). Growth is stimulated by succinate and acetate (1.0 g L^{-1}), and the following substrates can be oxidized in the presence of 0.1 g L^{-1} yeast extract: 3-methyl-D-glucose, citric acid, D-arabitol, D-fructose, D-fucose, D-galactose, D-mannose, D-melibiose, D-psicose, D-sorbitol, glucuronamide, hydroxybutyric acid, L-arabinose, L-proline, methyl pyruvate, *p*-hydroxy-phenylacetic acid, Tween 40, α-D-glucose, α-D-glucose-1-phosphate, β-hydroxy-D,L-butyric acid, and β-methyl-D-glucoside. Gelatin, aesculin, and pectin are hydrolyzed. Sensitive to troleandomycin, rifamycin SV, minocycline, lincomycin, vancomycin, nalidixic acid, and aztreonam in Biolog GEN III. Alkaline phosphatase, esterase (C4), esterase lipase (C8), leucine arylamidase, valine arylamidase, acid phosphatase, and naphthol-AS-BI-phosphohydrolase are produced. The mol % G + C content of the DNA is 70.6 mol %.

The type-strain, MS-P2^T (= JCM33060^T = ATCC TSD-174^T), was isolated from Mushroom Spring in the Lower Geyser Basin of Yellowstone National Park, WY, USA (GPS coordinates: Lat.: 44.5387, Long.: −110.798).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The GenBank/EMBL/DDBJ accession numbers for the 16S rRNA gene sequences of strain MS-P2^T are MN600983 and MZ358392. The genome sequence data for the type-strain are available under the GenBank/EMBL/DDBJ accession number CP076448. *Elioraea tepida* strain MS-P2^T has been deposited in the Japan Collection of Microorganisms (=JCM 33060^T) and in the American Type Culture Collection (=ATCC TSD-174^T).

Conflicts of Interest: Author M.T. is employed by DSMZ—German Culture Collection of Microorganisms and Cell Cultures, GmbH. The other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Yurkov, V.; Beatty, J.T. Aerobic anoxygenic phototrophic bacteria. *Microbiol. Mol. Biol. Rev.* **1998**, *62*, 695–724. [CrossRef]
2. Yurkov, V.; Hughes, E. Aerobic anoxygenic phototrophs: Four decades of mystery. In *Modern Topics in the Phototrophic Prokaryotes: Environmental and Applied Aspects*; Hallenbeck, P., Ed.; Springer International Publishing: Cham, Switzerland, 2017; pp. 193–214. [CrossRef]
3. Koblížek, M. Ecology of aerobic anoxygenic phototrophs in aquatic environments. *FEMS Microbiol. Rev.* **2015**, *39*, 854–870. [CrossRef]
4. Hanada, S.; Kawase, Y.; Hiraishi, A.; Takaichi, S.; Matsuura, K.; Shimada, K.; Nagashima, K.V. *Porphyrobacter tepidarius* sp. nov., a moderately thermophilic aerobic photosynthetic bacterium isolated from a hot spring. *Int. J. Syst. Bacteriol.* **1997**, *47*, 408–413. [CrossRef]
5. Rainey, F.A.; Silva, J.; Nobre, M.F.; Silva, M.T.; da Costa, M.S. *Porphyrobacter cryptus* sp. nov., a novel slightly thermophilic, aerobic, bacteriochlorophyll *a*-containing species. *Int. J. Syst. Evol. Microbiol.* **2003**, *53*, 35–41. [CrossRef]
6. Alarico, S.; Rainey, F.; Empadinhas, N.; Schumann, P.; Nobre, M.; da Costa, M.S. *Rubritepida flocculans* gen. nov., sp. nov., a new slightly thermophilic member of the alpha-1 subclass of the *Proteobacteria*. *Syst. Appl. Microbiol.* **2002**, *25*, 198–206. [CrossRef]
7. Thiel, V.; Tank, M.; Bryant, D.A. Diversity of chlorophototrophic bacteria revealed in the omics era. *Annu. Rev. Plant Biol.* **2018**, *69*, 21–49. [CrossRef]
8. Yurkov, V.; Stackebrandt, E.; Holmes, A.; Fuerst, J.A.; Hugenholtz, P.; Golecki, J.; Gad'on, N.; Gorlenko, V.M.; Kompantseva, E.I.; Drews, G. Phylogenetic positions of novel aerobic, bacteriochlorophyll-*a*-containing bacteria and description of *Roseococcus thiosulfatophilus* gen. nov., sp. nov., *Erythromicrobium ramosum* gen. nov., sp. nov., and *Erythrobacter litoralis* sp. nov. *Int. J. Syst. Evol. Microbiol.* **1994**, *44*, 427–434. [CrossRef]
9. Margesin, R.; Zhang, D.C. *Humitalea rosea* gen. nov., sp. nov., an aerobic bacteriochlorophyll-containing bacterium of the family *Acetobacteraceae* isolated from soil. *Int. J. Syst. Evol. Microbiol.* **2013**, *63*, 1411–1416. [CrossRef]
10. Saitoh, S.; Suzuki, T.; Nishimura, Y. Proposal of *Craurococcus roseus* gen. nov., sp. nov. and *Paracraurococcus ruber* gen. nov., sp. nov., novel aerobic bacteriochlorophyll *a*-containing bacteria from soil. *Int. J. Syst. Bacteriol.* **1998**, *48*, 1043–1047. [CrossRef]
11. Venkata Ramana, V.; Sasikala, C.; Takaichi, S.; Ramana, C.V. *Roseomonas aestuarii* sp. nov., a bacteriochlorophyll-*a* containing alphaproteobacterium isolated from an estuarine habitat of India. *Syst. Appl. Microbiol.* **2010**, *33*, 198–203. [CrossRef]
12. Wakao, N.; Yokoi, N.; Isoyama, N.; Hiraishi, A.; Shimada, K.; Kobayashi, M.; Kise, H.; Iwaki, M.; Itoh, S.; Takaichi, S. Discovery of natural photosynthesis using Zn-containing bacteriochlorophyll in an aerobic bacterium *Acidiphilium rubrum*. *Plant Cell Physiol.* **2012**, *37*, 889–893. [CrossRef]
13. Albuquerque, L.; Rainey, F.A.; Nobre, M.F.; da Costa, M.S. *Elioraea tepidiphila* gen. nov., sp. nov., a slightly thermophilic member of the *Alphaproteobacteria*. *Int. J. Syst. Evol. Microbiol.* **2008**, *58*, 773–778. [CrossRef]

14. Habib, N.; Khan, I.U.; Xiao, M.; Hejazi, M.S.; Tarhiz, V.; Zhi, X.-Y.; Li, W.-J. *Elioraea thermophila* sp. nov., a thermophilic bacterium from hot spring of the class *Alphaproteobacteria*, emended description of the genus *Elioraea* and proposal of *Elioraeaceae* fam. nov. *Int. J. Syst. Evol. Microbiol.* **2020**, *70*, 1300–1306. [CrossRef]

15. Hördt, A.; Lopez, M.G.; Meier-Kolthoff, J.P.; Schleuning, M.; Weinhold, L.M.; Tindall, B.J.; Gronow, S.; Kyrpides, N.C.; Woyke, T.; Göker, M. Analysis of 1000+ type-strain genomes substantially improves taxonomic classification of *Alphaproteobacteria*. *Front. Microbiol.* **2020**, *11*, 468. [CrossRef]

16. Oren, A.; Garrity, G.M. List of changes in taxonomic opinion no. 32. Notification of changes in taxonomic opinion previously published outside the IJSEM. *Int. J. Syst. Evol. Microbiol.* **2020**, *70*, 4061–4090. [CrossRef]

17. Thiel, V.; Wood, J.M.; Olsen, M.T.; Tank, M.; Klatt, C.G.; Ward, D.M.; Bryant, D.A. The dark side of the mushroom spring microbial mat: Life in the shadow of chlorophototrophs. I. Microbial diversity based on 16S rRNA gene amplicons and metagenomic sequencing. *Front. Microbiol.* **2016**, *7*, 919. [CrossRef]

18. Tank, M.; Thiel, V.; Ward, D.M.; Bryant, D.A. A panoply of phototrophs: An overview of the thermophilic chlorophototrophs of the microbial mats of alkaline siliceous hot springs in Yellowstone National Park, WY, USA. In *Modern Topics in the Phototrophic Prokaryotes: Environmental and Applied Aspects*; Hallenbeck, P., Ed.; Springer International Publishing: Zürich, Switzerland, 2017; pp. 87–137. [CrossRef]

19. Lee, H.-J.; Whang, K.-S. *Elioraea rosea* sp. nov., a plant promoting bacterium isolated from floodwater of a paddy field. *Int. J. Syst. Evol. Microbiol.* **2020**, *70*, 2132–2136. [CrossRef]

20. Robertson, S.; Ramaley, R.F.; Meyer, T.; Kyndt, J.A. Whole-genome sequence of a unique *Elioraea* species strain isolated from a Yellowstone National Park hot spring. *Microbiol. Res. Announc.* **2019**, *8*, e0907-19. [CrossRef]

21. Tank, M.; Bryant, D.A. *Chloracidobacterium thermophilum* gen. nov., sp. nov.: An anoxygenic microaerophilic chlorophoto-heterotrophic acidobacterium. *Int. J. Syst. Evol. Microbiol.* **2015**, *65*, 1426–1430. [CrossRef]

22. Tank, M.; Bryant, D.A. Nutrient requirements and growth physiology of the photoheterotrophic Acidobacterium, *Chloracidobacterium thermophilum*. *Front. Microbiol.* **2015**, *6*, 226. [CrossRef]

23. Saini, M.K.; Sebastian, A.; Siratori, Y.; Soulier, N.T.; Garcia Costas, A.M.; Drautz-Moses, D.I.; Schuster, S.C.; Albert, I.; Harruta, S.; Hanada, S.; et al. Genomic and phenotypic characterization of *Chloracidobacterium* species isolates provides evidence for multiple species. *Front. Microbiol.* **2021**, *12*, 704168. [CrossRef] [PubMed]

24. Frigaard, N.-U.; Maresca, J.A.; Yunker, C.E.; Jones, A.D.; Bryant, D.A. Genetic manipulation of carotenoid biosynthesis in the green sulfur bacterium *Chlorobium Tepidum*. *J. Bacteriol.* **2004**, *186*, 5210–5220. [CrossRef] [PubMed]

25. Hanada, S.; Takaichi, S.; Matsuura, K.; Nakamura, K. *Roseiflexus castenholzii* gen. nov., sp. nov., a thermophilic, filamentous, photosynthetic bacterium that lacks chlorosomes. *Int. J. Syst. Evol. Microbiol.* **2002**, *52*, 187–193. [CrossRef] [PubMed]

26. Minnikin, D.E.; Collins, M.D.; Goodfellow, M. Fatty acid and polar lipid composition in the classification of *Cellulomonas*, *Oerskovia* and related taxa. *J. Appl. Bacteriol.* **1979**, *47*, 87–95. [CrossRef]

27. Bligh, E.G.; Dyer, W.J. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* **1959**, *37*, 911–917. [CrossRef]

28. Saini, M.K.; Weng, C.-C.; Soulier, N.; Sebastian, A.; Albert, I.; Thiel, V.; Bryant, D.A.; Hanada, S.; Tank, M. *Caldichromatium japonicum* gen. nov., sp. nov., a novel thermophilic phototrophic purple sulphur bacterium of the *Chromatiaceae* isolated from Nakabusa hot springs, Japan. *Int. J. Syst. Evol. Microbiol.* **2020**, *70*, 5701–5710. [CrossRef]

29. Parks, D.H.; Imelfort, M.; Skennerton, C.T.; Hugenholtz, P.; Tyson, G.W. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **2015**, *25*, 1043–1055. [CrossRef]

30. Aziz, R.K.; Bartels, D.; Best, A.A.; DeJongh, M.; Disz, T.; Edwards, R.A.; Formsma, K.; Gerdes, S.; Glass, E.M.; Kubal, M.; et al. The RAST Server: Rapid annotations using subsystems technology. *BMC Genom.* **2008**, *9*, 75. [CrossRef]

31. Overbeek, R.; Olson, R.; Pusch, G.D.; Olsen, G.J.; Davis, J.J.; Disz, T.; Edwards, R.A.; Gerdes, S.; Parrello, B.; Shukla, M.; et al. The SEED and the Rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res.* **2014**, *42*, D206–D214. [CrossRef]

32. Brettin, T.; Davis, J.J.; Disz, T.; Edwards, R.A.; Gerdes, S.; Olsen, G.J.; Olson, R.; Overbeek, R.; Parrello, B.; Pusch, G.D.; et al. RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* **2015**, *5*, 8365. [CrossRef]

33. Tatusova, T.; Di Cuccio, M.; Badretdin, A.; Chetvernin, V.; Nawrocki, E.P.; Zaslavsky, L.; Lomsadze, A.; Pruitt, K.D.; Borodovsky, M.; Ostell, J. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **2016**, *44*, 6614–6624. [CrossRef]

34. Haft, D.H.; DiCuccio, M.; Badretdin, A.; Brover, V.; Chetvernin, V.; Badretin, A.; Coulouris, G.; Chitsaz, F.; Derbyshire, M.K.; Durkin, A.S.; et al. RefSeq: An update on prokaryotic genome annotation and curation. *Nucleic Acids Res.* **2018**, *46*, D851–D860. [CrossRef]

35. Li, W.; O'Neill, K.R.; Haft, D.H.; DiCuccio, M.; Chetvernin, V.; Badretdin, A.; Coulouris, G.; Chitsaz, F.; Derbyshire, M.K.; Durkin, A.S.; et al. RefSeq: Expanding the Prokaryotic genome annotation pipeline reach with protein family model curation. *Nucleic Acids Res.* **2021**, *49*, D1020–D1028. [CrossRef]

36. Lee, I.; Kim, Y.O.; Park, S.C.; Chun, J. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. *Int. J. Syst. Evol. Microbiol.* **2015**, *66*, 1100–1103. [CrossRef]

37. Meier-Kolthoff, J.P.; Auch, A.F.; Klenk, H.-P.; Göker, M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinform.* **2013**, *14*, 60. [CrossRef]

38. Meier-Kolthoff, J.P.; Göker, M.; Spröer, C.; Klenk, H.-P. When should a DDH experiment be mandatory in microbial taxonomy? *Arch. Microbiol.* **2013**, *195*, 413–418. [CrossRef] [PubMed]

39. Meier-Kolthoff, J.P.; Göker, M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat. Commun.* **2019**, *10*, 2182. [CrossRef] [PubMed]

40. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2 approximately maximum-likelihood trees for large alignments. *PLoS ONE* **2010**, *5*, e9490. [CrossRef] [PubMed]

41. Bhaya, D.; Grossman, A.R.; Steunou, A.S.; Khuri, N.; Cohan, F.M.; Hamamura, N.; Melendrez, M.C.; Bateson, M.M.; Ward, D.M.; Heidelberg, J.F. Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. *ISME J.* **2007**, *1*, 703–713. [CrossRef]

42. Klatt, C.G.; Wood, J.M.; Rusch, D.B.; Bateson, M.M.; Hamamura, N.; Heidelberg, J.F.; Grossman, A.R.; Bhaya, D.; Cohan, F.M.; Kühl, M.; et al. Community ecology of hot spring cyanobacterial mats: Predominant populations and their functional potential. *ISME J.* **2011**, *5*, 1262–1278. [CrossRef] [PubMed]

43. Liu, Z.; Klatt, C.G.; Wood, J.M.; Rusch, D.B.; Ludwig, M.; Wittekindt, M.; Tomsho, L.P.; Schuster, S.C.; Ward, D.M.; Bryant, D.A. Metatranscriptomic analyses of chlorophototrophs of a hot-spring microbial mat. *ISME J.* **2011**, *5*, 1279–1290. [CrossRef]

44. Klatt, C.G.; Liu, Z.; Ludwig, M.; Kühl, M.; Jensen, S.I.; Bryant, D.A.; Ward, D.M. Temporal metatranscriptomic patterning in phototrophic *Chloroflexi* inhabiting microbial mat in a geothermal spring. *ISME J.* **2013**, *7*, 1775–1789. [CrossRef] [PubMed]

45. Liu, Z.; Klatt, C.G.; Ludwig, M.; Rusch, D.B.; Jensen, S.I.; Kühl, M.; Ward, D.M.; Bryant, D.A. "*Candidatus* Thermochlorobacter aerophilum": An aerobic chlorophotoheterotrophic member of the phylum *Chlorobi*. *ISME J.* **2012**, *6*, 1869–1882. [CrossRef] [PubMed]

46. Bryant, D.A.; Hunter, C.N.; Warren, M.J. Biosynthesis of the modified tetrapyrroles—The pigments of life. *J. Biol. Chem.* **2020**, *295*, 6888–6925. [CrossRef] [PubMed]

47. Richter, M.; Rosselló-Móra, R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 19126–19131. [CrossRef] [PubMed]

48. Yoon, S.H.; Ha, S.M.; Kwon, S.; Lim, J.; Kim, Y.; Seo, H.; Chun, J. Introducing EzBioCloud: A taxonomically united database of 16S rRNA and whole genome assemblies. *Int. J. Syst. Evol. Microbiol.* **2017**, *67*, 1613–1617. [CrossRef]

49. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [CrossRef]

50. Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **1980**, *16*, 111–120. [CrossRef]

51. Keswani, J.; Whitman, W.B. Relationship of 16S rRNA sequence similarity to DNA hybridization in prokaryotes. *Int. J. Syst. Evol. Microbiol.* **2001**, *51*, 667–678. [CrossRef] [PubMed]

52. Ward, D.M.; Bateson, M.M.; Weller, R.; Ruff-Roberts, A.L. Ribosomal RNA analysis of microorganisms as they occur in nature. In *Advances in Microbial Ecology*; Arshall, K., Ed.; Springer: Boston, MA, USA, 1992; Volume 12, pp. 219–286. [CrossRef]

53. Thiel, V.; Hügler, M.; Ward, D.M.; Bryant, D.A. The dark side of the Mushroom Spring microbial mat: Life in the shadow of chlorophototrophs. II. Metabolic functions of abundant community members predicted from metagenomic analyses. *Front. Microbiol.* **2017**, *8*, 943. [CrossRef]

54. Thiel, V.; Garcia Costas, A.M.; Fortney, N.W.W.; Martinez, J.N.; Roden, E.E.; Boyd, E.S.; Ward, D.M.; Bryant, D.A. "*Candidatus* Thermonerobacter thiotrophicus," a non-phototrophic, sulfate-reducing member of the phylum *Chlorobi* that inhabits hot-spring communities. *Front. Microbiol.* **2019**, *9*, 3159. [CrossRef] [PubMed]

55. Becraft, E.; Wood, J.M.; Rusch, D.B.; Kühl, M.; Jensen, S.I.; Bryant, D.A.; Roberts, D.W.; Cohan, F.M.; Ward, D.M. The molecular dimension of microbial species. 1. Ecological distinctions among, and homogeneity within, putative ecotypes of *Synechococcus* inhabiting the cyanobacterial mat of Mushroom Spring, Yellowstone National Park. *Front. Microbiol.* **2015**, *6*, 590. [CrossRef] [PubMed]

Article

Genomic Comparison, Phylogeny and Taxonomic Reevaluation of the *Ectothiorhodospiraceae* and Description of *Halorhodospiraceae* fam. nov. and *Halochlorospira* gen. nov.

Johannes F. Imhoff [1,*], John A. Kyndt [2] and Terrance E. Meyer [3]

1 GEOMAR Helmholtz Centre for Ocean Research Kiel, Düsternbrooker Weg 20, 24105 Kiel, Germany
2 College of Science and Technology, Bellevue University, Bellevue, NE 68005, USA; jkyndt@bellevue.edu
3 Department of Biochemistry, University of Arizona, Tucson, AZ 85721, USA; temeyer@email.arizona.edu
* Correspondence: jimhoff@geomar.de

Abstract: The *Ectothiorhodospiraceae* family represents purple sulfur bacteria of the *Gammaproteobacteria* found primarily in alkaline soda lakes of moderate to extremely high salinity. The main microscopically visible characteristic separating them from the *Chromatiaceae* is the excretion of the intermediate elemental sulfur formed during oxidation of sulfide prior to complete oxidation to sulfate rather than storing it in the periplasm. We present a comparative study of 38 genomes of all species of phototrophic *Ectothiorhodospiraceae*. We also include a comparison with those chemotrophic bacteria that have been assigned to the family previously and critically reevaluate this assignment. The data demonstrate the separation of *Halorhodospira* species in a major phylogenetic branch distant from other *Ectothiorhodospiraceae* and support their separation into a new family, for which the name *Halorhodospiraceae* fam. nov. is proposed. In addition, the green-colored, bacteriochlorophyll-containing species *Halorhodospira halochloris* and *Halorhodospira abdelmalekii* were transferred to the new genus *Halochlorospira* gen. nov. of this family. The data also enable classification of several so far unclassified isolates and support the separation of *Ectothiorhodospira shaposhnikovii* and *Ect. vacuolata* as well as *Ect. mobilis* and *Ect. marismortui* as distinct species.

Keywords: purple sulfur bacteria; genomic phylogeny; *Ectothiorhodospiraceae*; *Halorhodospiraceae*; new family and genus

1. Introduction

The phototrophic purple sulfur bacteria are *Gammaproteobacteria* that use sulfide and other reduced sulfur sources as photosynthetic electron donors and oxidize these to sulfate as the final oxidation product. The genus *Ectothiorhodospira* originally was included in the *Chromatiaceae* family [1]. Phenotypic differences and the distinction of *Ectothiorhodospira* from *Chromatiaceae* by oligonucleotide patterns of 16S rRNA molecules [2,3] led to their separation in a distinct family, the *Ectothiorhodospiraceae* [4].

The two families of purple sulfur bacteria can be distinguished by the most obvious differences in the oxidation of sulfide which are microscopically visible, the formation of elemental sulfur globules inside the cells in *Chromatiaceae* and outside the cells in *Ectothiorhodospiracae* [5–8]. Other characteristic properties that distinguish species of both families are different internal membrane systems, which are in the form of vesicles in *Chromatiaceae* and in the form of membrane stacks in *Ectothiorhodospiraceae* [5–7], and by a number of chemotaxonomic properties, including the quinone, lipid and fatty acid composition [9,10] and lipopolysaccharide structures [11–13]. Sequence analysis of the 16S rRNA gene from available type strains of *Ectothiorhodospiraceae* had demonstrated the clear divergence of two groups within this family which had been recognized as two different genera: slightly to moderately halophilic species of *Ectothiorhodospira* and extremely halophilic species, which were transferred to the new genus *Halorhodospira* [14].

Further comprehensive phylogenetic studies, including large numbers of phototrophic purple bacteria of *Alpha-*, *Beta-* and *Gammaproteobacteria*, comparing sequences of 16S rRNA genes and *bchXYZ* genes [15], *pufLM* genes [16] and the glycine/sarcosine methyltransferase essential for glycine betaine biosynthesis [17] all supported the recognition of three major groups of purple sulfur bacteria, the *Chromatiaceae*, the *Ectothiorhodospiraceae* (except *Halorhodospira* species) and the extremely halophilic *Halorhodospira* species.

In this communication we compare the genome information of a large number of available strains and species of the *Ectothiorhodospiraceae*, highlight some properties related to sulfur metabolism, electron transport and photosynthesis and discuss the phylogenetic and taxonomic status of these bacteria. The data are in support of a distinction of the *Halorhodospira* species from *Ectothiorhodospiraceae* and *Chromatiaceae* on a family level. The name *Halorhodospiraceae*, fam. nov. is proposed for this new family. *Ectothiorhodospiraceae* and *Halorhodospiraceae* families are defined. In addition, the separation of *Halo-rhodospira halochloris* and *Halorhodospira abdelmalekii* into the new genus *Halochlorospira* gen. nov. as *Halochlorospira halochloris* and *Halochlorospira abdelmalekii* is proposed.

According to genome information, including ANI, most of the chemotrophic genera, which have been assigned in recent years in the literature and in databases to the *Ectothiorhodospiraceae*, should be excluded from this family due to their large phylogenetic distance and significantly different phenotypic properties.

2. Material and Methods

2.1. Genome Sequences Used in this Study

Several genome sequences from phototrophic *Ectothiorhodospiraceae* and *Halo-rhodospiraceae* species have been published before [17–23] and a number of additional genome sequences have been established during this study (Table 1). The genome sequences from other strains of phototrophic bacteria used in this study are the following: *Allochromatium humboldtianum* DSM 21881 (JABZEO000000000), *Allochromatium vinosum* DSM 180 (CP001896), *Chromatium okenii* DSM 169 (NRRQ01000000), *Chromatium okenii* LaCa (PPGH01000000), *Halochromatium glycolicum* DSM 11080 (NRSJ01000000), *Halochromatium salexigens* DSM 4395 (NHSF01000000), *Marichromatium bheemlicum* DSM 18632 (JAAXKX000000000), and *Marichromatium gracile* DSM 203 (SMDC01000000). Those from chemotrophic species are: *Acidiferrobacter thiooxydans* m-1 (PSYR01000000), *Alkalilimnicola ehrlichii* MLHE-1 (CP000453), *Alkalispirillum mobile* DSM 12769 (RCDA01000000), *Aquisalimonas asiatica* CGMCC 1.6291 (FOEG01000000), *Arhodomonas aquaeolei* DSM 8974 (ARGF00000000), *Halofilum ochraceum* XJ16 (LVEG02000000), *Halopeptonella vilamensis* DSM 21056 (VMKO01000000), *Inmirania thermothiophila* DSM 100275 (RJVI01000000), *Nitrococcus mobilis* Nb-231 (AAOF00000000), *Oceanococcus atlanticus* 22II-S10r2 (AQQV00000000), *Spiribacter salinus* M19-40 (CP005963), *Thioalbus denitrificans* DSM 26407 (QPJY01000000), *Thioalkalivibrio sulfidophilus* HL-EbGr7 (CP001339), *Thioalkalivibrio versutus* AL2 (MVAR00000000), *Acidihalobacter prosperus* DSM 5130 (JQSG00000000), *Thiogranum longum* DSM 19610 (SMFX01000000), *Thiohalomonas denitrificans* HLD2 (FMWD01000000) and *Thiohalospira halophila* DSM 15071 (FOMJ01000000).

Table 1. Comparison of genome features of the *Ectothiorhodospiraceae* genomes in this study. The genomes are colored by groups as described in the text.

Genome Name	Strain	Alternate Strain	Type Strain	GenBank Accessions	Contigs	Genome Length	GC mol%	quinone	CDS	Coarse Consis-tency	Fine Consis-tency	Sequencing Depth	Reference
Ectothiorhodospira haloalkaliphila	ATCC 51935		yes	AJUE00000000	36	3,445,226	63.0	MK-7/Q-8	3155	95.3	93.9	Unknown	[20]
Ectothiorhodospira haloalkaliphila	9902	strain C	no	JAJNQR000000000	243	3,666,279	62.7		3602	95.4	93.7	106x	this study
Ectothiorhodospira variabilis	DSM 21381	WN 22	yes	JAJNAO000000000	84	3,416,336	63.0		3250	94.3	92.8	119x	this study
Ectothiorhodospira variabilis		WN 21Y	no	JAJNAP000000000	84	3,452,695	63.0		3302	94.3	92.9	145x	this study
Ectothiorhodospira variabilis		WN 21R	no	JAJNAQ000000000	71	3,484,384	63.0		3341	94.3	92.9	119x	this study
Ectothiorhodospira variabilis		WN 2R	no	JAJNAR000000000	73	3,484,893	63.0		3341	94.3	92.9	132x	this study
Ectothiorhodospira sp.	BSL_9		no	CP011994	1	3,550,080	63.0		3350	95.7	93.1	298x	[21]
Ectothiorhodospira sp.	9100	YC6.1	no	JAJNAN000000000	161	3,238,183	62.3		3169	93.3	92	103x	this study
Ectothiorhodospira sp.	9905		no	JAJNAM000000000	180	3,202,442	62.3		3120	94.8	93.4	52x	this study
Ectothiorhodospira marina	DSM 241	9914	yes	FOAA00000000	41	3,185,852	62.3	MK-7/Q-8	3059	94.8	93.9	378x	[17]
Ectothiorhodospira mobilis	DSM 237		yes	NRSK01000000	119	2,796,053	68.3	MK-7/Q-8	2714	95.4	93.1	539x	[17]
Ectothiorhodospira marismortui	DSM 4180	9410	yes	FOUO00000000	35	2,624,954	68.2	MK-7/Q-8	2474	94.6	93.7	363x	[17]
Ectothiorhodospira marismortui	DG9		no	JAJOZD000000000	40	2,682,160	68.4		2522	94.7	93.7	139x	this study
Ectothiorhodospira vacuolata	DSM 2111	Beta 1	yes	JAJMLZ000000000	40	3,341,893	63.4	MK-7/Q-7	3100	93.7	92.5	125x	this study
Ectothiorhodospira sp.	PHS-1		no	AGBG00000000	114	2,943,210	63.7		2725	95.6	94.2	42x	[19]
Ectothiorhodospira shaposhnikovii	DSM 243	9710	yes	NRSM01000000	154	3,788,226	62.4	MK-7/Q-7	3743	96	93.9	160x	[17]
Ectothiorhodospira sp.	A-7R		no	JAJNQO000000000	47	3,282,361	63.0		3009	96.1	94.4	84x	this study
Ectothiorhodospira sp.	A-7Y		no	JAJNQN000000000	50	3,282,916	63.0		3019	95.9	94.3	52x	this study
Ectothiorhodospira sp.	B14B		no	JAJNQM000000000	164	3,422,169	62.9		3209	95.5	93	10x	this study
Ectothiorhodospira magna	DSM 22250	B7-7	yes	FOFO00000000	58	2,721,342	60.9		2559	94.7	93.6	438x	[17]
Ectothiorhodosinus mongolicus	DSM15479	M9	yes	FTPK00000000	6	1,990,961	55.6		1867	95.2	94.5	593x	[17]
Thiorhodospira sibirica	ATCC 700588		yes	AGFD00000000	186	3,188,869	56.7		2842	93	92.1	30x	[18]
Halorhodospira abdelmalekii	DSM 2110		yes	NRRN01000000	158	3,090,835	62.9	MK-4	2914	94.4	93.2	122x	[17]
Halorhodospira halochloris	DSM 1059	9850	yes	AP017372	1	2,819,782	55.8	MK-4	2699	94.4	93.7	123x	[23]
Halorhodospira halochloris		9851	no	JAJNRA000000000	141	2,883,546	56.1	MK-4	2847	94.1	93.3	136x	this study
Halorhodospira halochloris		9852	no	JAJNRB000000000	188	2,888,425	56.1		2861	93.6	92.8	91x	this study
Halorhodospira neutriphila	DSM 15116		yes	NRSH01000000	287	2,394,163	72.0		2535	95.5	92.3	46x	[17]
Halorhodospira halophila	DSM 244	SL1	yes	CP000544	1	2,678,452	68.0	MK-8	2414	99.7	99.3	Unknown	[22]
Halorhodospira halophila		9630	no	NRRO00000000	45	2,682,427	67.9	MK-8	2522	99.6	98.9	220x	[17]
Halorhodospira halophila		9620	no	NHSE01000000	99	2,703,685	68.4		2600	99.2	98.3	49x	this study
Halorhodospira halophila		9621	no	JAJNRD000000000	87	2,912,850	67.9		2705	99.6	99	70x	this study
Halorhodospira halophila		9622	no	JAJNRE000000000	121	2,909,722	67.9	MK-8	2776	99.6	99	80x	this study
Halorhodospira halophila		9624	no	JAJNRC000000000	20	2,779,891	68.1		2599	97.6	96.6	117x	this study
Halorhodospira halophila		9626	no	NHSN01000000	45	2,857,375	68.2		2720	99.5	98.9	72x	[17]
Halorhodospira halophila		9628	no	JAJNRF000000000	27	2,787,144	68.1		2598	99.4	99	68x	this study
Halorhodospira halophila	M38		no	JAJNQP000000000	101	2,748,046	68.3		2572	99.4	98.8	143x	this study
Halorhodospira halophila	M39old		no	JAJNQQ000000000	105	2,747,719	68.3		2571	99.4	98.8	135x	this study

2.2. Genomic DNA Extraction and Sequencing

For the strains sequenced in this study, genomic DNA was prepared from frozen cells, using the GeneJET DNA purification kit (Thermo Fisher Scientific, Waltham, MA, USA). The quantity and purity of DNA were determined using Qubit and NanoDrop instruments and showed absorbance 260/280 ratios between 1.67 and 2.12. The DNA libraries were prepared with the Nextera DNA flex library prep kit (Illumina, Inc., San Diego, CA, USA). All genomes were sequenced using 500 µL of a 1.8pM library with an Illumina MiniSeq instrument, using paired-end sequencing (2 × 150 bp). Quality control of the reads was performed using FASTQC in BaseSpace (Illumina, version 1.0.0), using a kmer size of 5 and contamination filtering. The data for each was assembled de novo using Unicycler [24] in PATRIC [25]. The genome sequences were annotated using RAST (Rapid Annotations using Subsystem Technology; version 2.0; [26]).

2.3. Whole Genome Comparison

Average percentage nucleotide identity (ANIb) between the whole genomes was calculated using JSpecies [27]. JSpecies uses a pairwise genome comparison algorithm to measure the probability of genomes belonging to the same species, with an arbitrary species cutoff of 95%. The whole genome-based phylogenetic tree was generated within PATRIC [25], using the CodonTree pipeline which uses PGFams as homology groups. Among these selected genomes, 141 PGFams were found using the CodonTree analysis, and the aligned proteins and coding DNA from single-copy genes were used for RAxML analysis [28,29]. 100 rounds of bootstrapping were performed using the 'Rapid bootstrapping option. The resulting Newick file was used in iTOL for tree visualization [30]. Average amino acid identity (AAI) values were calculated from the proteome comparison in PATRIC. Only bi-directional hits were used for this analysis and *Hlr. halophila* SL1^T and *Hlr. halochloris* DSM1059^T (BN9850) as the reference strains. Pairwise 16S rRNA comparisons were performed using LALIGN (EMBL-EBI), using the genome-derived 16S rRNA sequences.

The alignment for the 16S rRNA comparisons was performed using Clustal Omega [31] which uses seeded guide trees and Hidden Markov Model (HMM) profiles to generate multiple sequence alignments. The phylogenetic tree was calculated by the neighbour-joining (NJ) method [32] in JALVIEW [33] and a Newick file was generated. The Jalview NJ method uses the BLOSUM62 substitution matrix to compute a sum of scores for the residue (base) pairs at each aligned position. iTOL was used to draw the phylogenetic trees expressed in the Newick phylogenetic tree format [30]. Due to the lack of complete 16S rRNA sequences from whole genomes, Genbank 16S rRNA sequences were used instead for *Ect. salini* (FM244738.1), *Ect. imhoffii*, *Arhodomonas aquaeoli* (M26631.2) and *Nitrococcus mobilis* (NR_104912.1), for the 16S rRNA phylogenetic tree.

For the synteny analysis, comparative genome regions were generated using global PATRIC PGFam families to determine a set of genes that match a focus gene. The gene set is compared to the focus gene using BLAST and sorted by BLAST scores within PATRIC [25]. The Compare Region Viewer in PATRIC displays the focus gene along with the other genes in the same family and their flanking regions in their genomes. The *bch*B gene was used as a focus gene to analyze synteny of the photosynthetic and bacteriochlorophyll gene cluster.

3. Results and Discussion

With the availability of large numbers of genome sequences, the average nucleotide identity (ANI) of genomes became an established measure to compare the similarity/relatedness of bacteria and is an accepted alternative to the classical DNA–DNA hybridization method, which only provided reliable results if performed by experts familiar with this method. It has been suggested to recognize bacteria with ANI >95% as belonging to the same species [27], while bacteria with ANI <90% would be recognized in most cases as separate species. Those with values between 90 and 95% identity may be argued either way depending on other properties. As ANI is apparently more precise in the differentiation of closely related bacteria compared to the 16S rRNA gene sequences [27], we have analyzed

ANI from all phototrophic *Ectothiorhodospiraceae* species with available genome sequences with representatives of the genera *Ectothiorhodospira*, *Ectothiorhodosinus*, *Thio-rhodospira* and *Halorhodospira*. In addition, representative chemotrophic *Ectothiorhodospiraceae* with available genome sequences were included.

As the species of *Halorhodospira* are clearly distinct from other *Ectothiorhodospiraceae* genera and the separation of *Halorhodospira* species on a family level is proposed, in the following, the term *Ectothiorhodospiraceae* will be used in the strict sense with the exclusion of *Halorhodospira* species and species of *Halorhodospira* will be treated separately.

3.1. Genomes of Ectothiorhodospiraceae

The *Ectothiorhodospiraceae* (excluding *Halorhodospira*) currently contain 10 named phototrophic species. The genome sequences of *Trs. sibirica* ATCC 700588[T] [18], *Ect.* strain PHS-1 [19], *Ect. haloalkaliphila* ATCC 51935[T] [20] and *Ect.* strain BSL-9 [21] have been previously published. In addition, genome sequence information for *Ect. magna* DSM 22250[T], *Ect. marina* DSM 241[T], *Ect. marismortui* DSM 4180[T], *Ect. mobilis* DSM 237[T], *Ect. shaposhnikovii* DSM 243[T] and *Ectothiorhodosinus mongolicus* DSM 15479[T] was presented in earlier publications [15–17]. Here we include additional genome sequences of type strains of *Ect. vacuolata* DSM 2111[T] and *Ect. variabilis* DSM 21381[T] and additional isolates (DG9, A-7R, A-7Y, B14B, WN21Y, WN21R, WN2R, BN9902 = C, BN9100 = YC6.1 and BN9905 = ScotB) and use this information for a comparison of species and strains (Table 1). Based on comparison of genome length and its G + C content (mol%), several groups of strains/species can be clearly distinguished (Table 1). The group containing *Ect. mobilis* DSM 237 and both *Ect. marismortui* strains (DSM 4180 and DG9) has a smaller genome size (2.62–2.80 Mbp) and higher percentage of 68.2–68.4 mol% G + C as compared to the majority of the genomes that have genome sizes from 3.20 Mbp to 3.79 Mbp (with the exception of *Ect.* spec. strain PHS-1 with 2.94 Mbp) and a G + C content of 62.3 to 63.7 mol%. *Ectothiorhodosinus mongolicus*, *Thiorhodospira sibirica* and *Ect. magna* apparently are more distinct and also have a lower G + C content (Table 1).

The ANI comparison of the phototrophic *Ectothiorhodospiraceae* (Table 2) clearly identifies three groups of species, with ANI values >80%: (i) including the type strains of *Ect. vacuolata*, *Ect. shaposhnikovii* and *Ect. magna* (colored orange in Table 2), with *Ect. magna* being clearly more distant to the others with ANI values between 83–87%; (ii) including the type strains of *Ect. haloalkaliphila*, *Ect. variabilis* and *Ect. marina* (colored yellow in Table 2); (iii) including the type strains of *Ect. mobilis* and *Ect. marismortui* (colored green in Table 2). More distant to these groups are *Ectothiorhodosinus mongolicus* and *Thiorhodospira sibirica*, with ANI values of <70% relative to the others.

The whole genome-based phylogenetic tree (Figure 1), constructed using unique PGFams as homology groups, supports the ANI analysis and confirms the three groups of strains and species as separate clades in the tree. It also includes representative genomes of chemotrophic species that had previously been considered as belonging to *Ectothiorhodospiraceae* but are phylogenetically at a distance, which suggests their grouping in a separate new family. The only chemotrophic bacteria that are especially closely related to the phototrophic *Ectothiorhodospiraceae* species, in particular to *Thiorhodospira sibirica*, are species of the two groups of *Thioalkalivibrio* (Figure 1). Of the two distinct groups, one includes, among others, the type species *Thioalkalivibrio versutus* and *Thioalkalivibrio halo-philus*, and the other, among others, *Thioalkalivibrio sulfidophilus* and *Thioalkalivibrio denitrificans* [34].

Table 2. Average percentage nucleotide identity (ANI) between pairs of genomes. ANI values >80% are colored and values above the species cutoff (>95%) are presented in bold.

Ect. vacuolata DSM 2111																		
96.3	*Ect.* PHS-1																	
90.5	90.3	*Ect. shaposhnikovii* DSM 243																
89.4	89.1	88.1	*Ect.* B14B															
89.2	88.7	87.8	**99**	*Ect.* A-7Y														
88.9	88.7	87.8	**99.1**	**99.9**	*Ect.* A-7R													
86.2	85.9	84.6	83.4	83.3	83.6	*Ect. magna* DSM 22250												
76.8	76.8	76.5	76.4	76.4	76.5	76	*Ect.* BSL-9											
76.7	76.6	76.4	76	76.1	76.3	75.9	91.5	*Ect. haloalkaliphila* ATCC 51935										
76.5	76.5	76.2	76	76.1	76.1	76	91.3	**98**	*Ect. haloalkaliphila* 9902									
76.6	76.6	76.6	75.9	76	76.2	75.8	91.2	**98.3**	**98.2**	*Ect. variabilis* WN 22								
76.1	76.1	76	75.7	75.9	76.1	75.5	86.9	86.8	86.8	86.6	*Ect. marina* DSM 241							
76	76.3	75.6	75.5	75.7	75.9	75.5	87.3	86.9	86.8	86.7	**96.7**	*Ect.* YC6.1						
75.9	76.1	75.7	75.7	75.7	75.8	75.5	87.2	86.8	86.8	86.7	**96.8**	**99.9**	*Ect.* 9905					
76	75.4	75.4	75.1	75	75.3	74.9	76.7	76.2	76.3	75.9	76.1	75.9	75.9	*Ect. mobilis* DSM 237				
75.9	75.8	75.4	75	75	75.3	74.9	76.7	76.4	76.4	76.1	76.1	76	76	92.6	*Ect. marismortui* DSM 4180			
75.5	75.7	75	75	75	75.3	74.9	76.2	75.9	76.2	75.9	76	76	76	92.8	**98**	*Ect. marismortui* DG9		
70	70	69.8	69.9	69.8	69.9	69.9	70.2	70.2	70.1	70.1	70.1	70	70	69.5	69.3	69.5	*Ets. mongolicus* DSM 15479	
69.9	69.8	69.8	69.7	69.5	69.7	70	69.5	69.3	69.3	69.3	69.2	69.5	69.4	69.5	69.3	69.2	68.3	*Trs. siberica* ATCC700588

In addition to the whole genome-based tree, we also constructed a 16S rRNA-based phylogenetic tree that includes a comparison of all sequenced phototrophic *Ectothiorhodospiraceae* (Figure 2), including *Ect. salini* and *Ect. imhoffii*, for which there is no genome sequence available. Included for comparison were 16S rRNAs of the chemotrophic species previously assigned under *Ectothiorhodospiraceae*. The results of the 16S rRNA tree are consistent with the whole genome-based tree and ANI comparisons.

These data show that the type strains of *Ect. variabilis* DSM 21381[T] (WN 22[T]) and *Ect. haloalkaliphila* ATCC 51935[T] are highly similar. With an ANI of 98.3%, which is above the suggested 95% species cutoff, they may be regarded belonging to the same species. Both strains originate from soda lakes of the Wadi el-Natrun in Egypt [14,35,36]. Additional isolates from the Wadi el-Natrun (BN9902, WN21Y, WN21R, WN2R) are also highly similar to this species (Table 1 [36]). Despite some earlier results [14,36], 16S rRNA sequences support this high similarity, with 97.2% similarity between the two type strains, but may not necessarily demark species identity. This issue will have to be resolved in more detailed studies.

The results also support the recognition of the following closely related but distinct species. *Ect. marismortui* DSM 4180[T] is close to *Ect. mobilis* DSM 237[T] and *Ect. shaposhnikovii* DSM 243[T] close to *Ect. vacuolata* DSM 2111[T]. Previously, the taxonomic status of *Ect. marismortui* DSM 4180[T] and *Ect. vacuolata* DSM 2111[T] were disputed. They were recognized as distinct species by Imhoff [6] and Imhoff and Süling [14], though it was suggested by Ventura et al. [37,38] to combine *Ect. vacuolata* with *Ect. shaposhnikovii* and *Ect. marismortui* with *Ect. mobilis*, primarily based on data on DNA–DNA hybridization and ribotyping. The results of ribotyping, however, did not clearly support this conclusion. They revealed identity between two strains of *Ect. mobilis* (DSM 237[T] and DSM 240) but clear differences of these to *Ect. marismortui* DSM 4180[T]. They also revealed the identity of two strains of *Ect. vacuolata* (DSM 2111[T] and B3) with clear differences to *Ect. shaposhnikovii* DSM 243[T] [37]. In addition, the values of DNA–DNA hybridization were only 80–84% [37]. In contradiction to this suggestion and based on genetic studies of Rubisco and nitrogenase genes, Tourova et al. [39] recognized the four species as distinct species. The presented data on ANI and whole genome-based comparisons are in good agreement with considerations by Imhoff [6] and Tourova et al. [39], recognizing *Ect. vacuolata* DSM 2111[T] and *Ect. shaposhnikovii* DSM 243[T] as distinct species with only 90% ANI and also recognizing *Ect. marismortui* DSM 4180[T] and *Ect. mobilis* DSM 237[T] as distinct species with ANI values of 93%. Therefore, ANI supports the distinction of these four bacteria at the species level and also confirms the close relationships of the two couples of species.

Figure 1. Whole genome-based phylogenetic tree of the *Ectothiorhodospiraceae*. The support values for the phylogenetic tree are generated using 100 rounds of the 'Rapid bootstrapping' option of RaxML. The tree was rooted at midpoint and the branch length tree scale is defined as the mean number of substitutions per site, which is an average across both nucleotide and amino acid changes. *Oceanococcus* and *Acidiferrobacter* were found not to belong to any of the three families but were included in the tree as outgroups.
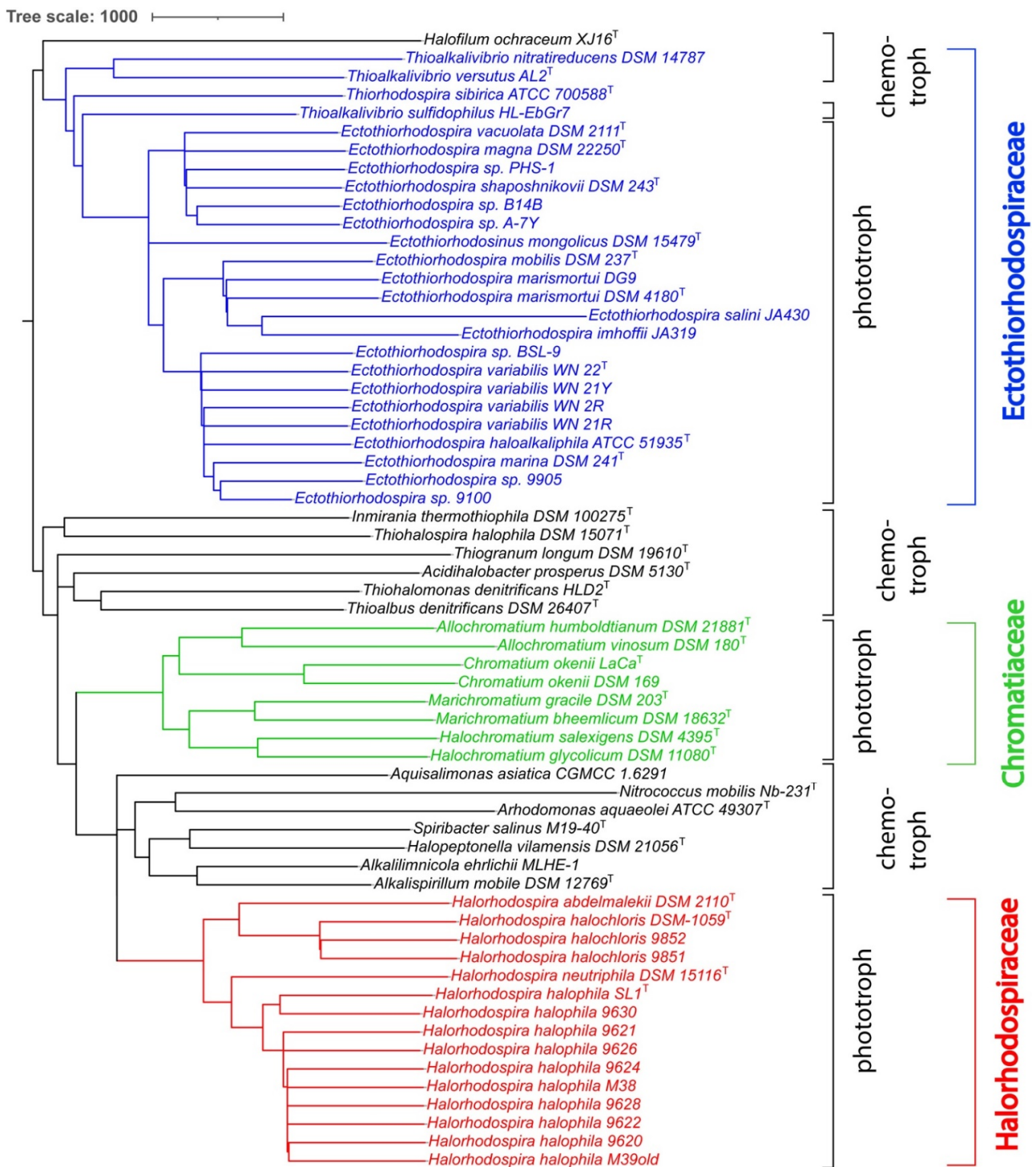
Figure 2. 16S rRNA-derived phylogenetic tree for *Ectothiorhodospiraceae* species. The phylogenetic tree was calculated by the neighbour-joining (NJ) method [32] in Jalview [33]. The Jalview NJ method uses the BLOSUM62 substitution matrix to compute a sum of scores for the residue (base) pairs at each aligned position. The length of the branches is proportional to the number of nucleotide substitutions per site. iTOL was used to draw the phylogenetic trees expressed in the Newick phylogenetic tree format [30].

The results also suggest the assignment of several new isolates to one or the other of the established species. Of two strains that were isolated as arsenite-oxidizing bacteria [40], strain BSL-9 originating from Big Soda Lake represents a distinct new species closely related

to *Ect. haloalkaliphila* and *Ect. marina*, while strain PHS-1 originating from Mono Lake is a strain of *Ect. vacuolata* (Figure 1 and Table 2). Strain DG9 is an additional isolate of *Ect. marismortui* from Berikei Sulfur Springs in Dagestan (Russia, [41]). Strains BN9100 (=YC6.1), an isolate from Solar Lake by H. Biebl (Braunschweig, Germany) and BN9905 (=ScotB), isolated from the seashore near Inverary (Scotland) by one of us (JFI), are new isolates of *Ect. marina.* Strains A-7R, A-7Y and B14B form a separate clade on the whole genome phylogenetic tree and have ANI values <90% with respect to any of the other strains, but high ANI values amongst themselves, and therefore represent a new species related to *Ect. shaposhnikovii* and *Ect. vacuolata* (Figure 1 and Table 2).

The present data support the distinction of *Ectothiorhodosinus mongolicus* [42] and also *Thiorhodospira sibirica* [43] in genera separate from *Ectothiorhodospira*. *Ect. magna* [44], though distantly related to other species of the genus, is placed inside a group of species together with *Ect. shaposhnikovii* and *Ect. vacuolata*.

3.2. Genomes of Halorhodospira Species

Four *Halorhodospira* species are currently recognized: *Hlr. halophila* [45,46], *Hlr. halochloris* [47], *Hlr. abdelmalekii* [48] and *Hlr. neutriphila* [49]. The genome sequence of *Hlr. halophila* SL1[T] was previously determined [22] and information on genome sequences of *Hlr. halophila* BN9626, BN9620, BN9630, *Hlr. neutriphila* DSM 15116[T], *Hlr. halochloris* DSM 1059[T] and *Hlr. abdelmalekii* DSM 2110[T] were reported earlier [15–17].

We have now added genome sequences of additional strains of *Hlr. halochloris* and *Hlr. halophila*, which originate from the Wadi el-Natrun in Egypt [35,50] and from Mongolian soda lakes (Gorlenko, personal communication) (Table 1). The ANI, whole genome-based and also 16S rRNA-based phylogenetic tree comparison of the *Halorhodospira* species (Table 3, Figures 1 and 2) show a clear distinction of five different species.

Table 3. ANI, average percentage nucleotide identity between pairs of genomes. ANI values >90% are colored and values above the species cutoff (>95%) are presented in bold.

SL1	BN9630	BN9626	BN9620	BN9628	BN9624	BN9622	BN9621	M39old	M38	neutriphila	abdelmalekii	BN9852	BN9850	BN9851	Strain
															Hlr. halophila SL1
99.99															*Hlr. halophila* BN9630
85.8	85.8														*Hlr. halophila* BN9626
85.8	85.8	**98**													*Hlr. halophila* BN9620
85.7	85.7	**98.5**	**98.1**												*Hlr. halophila* BN9628
85.7	85.7	**98.5**	**98.1**	**100**											*Hlr. halophila* BN9624
85.7	85.7	**98.1**	**98.5**	**98.4**	**98.4**										*Hlr. halophila* BN9622
85.7	85.7	**98.2**	**97.9**	**98.7**	**98.7**	**98.1**									*Hlr. halophila* BN9621
85.6	85.5	**97.4**	**97.9**	**97.4**	**97.4**	**98**	**97.3**								*Hlr. halophila* M39old
85.6	85.5	**97.3**	**97.8**	**97.4**	**97.4**	**98**	**97.3**	**100**							*Hlr. halophila* M38
77.4	77.4	77.4	77.4	77.6	77.6	77.6	77.6	77.7	77.6						*Hlr. neutriphila* DSM 15116
73.3	73.3	73.7	73.7	73.6	73.6	73.6	73.5	73.4	73.4	74.2					*Hlr. abdelmalekii* BN9840
70.6	70.5	70.9	71.1	71.2	71.2	71.3	71	70.8	70.8	70.7	72.7				*Hlr. halochloris* BN9852
70.6	70.5	70.7	70.9	71	71	71	70.2	70.7	70.8	70.7	72.5	**97.6**			*Hlr. halochloris* BN9850
70.5	70.4	70.7	71	70.9	70.9	70.8	71	70.7	70.8	70.6	72.5	**97.6**	**99.94**		*Hlr. halochloris* BN9851

The first group consists of *Hlr. halophila* SL1[T] and *Hlr. halophila* BN9630. The close relationship of strain BN9630 to the *Hlr. halophila* type strain was demonstrated previously by 16S rRNA sequence analyses and fatty acid analyses [9,14] and is now supported by near identity in terms of the ANI (>99%) (Table 3). The 16S rRNA sequences delineated from the genome sequence actually are only two bases different between the two strains.

A second group is formed by strain BN9626 and a number of other isolates (BN9620, BN9621, BN9622, BN9624, BN9628, M38 and M39old), which differed from strain SL1[T] with respect to fatty acid composition [9] and revealed only 85–86% ANI identity to *Hlr. halophila* SL1[T] (Table 3). This indicates that these strains could be recognized as strains of a distinct species. The average amino acid identity (AAI) is 89.3% for 2414 orthologues of strains SL1[T] and BN9626, which is consistent with the ANI. The 16S rRNA identity (obtained from the genome sequences) between *Hlr. halophila* BN9626 and SL1[T] with 18 differences out of 1537 bp is 98.8%, which is within a range allowing species distinction according to [51] (see systematic conclusions, below). More detailed studies are required to clarify the situation.

Third, according to the whole genome phylogenetic tree (Figure 1) *Hlr. neutriphila* DSM 15166[T] [49] is clearly distinct from *Hlr. halophila*. The ANI to strains of *Hlr. halophila* is

77.4–77.6%, indicating considerably greater divergence. The 16S rRNA obtained from the genome sequence also averages 54 bases different from that of strains SL1^T and BN9626 (96.5% identity), which places *Hlr. neutriphila* on a separate clade on the 16S rRNA tree (Figure 2). The AAI for *Hlr. neutriphila* and strains SL1^T and BN9626 is 72.9% (1730 orthologs). The G + C content of the genome sequence from *Hlr. neutriphila* is 72 mol% and thus significantly higher than all other *Halorhodospira* species (Table 1).

Further, according to the genome-based phylogenetic tree, the green-colored, bacteriochlorophyll *b* containing species *Hlr. halochloris* and *Hlr. abdelmalekii* form a major branch distinct from the groups of red-colored species, which have bacteriochlorophyll *a* (Figure 1). The complete genome sequence of *Hlr. halochloris* DSM 1059^T (=BN9850^T) (APO17372) was established by [23] and we have added a second genome sequence of this strain (NRRM00000000). The genome sequence supposedly determined from the *Hlr. halochloris* strain A^T (=DSM 1059^T) by the authors of [52] turned out to be almost identical to the one from *Ect. haloalkaliphila* ATCC 51935^T (AJUE00000000) and is not from *Hlr. halochloris*. It could originate from a culture contaminant or a mislabeled culture.

The genomes of the three *Hlr. halochloris* strains (BN9850^T, BN9851 and BN9852) have a significantly lower G + C content (55.8–56.1 mol%) compared to *Hlr. abdelmalekii* (62.9 mol%) and other *Halorhodospira* species (>67.9 mol%; Table 1). The genome sequence of *Hlr. abdelmalekii* BN9840^T (= DSM 2110^T) is significantly different from *Hlr. halochloris* (ANI of 72.5%), *Hlr. halophila* and *Hlr. neutriphila* (ANI of approximately 73–74%). There are 65 differences (95.8% identity out of 1551 bp) in the 16S rRNA between *Hlr. abdelmalekii* and *Hlr. halochloris*, as obtained from the genome sequences. The 16S rRNA of the green-colored species, *Hlr. halochloris* and *Hlr. abdelmalekii*, on average has 87 differences to the red-colored species. The AAI between the two green-colored species is 71.6% (1856 orthologs) and that for the two green-colored species and the three red-colored species is 68.6%.

3.3. Genome-Delineated Properties

3.3.1. Sulfur and Thiosulfate Oxidation

In the *Chromatiaceae,* elemental sulfur is stored in the periplasm inside a proteinaceous membrane made up of one to five sulfur globule proteins rich in alternating Gly and Tyr residues [53,54]. None of the *Ectothiorhodospiraceae* and the *Halorhodospira* species, for that matter, store sulfur intracellularly and contain sulfur globule proteins; the sulfur is transported outside the cells and must be transported back inside for further oxidation, once all the sulfide and thiosulfate are exhausted. It is only then that it is further oxidized to sulfate in the cytoplasm. Detailed studies on processes of sulfur oxidation in purple sulfur bacteria have been made by Dahl and coworkers [55,56].

Most of the *Ectothiorhodospiraceae* species contain the thiosulfate-oxidizing enzymes SoxAXYZB, with the apparent exception of *Ect. magna* and *Trs. sibirica,* that only contain SoxY and Z. The genes are organized in two separate operons. As is apparent from the complete genome of *Ect.* BSL-9 containing a single chromosome, the genes *soxA* and *soxX* are in one location and *soxYZB* in another. In those *Halorhodospira* species that use thiosulfate, the diheme SoxA and SoxX are fused into the single triheme protein, SoxXA. *Hlr. abdelmalekii* and *Hlr. halochloris* do not use thiosulfate as a growth substrate and do not contain Sox enzymes. In *Chromatiaceae*, genes of the thiosulfate-oxidizing enzymes are also located in two operons; however, *soxB* is associated with *soxAX* and the two gene clusters are *soxAXB* and *soxYZ*.

3.3.2. Glutathione

Glutathione is a common small molecule reductant in bacteria, which has several important functional roles [57]. It is present in most bacteria, though in some phototrophic purple bacteria it is replaced by the glutathione amide, in which the terminal glycine carboxylate is amidated, as was first shown for *Marichromatium gracile* [58]. In most bacteria, glutathione reductase has an Arg21 that, according to three-dimensional structural analysis, contributes to the binding of glutathione via a salt bridge to the carboxyl group of the

glutathione glycine residue [59]. As found for the glutathione reductase of *Marichromatium gracile* and other *Chromatiaceae*, the Arg21 is substituted by Glu21 and in consequence glutathione, but not its amide, is repelled from the binding site [57]. Thus, glutathione reductase Glu/Arg21 can be used as a proxy for the presence or absence of glutathione amide. We have found through genome sequencing that all *Chromatiaceae* and *Ectothiorhodospiraceae* (with the exceptions of *Ect. mobilis* DSM 237[T] and both strains of *Ect. marismortui*, DSM 4180[T] and DG9, where a glutathione reductase has not been found) have Glu21 and are thus likely to have glutathione amide. However, glutathione reductases of all *Halorhodospira* species have Arg21 like the majority of bacteria and are likely to have the normal glutathione.

3.3.3. Carboxysome Genes

The carboxysome is an important structure in many autotrophic bacteria, such as the cyanobacteria [60] and many autotrophic proteobacteria [61]. The carboxysome forms a proteinaceous membrane that encloses Rubisco and carbonic anhydrase and its genes are associated with bicarbonate transporters. The carboxysome shields Rubisco from nonfunctional reactions with oxygen and provides CO_2 where it is needed. Though a definite proof for the physical structures in purple sulfur bacteria is lacking, we have found genes of putative carboxysome peptides A and B and carboxysome shell proteins *csoS1*, *csoS2* and *csoS3* in many *Chromatiaceae* and *Ectothiorhodospiraceae*. They are present in seven species of *Ectothiorhodospiraceae* but not in in *Ect. mobilis* DSM 237[T] and both strains of *Ect. marismortui* (DSM 4180[T] and DG9), *Ect.* PHS-1 and *Trs. Sibirica*. Carboxysome genes appear to be absent from *Halorhodospiraceae*, though ribulose bisphosphate carboxylase, phosphoribulose kinase and enzymes involved ini the Calvin Cycle are present.

3.3.4. Nitric Oxide Reduction

As mentioned above, *Ectothiorhodospira* strains A-7Y, A-7R and B14B are closely related but different from *Ect. shaposhnikovii* DSM 243[T] and *Ect. vacuolata* DSM 2111[T]. These three strains have 15 unique PGFams that are not present in any of the other *Ectothiorhodospiraceae* genomes and all of them contain a gene cluster for nitric oxide reductase (NOR) that seems to be missing from all the other species. The *norCBQD* cluster contains two NO reductase activation proteins and both the small and large NOR subunits (C and B), which show similarity to cytochrome oxidases. There have been few studies of nitric oxide reduction in the *Ectothiorhodospiraceae*, but these three strains appear to at least have the genetic capability for nitric oxide reduction. In addition to the overall genomic distinction, these genetic differences are another reason to consider these strains as a separate species.

3.3.5. HIPIP

In *Chromatiaceae* the iron–sulfur protein HiPIP was shown to be the electron donor to the photosynthetic reaction center PufLM and the electron acceptor from the BC1 complex [62,63]. The majority of *Ectothiorhodospiraceae* contain either three or four HiPIP isozymes, with the exception of *Ets. mongolicus* and *Trs. sibirica*, which have only one. In addition, *Hlr. halophila* has genes for four HiPIP isoenzymes, two of which are abundant in cell extracts [64]. The gene for one of these isozymes is mixed in with the photosynthetic genes, suggesting that it is the electron donor in this species as it is in *Chromatiaceae* and presumably *Ectothiorhodospiraceae*. It is likely that it acts as a mediator of electrons between the bc_1 complex and photosynthetic reaction center in these bacteria. HiPIP II actually forms a strong complex with the reaction center and rapidly reacts with soluble HiPIP [65]. On the other hand, the green-colored *Halorhodospira* species do not have HiPIP and are likely to utilize cytochrome c_5, which is present in all *Halorhodospira* species characterized to date, but not in the *Ectothiorhodospiraceae*. Cytochrome c_5 is known as a soluble mediator in the green sulfur bacteria in the family *Chlorobiaceae* and the membrane bound c_5 is a possible component of the bc complex [66].

In the green-colored *Halorhodospira* species, there are no HiPIP genes at all, which is an important difference to the red-colored *Halorhodospira* species. In this case, the electron mediator between the bc_1 complex and photosynthetic reaction center must be different. Perhaps due to the difficulty in growing *Hlr. halochloris* and *Hlr. abdelmalekii*, no experimental studies of electron transfer have been published. However, all of the *Halorhodospira* species produce a soluble cytochrome c_5 similar to those found in *Chlorobiaceae* species and which is thought to couple the bc complex to the photosynthetic reaction center. It is proposed that this cytochrome c_5 is the electron mediator in these two species.

3.3.6. Cytochrome b_5

Cytochrome b_5, which is a well-known constituent of eukaryotic cells, has for the first time found in *Ect. vacuolata*, where its function is unknown [67]. It differs from its eukaryotic counterparts in having a cysteine disulfide and the presence of a signal peptide, which suggests that it is located in the periplasm. At least seven examples are now known from the genome sequences of *Ectothiorhodospiraceae*. These are *Ect. vacuolata* DSM 2111^T, *Ect.* PHS-1, *Ect. shaposhnikovii* DSM 243^T, *Ect. mobilis* DSM 237^T, *Ect.* BSL-9, *Ect. haloalkaliphila* ATCC 51935^T and *Ect. marina* DSM 241^T.

The 10-heme cytochromes, MtrA and MtrF (or OmcA), as well as the outer membrane porin MtrB, are involved in the reduction and solubilization of insoluble iron and manganese minerals primarily in *Shewanella* species, as well as in other bacteria [68]. MtrAB are occasionally found in the purple sulfur bacteria, such as *Ect. vacuolata, Ect. shaposhnikovii, Ect. haloalkaliphila, Ect.* BSL-9 and *Ect. marina*. In the species of *Ectothiorhodospiraceae*, these two genes are also associated with OmcA, as in *Shewanella*. Cytochrome c_4 or HiPIP are sometimes found associated with them, suggesting oxidation of FeII rather than reduction of FeIII, since they have high redox potentials. A HiPIP gene is adjacent to the MtrABF cluster in *Ect. vacuolata* DSM 2111^T and *Hrs. halophila*, although not in other species considered herein.

3.3.7. Arsenic Oxidation

The *Ectothiorhodospiraceae* and the *Halorhodospiraceae* are generally capable of oxidizing arsenic (III) to arsenic (V), but only a few species are capable of using As(III) as sole electron donor for growth, including *Ect.* species strain PHS-1 and *Ect.* species strain BSL-9 [40]. The enzymes involved are the molybdopterin proteins ArxA and ArrA, which are closely related. Most *Ectothiorhodospiraceae* have *arrABC* genes, but only *Ect.* PHS-1, *Ect.* BSL-9, *Ect. shaposhnikovii* DSM 243^T, *Ect.* B14B, *Ect.* A-7Y, *Ect.* A-7R, and *Hrs. halophila* SL1^T and BN9630, have five to eight *arx* genes. The authors of [40] examined four of these species plus *Ect. vacuolata* for oxidation of As(III) and checked for As(III) dependent growth and found that only PHS-1 and BSL-9 would grow with arsenic. Therefore, there has to be more to the story because *Ect. shaposhnikovii* and *Hrs. halophila* have the requisite genes but failed to grow.

3.3.8. Photoactive Yellow Protein—PYP

Another interesting difference between the red- and green-colored *Halorhodospira* strains is that only the red strains have a pair of photoactive yellow proteins (PYP) [69]. It is completely absent in the green-colored strains but has been discovered in the *Thio-rhodospira sibirica* genome as well as *Halochromatium salexigens* (*Chromatiaceae*). The only proven role of PYP in purple bacteria is to reverse the effects of red light on the bacteriophytochrome in the hybrid protein PPR in *Rhodocista centenaria*, which is a PYP/bacteriophytochrome/histidine kinase [70]. The *Chromatiaceae* do not have PPR, but some species do have a hybrid PYP/bacteriophytochrome/diguanylate cyclase/phosphodiesterase, which is called PPD [71]. Interaction among the separate domains has not been demonstrated in PPD and the functional role of the PYP domain is likely to be different than it is in PPR.

3.3.9. Photosynthesis Gene Clusters

As other phototrophic *Proteobacteria*, *Ectothiorhodospiraceae* and *Halorhodospira* species have photosynthesis gene clusters, including genes for carotenoid and bacteriochlorophyll biosynthesis, the photosynthetic reaction center and antenna proteins, as well as regulatory and sensory proteins. The structure of the gene clusters and the arrangement of genes show clear differences between *Ectothiorhodospira* and *Halorhodospira* species (Figure 3).

1. BchB: protochlorophyllide reductase subunit B (EC 1.3.7.7)
2. BchH: Protoporphyrin IX Mg-chelatase subunit H (EC 6.6.1.1)
3. BchL: Protochlorophyllide reductase (EC 1.3.7.7)
4. BchM: Mg-protoporphyrin O-methyltransferase (EC 2.1.1.11)
5. PucC protein
6. PuhA: photosynthetic reaction center H subunit
7. PuhB: Photosynthetic complex assembly protein
8. PuhC: Photosynthetic complex assembly protein
9. PuhE: Photosynthetic complex assembly protein
10. BchN: protochlorophyllide reductase subunit N (EC 1.3.7.7)
11. BchF: 2-vinyl bacteriochlorophyllide hydratase (EC 4.2.1.-)
12. PpsR: Transcriptional regulator
13. PpaA: Transcirptional regulator (cobalamin binding protein)
14. Chlorophyll a synthase ChlG (EC 2.5.1.62)
15. Tryptophan synthase (indole-salvaging) (EC 4.2.1.122)

Figure 3. Comparison of the *bch*B genomic region between representatives of *Halorhodospira* and *Ectothiorhodospira* species. Genes are colored based on their family membership.

While in *Halorhodospira* species a cluster with *ppsR-bchFNBHLM* genes (the regulator gene *ppsR* is absent from *Hlr. abdelmalekii* and *Hlr. halochloris*) is present, in *Ectothiorhodospira* species an additional regulatory *ppaA* gene is combined with the *ppsR* gene (genes 12 and 13 in Figure 3).

The presence of the two regulatory genes *ppsR* and *ppaA* in the photosynthetic gene cluster is common to many phototrophic *Alpha-* and *Betaproteobacteria* as well as *Gemmatimonas*. The *ppaA* gene is absent from *Chromatiaceae* and *Halorhodospiraceae* and is found among phototrophic *Gammaproteobacteria* only in the genus *Ectothiorhodospira* (Figure 3). Both regulatory genes are absent from *Ets. Mongolicus.* Quite characteristic for *Ectothio-*

rhodospiraceae, and different to most other purple bacteria, including the *Halorhodospiraceae*, is a gene cluster *ChlG-ppsR-ppaA-bchFNB*, with the exclusion of *bchL* and *bchH* genes from the common *bchFNBHLM* gene cluster, as shown for representatives of the genus *Ectothiorhodospira* (Figure 3). Both *bchH* and *bchL* genes are at separate locations in these bacteria. All *Halorhodospira* species have an additional regulator gene *pufQ* which is located between *bchZ* and *pufB* (*bchCXYZ-pufQBALMC*). This regulator is found in many phototrophic *Alphaproteobacteria* but is absent from all *Ectothiorhodospiraceae* and from *Chromatiaceae* as well. All *Halorhodospira* species and *Ectothiorhodospiraceae* lack the aerobic Mg-protoporphyrin IX monomethylester oxidative cyclase (*acsF*) and therefore depend on anoxic conditions for bacteriochlorophyll biosynthesis using the anaerobic form of the enzyme (encoded by *bchlE*).

3.4. Habitats and Environmental Distribution

Quite remarkably, species of phototrophic *Ectothiorhodospiraceae* and *Halorhodospiraceae* (*Ectothiorhodospira, Halorhodospira*, *Thiorhodospira* and *Ectothiorhodosinus* species), including their chemotrophic relatives, are characteristic inhabitants of marine and saline waters worldwide and preferably develop in alkaline and saline soda lakes. They are phylogenetically related to alkaliphilic chemotrophic sulfur-oxidizing bacteria of the genera *Thioalkalivibrio*, *Alkalilimnicola*, *Alkalispirillum* and are important sulfur-oxidizing chemotrophic bacteria in many alkaline soda lakes [72,73]. A recent review summarizes their occurrence in various types of salt and soda lakes in different geographic regions [74].

Although *Halorhodospira halophila* was first isolated from Summer Lake, Oregon [45,46] and *Hlr. neutriphila* originates from a marine saltern [49], the great majority of the studied strains originate from African soda lakes, most prominently those of the Wadi el-Natrun. This contains all strains of the green-colored species and a second strain of *Hlr. halophila* (9630). Several other strains assigned to this species (including, among others, BN9620, BN9622 and BN9624) are likely distinct on the species level from *Hlr. halophila*. In addition to strains from the Wadi el-Natrun, two isolates from Mongolian soda lakes (M38 and M39old) belong to this presumably new species.

The first intensively studied soda lakes with mass developments of red-colored and green-colored *Halorhodospira* species were those in the Wadi el-Natrun in Egypt [35,50,75]. While pH-optima of the type strain of *Hlr. halophila* SL1^{T} (isolated from the highly alkaline and saline Summer Lake in Oregon) were found to be at 7.4–7.9 [45,46], isolates from the Wadi el-Natrun had pH optima at 8.5–9.0. Two more haloalkaliphilic species, the green-colored, bacteriochlorophyll *b*-producing *Hlr. halochloris* [47] and *Hlr. abdelmalekii* [48], originate from these soda lakes.

In addition, isolates of the moderate halophilic and alkaliphilic *Ectothiorhodospira* species *Ect. haloalkaliphila* [14,35] and *Ect. variabilis,* which is most closely related to *Ect. halo-alkaliphila* [36], were found in soda lakes of the Wadi el-Natrun and also in soda lakes from Siberia and Mongolia [36].

Halorhodospira halophila and two morphological distinct bacteria assigned to the genus *Ectothiorhodospira* were present in Mongolian soda lakes with high salinities (>15% salts), while other species of *Ectothiorhodospira* and *Thiorhodospira* were isolated from lakes with lower salinities of up to 5.5–6.0% salts. For the first time the moderately halophilic *Ectothiorhodosinus mongolicus* (salt optimum 1–7%) was isolated from one of these lakes [42]. Additionally, *Thiorhodospira sibirica* [43] and *Ect. magna* [44] were isolated from soda lakes of remote areas in Asia in Siberia, Mongolia and the Transbaikal region, but have so far not been found in other locations. Possibly the remote origin of these bacteria indicates their separate evolution and is a reason for their distant relationship to other strains of the *Ectothio-rhodospiraceae/Halorhodospiraceae* families.

Other *Ectothiorhodospira* species, especially *Ect. marina* and *Ect. mobilis*, prefer marine habitats, where they have been regularly observed and repeatedly been isolated from.

In recent years, proof of the presence of *Ectothiorhodospiraceae* in alkaline and saline lakes worldwide has been obtained by analysis of clone libraries. The analysis of 16S rRNA

gene clone libraries from several lakes of the Wadi el-Natrun (Lake Fazda, Lake Hamra and Lake UmRisha) revealed a great diversity of sequences related to *Ectothio-rhodospiraceae* species, including *Hlr. halochloris*, and *Ect. haloalkaliphila* [76]. In addition, communities of phototrophic purple bacteria in Chilean salt lakes of the Atacama Desert, Laguna Chaxa and Laguna Tebenquiche that were studied by clone libraries of *pufLM* gene sequences were found to contain bacteria related to *Ect. mobilis, Ect. variabilis* and *Hlr. halophila* as closest relatives [77,78].

4. Systematic Conclusions

The comparison of a large number of genome sequences of phototrophic purple sulfur bacteria in the present study clearly demonstrated the need to separate *Halorhodospira* species and relatives from the *Ectothiorhodospiraceae* and to place them in the new family *Halorhodospiraceae*. The comparison of the genomic phylogeny, including considerations of ANI, demonstrated that the phototrophic *Ectothiorhodospiraceae* as currently known represent two separate families, being almost equally distant from the *Chromatiaceae* family. Based on the significant differences of the extremely halophilic species from all other *Ectothiorhodospiraceae*, we propose to recognize these species as members of a new family, the *Halorhodospiraceae* fam. nov.

The herein proposed reclassification of the families of phototrophic purple sulfur bacteria with the separation of *Halorhodospira* species from the *Ectothiorhodospiraceae* family and the existence of three families of purple sulfur bacteria within the *Gammaproteobacteria* requires a careful reconsideration of species assignment to these families and sheds new light on the assignment of a number of chemotrophic species and genera to the *Ectothiorhodospiraceae* which have been made in recent years but do not warrant to be included within this family.

Among those bacteria that have been assigned to the *Ectothiorhodospiraceae* which without doubt cannot be considered as members of this family are *Acidiferrobacter thiooxydans* [79], now classified with *Acidiferribacteraceae* and *Acidiferribacterales* [80,81] and *Oceanococcus atlanticus* [82]. They do not fit into any of the three phototrophic *Chromatiales* families (Figure 1). In addition, *Halofilum ochraceum* [83], with ANI values to all considered *Ectothiorhodospiraceae* of only 65–68%, should not be included in *Ectothiorhodospiraceae* or *Halorhodospiraceae*. In addition, a few bacteria for which genome sequences at present are not available, should not be included in *Ectothiorhodospiraceae* or *Halorhodospiraceae* based on information available for 16S rRNA gene sequences and other properties. One of these species is *Natronocella acetinitrilica* [84]. Two other species are *Methylonatrum kenyense*, which is a gammaproteobacterium considered to be of unknown affiliation (genus incertae sedis), and *Methylohalomonas lacus* [85], which may be related to *Thioalkali-spira* species of the *Thioalkalispiraceae* [86].

Other species/genera that have been assigned to the *Ectothiorhodospiraceae* are distantly related to the phototrophic *Chromatiaceae, Ectothiorhodospiraceae* or *Halorhodospiraceae* by deep-branching lineages both in genome-based and 16S rRNA-based phylogenetic trees (Figures 1 and 2). These are *Thiohalospira halophila, Thiohalomonas denitrificans, Thiogranum longum, Thioalbus denitrificans, Inmirania thermothiophila* and the acidiphilic *Acidihalobacter prosperous* and related species, which therefore should not be considered as belonging to either *Ectothiorhodospiraceae* or *Halorhodospiraceae*. Their placement in one or more new separate families needs to be evaluated and these bacteria will not be further considered in the present discussion.

Halophilic and alkaliphilic chemotrophic bacteria that are most closely related to phototrophic *Ectothiorhodospiraceae* are species of the genus *Thioalkalivibrio*, particularly those belonging to the *Thioalkalivibrio sulfidophilus* cluster (Figures 1 and 2). Species of both clusters of *Thioalkalivibrio* are the only known chemotrophic bacteria that can be con-sidered to be included in the *Ectothiorhodospiraceae*.

The new family *Halorhodospiraceae* is represented by two major groups of phototrophic bacteria, the red-colored *Hlr. halophila* and relatives and the green-colored *Hlr. halochloris*

and relatives. Phenotypic and phylogenetic differences suggest a separation of the green-colored species in a separate genus, with ANI values of the type species of 70.4–71.3 to strains of *Hlr. halophila*. Although *Hlr. neutriphila* appears phylogenetically and phenotypically (characteristically by the preference for neutral pH and high G + C content of 72 mol%) distinct from *Hlr. halophila*, an ANI >77% to *Hlr. halophila* (Table 3) and 16S rRNA identity of 96.5% to the type strain of *Hlr. halophila* SL1 support recognizing this bacterium as a distinct species of the genus *Halorhodospira*. This is in line with proposals made by others suggesting genus delineations at ANI values close to 74% [51,87].

According to both the phylogenomic and the 16S rRNA tree, a phylogenetic cluster distinct to these phototrophic bacteria is represented by a group of chemotrophic bacteria, including chemotrophic alkaliphilic and halotolerant species, the nitrifying *Nitrococcus mobilis*, *Aquisalimonas asiatica*, *Spiribacter salinus*, *Spiribacter (Halopeptonella) vilamensis* and *Arhodomonas aquaeolei* (Figures 1 and 2). According to 16S rRNA phylogeny, this cluster includes *Alkalispirillum mobilis* and *Alkalilimnicola ehrlichii* and relatives as well, which, according to the genomic tree, form a separate lineage closer to *Halorhodospira*. The phylogenetic distance of the whole group of chemotrophic relatives suggests the exclusion of these bacteria from the *Halorhodospiraceae* and placing them within one or more separate new families. Accordingly, the *Halorhodospiraceae* are represented exclusively by phototrophic bacteria.

Based on the present data, *Ectothiorhodosinus mongolicus* [42] and *Thiorhodospira sibirica* [43] represent genera distinct from *Ectothiorhodospira*. *Ect. magna* [44], though distantly related to other species of the genus, is placed inside a group of species together with *Ect. shaposhnikovii* and *Ect. vacuolata*. ANI values among all studied strains/species of the genus *Ectothiorhodospira* are >74.9% which is in line with a proposed genus demarcation of approximately 74% [88]. Other strains (strains A-7Y, A-7R and B14B) with ANI values to related *Ectothiorhodospira* species below 90% (87.8–89.4) presumably represent a new species of this genus. Although little information is known, and genome sequences are not available for *Ect. salini* and "*Ect. imhoffii*", both species appear most closely related to *Ect. mobilis* and *Ect. marismotui* according to 16S rRNA gene sequences (Figure 2; [88,89]).

Characteristics of the three families.

There are three families of the *Chromatiales* that are characterized by the presence of phototrophic purple sulfur bacteria forming coherent clusters that are, with very few exceptions, clearly distinct from their chemotrophic relatives. These families can be distinguished by a number of phenotypic properties as well as on the basis of genomic properties.

The <u>*Chromatiaceae*</u> primarily live in fresh water or marine habitats; they are phototrophic and primarily autotrophic. When oxidizing sulfide and thiosulfate, intermediate sulfur is stored in the periplasmic space, enclosed within a proteinaceous membrane containing one to five small, related sulfur globule proteins. The thiosulfate dehydrogenase, SoxA, and its electron-acceptor, SoxX, exist as separate subunits. The small protective thiol, glutathione, is present as the amide.

<u>*Ectothiorhodospiraceae*</u> primarily live in shallow alkaline desert soda lakes and also marine shallow waters and tolerate up to about 5–7% salt. They are primarily phototrophic and autotrophic. When oxidizing sulfide and thiosulfate, the intermediate sulfur is excreted into the growth medium without a membrane. SoxA and SoxX exist as separate subunits. Glutathione is present as the amide.

The <u>*Halorhodospiraceae*</u> are found in soda lakes, require elevated salt concentrations for growth and some species can even live in saturated brines. Intermediate elemental sulfur is excreted into the growth medium. SoxX is fused to the N-terminus of SoxA. Glutathione is not amidated. They are phototrophic bacteria forming two distinct groups of species, red-colored species producing bacteriochlorophyll *a* and green-colored species producing bacteriochlorophyll *b*. The red-colored species likely use the small iron–sulfur protein HiPIP as mediator between the bc_1 complex and the photosynthetic reaction center PufLM. On the other hand, the green species are likely to use the small soluble cytochrome c_5 as mediator.

5. Emended Description of the Family *Ectothiorhodospiraceae* Imhoff 1984a, 33^{VP}

Ec.to.thi'o.rho.do.spi.ra'ce.ae. M.L. fem. n. *Ectothiorhodospira* type genus of the family; -aceae ending to denote a family; M. L. fem. pl. n. Ectothiorhodospiraceae the *Ectothiorhodospira* family.

The family constitutes slightly or moderately halophilic phototrophic bacteria growing under alkaline conditions and their chemotrophic relatives. They are Gram-negative and belong to the *Gammaproteobacteria*. Cells are spiral-, vibrioid- or rod-shaped, motile by means of polar flagella and divide by binary fission. They are either phototrophic purple sulfur bacteria that perform anoxygenic photosynthesis with bacteriochlorophylls and carotenoids as photosynthetic pigments or aerobic chemolithotrophic sulfur-oxidizing bacteria, some of which may use nitrate as alternative electron acceptors. Growth of phototrophic representatives is preferably anaerobic in the light, with reduced sulfur compounds as electron donors. Sulfide is oxidized to elemental sulfur and is deposited outside the cells, eventually also in the peripheral periplasmic space of the cell body. The final oxidation product is sulfate. *Ectothiorhodospiraceae* are found in marine and moderately saline environments containing sulfide and having an alkaline to extremely alkaline pH. Glycine betaine, but not ectoine, is the major compatible solute in these bacteria.

The mol% G + C of the DNA is from 53.8–68.4 (genome sequence).

Type genus: *Ectothiorhodospira* Pelsh 1936, 120.

6. Description of the Family *Halorhodospiraceae*. fam. nov.

Ha.lo.rho.do.spi.ra'ce.ae. M.L. fem. n. *Halorhodospira* type genus of the family; -aceae ending to denote a family; M. L. fem. pl. n. *Halorhodospiraceae* the family of *Halorhodospira*.

The family constitutes moderately or extremely halophilic and extremely halotolerant phototrophic bacteria growing under alkaline conditions. They are Gram-negative and belong to the *Gammaproteobacteria*. Cells are spiral-, vibrioid- or rod-shaped, motile by means of polar flagella and divide by binary fission. They are phototrophic purple sulfur bacteria that perform anoxygenic photosynthesis with bacteriochlorophylls and carotenoids as photosynthetic pigments and have internal photosynthetic membranes as lamellar stacks continuous with the cytoplasmic membrane. Photosynthetic pigments are bacteriochlorophyll *a* or *b* and carotenoids. Growth of phototrophic representatives is preferably anaerobic in the light, with reduced sulfur compounds as electron donors. Sulfide is oxidized to elemental sulfur and is deposited outside the cells. The final oxidation product is sulfate. *Halorhodospiraceae* are found in saline, preferably extremely saline environments containing sulfide and having an alkaline to extremely alkaline pH. They are regular inhabitants and represent major groups of the bacterial populations of soda lakes of various salinities. Species of this family are the most halophilic eubacteria. Glycine betaine, ectoine and trehalose accumulate as compatible solutes in response to salt and osmotic stress.

The mol% G + C of the DNA is 55.8–72.0 (genome sequence).

Type genus: *Halorhodospira* Imhoff and Süling 1996, 112; Imhoff and Süling 1997, 915.VP.

6.1. Emended Description of the Genus Halorhodospira

Ha'lo.rho'do. spi'ra. Gr.gen. n. *halos* of the salt; Gr. n. *rhodon* the rose; Gr. n. *spira* the spiral; M.L fem. n. *Halorhodospira*, the spiral rose from salt lakes.

Cells are spiral- or rod-shaped, 0.6–1.2 μm in diameter, motile by bipolar flagella and multiply by binary fission. They are Gram-negative and belong to the *Gammaproteobacteria* and grow photoautotrophically under anoxic conditions with reduced sulfur compounds as electron donors, or photoheterotrophically with a limited number of simple organic compounds. Sulfide is oxidized to elemental sulfur, which is deposited outside the cells and may be further oxidized to sulfate. Internal photosynthetic membranes appear as lamellar stacks continuous with the cytoplasmic membrane. Photosynthetic pigments are bacteriochlorophyll *a* and carotenoids. Growth is dependent on highly saline and alkaline

conditions. Greater than 10% (*w/v*) total salt concentration is required for optimal growth in all known species, some of which even grow in saturated salt solutions. Glycine betaine, ectoine and trehalose accumulate as compatible solutes in response to salt and osmotic stress. Growth factors are not required. Storage products are polysaccharides, poly-β-hydroxybutyrate and polyphosphate. *Halorhodospira* species are found in hypersaline and extremely saline environments, preferably with moderately to extremely alkaline pH (up to pH 11–12), that contain sulfide and are exposed to light, such as salt flats, salt lakes and soda lakes, but some species may also inhabit salterns and coastal lagoons.

The mol% G + C of the DNA is from 67.9–72.0 (genome sequence).

Type species: *Halorhodospira (Hlr.) halophila* (Raymond and Sistrom) Imhoff and Süling 1996, 110.

6.2. Description of the Genus Halochlorospira gen. nov.

Ha'lo.chlo'ro. spi'ra. Gr.gen. n. *halos* of the salt; Gr. n. *chloros* green; Gr. n. *spira* the spiral; M.L fem. n. *Halochlorospira*, the green spiral from salt lakes.

Cells are spiral- or rod-shaped, 0.5–1.2 μm in diameter, motile by bipolar flagella and multiply by binary fission. They are Gram-negative and belong to the *Gammaproteobacteria* and grow photoautotrophically under anoxic conditions with reduced sulfur compounds as electron donors, or photoheterotrophically with a limited number of simple organic compounds. Internal photosynthetic membranes appear as lamellar stacks continuous with the cytoplasmic membrane. Photosynthetic pigments are bacteriochlorophyll *b* and carotenoids. Sulfide is oxidized to elemental sulfur, which is deposited outside the cells and may be further oxidized to sulfate. Growth is dependent on highly saline and alkaline conditions. Greater than 10% (*w/v*) total salt concentration is required for optimal growth by all known species, some of which grow in saturated salt solutions. Glycine betaine, ectoine and trehalose accumulate as compatible solutes in response to salt and osmotic stress. Growth factors not required. Storage products are polysaccharides, poly-β-hydroxybutyrate and polyphosphate. *Halochlorospira* species are found in hypersaline and extremely saline environments with slightly to extremely alkaline pH (up to pH 11–12) that contain sulfide and are exposed to light, such as salt flats, salt lakes and soda lakes.

The mol% G + C of the DNA is from 55.8–62.9 (genome sequence).

Type species: *Halochlorospira (Hcs.) halochloris* (Imhoff and Trüper, 1977).

6.3. Description of Halochlorospira halochloris comb. nov.

The description is entirely the same as for *Halorhodospira halochloris*.

6.4. Description of Halochlorospira abdelmalekii comb. nov.

The description is entirely the same as for *Halorhodospira abdelmalekii*.

References

1. Pfennig, N.; Trüper, H.G. Higher Taxa of the Phototrophic Bacteria. *Int. J. Syst. Bacteriol.* **1971**, *21*, 17–18. [CrossRef]
2. Fowler, V.J.; Pfennig, N.; Schubert, W.; Stackebrandt, E. Towards a phylogeny of phototrophic purple sulfur bacteria? 16S rRNA oligonucleotide cataloguing of 11 species of *Chromatiaceae*. *Arch. Microbiol.* **1984**, *139*, 382–387. [CrossRef]
3. Stackebrandt, E.; Fowler, V.J.; Schubert, W.; Imhoff, J.F. Towards a phylogeny of phototrophic purple sulfur bacteria—The genus *Ectothiorhodospira*. *Arch. Microbiol.* **1984**, *137*, 366–370. [CrossRef]
4. Imhoff, J.F. Reassignment of the Genus *Ectothiorhodospira* Pelsh 1936 to a New Family, *Ectothiorhodospiraceae* fam. nov., and Emended Description of the *Chromatiaceae* Bavendamm 1924. *Int. J. Syst. Bacteriol.* **1984**, *34*, 338–339. [CrossRef]
5. Imhoff, J.F. Family *Ectothiorhodospiraceae*. In *Bergey's Manual of Systematic Bacteriology*, 2nd ed.; Brenner, D.J., Krieg, N.R., Staley, J.T., Eds.; Springer: New York, NY, USA, 2005a; Volume 2, Part B; pp. 41–43.
6. Imhoff, J.F. Genus *Ectothiorhodospira*. In *Bergey's Manual of Systematic Bacteriology*, 2nd ed.; Brenner, D.J., Krieg, N.R., Staley, J.T., Eds.; Springer: New York, NY, USA, 2005b; Volume 2, Part B, pp. 43–48.
7. Imhoff, J.F. Genus *Halorhodospira*. In *Bergey's Manual of Systematic Bacteriology*, 2nd ed.; Brenner, D.J., Krieg, N.R., Staley, J.T., Eds.; Springer: New York, NY, USA, 2005c; Volume 2, Part B, pp. 49–52.
8. Imhoff, J.F. The family *Ectothiorhodospiraceae*. In *The Prokaryotes. An Evolving Electronic Resource for the Microbiological Community*, 3rd ed.; Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., Stackebrandt, E., Eds.; Springer Verlag: New York, NY, USA, 1999.
9. Thiemann, B.; Imhoff, J.F. Differentiation of *Ectothiorhodospiraceae* Based on Their Fatty Acid Composition. *Syst. Appl. Microbiol.* **1996**, *19*, 223–230. [CrossRef]
10. Imhoff, J.F.; Bias-Imhoff, U. Lipids, Quinones and Fatty Acids of Anoxygenic PHOTOTROPHIC bacteria. In *Anoxygenic Photosynthetic Bacteria*; Blankenship, R.E., Madigan, M.T., Bauer, C.E., Eds.; Kluwer Academic Publ.: Dordrecht, The Netherlands, 1995; pp. 179–205.
11. Zahr, M.; Fobel, B.; Mayer, H.; Imhoff, J.F.; Campos, V.P.; Weckesser, J. Chemical composition of the lipopolysaccharides of *Ectothiorhodospira shaposhnikovii*, *Ectothiorhodospira mobilis* and *Ectothiorhodospira halophila*. *Arch. Microbiol.* **1992**, *157*, 499–504.
12. Weckesser, J.; Drews, G.; Mayer, H. Lipopolysaccharides of photosynthetic bacteria. *Ann. Rev. Microbiol.* **1979**, *33*, 215–239. [CrossRef]
13. Weckesser, J.; Mayer, H.; Schulz, G. Anoxygenic Phototrophic Bacteria: Model Organisms for Studies on Cell Wall Macromolecules. In *Anoxygenic Photosynthetic Bacteria*; Blankenship, R.E., Madigan, M.T., Bauer, C.E., Eds.; Kluwer Academic Publ.: Dordrecht, The Netherlands, 1995; pp. 207–230.
14. Imhoff, J.F.; Süling, J. The phylogenetic relationship among *Ectothiorhodospiraceae*: A reevaluation of their taxonomy on the basis of 16S rDNA analyses. *Arch. Microbiol.* **1996**, *165*, 106–113. [CrossRef]
15. Imhoff, J.F.; Rahn, T.; Künzel, S.; Neulinger, S.C. Phylogeny of Anoxygenic Photosynthesis Based on Sequences of Photosynthetic Reaction Center Proteins and a Key Enzyme in Bacteriochlorophyll Biosynthesis, the Chlorophyllide Reductase. *Microorganisms* **2019**, *7*, 576. [CrossRef]
16. Imhoff, J.F.; Rahn, T.; Künzel, S.; Neulinger, S.C. Photosynthesis Is Widely Distributed among Proteobacteria as Demonstrated by the Phylogeny of PufLM Reaction Center Proteins. *Front. Microbiol.* **2018**, *8*, 2679. [CrossRef]
17. Imhoff, J.F.; Rahn, T.; Künzel, S.; Keller, A.; Neulinger, S.C. Osmotic Adaptation and Compatible Solute Biosynthesis of Phototrophic Bacteria as Revealed from Genome Analyses. *Microorganisms* **2019**, *9*, 46. [CrossRef] [PubMed]
18. Lucas, S.; Han, J.; Lapidus, A.; Cheng, J.F.; Goodwin, L.; Pitluck, S.; Peters, L.; Land, M.L.; Hauser, L.; Vogl, K.; et al. The draft genome of *Thiorhodospira sibirica* ATCC 700588. 2011, unpublished.
19. Saltikov, C.W.; Zargar, K.; Conrad, A.; Bernick, D.; Lowe, T.M.; Stolc, V.; Hoeft, S.; Oremland, R.S.; Stolz, J. *Ectothiorhodospira* sp. PHS-1, whole genome shotgun sequence. 2012, unpublished.
20. Bryant, D.A.; Huntemann, M.; Han, J.; Chen, A.; Kyrpides, N.; Mavromatis, K.; Markowitz, V.; Palaniappan, K.; Ivanova, N.; Schaumberg, A.; et al. *Ectothiorhodospira haloalkaliphila* ATCC 51935, whole genome shotgun sequence. DOE Joint Genome Institute. 2013, unpublished.
21. Hernandez-Maldonado, J.; Stoneburner, B.; Boren, A.; Miller, L.; Rosen, M.; Oremland, R.S.; Saltikov, C.W. Genome sequence of the *photoarsenotrophic* bacterium *Ectothiorhodospira* sp. strain BSL-9, isolated from a hypersaline alkaline arsenic-rich extreme environment. *Genome Announc.* **2016**, *4*, e01139-16. [CrossRef]
22. Challacombe, J.F.; Majid, S.; Deole, R.; Brettin, T.S.; Bruce, D.; Delano, S.F.; Detter, J.C.; Gleasner, C.D.; Han, C.S.; Misra, M.; et al. Complete genome sequence of *Halorhodospira halophila* SL1. *Stand. Genom. Sci.* **2013**, *8*, 206–214. [CrossRef]
23. Tsukatani, Y.; Hirose, Y.; Harada, J.; Yonekawa, C.; Tamiaki, H. Unusual features in the photosynthetic machinery of *Halorhodospira halochloris* DSM 1059 revealed by complete genome sequencing. *Photosynth. Res.* **2019**, *140*, 311–319. [CrossRef]
24. Wick, R.R.; Judd, L.M.; Gorrie, C.L.; Holt, K.E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* **2017**, *13*, e1005595. [CrossRef]
25. Wattam, A.R.; Davis, J.J.; Assaf, R.; Boisvert, S.; Brettin, T.; Bun, C.; Conrad, N.; Dietrich, E.M.; Disz, T.; Gabbard, J.L.; et al. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Res.* **2017**, *45*, D535–D542. [CrossRef]
26. Aziz, R.K.; Bartels, D.; Best, A.A.; DeJongh, M.; Disz, T.; Edwards, R.A.; Formsma, K.; Gerdes, S.; Glass, E.M.; Kubal, M.; et al. The RAST server: Rapid annotations using subsystems technology. *BMC Genom.* **2008**, *9*, 75. [CrossRef]

27. Richter, M.; Rosselló-Móra, R.; Glöckner, F.O.; Peplies, J. JSpeciesWS: A web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* **2016**, *32*, 929–931. [CrossRef] [PubMed]

28. Stamatakis, A.; Hoover, P.; Rougemont, J. A Rapid Bootstrap Algorithm for the RAxML Web Servers. *Syst. Biol.* **2008**, *57*, 758–771. [CrossRef] [PubMed]

29. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [CrossRef] [PubMed]

30. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **2019**, *47*, 256–259. [CrossRef]

31. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539. [CrossRef]

32. Saitou, N.; Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **1987**, *4*, 406–425.

33. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.A.; Clamp, M.; Barton, G.J. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009**, *25*, 1189–1191. [CrossRef]

34. Sorokin, D.Y.; Muntyan, M.S.; Pantaleeva, A.N.; Muyzer, G. *Thioalkalivibrio sulfidophilus* sp. nov., a haloalkaliphilic, sulfur-oxidizing gammaproteobacterium from alkaline habitats. *Int. J. Syst. Evol. Microbiol.* **2012**, *62*, 1884–1889. [CrossRef]

35. Imhoff, J.F.; Hashwa, F.; Trüper, H.G. Isolation of extremely halophilic phototrophic bacteria from the alkaline Wadi Natrun, Egypt. *Arch. Hydrobiol.* **1978**, *84*, 381–388.

36. Gorlenko, V.M.; Bryantseva, I.A.; Rabold, S.; Tourova, T.P.; Rubtsova, D.; Smirnova, E.; Thiel, V.; Imhoff, J.F. A novel alkaliphilic and halophilic purple sulfur bacterium *Ectothiorhodospira variabilis* from soda lakes. *Int. J. Syst. Evol. Microbiol.* **2009**, *59*, 658–664. [CrossRef]

37. Ventura, S.; Viti, C.; Pastorelli, R.; Giovannetti, L. Revision of species delineation in the genus *Ectothiorhodospira*. *Int. J. Syst. Evol. Microbiol.* **2000**, *50*, 583–591. [CrossRef] [PubMed]

38. Ventura, S.; Giovanetti, L.; Gori, A.; Viti, C.; Materassi, R. Total DNA restriction pattern and quinone composition of members of the family *Ectothiorhodospiraceae*. *System. Appl. Microbiol.* **1993**, *16*, 405–410. [CrossRef]

39. Tourova, T.P.; Spiridonova, E.M.; Berg, I.; Slobodova, N.V.; Boulygina, E.S.; Sorokin, D.Y. Phylogeny and evolution of the family *Ectothiorhodospiraceae* based on comparison of 16S rRNA, cbbL and nifH gene sequences. *Int. J. Syst. Evol. Microbiol.* **2007**, *57*, 2387–2398. [CrossRef] [PubMed]

40. McCann, S.H.; Boren, A.; Hernandez-Maldonado, J.; Stoneburner, B.; Saltikov, C.W.; Stolz, J.F.; Oremland, R.S. Arsenite as an Electron Donor for Anoxygenic Photosynthesis: Description of Three Strains of *Ectothiorhodospira* from Mono Lake, California and Big Soda Lake, Nevada. *Life* **2016**, *7*, E1. [CrossRef] [PubMed]

41. Gorlenko, V.M.; Burganskaya, E.I.; Bryantseva, I.A. Phototrophic communities of the Berikei highly mineralized mesothermal sulfide springs (Dagestan, Russia). *Microbiology* **2019**, *88*, 146–155. [CrossRef]

42. Gorlenko, V.M.; Bryantseva, I.A.; Panteleeva, E.E.; Turova, T.P.; Kolganova, T.V.; Makhneva, Z.K.; Moskalenko, A.A. *Ectothiorhodosinus mongolicum* gen. nov., sp. nov., a new purple sulfur bacterium from a soda lake in Mongolia. *Microbiology (Moscow)* **2003**, *73*, 66–73. [CrossRef]

43. Bryantseva, I.A.; Gorlenko, V.M.; Kompantseva, E.I.; Imhoff, J.F.; Süling, J.; Mityushina, L. *Thiorhodospira sibirica* gen. nov., sp. nov., a new alkaliphilic purple sulfur bacterium from a Siberian soda lake. *Int. J. Syst. Evol. Microbiol.* **1999**, *49*, 697–703. [CrossRef]

44. Bryantseva, I.A.; Tourova, T.P.; Kovaleva, O.L.; Kostrikina, N.A.; Gorlenko, V.M. *Ectothiorhodospira magna* sp. nov., a new large alkaliphilic purple sulfur bacterium. *Microbiology* **2010**, *79*, 780–790. [CrossRef]

45. Raymond, J.C.; Sistrom, W.R. The isolation and preliminary characterization of a halophilic photosynthetic bacterium. *Arch. Microbiol.* **1967**, *59*, 255–268. [CrossRef]

46. Raymond, J.C.; Sistrom, W.R. *Ectothiorhodospira halophila*: A new species of the genus *Ectothiorhodospira*. *Archiv. Mikrobiol.* **1969**, *69*, 121–126. [CrossRef]

47. Imhoff, J.F.; Trüper, H.G. *Ectothiorhodospira halochloris* sp. nov., a new extremely halophilic phototrophic bacterium containing bacteriochlorophyll b. *Arch. Microbiol.* **1977**, *114*, 115–121. [CrossRef]

48. Imhoff, J.F.; Truper, H.G. *Ectothiorhodospira abdelmalekii* sp. nov., a new halophilic and alkaliphilic phototrophic bacterium. *Zentbl. Bakteriol.* **1981**, *2*, 228–234. [CrossRef]

49. Hirschler-Réa, A.; Matheron, R.; Riffaud, C.; Mouné, S.; Eatock, C.; Herbert, R.A.; Willison, J.C.; Caumette, P. Isolation and characterization of spirilloid purple phototrophic bacteria forming red layers in microbial mats of Mediterranean salterns: Description of *Halorhodospira neutriphila* sp. nov. and emendation of the genus *Halorhodospira*. *Int. J. Syst. Evol. Microbiol.* **2003**, *53*, 153–163. [CrossRef] [PubMed]

50. Imhoff, J.F.; Sahl, H.G.; Soliman, G.S.; Trüper, H.G. The Wadi Natrun: Chemical composition and microbial mass developments in alkaline brines of Eutrophic Desert Lakes. *Geomicrobiol. J.* **1979**, *1*, 219–234. [CrossRef]

51. Stackebrandt, E.; Ebers, J. Taxonomic parameters revisited: Tarnished gold standards. *Microbiol. Today* **2006**, *33*, 152–155.

52. Singh, K.S.; Kirksey, J.; Hoff, W.D.; Deole, R. Draft Genome Sequence of the Extremely Halophilic Phototrophic Purple Sulfur Bacterium *Halorhodospira halochloris*. *J. Genom.* **2014**, *2*, 118–120. [CrossRef]

53. Brune, D.C. Isolation and characterization of sulfur globule proteins from *Chromatium vinosum* and *Thiocapsa roseopersicina*. *Arch. Microbiol.* **1995**, *163*, 391–399. [CrossRef]

54. Pattaragulwanit, K.; Brune, D.C.; Trüper, H.G.; Dahl, C. Molecular genetic evidence for extracytoplasmic localization of sulfur globules in *Chromatium vinosum*. *Arch. Microbiol.* **1998**, *169*, 434–444. [CrossRef]

55. Dahl, C.; Franz, B.; Hensen, D.; Kesselheim, A.; Zigann, R. Sulfite oxidation in the purple sulfur bacterium *Allochromatium vinosum*: Identification of SoeABC as a major player and relevance of SoxYZ in the process. *Microbiology* **2013**, *159*, 2626–2638. [CrossRef]

56. Dahl, C. Cytoplasmic sulfur trafficking in sulfur-oxidizing prokaryotes. *IUBMB Life* **2015**, *67*, 268–274. [CrossRef]

57. Fahey, R.C. Novel Thiols of Prokaryotes. *Annu. Rev. Microbiol.* **2001**, *55*, 333–356. [CrossRef] [PubMed]

58. Bartsch, R.G.; Newton, G.L.; Sherrill, C.; Fahey, R.C. Glutathione amide and its perthiol in anaerobic sulfur bacteria. *J. Bacteriol.* **1996**, *78*, 4742–4746. [CrossRef] [PubMed]

59. van Petegem, F.; De Vos, D.; Savvides, S.; Vergauwen, B.; van Beeumen, J. Understanding nicotinamide dinucleotide cofactor and substrate specificity in class I flavoprotein disulfide oxidoreductases: Crystallographic analysis of a glutathione amide reductase. *J. Mol. Biol.* **2007**, *374*, 883–889. [CrossRef] [PubMed]

60. Kerfeld, C.A.; Melnicki, M.R. Assembly, function and evolution of cyanobacterial carboxysomes. *Curr. Opin. Plant Biol.* **2016**, *31*, 66–75. [CrossRef]

61. MacCready, J.S.; Tran, L.; Basalla, J.L.; Hakim, P.; Vecchiarelli, A.G. The McdAB system positions α-carboxysomes in proteobacteria. *Mol. Microbiol.* **2021**, *116*, 277–297. [CrossRef]

62. Venturoli, G.; Mamedov, M.D.; Mansy, S.S.; Musiani, F.; Strocchi, M.; Francia, F.; Semenov, A.Y.; Cowan, J.A.; Ciurli, S. Electron transfer from HiPIP to the photooxidized tetraheme cytochrome subunit of *Allochromatium vinosum* reaction center: New insights from site-directed mutagenesis and computational studies. *Biochemistry* **2004**, *43*, 437–445. [CrossRef]

63. Nagashima, K.V.P.; Matsuura, K.; Shimada, K.; Verméglio, A. High-Potential Iron−Sulfur Protein (HiPIP) Is the Major Electron Donor to the Reaction Center Complex in Photosynthetically Growing Cells of the Purple Bacterium *Rubrivivax gelatinosus*. *Biochemistry* **2002**, *41*, 14028–14032. [CrossRef]

64. Przysiecki, C.T.; Meyer, T.E.; Cusanovich, M.A. Circular dichroism and redox properties of high redox potential ferredoxins. *Biochemistry* **1985**, *24*, 2542–2549. [CrossRef]

65. Lieutaud, C.; Alric, J.; Bauzan, M.; Nitschke, W.; Schoepp-Cothenet, B. Study of the high-potential iron sulfur protein in *Halorhodospira halophila* confirms that it is distinct from cytochrome c as electron carrier. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 3260–3265. [CrossRef]

66. Azai, C.; Tsukatani, Y.; Itoh, S.; Oh-Oka, H. C-type cytochromes in the photosynthetic electron transfer pathways in green sulfur bacteria and heliobacteria. *Photosynth. Res.* **2010**, *104*, 189–199. [CrossRef]

67. Kostanjevecki, V.; Leys, D.; van Driessche, G.; Meyer, T.E.; Cusanovich, M.A.; Fischer, U.; Guisez, Y.; van Beeumen, J. Structure and characterization of *Ectothiorhodospira vacuolata* cytochrome b(558), a prokaryotic homologue of cytochrome b(5). *J. Biol. Chem.* **1999**, *274*, 35614–35620. [CrossRef] [PubMed]

68. Richter, K.; Schicklberger, M.; Gescher, J. Dissimilatory reduction of extracellular electron acceptors in anaerobic respiration. *Appl. Environ. Microbiol.* **2012**, *78*, 913–921. [CrossRef] [PubMed]

69. Meyer, T.E.; Kyndt, J.A.; Memmi, S.; Moser, T.; Colón-Acevedo, B.; Devreese, B.; Van Beeumen, J.J. The growing family of photoactive yellow proteins and their presumed functional roles. *Photochem. Photobiol. Sci.* **2012**, *11*, 1495–1514. [CrossRef] [PubMed]

70. Kyndt, J.A.; Fitch, J.C.; Meyer, T.E.; Cusanovich, M.A. The Photoactivated PYP Domain of *Rhodospirillum centenum* Ppr Accelerates the Recovery of the Bacteriophytochrome Domain after White Light Illumination. *Biochemistry* **2007**, *46*, 8256–8262. [CrossRef] [PubMed]

71. Kyndt, J.A.; Fitch, J.C.; Meyer, T.E.; Cusanovich, M.A. *Thermochromatium tepidum* Photoactive Yellow Protein/Bacteriophytochrome/Diguanylate Cyclase: Characterization of the PYP Domain. *Biochemistry* **2005**, *44*, 4755–4764. [CrossRef]

72. Tourova, T.P.; Slobodova, N.V.; Bumazhkin, B.K.; Kolganova, T.V.; Muyzer, G.; Sorokin, D.Y. Analysis of community composition of sulfur-oxidizing bacteria in hypersaline and soda lakes using soxB as a functional marker. *FEMS Microbiol. Ecol.* **2013**, *84*, 280–289. [CrossRef]

73. Sorokin, D.Y.; Gorlenko, V.M.; Namsaraev, B.B.; Namsaraev, Z.B.; Lysenko, A.M.; Eshinimaev, B.T.; Khmelenina, V.N.; Trotsenko, Y.A.; Kuenen, J.G. Prokaryotic communities of the north-eastern Mongolian soda lakes. *Hydrobiologia* **2004**, *522*, 235–248. [CrossRef]

74. Imhoff, J.F. Anoxygenic Phototrophic Bacteria from Extreme Environments. In *Modern Topics in the Phototrophic Prokaryotes: Environmental and Applied Aspects*; Hallenbeck, P.C., Ed.; Springer: Cham, Switzerland, 2017; pp. 427–480.

75. Jannasch, H.W. Die bakterielle Rotfärbung der Salzseen des Wadi Natrun. *Arch. Hydrobiol.* **1957**, *53*, 425–433. [CrossRef]

76. Mesbah, N.M.; Abou-El-Ela, S.H.; Wiegel, J. Novel and unexpected prokaryotic diversity in water and sediments of the alkaline, hypersaline lakes of the Wadi an Natrun, Egypt. *Microbiol. Ecol.* **2007**, *54*, 598–617. [CrossRef]

77. Tank, M.; Thiel, V.; Imhoff, J.F. Phylogenetic relationship of phototrophic purple sulfur bacteria according to *pufL* and *pufM* genes. *Int. Microbiol.* **2009**, *12*, 175–185.

78. Thiel, V.; Tank, M.; Neulinger, S.C.; Gehrmann, L.; Dorador, C.; Imhoff, J.F. Unique communities of anoxygenic phototrophic bacteria in saline lakes of Salar de Atacama (Chile). Evidence for a new phylogenetic lineage of phototrophic *Gammaproteobacteria* from *pufLM* gene analyses. *FEMS Microbiol. Ecol.* **2010**, *74*, 510–522. [CrossRef] [PubMed]

79. Hallberg, K.B.; Hedrich, S.; Johnson, D.B. *Acidiferrobacter thiooxydans*, gen. nov. sp. nov.; an acidophilic, thermo-tolerant, facultatively anaerobic iron- and sulfur-oxidizer of the family *Ectothiorhodospiraceae*. *Extremophiles* **2011**, *15*, 271–279. [CrossRef] [PubMed]

80. Kojima, H.; Shinohara, A.; Fukui, M. *Sulfurifustis variabilis* gen. nov., sp. nov., a sulfur oxidizer isolated from a lake, and proposal of *Acidiferrobacteraceae* fam. nov. and *Acidiferrobacterales* ord. nov. *Int. J. Syst. Evol. Microbiol.* **2015**, *65*, 3709–3713. [CrossRef]

81. Issotta, F.; Moya-Beltran, A.; Mena, C.; Covarrubias, P.C.; Thyssen, C.; Bellenberg, S.; Sand, W.; Quatrinia, R.; Vera, M. Insights into the biology of acidophilic members of the *Acidiferrobacteraceae* family derived from comparative genomic analyses. *Res. Microbiol.* **2018**, *169*, 608–617. [CrossRef] [PubMed]

82. Li, G.; Lai, Q.; Liu, X.; Sun, F.; Du, Y.; Li, G.; Shao, Z. *Maricoccus atlantica* gen. nov. sp. nov., isolated from deep sea sediment of the Atlantic Ocean. *Antonie van Leeuwenhoek* **2013**, *104*, 1073–1081, Erratum in *Antonie van Leeuwenhoek* **2014**, *105*, 439. [CrossRef] [PubMed]

83. Xia, J.; Zhao, J.-X.; Sang, J.; Chen, G.-J.; Du, Z.-J. *Halofilum ochraceum* gen. nov., sp. nov., a gammaproteobacterium isolated from a marine solar saltern. *Int. J. Syst. Evol. Microbiol.* **2017**, *67*, 932–938. [CrossRef] [PubMed]

84. Sorokin, D.Y.; van Pelt, S.; Tourova, T.P.; Takaichi, S.; Muyzer, G. Acetonitrile degradation under haloalkaline conditions by *Natronocella acetinitrilica* gen. nov., sp.no. *Microbiology* **2007**, *153*, 1157–1164. [CrossRef]

85. Sorokin, D.Y.; Trotsenko, Y.A.; Doronina, N.V.; Tourova, T.P.; Galinski, E.A.; Kolganova, T.V.; Muyzer, G. *Methylohalomonas lacus* gen. nov., sp. nov. and *Methylonatrum kenyenese* gen. nov., sp. nov., methylotrophic *Gammaproteobacteria* from hypersaline lakes. *Int. J. Syst. Evol. Microbiol.* **2007**, *57*, 2762–2769. [CrossRef]

86. Mori, K.; Suzuki, K.-I.; Urabe, T.; Sugihara, M.; Tanaka, K.; Hamada, M.; Hanada, S. *Thioprofunndum hispidum* sp. nov., an obligately chemolithoautotrophic sulfur-oxidizing gammaproteobacterium isolated from the hydrothermal field on Suiyo Seamount, and proposal of *Thioalkalispiraceae* fam. nov. in the order *Chromatiales*. *Int. J. Syst. Evol. Microbiol.* **2011**, *61*, 2412–2418. [CrossRef]

87. Barco, R.A.; Garrity, G.M.; Scott, J.J.; Amend, J.P.; Nealson, K.H.; Emerson, D. A genus definition for Bacteria and Archaea based on a standard genome relatedness index. *mBio* **2020**, *11*, e02475-19. [CrossRef]

88. Cai, J.; Guan, Y.; Li, F.; Zhao, Y.; Feng, C.; Tang, N. Biomass and pigments production of a newly isolated photosynthetic bacterium *Ectothiorhodospira shaposhnikovii* P2 from saline wastewater. *Int. J. Environ. Sci. Technol.* **2018**, *16*, 7487–7496. [CrossRef]

89. Ramana, V.V.; Sasikala, C.; Ramaprasad, E.V.V.; Ramana, C.V. Description of *Ectothiorhodospira salini* sp. nov. *J. Gen. Appl. Microbiol.* **2010**, *56*, 313–319. [CrossRef] [PubMed]

Article

Seasonal Dynamics of Lake Winnipeg's Microbial Communities Reveal Aerobic Anoxygenic Phototrophic Populations Coincide with Sunlight Availability

Steven B. Kuzyk, Xiao Ma and Vladimir Yurkov *

Department of Microbiology, University of Manitoba, Winnipeg, MB R3T 2N2, Canada
* Correspondence: vladimir.yurkov@umanitoba.ca

Abstract: In this first comprehensive study of Lake Winnipeg's microbial communities, limnetic and littoral euphotic zones were examined during each season from 2016 through 2020. Classical cultivation and modern high-throughput sequencing techniques provided quantification and identification of key phototrophic populations, including aerobic anoxygenic phototrophs (AAP). Annual dynamics found total heterotrophs reached 4.23×10^6 CFU/g in littoral sands, and 7.69×10^4 CFU/mL in summer littoral waters on oligotrophic media, higher counts than for copiotrophic compositions. Limnetic numbers inversely dipped to 4.34×10^3 CFU/mL midsummer. Cultured AAP did not follow heterotrophic trends, instead peaking during the spring in both littoral and limnetic waters as 19.1 and 4.7% of total copiotrophs, or 3.9 and 4.9% of oligotrophs, decreasing till autumn each year. Complementary observations came from environmental 16S V4 rRNA gene analysis, as AAP made up 1.49 and 1.02% of the littoral and limnetic sequenced communities in the spring, declining with seasonal progression. Spatial and temporal fluctuations of microbes compared to environmental factors exposed photosynthetic populations to independently and regularly fluctuate in the ecosystem. Oxygenic phototrophic numbers expectantly matched the midsummer peak of Chl *a* and *b*, oxygenic photosynthesis related carbon fixation, and water temperature. Independently, AAP particularly colonized spring littoral areas more than limnetic, and directly corresponded to habitat conditions that specifically promoted growth: the requirement of light and organic material.

Keywords: microbial ecology of lakes; aerobic anoxygenic phototrophs; bacterial community; Lake Winnipeg; food web dynamics; picoplankton; bacterioplankton

1. Introduction

Lakes are the largest accumulations of liquid freshwater on Earth yet make up only 1% of the world's surface. These habitats have maintained incredible importance as sources for fisheries, drinking water and irrigation supply. The North American continent contains many of the highest volume natural water reservoirs, where the detrimental effects of anthropogenic eutrophication have caused the Laurentian Great Lake ecosystems to be rigorously monitored [1–4]. In contrast, the isolated northern Great Bear, Great Slave, and Lake Winnipeg have received less attention [5–7]. Of these three, Great Bear Lake has incurred the least human exposure, and Great Slave similarly maintained low land usage within its watershed. Inversely, Lake Winnipeg has been subjected to considerable nutrient loading due to its massive drainage basin spanning over 1,000,000 km^2 across every Canadian prairie province and some American states [8]. All significant agricultural, industrial, municipal, and urban runoff eventually enters its waters prior to draining into the Hudson's Bay of the Arctic Ocean via a single outlet, the Nelson River. In response, studies have predominantly focused on chemical and environmental fluctuations, as well as the impact to fish populations and other eukaryotes, presented thoroughly in recent collaborative efforts by academic researchers and provincial and federal governments [9,10].

Nonetheless, some knowledge gaps have remained, including a lack of information regarding the primary producers, consumers, and recyclers of nutrients and biomass. The organisms at the base of food web dynamics, the microbial community, were thus far largely unexplored.

Algae and cyanobacteria have been commonly characterized in freshwaters, acting as aerobic oxygenic photoautotrophs (AOP) that use chlorophyll (Chl)-based photosynthesis, capable of fixing CO_2 into organic compounds consumable throughout the food chain [11,12]. However, as published for other great lakes [13], or in our cursory study of Lake Winnipeg [14,15], AOP made up only a portion of the microbial community, where many other photoheterotrophs, photoautotrophs, chemoheterotrophs, and chemoautotrophs also inhabited and influenced the ecosystem. Phototrophic populations have been particularly diverse in aquatic habitats, with predominant groups differentiated by their photosynthetic systems. Anaerobic anoxygenic phototrophs (AnAnP) produce biological energy from light using bacteriochlorophyll (BChl) mediated photophosphorylation in the absence of oxygen, typical of stratified or benthic zones [16]. Aerobic anoxygenic phototrophs (AAP) similarly use BChl *a* as a primary pigment, but require oxic euphotic habitats rather than anoxic ones. Photosynthesis has likely evolved in AAP to act as an auxiliary energy production strategy to outcompete typical cohabitating heterotrophs by consuming organics faster due to the supplemental use of light energy [17,18]. In addition to assorted photosynthesizing populations, numerous other trophic groups co-exist with harmonious and complex interactions occurring on the microscopic scale in each ecosystem [19,20].

Regarding aquatic microbial ecology, many methods have been developed to study either individual populations or entire communities, uncovering a wide spectrum in microbial species composition among freshwater systems [13]. While variability could indeed be present, results have been difficult to compare due to dissimilar sampling methods/approaches. Furthermore, limited sampling locations and infrequent return visits have yet to address spatial and temporal system dynamics and stabilities [1–4]. To determine such ecosystem patterns, Lake Winnipeg's microbial communities were investigated at multiple euphotic zones, over all four seasons during the past 5 years. Chosen limnetic and littoral sites were analyzed with classical microbial cultivation and modern sequencing techniques in addition to physico-chemical parameter collection, a crucial combination which allowed for both diversity and functionality to be monitored in unison. Among the phototrophic groups, AAP were specifically regarded as they have been discovered in a vast array of environments at high numbers [17,20–22]. However, their role in ecology and biogeochemical cycling of elements has remained unclear. The abundance, fluctuations, and relation of AAP to other populations of the community were therefore highlighted in this work.

2. Materials and Methods

2.1. Sampling and Environmental Parameters

Lake Winnipeg was visited during the spring in May-June, summer between July-August, fall amid September-October, and winter among February-March, with 4 years' worth of seasons occurring from 2016 to 2020 (Table S1). Euphotic regions were sampled 0.25 m below the surface at 5 littoral areas accessed by land, and 5 limnetic zones retrieved using a rosette canister deployed off the research vessel M.V. *Namao* during open-water months (Figure 1A), or from helicopter transport and ice coring equipment when the lake was frozen (Figure 1B). Each littoral site was located where lake water was 0.5 m deep, light reached the bottom, and had sediment samples collected.
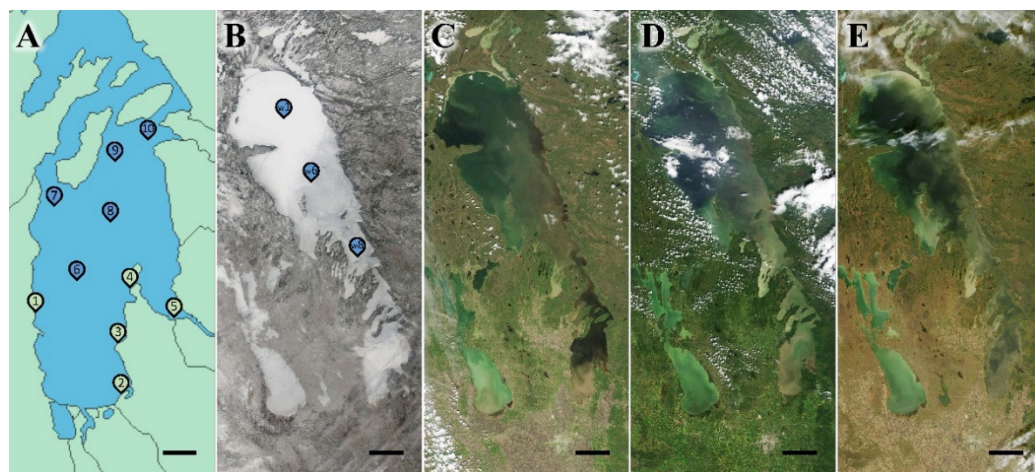
Figure 1. Lake Winnipeg seasonal progression and sampling locations. (**A**) Map of the South Basin of Lake Winnipeg, with littoral (green) and limnetic (blue) sites selected for study: S1–5, S6–10 (W1, W6, and W8 alternatively), respectively; Scale bar, 10 km. Satellite images of entire waterbody depicting seasonal change to lake habitat and surrounding lands during (**B**) Winter, 26 March 2017 indicating additional under-ice sites, (**C**) Spring, 31 May 2017, (**D**) Summer, 3 August 2017, and (**E**) Fall, 11 October 2017; Scale bars, 40 km. Photographs provided by Dr. K. Scott and the Lake Winnipeg Research Consortium (LWRC), in partnership with the Rapid Response Imagery from the Land Atmosphere Near-real time Capability for EOS (LANCE) system operated by the NASA/GSFC/Earth Science Data and Information System (ESDIS).

Littoral areas included Site 1 at Gimli Beach (50°38′2″ N, 96°59′2″ W); Site 2, Patricia Beach (50°25′25″ N, 96°37′2″ W); Site 3, Grand Beach (50°34′11″ N 96°35′59″ W); Site 4, Victoria Beach (50°42′16″ N, 96°33′49″ W); and Site 5 near the Sagkeeng First Nation Community (50°37′11″ N, 96°18′55″ W). Limnetic locations were spread across the South Basin, chosen from the expansive list that the M.V. *Namao* frequently visited, denoted here as Sites 6–10, with the catalog names provided in quotations. Site 6 "W10" (50°50.799′ N, 96°46.157′ W), Site 7 "57B" (50°58.529′ N, 96°52.551′ W), Site 8 "W9" (51°01.334′ N, 96°35.038′ W), Site 9 "12B" (51°08.014′ N, 96°37.159′ W), and Site 10 "46S" (51°08.512′ N, 96°25.270′ W) were all kilometers apart from each other as well as from the nearest shoreline. Only Sites 6 and 8 could be sampled in the winter, and during 2019–2020, 3 additional sites were included, W1 (53°22′34.788″, −98°23′25.188″), W6 (52°38′34.8″, −97°44′5.1″), and W8 (51°45′47.4474″, −96°50′21.6954″). While Sites 1–4, 6, and 7 were primarily fed by the waters of the Red River, Site 5 was downstream of the Winnipeg River, and 8–10 likely contained a mixture of the two inflows. Since Lake Winnipeg predominantly had a soft lakebed [23], sandy-silty shorelines were common.

Water temperature and pH were recorded with a calibrated Beckman Φ255 pH/Temp/mV meter, clarity by Secchi disk test, and light intensity (Lux) using a VWR Light Meter 21800-014. Data for daily average ambient temperatures, cloud coverage, and hours of sunlight were received from government sources. Major elemental compositions of limnetic water were determined via collision/reaction cell-based inductively coupled plasma mass spectrometry (CRC ICPMS) [15]. All 43 elements and their detection limits were collected by the Water Quality Management Section of Manitoba Environment, Climate and Parks Department, Canada and averages per season listed in Table S2. Concentrations of Chl *a*, Chl *b*, and BChl *a* from microbial communities were quantified as follows: A defined volume of liquid sample from each site was filtered through a 0.22 μm pore size 47 mm nitrocellulose membrane (Millipore-Sigma, Burlington, MA, USA) to concentrate cells, prior to the addition of 7:2 acetone:methanol (5 mL) for the extraction of pigments overnight in the dark at 4 °C. A Hitachi U-2010 spectrophotometer recorded absorption spectra from 350 to 1100 nm using a 50 mm path length quartz cuvette. Peaks at 663, 645, and 770 nm were

used to calculate the concentration of Chl *a*, Chl *b*, and BChl *a*, respectively, per volume of water sample [24], based on classical approaches [25,26].

2.2. Primary Productivity Assay

Community photo- and chemosynthetic carbon fixation was determined during the summer and fall of 2017, as well as the spring, summer, and fall of 2018 and 2019 by a ^{14}C-labelled NaHCO$_3$ accumulation assay [27], with some minor procedural modifications [21]. Winter in situ analysis was not possible due to the extreme cold temperatures, where shallow littoral areas were frozen solid, and limnetic zones were not accessible twice during the same day. Measurement involved a 3 to 24 h in situ incubation of 3 sealed glass Balch vials each containing 9 mL of natural lake water and 5 μCi ^{14}C-labelled NaHCO$_3$. The first of three translucent vials contained only the two components, while the second was additionally wrapped in aluminum foil. The third uncovered vial contained the two components listed as well as 7 μmol diuron, a photosystem II (PSII) inhibitor. After residing in the lake, 400 μL of 40% formaldehyde (final as ~2%) was added, stopping reactions and "fixing" them for storage. A 0.1 mL aliquot was later filtered through a 0.22 μm pore size filter (Millipore-Sigma, Burlington, MA, USA) until dry, prior to washing off unfixed ^{14}C with 5 mL of distilled water. Filters were placed in 0.5 mL ScintiSafe Plus 50% (Fisher Scientific, Waltham, MA, USA) scintillation fluid and measured for 5 min in a Beckman LS 6500. Counts were compared to 0.1 mL of unfiltered liquid for each sample. Carbonates and bicarbonates were determined by titration via a total alkalinity test [28]. Rates of carbon fixation were calculated and equalized to a 24 h incubation using the equation:

$$Ca = \frac{r \cdot C_{\text{in water}}}{R \cdot t} \tag{1}$$

Here, Ca was the amount of fixed carbon (mg/L); $C_{\text{in water}}$, the total carbon present in water; R, the counts in control (unfiltered liquid sample); r, was counts in filtered material; and t, time of incubation. Chemosynthesis, anoxygenic or oxygenic photosynthesis related carbon fixation were calculated as the rates of vial 2, vial 3 minus 2, or vial 1 minus 3, respectively [21].

2.3. Cultivation and Enumeration of Targeted Populations

Water and sediment samples were decimally diluted to 10^{-7} and 0.1 mL aliquots were spread in triplicates on 2% agar plates. An assortment of media was designed (Table S3) to isolate and enumerate several specific physiological groups of microorganisms, with a main focus on phototrophs. Heterotrophs, including photoheterotrophic AAP, were cultivated with a standard rich organic (RO) medium [14,29]. Oligotrophs were selected on an organic-limited medium (OM), which had all compounds as RO, with only 10% of the carbon sources. A comparable R2A composition has been successful for the cultivation of some AAP [30], where recent works have diluted its contents to improve isolation attempts [31,32]. With such considerations, RO was implemented as a copiotrophic carbon source, OM as oligotrophic, whereas R2A was an intermediate with carbon content between RO and OM. Potato-broth based medium (PM) was included as it usually stimulated intense cellular pigmentation as a very rich nutrient source [17], and helped to discover rare colony colorations. Purple non-sulfur bacteria and other AnAnP were enriched on a specific medium, PNSM, while AOP were selected on BG-11 [21]. Heterotrophs capable of resisting toxic metalloid oxides were counted on RO supplemented with 100 μg/mL K$_2$TeO$_3$ (RO$_T$), as some AAP have been found resistant to high levels [15,33,34]. Media were autoclaved at pH 5.9 and adjusted to 7.5 before agar solidification. Cultures were incubated for 7 days at 28 °C on RO, PM, R2A, and RO$_T$ in the dark, and BG-11 and PNSM, under constant illumination. All were grown aerobically, except cultures on PNSM which developed in translucent oxygen-limited Gas-Pak chambers.

Pigmented and total (including pigmented and colorless) colony forming units (CFU) were counted as CFU/g of sediments, and CFU/mL in water. To determine numbers of

AAP, numerous colored colonies were re-plated onto medium of isolation to verify growth and pigmentation, except colonies from RO$_T$, which were instead passaged onto RO [15]. Patched strains were later suspended in 0.2 mM TRIS-HCl buffer, pH 7.8 containing 70% glycerol to measure absorption spectra from 350 to 1100 nm using a HITACHI U2010 spectrophotometer. Strict aerobes containing BChl *a* bound to reaction center and light harvesting complexes were identified as AAP [17]. The numerical presence of AAP among pigmented heterotrophs was established (Table S4), and used to estimate their proportion among the total cultured copiotrophic or oligotrophic heterotrophic bacteria, on RO and OM, respectively [15].

2.4. Microbial Community eDNA Sequencing

Environmental DNA (eDNA) was extracted from liquid or sediment samples using a QIAGEN DNeasy® PowerWater® Kit with minor modifications, and bacterial 16S V4 rRNA genes were sequenced by MiSeq (Illumina) with primer set 515F, 806R [15]. Community structures were computationally analyzed with Qiime2 11.3, a Python based framework [35,36], and BIOM formatting [37]. Reads were joined via paired-ends, trimmed to 292 bp (kept > 99%), then deblurred [38] and denoised [39], prior to de novo clustering sequences by 99% identity. Operational "taxonomic" unit (OTU) alignment were applied using 99% homology [40,41] to the Silva132 database [42], with a final data set that included "borderline-chimeras". Phenetic identities were aligned as feature tables [43] at all hierarchical levels of the phylogenetic tree [44,45]. Both α- and β-diversity were statistically quantified, and graphed as boxplots or principle coordinate analysis (PCoA) [46,47]. Qiime2 project code text file describing entire pipeline and final output in both Qiime2 and BIOM formats were provided as Supplemental Materials. The AOP, AAP, and AnAnP proportions of the sequenced communities were analyzed by grouping all representative OTUs to known type species, Tables S5 and S6.

3. Results

3.1. Environmental Characteristics

Seasonal observations of sites over 4 years (Table S1) revealed trends in physicochemical environmental factors (Figure 2). Daylight was the longest during the spring with marginal variation, shortening into the fall (Figure 2). In comparison, lake water did not warm up until the peak of summer, decreasing to near 0 °C just below the ice surface at limnetic zones during winter. Littoral areas reached an average high of 21.5 °C, with some sites up to 26.0 °C, while limnetic regions had a moderate peak around 20.2 °C. Littoral warmed faster in the spring and summer, but was also colder than limnetic in the fall, which represented heat retention of open water. General trends in pH also followed seasonal changes, with winter limnetic zones being the most neutral at ~pH 7.5, whereas an increased pH ~8.2 was maintained at any other time. Littoral sites had a tendency to reach a slightly higher pH 8.5–8.7 in spring to summer, prior to dropping in the fall and holding a value similar to that offshore. Of the 43 elements estimated down to ppb in limnetic regions, most were invariably low (Table S2). Considering metals and metal (oids), only Al and Fe were above 1 ppm, where all others were closer to ppb concentrations. Extracted photosynthesis pigments Chl *a* and *b* had highest concentrations during the summer of most years, up to 16.5 and 5.3 µg/L on average, respectively (Figure A1). In comparison, levels of BChl *a* were lower and reached an average peak of 0.2 µg/L during the spring and decreased seasonally.
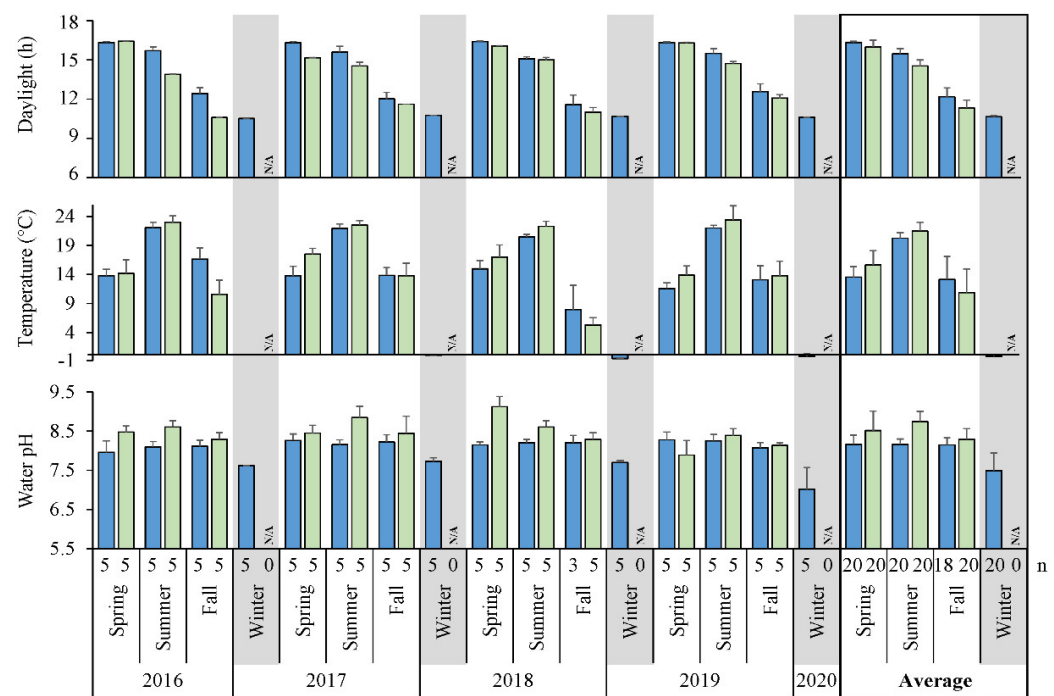
Figure 2. Periodic fluctuations of lake environmental parameters. Average light availability, water temperature and pH for all limnetic (blue bars) or littoral (light green bars) samples depicted for each season. N/A, not analyzed; n, number of samples per season.

3.2. Primary Productivity Rates

Total carbon fixation observed in situ at littoral sites between 2017 and 2019 had a repeatable yearly tendency for highest rates during the spring averaging 2.00 g C/m^2/day, diminishing throughout the summer to become 0.92 g C/m^2/day by the fall (Figure 3). Limnetic zones maintained a different pattern, where averaged offshore rates were at a minimum in the spring (0.28 g C/m^2/day), gradually increasing throughout the summer to reach a fall maximum of merely 0.75 g C/m^2/day. Independently, oxygenic phototrophs had a notable peak in fixation during the summer regardless of location (0.48–0.84 g C/m^2/day), with reduced rates in both spring (0.15–0.57) and fall (0.28–0.40). When considering zones, shorelines maintained higher rates than offshore. Anoxygenic photosynthesis derived fixation was exceptionally low across all limnetic areas (0.05–0.17 g C/m^2/day), and was only marginally higher in the littoral regions (0.11–0.23 g C/m^2/day). Limnetic dark reactions (chemosynthesis) increased slightly from spring through fall, 0.08 to 0.30 g C/m^2/day, respectively, whereas littoral zones had the highest rates in the spring (1.21 g C/m^2/day) that would later decrease, with some annual inconsistency (Figure 3).
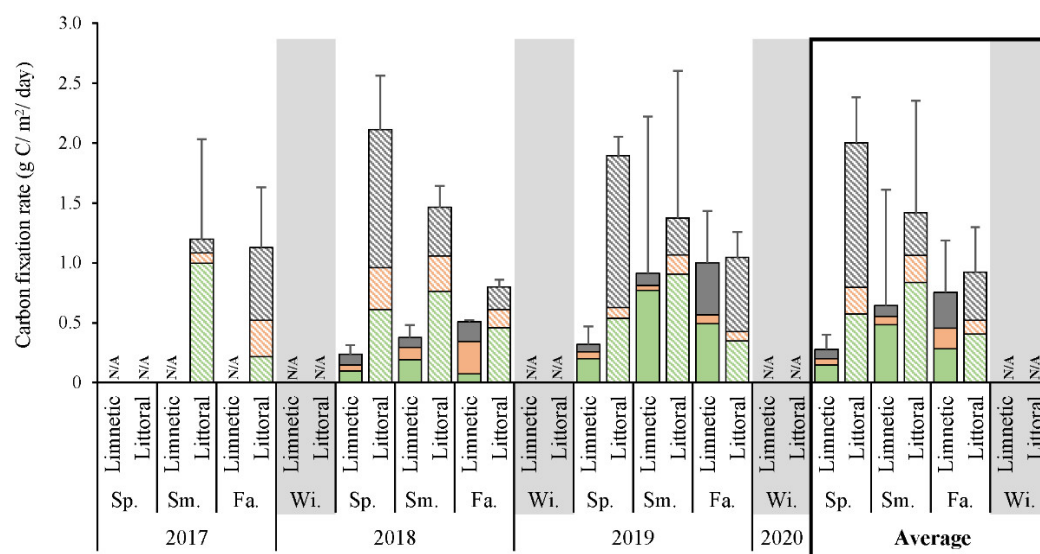
Figure 3. Primary productivity of Lake Winnipeg. [14]C fixation measured in situ with total rates composed of oxygenic (green) or anoxygenic (orange) photosynthesis related fixation, and from dark chemosynthetic reactions (grey bars). N/A, not analyzed; Sp., Spring; Sm., Summer; Fa., Fall; Wi., Winter.

3.3. Seasonal Fluctuations in Cultivated Microbial Counts

Total cultured bacteria on varied media were listed (Table S4), with patterns in RO grown copiotrophs and oligotrophs on OM evident after repetitive enumeration (Figure 4). Specifically, when grouped by zone and sample type for each season, littoral cultured bacteria peaked during the summer, while limnetic counts decreased at the same period (Figure 4A). Overall, littoral sediments maintained the highest numbers on OM (4.23×10^6 CFU/g), shoreline waters had $100\times$ less (7.69×10^4 CFU/mL), and offshore counts were a further $10\times$ decreased (4.34×10^3 CFU/mL). As a general trend, more bacteria were culturable on OM than RO, regardless of season or site. Furthermore, PM and R2A media used as alternative complex organic sources to cultivate heterotrophs also maintained lesser counts than OM (Table S4).

AOP grown on BG-11 had a general apex during the summer of most years, to a maximum average of 9.37×10^3 CFU/g and 4.13×10^3 CFU/mL in littoral sediment and water, respectively, and would typically only reach a peak of 1.08×10^2 CFU/mL offshore (Table S4). Anaerobes were often below detection limits and AnAnP hovered around 5.57×10^1 CFU/mL when present. In comparison, AAP isolated and counted from both RO and OM displayed trends independent of summer maximal heterotroph numbers (Figure 4B, C). The most AAP were detected in the spring of each year on either media, diminishing into the fall. The proportion of the pigmented bacteria in littoral locations reached a height of 32.9% on RO, at which time AAP also made up to 4.7, 19.1, and 11.2% of the entire cultured community of limnetic waters, littoral liquid, and sediments, respectively (Figure 4B). In general, these photoheterotrophs were cultivated in highest numbers on OM in comparison to $10\times$ nutrient enhanced RO or even higher carbon containing PM (Figure S1). They reached a peak of 2.39×10^4 CFU/g in the spring littoral sediments. However, AAP made up a larger proportion of the heterotrophic numbers on RO, and PM, than they did on OM regardless of sampling site (Figure 4B,C, Table S3). Furthermore, RO_T was the most selective medium as 47.5% of cultured bacteria in sediments challenged with 100 µg/mL tellurite were AAP.
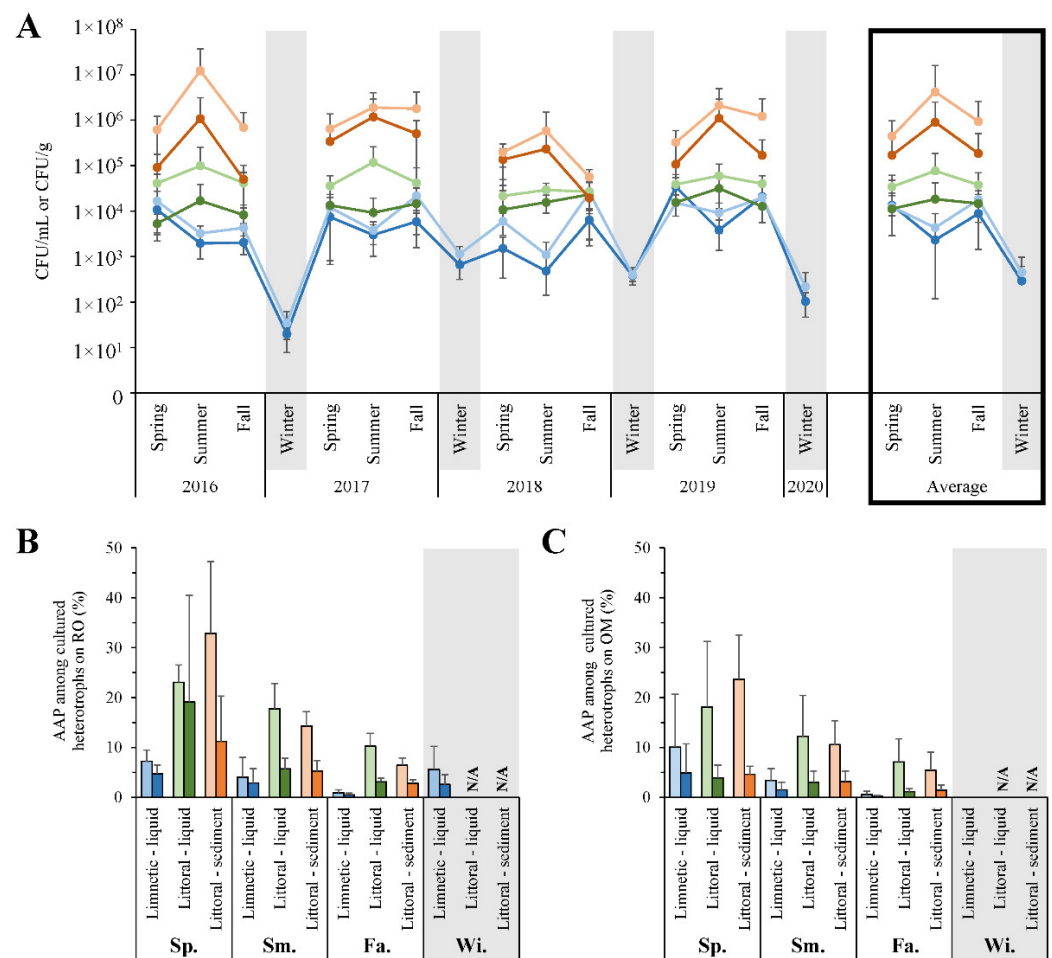
Figure 4. Enumeration of heterotrophs and AAP through all seasons and years. (**A**) Counts of total cultured heterotrophs from limnetic water (blue), littoral water (green), or sediment (orange) samples on copiotrophic RO (dark colors) or oligotrophic OM (lighter shade of each). (**B**) Copiotrophic and (**C**) oligotrophic AAP proportions of pigmented colonies (light shade) and total counts (dark shade) of seasonal averages from 2016–2020. N/A, not analyzed; Sp., Spring; Sm., Summer; Fa., Fall; Wi., Winter.

3.4. Sequenced Microbial Communities

Amplified DNA resulted in an average of 261,105 reads from each site. After trimming and binning, 21,877 features representing unique phylotypes were identified throughout, later distinguished via Silva132 database to 2266 taxa of species level or higher. All samples were similarly rarified when grouped by season (Figure S2A), where the year 2016 reached a plateau of the highest diversity followed by 2017, then 2018–2020 (Figure S2B). The dominant phyla detected were averaged by zone for each collection period (Figure 5A). Considering a generalized community structure of Lake Winnipeg, the mean phyla across all sites and seasons listed from major to minor were $40.9 \pm 10.4\%$ Actinobacteriota, $17.5 \pm 6.6\%$ Proteobacteria, $9.9 \pm 6.1\%$ Bacteroidota, $9.8 \pm 10.2\%$ Cyanobacteria, $5.7 \pm 4.0\%$ Verrucomicrobiota, $5.7 \pm 3.6\%$ Chloroflexi, $4.9 \pm 3.9\%$ Planctomycetota, $3.4 \pm 3.1\%$ Acidobacteriota, $0.8 \pm 0.6\%$ Gemmatimonadota, $0.7 \pm 1.6\%$ Armatimonadota, $0.3 \pm 0.7\%$ Patescibacteria, $0.2 \pm 0.5\%$ Nitrospirota, and $0.3 \pm 0.5\%$ minimal clades. Standard deviation revealed some phyla maintained stability in the lake, while others varied dependent on location and time. Community richness and collation were exemplified using a PCoA of differing Jaccard beta diversity (Figure 5C,D), where samples were found to group by season (Figure 5C). A shift in composition was also detected from 2016 and 2017 to 2018 through 2020 (Figure 5D). The limnetic and littoral bacteria were also differentiated as separate cohorts when visualized

in PCoA (Figure S3). Offshore samples were found at the extremities of the plot in tight groups, suggesting more stable communities. In comparison, shoreline representatives were both more centralized and dispersed, inferring a broad range of microbes was present with higher degrees of fluctuation.
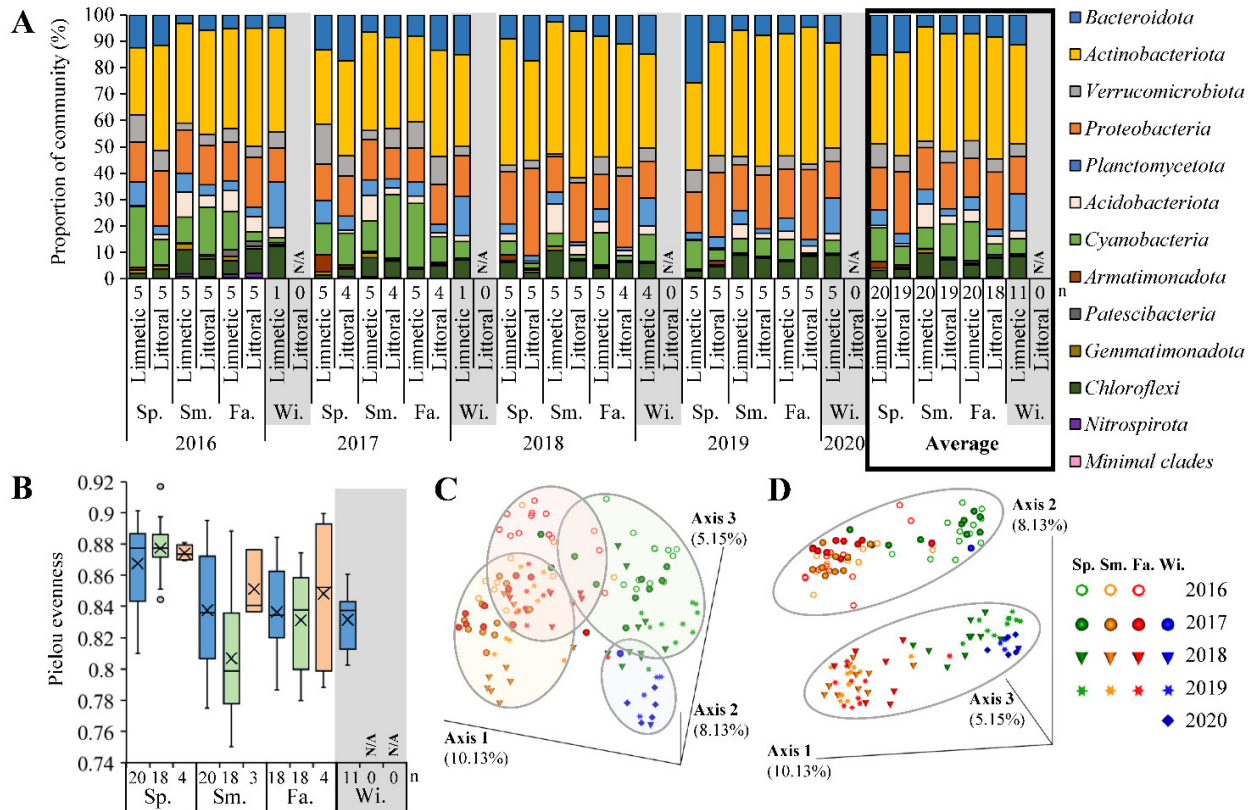
Figure 5. Bacterial community composition based on sequenced 16S V4 rRNA genes. (**A**) Phyla in water samples defined at limnetic vs. littoral sites. Minimal clades represented less than 1% per community with Firmicutes, Deinococcota, Bdellovibrionota, Crenarchaeota, Desulfobacterota, Hydrogenedentes, Margulisbacteria, Myxococcota, Sumerlaeota, SAR324, MBNT15, and NB1-j clades. (**B**) Total α-diversity of bacteria from limnetic water (blue), littoral water (green), or sediments (orange). (**C,D**) PCoA ordination based on Jaccard diversity distance matrix with percent variation as principal coordinates split on indicated axis. (**C**) From above axis 1 and 3 separate seasons, while (**D**) depicts major shift occurring along axis 2 by year with ellipses as visual cues. N/A, not analyzed; n, number of samples per season; Sp., Spring; Sm., Summer; Fa., Fall; Wi., Winter.

Considering alpha diversity, monthly changes later exposed an annual rhythmic pattern (Figure 5B). Overall, the highest number of different phylotypes was found during spring regardless of site. There was also the smallest variation in diversity at that time, suggesting all locations to have similarly complex microbial communities early in the year after ice break-up. By summer, bacterial heterogeneity decreased across the lake where littoral regions established the largest range in species present. While the number of individual taxa then increased at shorelines during the fall, the communities offshore steadily became more homogenous from spring to winter. Sediment bacteria had similar trends to those inhabiting the upper liquid layers at littoral zones (Figure S4), which could be explained by the proximity of sampling and the frequent mixing of sand and water. However, sediment communities were distinct due to the decrease in average diversity from spring to fall that increased in variation as winter drew near, suggesting a range in community structure formed in sediments as the year progressed (Figure 5B).

Based on 16S V4 rRNA gene sequences, the populations of AAP, AOP, and AnAnP were individually grouped to identify and assess their changing proportion in the bacterial community (Figure 6), using an updated list of known AAP or AnAnP (Tables S5 and S6). Seasonal patterns were found for each group, where AAP averaged 1.02 and 1.51% of the limnetic and littoral sequenced communities in the spring, respectively, dropping to 0.83 and 0.90% during the summer, and further to 0.41 and 0.54% by fall (Figure 6). AOP that included algae and cyanobacteria, made up approximately 10% of the sequences from limnetic sites at any given time, dipping during summer and winter, while inversely increasing at littoral zones midsummer (Figure 6). AOP numbers fluctuated substantially between sites and seasons, as depicted by large variations in standard deviation, whereas AAP were present and distributed more stably. The spring of 2018 was an exception, when records for Blastomonas related AAP were exceptionally high. AnAnP were quite nominal in comparison to other phototrophs during open water months, and only saw an increase in winter, dissimilar to the other populations (Figure 6).

While all specific clades and their relative proportions were provided in Supplemental Material, the three most prevalent AAP-associated genera in sequenced communities were Gemmatimonas, Roseomonas, then Blastomonas. For AOP, Cryptomonadaceae, Aphanizomenon, and Aulacoseira were predominant, whereas the highest AnAnP were Rhodoferax, Roseiflexus, and Rhodoplanes. Among known AAP (Table S5), Rhodobacteraceae, Comamonadaceae, and Acidobacteriaceae were not detected. AnAnP were exceedingly rare, with no examples of Heliobacteria, Chlorobiaceae, Ectothiorhodospiraceae, Chromatiaceae, or Halorhodospiraceae found.
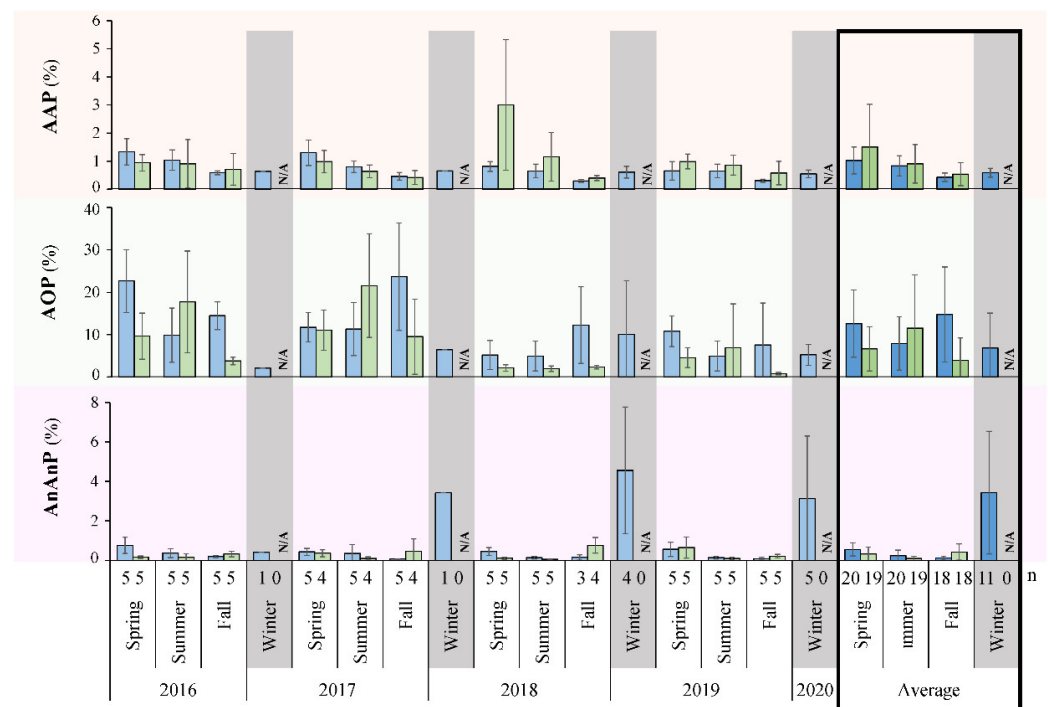
Figure 6. Relative proportions of phototrophs in sequenced communities. All 16S V4 rRNA genes of species and genera representing AAP, AOP, or AnAnP at limnetic (blue) and littoral (green) sites summed for each season. N/A, not analyzed; n, number of samples per season.

4. Discussion

4.1. Habitat Features

Lake Winnipeg has been regarded as atypical due to its shape and dimensions as the 11th largest freshwater reservoir in surface area, yet 25th by volume. Shallow in comparison to other great lakes, it has maintained a well-mixed state, where oxygen often penetrates to the bottom. As a result, thermocline-based stratification has been infrequent in the South

Basin [48]. Indeed, during our study oxygen was well dispersed during open-water seasons (Table S2). As Lake Winnipeg is north of the 49th parallel, the duration of daylight was dependent on the time of year (Figure 2). Water temperature lagged behind the spring peak of available sunlight, instead reaching height in the summer, likely due to the thermostability of the vast volume of water as a slow conductor of heat. The most significant pH change was observed at limnetic zones, where a drop to 7.5 occurred in winter. That decrease, and the slight additional alkalinity of pH 8.3–8.7 found in littoral areas over offshore values around 8.2 may have been due to respective water temperatures. Shallow waters had less exchange, and the warmer conditions likely increased solubility of carbonates and/or promoted activity of microbes, such as photosynthesis and biological carbon fixation, decreasing pH. Regardless, the yearly changes of these parameters (Figure 2), matched previously reported years [9,10]. Furthermore, most metal concentrations were typically below 1 ppm and detection limits, confirming no significant toxic influence on life (Table S2). Chl *a* and *b* curves indicating AOP followed that of temperature rather than daylight (Figure A1), and the averaged Chl *a* peak of 16.5 μg/L was similar to summer maxima observed via satellite imagery from 2002–2011 [49]. While BChl *a* from anoxygenic phototrophs was nearly undetectable and often below 1 μg/L, a common occurrence in oxic lakes [50], its low levels were still recognizable and matched daylight more so than temperature or the other chlorophylls.

4.2. Primary Productivity in Lake Winnipeg

The seasonal in situ carbon fixation analysis of whole community productivity found littoral microbes fixed at higher rates than those taken from limnetic zones (Figure 3). These increased values were likely due to the warmer nearshore waters during the spring and summer, as temperature has been shown to affect oxygenic phototrophs [51,52], and thus may affect all carbon fixation similarly. Moreover, the maximal littoral rates decreased each fall, correlating to light availability, confirming its additional impact to fixation levels. Independently, AOP fixation was predominant during summer months, tracing the measured Chl *a* and *b* concentrations in addition to general temperature trends. In comparison, AnAnP associated productivity were marginal regardless of season, likely inhibited by the oxic nature of the habitat. Furthermore, AnAnP related rates did not correlate with BChl *a* data, suggesting this pigment was unlinked to primary productivity.

The total fixation rates at shorelines reached a spring average high of 2.0 ± 0.1 g C/m^2/day (Figure 3), matching recently reported values for Lake Winnipeg calculated via oxygen production rather than using carbon isotopes [53]. When compared to satellite imagery-based findings from the other 11 largest freshwater lakes, Lake Winnipeg rates determined here were greater than all, yet were most equivalent to Lake Erie's estimated mean fixation of 1.2 g C/m^2/day [52]. Due to the well-known eutrophication of Lake Winnipeg, and confirmed Chl *a* levels that matched recent measurements, it may indeed currently have the highest rates among all great lakes. However, the relatively elevated numbers in Lake Winnipeg may simply be a result of underestimated values calculated in the recent satellite survey of other lakes, which was based solely on Chl *a* estimate without considering other types of carbon fixation. Indeed, a better match was found to earlier isotope-measured records of the Laurentian Great Lakes [54]. Specifically, Lake Ontario monthly dynamics aligned well with our three time points taken yearly, where littoral regions had maximal fixation rates in the spring, which occurred prior to the swell apparent at limnetic zones. Inversely, limnetic areas were slower to warm and cool, likely the cause of delayed fixation rate peaks. However, since Lake Ontario has also suffered from nutrient loading in its recent past, and primary production may have fluctuated, future work should include a combination of satellite imagery on Lake Winnipeg in addition to chemical isotope analyses to better support our conclusions and comparisons to other great lakes.

4.3. Trends in Cultivated Microbial Numbers

The height of heterotrophic cellular counts matched the peak of temperature during summer at all nearshore sites (Figure 4A). Littoral sands had 4.23×10^6 and 9.05×10^5 CFU/g, while its waters maintained 7.69×10^4 and 1.84×10^4 CFU/mL on oligotrophic and copiotrophic complex media, respectively, which matched previously reported values in freshwaters and sediments [55]. In comparison, limnetic numbers did not coordinate with temperature. This incongruence may have been due to the minimal organic carbon available offshore, where low organic production rates were observed (Figure 3). Since limnetic regions were slow to warm, autotrophic production of excess organic carbon was delayed, which in turn pushed back the bloom of heterotrophs till the fall. When considering concentrations of supplied organics, higher cell counts were found on oligotrophic OM rather than rich RO from all locations and times (Figure 4A), indicating that the natural lake bacteria were adapted to lower nutrient availability. In addition, cultivated numbers on PM were 10–100× lower than RO suggesting even fewer organisms preferred the highest level of organics provided (Table S4). R2A enumerated values fell between those counted on OM and RO, which also coincided with its organic content in between both custom-made media. Taken together, OM clearly provided the most ideal growth conditions by incurring the highest numbers of heterotrophs, that in-turn best represented the microbial community metabolism. Of note, carbon was observed at ~0.025 g/L in lake waters (Table S2), and OM had 0.1 g/L, about five times higher. This would suggest, that while common laboratory practices promote bacteria growth by supplying high levels of nutrients, more dilute concentrations closer to environmental conditions may be ideal for the cultivation of highest numbers of heterotrophs. While the amount of carbon in Lake Winnipeg were considered eutrophic, these findings also suggested that the current microbial community is accustomed to oligotrophy, and do not have preference to even higher levels of nutrients. Future work should include more dilute media for enumerations, as such composition may outperform the conditions provided in standard RO or previously considered oligotrophic OM. Follow-up microscopy may also shed light on the microbial content since cultivated counts fell short of microscopic observations from other Canadian lakes by 10–100× [22]. Later investigation could examine alive/dead cells, or if the additional microbial content had specific growth requirements not provided in laboratory.

Regarding phototrophs, few AnAnP and other anaerobes were detected (Table S4), presumably a result of the high inhibitory oxygen content. AOP were cultivated up to 9.37×10^3 CFU/mL on average, peaking in summer littoral sediments and traced both the Chl *a* and *b* levels as well as the AOP associated fixation rates. The curiosity about metals in the lake, particularly metalloids, led to the study of tellurite tolerance by the aerobic heterotrophic community. Since the toxic metalloid oxide tellurite was not present in significant amounts (Table S2), it was surprising to find ~5% of all heterotrophs at any given site or season capable of resisting at least 100 µg/mL, a substantially high level (Table S4). While it had been previously postulated that 1 µg/mL of tellurite was toxic to most tested bacteria [56], our findings suggested resistance to higher levels may be more common than previously considered. Furthermore, since 47.5% of heterotrophs cultivated on RO_T were found to be AAP, this suggested that the medium was selective for some phototrophs, and that they may have metal transformation capabilities as part of their niche.

Compared to other photosynthesizing populations, cultured AAP counts neither followed trends in temperature dynamics and total heterotrophs, nor did they match measurements of AOP. Instead, they reached a pinnacle during the spring as 19.1 and 4.7% of total copiotrophs (Figure 4B), or 3.9 and 4.9% of oligotrophs (Figure 4C), at littoral and limnetic locations, respectively. The proportion decreased till winter each year and directly coincided with the hours of available sunlight, BChl *a* concentration, and shoreline rates of total carbon fixation. Such correlations were reasonable to find for AAP as these bacteria presumably required sunlight, and their inability to fix CO_2 or act autotrophically meant they had to rely on other community members to provide organics [17]. In addition, AAP numbers did not match up with AnAnP associated fixation rates, dark (chemolithotrohic)

reactions, or AOP fixation rates independently. Rather, AAP numbers were highest at shorelines, and particularly coincided to the maximal rates of total fixed carbon productivity (Figure 3). AAP may thus be coordinated with organisms capable of fixing inorganic carbon, by consuming excess nutrients released by autotrophic neighbors in euphotic zones. Moreover, while more AAP were culturable on organic limiting OM, it was evident that they made up a larger proportion of bacteria capable of growth on the richer RO or PM (Figures 4 and S1). Obviously, they had a tolerance/ability to grow with excess nutrients, whereas several other heterotrophs did not. Hence, aerobic phototrophs were influenced by the environmental factor sunlight, and had advantage when organics were plentiful.

4.4. Sequenced Community Analysis

General Community Composition: Culture-independent analysis of 16S rRNA genes extracted and sequenced from each location indicated a rather stable bacterial community existed across the lake (Figure 5), which was reasonable as the habitat has been known to be well-mixed, therefore similarities between samples were expected [48]. While Lake Winnipeg contained a bacterial composition comparable to the average freshwater basin where *Actinobacteria* were dominant and *β-Proteobacteria* made up a large proportion [13], it was very similar to a particular few great lakes of North America. Lake Erie has been found analogous both in equivalent physical surface area and volume. In addition, it contained a similar distribution of bacterial phyla at limnetic sites during the summer [2]. However, there were vastly different cyanobacteria described. While only 0.2–0.8% of Lake Erie's microbial community belong to AOP, and predominantly contained *Planktothrix*, Lake Winnipeg mostly had *Cryptomonadaceae*, and 7.9–11.6% of its spring to summer community were AOP (Figure 6). This divergence may have been due to the irregular presence of AOP in the ecosystem, as blooms were well known to unevenly disperse yearly and spatially because of several factors including wind currents [11]. In Lake Winnipeg, a large variance in AOP indicators was found. Chl *a* and *b* concentrations (Figure A1), AOP cultivated numbers (Table S4), and proportions of sequenced communities (Figure 6) each had dispersed standard deviation, which could be explained by their tendency to move in clumps and not be homogenous in every liter of lake water. With such considerations, the Lake Erie analysis may have missed these populations by only sampling once at limited locations. Indeed, a more recent comprehensive study quantified cyanobacterial blooms in Lake Erie using specific 16S rRNA primers for AOP when sampled in the fall [57]. These findings were both coherent with ours and more comparable to other lakes [58], suggesting future work dedicated to AOP in Lake Winnipeg may be required to assess pelagic apart from aggregate forming populations.

Community Fluctuations: Repeating cyclical patterns were evident for Lake Winnipeg's bacterial populations (Figure 5C). Winter and spring samples individually had distinct groups, whereas summer and fall overlapped considerably. The warmer months likely developed homogenous compositions, as similar seasonal dynamics reaching a climax community has been observed in other locations with relatively comparable freeze-thaw cycles [58]. Lake Winnipeg's distinct winter structure was likely influenced by the drastic habitat transformation, that had low temperatures and an encapsulating frozen surface for several months (Figure 1B). While seasons returned to similar repeated conditions present in previous years (Figure 5C), a total diversity migration was observed from 2016 to 2020 (Figure 5D). Results for both 2016 and 2017 were similar, whereas 2018–2020 clustered apart. A noticeable transition occurred in 2017, as its rarified diversity was between the 2016 maximum and 2018–2020 minimum (Figure S2B). Since most monitored environmental factors remained consistent across all 5 years (Figure 2), it was doubtful they caused the drift in bacterial composition. Instead, such a shift may have been due to the recent infestation by zebra mussel bivalves, *Dreissena polymorpha*. These invasive species were first spotted in 2013, but then undoubtedly confirmed in 2015. Perhaps their filter-feeding activity and rapid consumption of bacteria caused the slow change of microbial community composition [15,23,55], and decreased microbial diversity via selective grazing. Further

monitoring of bacterial communities is therefore encouraged to see if this shift continues, or if natural decade-long fluctuations have been prone to occur in the habitat.

Coordination of Phototroph and Environmental Dynamics: Littoral AOP were observed maximally in the summer making up a significant proportion, 11.5% of the sequenced community (Figure 6). This followed the detected Chl *a* and *b*, highest AOP associated carbon fixation rates, as well as trends in water temperatures. They did not match the length of available daylight in spring, instead being delayed till summer. As suggested previously, temperature dependent growth and autotrophy contributed to this delay, rather than following available sunlight patterns [51]. As limnetic zones were even slower to warm than nearshore (Figure 2), lower proportions of AOP and associated fixation rates midsummer were expected and observed (Figures 3 and 6). As for AnAnP, they made up a minimal part of any sequenced community during open water seasons (<0.5%), but peaked in winter at limnetic sites (Figure 6). This may have been a result of the limited number of total active organisms throughout colder months, as indicated by the few cultivated heterotrophs and other bacteria (Figure 4A; Table S4). If all heterotrophs and AOP had a winter decrease, perhaps the persisting AnAnP represented a higher proportion. As oxygen was present in the lake, even under frozen ice conditions (Table S2), these results stimulate future research to understand why certain AnAnP clades made up larger winter values.

The sequenced AAP of the community maintained a seasonal pattern similar to that of the culture-based enumerations, maintaining the highest proportion in the spring, decreasing throughout the year (Figures 4B,C and 6). Up to 19.1% of the cultured heterotrophs on RO were AAP, whereas only 1.5–3.0% of the total sequences were from these phototrophs in spring shoreline waters. The diminished values estimated by sequencing suggested not all species present could grow on the provided RO or OM media. However, the corresponding trends between both techniques strongly inferred that cultivation reflected a fair vision of the microbial community, and that a majority of the bacteria were likely aerobic heterotrophs. In addition, the combination of applied methods permitted several comparisons to other lake studies. The highest cultured AAP count was 2.39×10^4 CFU/mL (Figure S1), a number similar to mid-summer microscopic observations of other Canadian lakes [22]. Seasonal fluctuations of AAP have been analyzed infrequently in large freshwater habitats, with notable studies from the small Ĉertovo and Vlkov Lakes, Czechia [50,59], and alpine Gossenköllesee Lake, Austria [60]. In each case, a culture-independent autofluorescence technique was used to detect anoxygenic phototrophs, reporting AAP to be prevalent in the warmest months. The latter two studies also correlated their high numbers to dissolved organic carbon trends. Dependency on available nutrients was further confirmed when AAP where additionally shown to follow diel changes of sunlight and carbon fixation in a natural setting [61]. Such associations were notable, as they supported our results produced from Lake Winnipeg. Here, AAP peaked during the spring, persisting throughout the open water seasons. Their numbers were directly associated with the amount of daylight available (Figure 2), which also aligned to when the most carbon was fixed at shorelines (Figure 3). These comparable trends strongly suggested that AAP followed seasonal fluctuations reflecting on each part of their namesake. When there is ample oxygen, light, and organic carbon, they thrive. A recent pertinent study of several freshwater lakes in Quebec, Canada reported that while AAP numbers followed light availability, they negatively correlated to water clarity [22]. This was likely due to their requirement of organic carbon, as translucent water has been indicative of an extremely oligotrophic environment, with few microbes and highly limited available nutrients.

In the area of Lake Winnipeg, the longest daylight occurred during spring, which was the same period for the highest rate of carbon fixation. This timing was therefore optimal for light-dependent phototrophs, as well as heterotrophs consuming nutrients produced by close neighbors. AAP were uniquely persistent during the early open-water season, as the conditions provided a selective advantage for them to thrive. Of note, some growth experiments have reported AAP to promote BChl *a* production only after incubation at refrigerant temperatures (4–8 °C) [62], which may suggest adaptation to the colder spring,

when sunlight was present. The low presence of AAP estimated in sequenced winter communities and cultivated on plates both revealed cold temperatures played a minor inhibitory role. The analysis of additional habitats far from the equator that maintain separate daylight and temperature shifts may help to confirm which parameters are most influential to AAP growth and photosynthesis. Furthermore, the analysis of isolates from such environments and their thermal tolerances in addition to pigment synthesis may elude how influential these factors are on growth.

AAP Diversity in Lake Winnipeg: Of the known taxa that contain AAP (Table S5), Lake Winnipeg samples had representative DNA from *Gemmatimonadota*: *Gemmmatimonas* and proteobacterial *Acetobacteraceae*: *Roseomonas*; *Methylobacteriaceae*: *Methylobacterium*, *Methylorubrum*; *Rhodospirillaceae*; *Sphingomonadaceae*: *Polymorphobacter*, *Blastomonas*; *Erythrobacteraceae*: *Porphyrobacter*; *Rhodocyclaceae*; and *Halieaceae*. Interestingly, no sequences were detected from AAP of the β-proteobacterial *Comamonadaceae* group, which included *Limnohabitans* and *Roseateles*. While *Limnohabitans* has been indicated in several lakes [63], perhaps its preference for oligotrophy, and the eutrophic nature of Lake Winnipeg let other organisms to out compete them. Furthermore, no AAP from *Rhodobacteraceae* were found in the sequenced communities. Likely, along with *Roseateles* [64], this was because most known examples such as *Roseobacter* [65], *Roseicyclus* [66], and *Charonomicrobium* [67] hailed from marine and salt water habitats. In addition, no representatives of the newly discovered saline-preferring prosthecate AAP *Photocaulis* of *Maricaulales* were detected [68]. Since Lake Winnipeg is freshwater, the presence of marine AAP was not expected. Finally, the unusual *Acidobacteriota*: *Chloracidobacterium* were also not observed. This was predicted since the representative type species was isolated from a hot spring, being thermophilic and microaerophilic [69], whereas Lake Winnipeg had quite moderate temperatures, and was well-aerated (Figure 2, Table S2). Future analysis of the Lake Winnipeg DNA dataset may reveal additional AAP species when reference databases are updated. Later work on isolated and purified strains from this habitat could also identify each taxon observed in the sequenced community, shedding light on their individuality and physiology of the diverse species present.

In summary, the well-mixed nature of Lake Winnipeg has maintained stable microbial communities that fluctuate seasonally with a common cyclical pattern, returning to an established composition each year. Littoral shorelines and limnetic open water had distinctly separate population dynamics, containing related yet specific diversity and culturable numbers of heterotrophs. While AOP were variably present, AAP made up a stable important component of the microbial community highest in the spring, decreasing throughout the year. Their levels clearly coincided with the available daylight in all areas, BChl *a* levels, and, particularly, with total carbon fixation rates at littoral zones, confirming that they were promoted when light and organic carbon is available. In comparison, temperature, pH, and physico-chemistry including metal concentrations may only have minor impacts on their population dynamics. Future research will corroborate if similar factors play important roles for AAP ecology in other habitats.

Author Contributions: Conceptualization, S.B.K. and V.Y.; methodology, S.B.K. and V.Y.; software, S.B.K.; validation, S.B.K. and V.Y.; formal analysis, S.B.K. and X.M.; investigation, S.B.K. and X.M.; resources, S.B.K. and V.Y.; data curation, S.B.K.; writing—original draft preparation, S.B.K.; writing—

review and editing, S.B.K. and V.Y.; visualization, S.B.K. and V.Y.; supervision, V.Y.; project administration, V.Y.; funding acquisition, V.Y. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The eDNA sequenced from each site via MiSeq (Illumina) was deposited on the National Center for Biotechnology Information Sequence Read Archive as BioProject PRJNA603112.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Figure A1. Photosynthesis primary pigment levels from each sampling location. Solvent extracts of collected biomass revealed BChl *a*, Chl *a*, and Chl *b* concentrations at limnetic (blue) or littoral (light green) zones. N/A, not analyzed; n, number of samples per season.

References

1. Reed, A.J.; Hicks, R.E. Microbial ecology of Lake Superior Bacteria and Archaea: An overview. *Aquat. Ecosyst. Health Manag.* **2011**, *14*, 386–395. [CrossRef]
2. Mou, X.; Jacob, J.; Lu, X.; Robbins, S.; Sun, S.; Ortiz, J.D. Diversity and distribution of free-living and particle-associated bacterioplankton in Sandusky Bay and adjacent waters of Lake Erie Western Basin. *J. Great Lakes Res.* **2013**, *39*, 352–357. [CrossRef]
3. Nakatsu, C.H.; Byappanahalli, M.N.; Nevers, M.B. Bacterial Community 16S rRNA Gene Sequencing Characterizes Riverine Microbial Impact on Lake Michigan. *Front. Microbiol.* **2019**, *10*, 996. [CrossRef]
4. Paver, S.F.; Newton, R.J.; Coleman, M.L. Microbial communities of the Laurentian Great Lakes reflect connectivity and local biogeochemistry. *Environ. Microbiol.* **2020**, *22*, 433–446. [CrossRef]
5. Evans, M. The large lake ecosystems of northern Canada. *Aquat. Ecosyst. Health Manag.* **2000**, *3*, 65–79. [CrossRef]
6. Boguski, D.A. Annotated Bibliography of Aquatic Research in the Great Slave Lake Region. *Biodive Sci. Manuscr. Rep.* **2015**, *1502*, 1–79.
7. Koehler, G.; Brua, R.B.; Enders, E.C. Foreword to the second Lake Winnipeg special issue. *J. Great Lakes Res.* **2021**, *47*, 553–555. [CrossRef]
8. Matisoff, G.; Watson, S.B.; Guo, J.; Duewiger, A.; Steely, R. Sediment and nutrient distribution and resuspension in Lake Winnipeg. *Sci. Total Environ.* **2017**, *575*, 173–186. [CrossRef]
9. (MWSD) Manitoba Water Stewardship Division. State of Lake Winnipeg: 1999 to 2007. *Manit. Steward. Board Environ. Can.*; 2011; pp. 1–222. Available online: https://publications.gc.ca/site/eng/407829/publication.html (accessed on 18 August 2022).
10. (MARD) Manitoba Agriculture and Resource Department. State of Lake Winnipeg 2nd edition. *Environ. Clim. Chang. Can.*; 2020; pp. 1–195. Available online: https://www.gov.mb.ca/water/pubs/water/lakes-beaches-rivers/state_lake_wpg_report_tech.pdf (accessed on 18 August 2022).
11. Kling, H.J.; Watson, S.B.; McCullough, G.K.; Stainton, M.P. Bloom development and phytoplankton succession in Lake Winnipeg: A comparison of historical records with recent data. *Aquat. Ecosyst. Health Manag.* **2011**, *14*, 219–224. [CrossRef]
12. Pick, F.R. Blooming algae: A Canadian perspective on the rise of toxic cyanobacteria. *Can. J. Fish. Aquat. Sci.* **2016**, *73*, 1149–1158. [CrossRef]
13. Newton, R.J.; Jones, S.E.; Eiler, A.; McMahon, K.D.; Bertilsson, S. A Guide to the Natural History of Freshwater Lake Bacteria. *Microbiol. Mol. Biol. Rev.* **2011**, *75*, 14–49. [CrossRef] [PubMed]
14. Kuzyk, S.B.; Pritchard, A.O.; Plouffe, J.; Sorensen, J.L.; Yurkov, V. Psychrotrophic violacein-producing bacteria isolated from Lake Winnipeg, Canada. *J. Great Lakes Res.* **2021**, *47*, 715–724. [CrossRef]
15. Kuzyk, S.B.; Wiens, K.; Ma, X.; Yurkov, V. Association of aerobic anoxygenic phototrophs and zebra mussels, *Dreissena polymorpha*, within the littoral zone of Lake Winnipeg. *J. Great Lakes Res.* **2021**, *47*, 567–582. [CrossRef]
16. Vila, X.; Abella, C.A. Effects of light quality on the physiology and the ecology of planktonic green sulfur bacteria in lakes. *Photosynth. Res.* **1994**, *41*, 53–65. [CrossRef] [PubMed]
17. Yurkov, V.; Beatty, J.T. Aerobic Anoxygenic Phototrophic Bacteria. *Microbiol. Mol. Biol. Rev.* **1998**, *62*, 695–724. [CrossRef] [PubMed]
18. Yurkov, V.; Hughes, E. Aerobic Anoxygenic Phototrophs: Four Decades of Mystery. In *Modern Topics in the Phototrophic Prokaryotes*; Springer International Publishing: Cham, Switzerland, 2017; pp. 193–214.
19. Vila, X.; Abella, C.; Figueras, J.; Hurley, J. Vertical models of phototrophic bacterial distribution in the metalimnetic microbial communities of several freshwater North-American kettle lakes. *FEMS Microbiol. Ecol.* **1998**, *25*, 287–299. [CrossRef]
20. Koblížek, M. Ecology of aerobic anoxygenic phototrophs in aquatic environments. *FEMS Microbiol. Rev.* **2015**, *39*, 854–870. [CrossRef]
21. Bilyj, M.; Lepitzki, D.; Hughes, E.; Swiderski, J.; Stackebrandt, E.; Pacas, C.; Yurkov, V. Abundance and Diversity of the Phototrophic Microbial Mat Communities of Sulphur Mountain Banff Springs and Their Significance to the Endangered Snail, *Physella johnsoni*. *Open J. Ecol.* **2014**, *4*, 488–516. [CrossRef]
22. Fauteux, L.; Cottrell, M.T.; Kirchman, D.L.; Borrego, C.M.; Garcia-Chaves, M.C.; del Giorgio, P.A. Patterns in Abundance, Cell Size and Pigment Content of Aerobic Anoxygenic Phototrophic Bacteria along Environmental Gradients in Northern Lakes. *PLoS ONE* **2015**, *10*, e0124035. [CrossRef]
23. Depew, D.C.; Krutzelmann, E.; Watchorn, K.E.; Caskenette, A.; Enders, E.C. The distribution, density, and biomass of the zebra mussel (*Dreissena polymorpha*) on natural substrates in Lake Winnipeg 2017–2019. *J. Great Lakes Res.* **2021**, *47*, 556–566. [CrossRef]
24. Ritchie, R.J. Measurement of chlorophylls *a* and *b* and bacteriochlorophyll *a* in organisms from hypereutrophic auxinic waters. *J. Appl. Phycol.* **2018**, *30*, 3075–3087. [CrossRef]
25. Gregersen, L.H.; Habicht, K.S.; Peduzzi, S.; Tonolla, M.; Canfield, D.E.; Miller, M.; Cox, R.P.; Frigaard, N.-U. Dominance of a clonal green sulfur bacterial population in a stratified lake. *FEMS Microbiol. Ecol.* **2009**, *70*, 30–41. [CrossRef] [PubMed]
26. SCOR-UNESCO. Determination of photosynthetic pigments. *Monogr. Oceanogr. Methodol.* **1966**, *1*, 11–18.
27. Steemann-Nielsen, E. The Use of Radioactive Carbon (^{14}C) for Measuring Organic Production in the Sea. *ICES J. Mar. Sci.* **1952**, *43*, 117–140. [CrossRef]
28. Franson, M. *Standard Methods for the Examination of Water and Wastewater*, 20th ed.; American Public Health Association: Washington, DC, USA, 1998.

29. Yurkov, V.; Van Gemerden, H. Abundance and salt tolerance of obligately aerobic, phototrophic bacteria in a marine microbial mat. *Neth. J. Sea Res.* **1993**, *31*, 57–62. [CrossRef]

30. Béjà, O.; Suzuki, M.T.; Heidelberg, J.F.; Nelson, W.C.; Preston, C.M.; Hamada, T.; Eisen, J.A.; Fraser, C.M.; DeLong, E.F. Unsuspected diversity among marine aerobic anoxygenic phototrophs. *Nature* **2002**, *415*, 630–633. [CrossRef]

31. Zervas, A.; Zeng, Y.; Madsen, A.M.; Hansen, L.H. Genomics of Aerobic Photoheterotrophs in Wheat Phyllosphere Reveals Divergent Evolutionary Patterns of Photosynthetic Genes in *Methylobacterium* spp. *Genome Biol. Evol.* **2019**, *11*, 2895–2908. [CrossRef]

32. Kopejtka, K.; Zeng, Y.; Kaftan, D.; Selyanin, V.; Gardian, Z.; Tomasch, J.; Sommaruga, R.; Koblížek, M. Characterization of the Aerobic Anoxygenic Phototrophic Bacterium *Sphingomonas* sp. AAP5. *Microorganisms* **2021**, *9*, 768. [CrossRef]

33. Hughes, E.; Head, B.; Maltman, C.; Piercey-Normore, M.; Yurkov, V. Aerobic anoxygenic phototrophs in gold mine tailings in Nopiming Provincial Park, Manitoba, Canada. *Can. J. Microbiol.* **2017**, *63*, 212–218. [CrossRef]

34. Maltman, C.; Yurkov, V. The Effect of Tellurite on Highly Resistant Freshwater Aerobic Anoxygenic Phototrophs and Their Strategies for Reduction. *Microorganisms* **2015**, *3*, 826–838. [CrossRef]

35. McKinney, W. Data Structures for Statistical Computing in Python. In Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28 June–3 July 2010; pp. 56–61.

36. Bolyen, E.; Rideout, J.R.; Dillon, M.R.; Bokulich, N.A.; Abnet, C.C.; Al-Ghalith, G.A.; Alexander, H.; Alm, E.J.; Arumugam, M.; Asnicar, F.; et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **2019**, *37*, 852–857. [CrossRef]

37. McDonald, D.; Clemente, J.C.; Kuczynski, J.; Rideout, J.R.; Stombaugh, J.; Wendel, D.; Wilke, A.; Huse, S.; Hufnagle, J.; Meyer, F.; et al. The Biological Observation Matrix (BIOM) format or: How I learned to stop worrying and love the ome-ome. *Gigascience* **2012**, *1*, 7. [CrossRef] [PubMed]

38. Amir, A.; McDonald, D.; Navas-Molina, J.A.; Kopylova, E.; Morton, J.T.; Zech Xu, Z.; Kightley, E.P.; Thompson, L.R.; Hyde, E.R.; Gonzalez, A.; et al. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *MSystems* **2017**, *2*, e00191-16. [CrossRef] [PubMed]

39. Bokulich, N.A.; Subramanian, S.; Faith, J.J.; Gevers, D.; Gordon, J.I.; Knight, R.; Mills, D.A.; Caporaso, J.G. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat. Methods* **2013**, *10*, 57–59. [CrossRef] [PubMed]

40. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2—Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* **2010**, *5*, e9490. [CrossRef]

41. Rognes, T.; Flouri, T.; Nichols, B.; Quince, C.; Mahé, F. VSEARCH: A versatile open source tool for metagenomics. *PeerJ* **2016**, *4*, e2584. [CrossRef]

42. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glöckner, F.O. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **2012**, *41*, D590–D596. [CrossRef] [PubMed]

43. Weiss, S.; Xu, Z.Z.; Peddada, S.; Amir, A.; Bittinger, K.; Gonzalez, A.; Lozupone, C.; Zaneveld, J.R.; Vázquez-Baeza, Y.; Birmingham, A.; et al. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* **2017**, *5*, 27. [CrossRef]

44. Stackebrandt, E.; Goodfellow, M. *Nucleic Acid Techniques in Bacterial Systematics*; Wiley: Hoboken, NJ, USA, 1991; pp. 115–175.

45. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef]

46. Faith, D.P.; Minchin, P.R.; Belbin, L. Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio* **1987**, *69*, 57–68. [CrossRef]

47. Vázquez-Baeza, Y.; Gonzalez, A.; Smarr, L.; McDonald, D.; Morton, J.T.; Navas-Molina, J.A.; Knight, R. Bringing the Dynamic Microbiome to Life with Animations. *Cell Host Microbe* **2017**, *21*, 7–10. [CrossRef] [PubMed]

48. Wassenaar, L.I. Dissolved oxygen status of Lake Winnipeg: Spatio-temporal and isotopic (δ^{18}O–O$_2$) patterns. *J. Great Lakes Res.* **2012**, *38*, 123–134. [CrossRef]

49. Binding, C.E.; Greenberg, T.A.; McCullough, G.; Watson, S.B.; Page, E. An analysis of satellite-derived chlorophyll and algal bloom indices on Lake Winnipeg. *J. Great Lakes Res.* **2018**, *44*, 436–446. [CrossRef]

50. Mašín, M.; Nedoma, J.; Pechar, L.; Koblížek, M. Distribution of aerobic anoxygenic phototrophs in temperate freshwater systems. *Environ. Microbiol.* **2008**, *10*, 1988–1996. [CrossRef]

51. Robarts, R.D.; Zohary, T. Temperature effects on photosynthetic capacity, respiration, and growth rates of bloom-forming cyanobacteria. *N. Z. J. Mar. Freshw. Res.* **1987**, *21*, 391–399. [CrossRef]

52. Sayers, M.; Bosse, K.; Fahnenstiel, G.; Shuchman, R. Carbon Fixation Trends in Eleven of the World's Largest Lakes: 2003–2018. *Water* **2020**, *12*, 3500. [CrossRef]

53. Yezhova, Y.; Capelle, D.; Stainton, M.; Papakyriakou, T. Carbon fixation by the phytoplankton community across Lake Winnipeg. *J. Great Lakes Res.* **2021**, *47*, 703–714. [CrossRef]

54. Vollenweider, R.A.; Munawar, M.; Stadelmann, P. A Comparative Review of Phytoplankton and Primary Production in the Laurentian Great Lakes. *J. Fish. Res. Board Can.* **1974**, *31*, 739–762. [CrossRef]

55. Frischer, M.E.; Nierzwicki-Bauer, S.A.; Parsons, R.H.; Vathanodorn, K.; Waitkus, K.R. Interactions between zebra mussels (*Dreissena polymorpha*) and microbial communities. *Can. J. Fish. Aquat. Sci.* **2000**, *57*, 591–599. [CrossRef]

56. Yurkov, V.; Jappe, J.; Vermeglio, A. Tellurite resistance and reduction by obligately aerobic photosynthetic bacteria. *Appl. Environ. Microbiol.* **1996**, *62*, 4195–4198. [CrossRef]

57. Jankowiak, J.; Hattenrath-Lehmann, T.; Kramer, B.J.; Ladds, M.; Gobler, C.J. Deciphering the effects of nitrogen, phosphorus, and temperature on cyanobacterial bloom intensification, diversity, and toxicity in western Lake Erie. *Limnol. Oceanogr.* **2019**, *64*, 1347–1370. [CrossRef]

58. Yan, Q.; Stegen, J.C.; Yu, Y.; Deng, Y.; Li, X.; Wu, S.; Dai, L.; Zhang, X.; Li, J.; Wang, C.; et al. Nearly a decade-long repeatable seasonal diversity patterns of bacterioplankton communities in the eutrophic Lake Donghu (Wuhan, China). *Mol. Ecol.* **2017**, *26*, 3839–3850. [CrossRef] [PubMed]

59. Kolářová, E.; Medová, H.; Piwosz, K.; Koblížek, M. Seasonal dynamics of aerobic anoxygenic phototrophs in freshwater lake Vlkov. *Folia Microbiol.* **2019**, *64*, 705–710. [CrossRef]

60. Čuperová, Z.; Holzer, E.; Salka, I.; Sommaruga, R.; Koblížek, M. Temporal Changes and Altitudinal Distribution of Aerobic Anoxygenic Phototrophs in Mountain Lakes. *Appl. Environ. Microbiol.* **2013**, *79*, 6439–6446. [CrossRef] [PubMed]

61. Fecskeová, L.K.; Piwosz, K.; Hanusová, M.; Nedoma, J.; Znachor, P.; Koblížek, M. Diel changes and diversity of pufM expression in freshwater communities of anoxygenic phototrophic bacteria. *Sci. Rep.* **2019**, *9*, 18766. [CrossRef] [PubMed]

62. Csotonyi, J.T.; Swiderski, J.; Stackebrandt, E.; Yurkov, V. Novel halophilic aerobic anoxygenic phototrophs from a Canadian hypersaline spring system. *Extremophiles* **2008**, *12*, 529–539. [CrossRef] [PubMed]

63. Kasalický, V.; Zeng, Y.; Piwosz, K.; Šimek, K.; Kratochvilová, H.; Koblížek, M. Aerobic Anoxygenic Photosynthesis Is Commonly Present within the Genus *Limnohabitans*. *Appl. Environ. Microbiol.* **2018**, *84*, e02116-17. [CrossRef]

64. Suyama, T.; Shigematsu, T.; Takaichi, S.; Nodasaka, Y.; Fujikawa, S.; Hosoya, H.; Tokiwa, Y.; Kanagawa, T.; Hanada, S. *Roseateles depolymerans* gen. nov., sp. nov., a new bacteriochlorophyll *a*-containing obligate aerobe belonging to the β-subclass of the *Proteobacteria*. *Int. J. Syst. Evol. Microbiol.* **1999**, *49*, 449–457. [CrossRef]

65. Shiba, T. *Roseobacter litoralis* gen. nov., sp. nov., and *Roseobacter denitrificans* sp. nov., Aerobic Pink-Pigmented Bacteria Which Contain Bacteriochlorophyll *a*. Syst. *Appl. Microbiol.* **1991**, *14*, 140–145.

66. Rathgeber, C.; Yurkova, N.; Stackebrandt, E.; Schumann, P.; Beatty, J.T.; Yurkov, V. *Roseicyclus mahoneyensis* gen. nov., sp. nov., an aerobic phototrophic bacterium isolated from a meromictic lake. *Int. J. Syst. Evol. Microbiol.* **2005**, *55*, 1597–1603. [CrossRef]

67. Csotonyi, J.T.; Stackebrandt, E.; Swiderski, J.; Schumann, P.; Yurkov, V. An alphaproteobacterium capable of both aerobic and anaerobic anoxygenic photosynthesis but incapable of photoautotrophy: *Charonomicrobium ambiphototrophicum*, gen. nov., sp. nov. *Photosynth. Res.* **2011**, *107*, 257–268. [CrossRef] [PubMed]

68. Kuzyk, S.B.; Jafri, M.; Humphrey, E.; Maltman, C.; Kyndt, J.A.; Yurkov, V. Prosthecate aerobic anoxygenic phototrophs *Photocaulis sulfatitolerans* gen. nov. sp. nov. and *Photocaulis rubescens* sp. nov. isolated from alpine meromictic lakes in British Columbia, Canada. *Arch. Microbiol.* **2022**, *204*, 444. [CrossRef] [PubMed]

69. Tank, M.; Bryant, D.A. Nutrient requirements and growth physiology of the photoheterotrophic *Acidobacterium*, *Chloracidobacterium thermophilum*. *Front. Microbiol.* **2015**, *6*, 226. [CrossRef] [PubMed]

70. Rathgeber, C.; Beatty, J.T.; Yurkov, V. Aerobic phototrophic bacteria: New evidence for the diversity, ecological importance and applied potential of this previously overlooked group. *Review. Photosynth. Res.* **2004**, *81*, 113–128. [CrossRef]

71. Yurkov, V.; Csotonyi, J. New light on aerobic anoxygenic phototrophs. Book chapter. In *The Purple Photosynthetic Bacteria*; Hunter, C.N., Daldal, F., Thurnauer, M.C., Beatty, J.T., Eds.; Springer: Dordrecht, The Netherlands, 2009; pp. 31–55.

72. Yurkov, V.; Hughes, E. Genes associated with the peculiar phenotypes of the aerobic anoxygenic phototrophs. In *Genome Evolution of Photosynthetic Bacteria*; Beatty, J.T., Jacquot, J.P., Gadal, P., Eds.; Elsevier: Amsterdam, The Netherlands, 2013; Volume 66, pp. 327–358.

73. Peyraud, R.; Schneider, K.; Kiefer, P.; Massou, S.; Vorholt, J.A.; Portais, J.C. Genome-scale reconstruction and system level investigation of the metabolic network of *Methylobacterium extorquens* AM1. *BMC Syst. Biol.* **2011**, *5*, 189. [CrossRef]

74. Tang, K.; Yang, L.H.; Chen, Y.P.; Tao, Y.; Feng, F.Y.; Meng, J.Y. *Aerophototrophica crusticola* gen. nov., sp. nov., isolated from desert biocrusts. *Int. J. Syst. Evol. Microbiol.* **2021**, *71*, 004677. [CrossRef]

75. Mehrshad, M.; Salcher, M.M.; Okazaki, Y.; Nakano, S.-i.; Šimek, K.; Andrei, A.-S.; Ghai, R. Hidden in plain sight—highly abundant and diverse planktonic freshwater Chloroflexi. *Microbiome* **2018**, *6*, 176. [CrossRef]

76. Imhoff, J.F.; Kyndt, J.A.; Meyer, T.E. Genomic Comparison, Phylogeny and Taxonomic Reevaluation of the *Ectothiorhodospiraceae* and Description of *Halorhodospiraceae* fam. nov. and *Halochlorospira* gen. nov. *Microorganisms* **2022**, *10*, 295. [CrossRef]

77. Kyndt, J.A.; Salama, D.M.; Meyer, T.E.; Imhoff, J.F. Phylogenetic relationship of phototrophic heliobacteria and systematic reconsideration of species and genus assignments based on genome sequences of eight species. *Int. J. Syst. Evol. Microbiol.* **2021**, *71*, 004729. [CrossRef]

78. Imhoff, J.F. Phylogenetic taxonomy of the family *Chlorobiaceae* on the basis of 16S rRNA and fmo (Fenna– Matthews–Olson protein) gene sequences. *Int. J. Syst. Evol. Microbiol.* **2003**, *53*, 941–951. [CrossRef]

79. Saini, M.K.; ChihChe, W.; Soulier, N.; Sebastian, A.; Albert, I.; Thiel, V.; Bryant, D.A.; Hanada, S.; Tank, M. *Caldichromatium japonicum* gen. nov., sp. nov., a novel thermophilic phototrophic purple sulphur bacterium of the *Chromatiaceae* isolated from Nakabusa hot springs, Japan. *Int. J. Syst. Evol. Microbiol.* **2020**, *70*, 5701–5710. [CrossRef] [PubMed]

80. Nupur, N.; Saini, M.K.; Singh, P.K.; Korpole, S.; Tanuku, N.R.S.; Takaichi, S.; Pinnaka, A.K. Imhoffiella gen. nov., a marine phototrophic member of the family *Chromatiaceae* including the description of *Imhoffiella purpurea* sp. nov. and the reclassification of *Thiorhodococcus bheemlicus* Anil Kumar et al. 2007 as *Imhoffiella bheemlica* comb. nov. *Int. J. Syst. Evol. Microbiol.* **2017**, *67*, 1949–1956. [CrossRef] [PubMed]

81. Imhoff, J.F.; Suling, J.; Petri, R. Phylogenetic relationships among the *Chromatiaceae*, their taxonomic reclassification and description of the new genera *Allochromatium, Halochromatium, Isochromatium, Marichromatium, Thiococcus, Thiohalocapsa* and *Thermochromatium*. *Int. J. Syst. Bacteriol.* **1998**, *48*, 1129–1143. [CrossRef] [PubMed]

82. Gorlenko, V.M.; Krassilnikova, E.N.; Kikina, O.G.; Tatarinova, N.Y. The new motile purple sulfur bacteria *Lamprobacter modesto-halophilus* nov. gen., nov. sp. with gas vacuoles. *Izv. Akad. Nauk SSSR Ser. Biol.* **1979**, *5*, 755–767.

83. Watson, S.W. Characteristics of a marine nitrifying bacterium, *Nitrosocystis oceanus* sp. n. *Limnol. Oceanogr.* **1965**, *10*, R274–R289. [CrossRef]

84. Tanuku, N.R.; Shinichi, T.; Pinnaka, A.K. *Phaeobacterium nitratireducens* gen. nov., sp. nov., a phototrophic gammaproteobacterium isolated from a mangrove forest sediment sample. *Int. J. Syst. Evol. Microbiol.* **2015**, *65*, 2357–2364.

85. Shivali, K.; Sasikala, C.; Ramana, C.V. MLSA barcoding of *Marichromatium* spp. and reclassification of *Marichromatium fluminis* (Sucharita et al., 2010) as *Phaeochromatium fluminis* gen. nov. comb. nov. *Syst. Appl. Microbiol.* **2012**, *35*, 221–225. [CrossRef]

86. Dilling, W.; Liesack, W.; Pfennig, N. *Rhabdochromatium marinum* gen. nom. rev., sp. nov., a purple sulfur bacterium from a salt marsh microbial mat. *Arch. Microbiol.* **1995**, *164*, 125–131. [CrossRef]

87. Bryantseva, I.A.; Gorlenko, V.M.; Kompantseva, E.I.; Imhoff, J.F. *Thioalkalicoccus limnaeus* gen. nov., sp. nov., a new alkaliphilic purple sulfur bacterium with bacteriochlorophyll *b*. *Int. J. Syst. Evol. Microbiol.* **2000**, *50*, 2157–2163. [CrossRef]

88. Rees, G.N.; Harfoot, C.G.; Janssen, P.H.; Schoenborn, L.; Kuever, J.; Lunsdorf, H. *Thiobaca trueperi* gen. nov., sp. nov., a phototrophic purple sulfur bacterium isolated from freshwater lake sediment. *Int. J. Syst. Evol. Microbiol.* **2002**, *52*, 671–678. [CrossRef]

89. Luedin, S.M.; Pothier, J.F.; Danza, F.; Storelli, N.; Frigaard, N.-U.; Wittwer, M.; Tonolla, M. Complete genome sequence of "*Thiodictyon syntrophicum*" sp. nov. strain Cad16^T, a photolithoautotrophic purple sulfur bacterium isolated from the alpine meromictic Lake Cadagno. *Genomic Sci.* **2018**, *13*, 14. [CrossRef]

90. Eichler, B.; Pfennig, N. Isolation and characteristics of *Thiopedia rosea* (neotype)*. *Arch. Microbiol.* **1991**, *155*, 210–216. [CrossRef]

91. Imhoff, J.F.; Meyer, T.E.; Kyndt, J.A. The genome sequence of the giant phototrophic gammaproteobacterium *Thiospirillum jenense* gives insight into its physiological properties and phylogenetic relationships. *Arch. Microbiol.* **2021**, *203*, 97–105. [CrossRef] [PubMed]

92. van Niel, C.B. The culture, general physiology, morphology, and classification of the non-sulfur purple and brown bacteria. *Bacteriol. Rev.* **1944**, *8*, 1–118. [CrossRef] [PubMed]

Article

Nitrogen Fixation Activity and Genome Analysis of a Moderately Haloalkaliphilic Anoxygenic Phototrophic Bacterium *Rhodovulum tesquicola*

Anastasia V. Komova [1,*], **Elizaveta D. Bakhmutova** [1], **Anna O. Izotova** [1], **Evelina S. Kochetova** [1], **Stepan V. Toshchakov** [1], **Zorigto B. Namsaraev** [1], **Maxim V. Golichenkov** [2,3] and **Aleksei A. Korzhenkov** [1]

[1]   Kurchatov Centre for Genome Research, NRC Kurchatov Institute, 123098 Moscow, Russia
[2]   Soil Biology Department, Soil Science Faculty, Moscow State University, 119991 Moscow, Russia
[3]   Department of Biology, Faculty of Pediatrics, Pirogov Russian National Research Medical University, 117997 Moscow, Russia
*   Correspondence: komovaav@gmail.com

Abstract: The genome of the moderately haloalkaliphilic diazotrophic anoxygenic phototrophic bacterium *Rhodovulum tesquicola* A-36s[T] isolated from an alkaline lake was analyzed and compared to the genomes of the closest species *Rhodovulum steppense* A-20s[T] and *Rhodovulum strictum* DSM 11289[T]. The genomic features of three organisms are quite similar, reflecting their ecological and physiological role of facultative photoheterotrophs. Nevertheless, the nitrogenase activity of the pure cultures of the studied bacteria differed significantly: the highest rate (4066 nmoles C_2H_2/mg of dry weight per hour) was demonstrated by *Rhodovulum strictum* while the rates in *Rhodovulum tesquicola* and *Rhodovulum steppense* were an order of magnitude lower (278 and 523 nmoles C_2H_2/mg of dry weight per hour, respectively). This difference can be attributed to the presence of an additional nitrogenase operon found exclusively in *R. strictum* and to the structural variation in nitrogenase operon in *R. tesquicola*.

Keywords: anoxygenic phototrophs; halophiles; alkaliphiles; nitrogen fixation; diazotroph; *Rhodovulum tesquicola*; genome sequence

1. Introduction

The genus *Rhodovulum* includes anoxygenic phototrophic bacteria that require sodium chloride for growth [1,2]. The members of the genus are widely distributed in marine and saline environments [3–5], and due to their metabolic capacities may be used in a variety of biotechnological applications [6–8]. Additionally, recent studies show a wider spread of photoferrotrophy and phototrophic extracellular electron uptake in *Rhodovulum* that previously thought [9]. Thus, the genome analysis of the representatives of the genus appears to provide new insights into the genetic and evolutionary basis of metabolic processes in halotolerant anoxygenic phototrophic *Proteobacteria*. To date, *Rhodovulum* comprises 20 validly described species, for eight of which no genome sequence has been published so far.

Bacteria of the genus *Rhodovulum* preferably grow photoheterotrophically and are able to fix nitrogen. Thus, they serve as an important mediator of the carbon and nitrogen cycles as it links the organic matter decomposition and nitrogen fixation. Saline and soda lakes are among the most productive ecosystems in the world [10,11]. The alkaline conditions favour the volatilization of ammonium from the lakes; therefore the regular supply of nitrogen is needed to provide the functioning of ecosystem [12].

The process of nitrogen fixation in saline and soda lakes was studied mostly for the natural ecosystems and mesocosms [13–17]. The nitrogen fixation in pure cultures of phototrophic *Alphaproteobacteria* was studied in details for the neutrophilic and some

of the marine and true halophilic representatives [18–20]. Nevertheless, little is known about this process in pure cultures of halotolerant and alkaliphilic anoxygenic phototrophic *Alphaproteobacteria* that are widespread in the lakes with relatively low salinity and increased alkalinity [5,21].

Rhodovulum tesquicola A-36s[T] was isolated from an alkaline lake Sul'fatnoe (Buryat Republic, Siberia, Russia; pH 9.2, mineralization 7.7 g/L) [5] and later described as a type strain of a new species within *Rhodovulum* genus [22]. It is a typical anoxygenic phototrophic bacterium able to grow photoheterotrophically and chemoheterotrophically with a range of organic substrates, and to grow photolithoautotrophically with sulfur compounds as electron donors. However, its ability to fix nitrogen has not been studied before. Here we present the analysis of genome of *R. tesquicola* in comparison with the genomes of the closest species, *R. steppense* A-20s[T] and *R. strictum* DSM 11289[T], and the data on the nitrogen fixation activity by the pure cultures of these organisms.

2. Materials and Methods

2.1. Microbial Cultivation

R. tesquicola A-36s[T] (=ATCC BAA-1573[T]), *R. steppense* A-20s[T] (=DSM 21153[T]) (both isolated previously by A.V.K. and kept at the laboratory collection) and *R. strictum* DSM 11289[T] obtained from the Leibniz-Institut DSMZ, Braunschweig, Germany were grown photoheterotrophically at 30 °C on the previously described medium [22] (pH 8.5–9).

2.2. Nitrogenase Activity Measurement

Nitrogenase activity measurement was performed by the acetylene reduction assay [23]. 1 mL of acetylene (about 10% of gaseous headspace in the vial) was injected into the hermetically sealed 23 mL vials containing liquid photoheterotrophically grown cultures of the investigated bacteria with argon filled headspace. The vials were incubated at 25 °C for 20 h. After that, 1 mL of gaseous phase was sampled and the detection of ethylene was performed on Crystal-2000 chromatograph (LLC NPF Meta-Chrom, Yoshkar-Ola, Russia) with a heated flame ionization detector (HFID). Instrument characteristics: column length—1 m, diameter—3 mm, filler—Porapak N 80/100, column temperature—60 °C, detector temperature—160 °C, evaporator temperature—100 °C, carrier gas flow (N_2)—50 mL/min, air—280 mL/min, hydrogen—28 mL/min.

2.3. Genome Sequencing and Assembly

Genomic DNA was isolated from cell culture of *R. tesquicola* A-36s[T] using the QIAamp® DNA Mini kit following the manufacturer's recommendations (Qiagen, Dusseldorf, Germany).

DNA was fragmented by ultrasound on a Covaris M220 (Covaris, Woburn, MA, USA). A paired-end genomic library (average insertion size 300 bp) was made using the NEBNext® Ultra[TM] II DNA Library Prep Kit (New England Biolabs, Ipswich, MA, USA). The DNA library was sequenced using an Illumina NovaSeq 6000 System (Illumina, San Diego, CA, USA) with a 2 × 250 bp paired-end read.

Genome assembly was performed using ZGA software pipeline v0.1 [24]. Sequencing reads were processed using BBTools software toolkit v38.96 [25]. Reads with low quality bases and adapter sequences were trimmed from sequencing reads with BBduk v38.96 (ktrim = r qtrim = rl trimq = 21 k = 20 mL = 33 trimpolyg = 4 trimpolya = 4). Trimmed reads were corrected with Tadpole (mode = correct, cecc = t). Corrected reads were normalized with BBnorm v38.96 to 250× depth coverage. Normalized reads were assembled to scaffolds with SPAdes v3.14 [26] using default parameters. The resulting assembly was polished with Pilon v1.24 [27] using trimmed reads.

Primary annotation was performed using DFAST v1.2.15 [28], CDS were predicted with Prodigal. Additional annotation was performed with NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v6.1 [29], proteins were predicted using GeneMarkS-2+ [30] and best-placed reference protein set. KEGG orthologs (KO) annotation was performed

using KofamKOALA [31]. Completeness for a given KEGG module [32] was calculated using the program anvi-estimate-metabolism from anvi'o toolkit v7.1 [33].

All available on the 1 July 2022 *Rhodovulum* genomes including the genomes of *R. steppense* A-20s^T and *R. strictum* DSM 11289^T were downloaded from NCBI Assembly database. Genome completeness was assessed with CheckM v1.2.0 [34] taxonomy workflow using *Rhodobacteraceae* marker set. Pairwise average nucleotide identity (ANI) was estimated using FastANI [35]. Pangenomic analysis was conducted using Proteinortho v6 [36] with BLASTP+ [37] algorithm. Nitrogen fixation genes in *Rhodovulum* genomes were searched using gene annotations in NCBI protein and Uniprot and BLASTP+ [37] search in NCBI nr and Swissprot databases. Phi29-like polymerase and viral single-stranded DNA-binding protein were searched using BLASTP+ [37] (e-value 0.001) with P03680 and Q38504 reference sequences, respectively.

Amino acid sequences of nifH gene homologs were found via BLASTP+ [37] search in NBCI nr database. The first 250 hits were downloaded and clustered using CD-HIT [38] with 95% identity threshold. Representative sequences of 16S rRNA and *nifH* genes were aligned using MAFFT v7.475 with the FFT-NS-i algorithm [39]. The resulting alignments were manually curated and phylogenetic trees were inferred using FastTree v2.1.10 [40] in auto mode. The trees were visualized using ITOL web-server [41].

3. Results and Discussion

3.1. General Characteristics of Genome

The sequencing yielded in 6,272,843 pair-end reads with a length of 251 bp, totaling to 41.39 Gbp. The draft genome of *R. tesquicola* A-36s^T consists of 80 sequences resulting in 3,521,822 bp with GC-content of 67.68% and N50 of 156,513 bp (Table 1). The most similar to *R. tesquicola* A-36s^T genomes of cultivated *Rhodovulum* species were *R. steppense*—ANI value 94.86% and *R. strictum*—88.91% (Figures 1 and 2). These values support the novelty of *R. tesquicola*.

Table 1. Genome features of *R. tesquicola*, *R. steppense* and *R. strictum*, according to NCBI database.

	Rhodovulum tesquicola **A-36s^T**	*Rhodovulum steppense* **A-20s^T**	*Rhodovulum strictum* **DSM 11289^T**
Genome size	3,521,822	3,665,949	3,825,097
Total number of genes	3490	3634	3756
Proteins	3375	3582	3636
rRNA genes	1, 1, 1 (5S, 16S, 23S)	1, 1, 2 (5S, 16S, 23S partial)	1, 1, 1 (5S, 16S, 23S)
tRNA genes	45	51	42
Genome completeness, %	99.41	99.12	98.83
Genome contamination, %	0	0.15	0.59

The completeness of the KEGG modules was assessed for *R. tesquicola*, *R. steppense* and *R. strictum* genomes. The completeness differs only in 28 out of 204 modules. Only 18 differing modules are complete for at least 50% for at least one of the genomes and differ for at least 10% (Figure 3, Supplementary Table S1).

Pangenome analaysis of three *Rhodovulum* genomes (Figure 4) showed that *R. tesquicola* has the smallest genome and more common protein clusters comparing to *R. steppense* and *R. strictum*. *R. tesquicola* and *R. steppense* have more in common than any other pair of genomes that emphasize their proximity. The core genome of the three *Rhodovulum* genomes consists of 2646 protein clusters corresponding to 79% of *R. tesquicola* genome. Among genes lacking in *R. tesquicola* genome there are mobile elements, genes of prophage loci and hypothetical proteins without any functional annotation. Phi29-like DNA polymerase and phage single-stranded DNA-binding protein genes were not found in any of the three genomes.
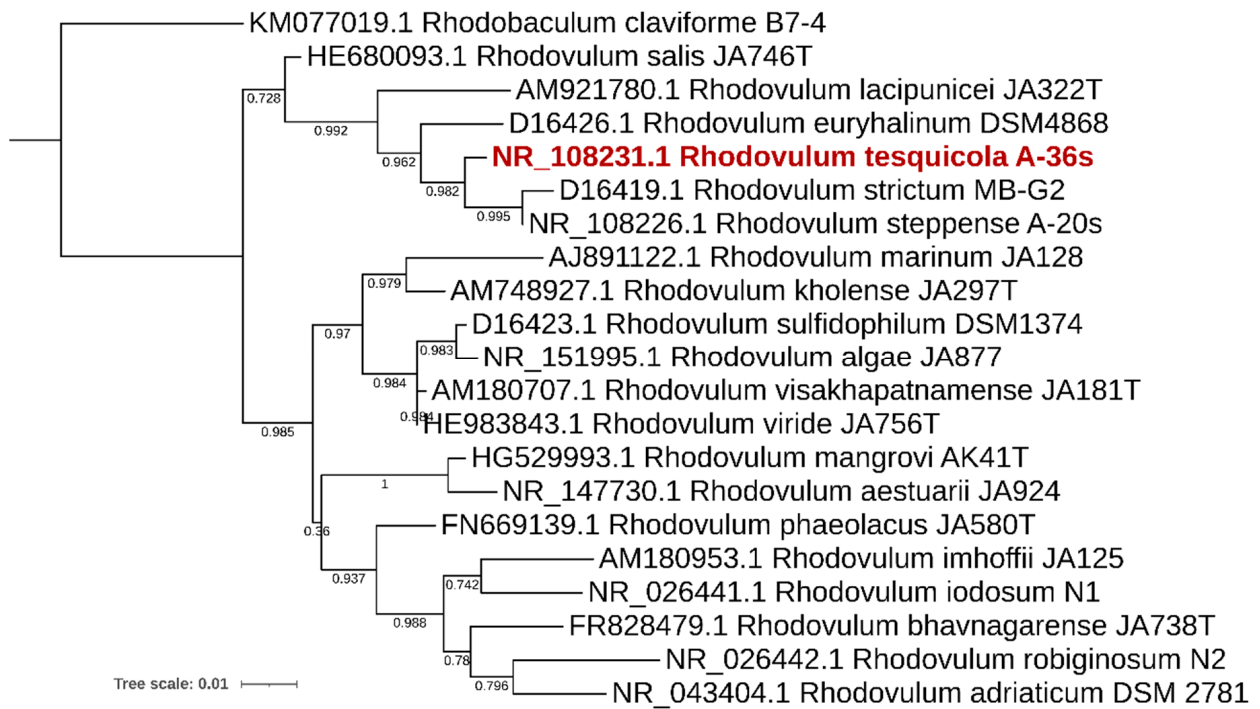
Figure 1. 16S rRNA gene phylogeny of the genus *Rhodovulum*. The tree is scaled by evolutionary distance (count of substitutions per site).
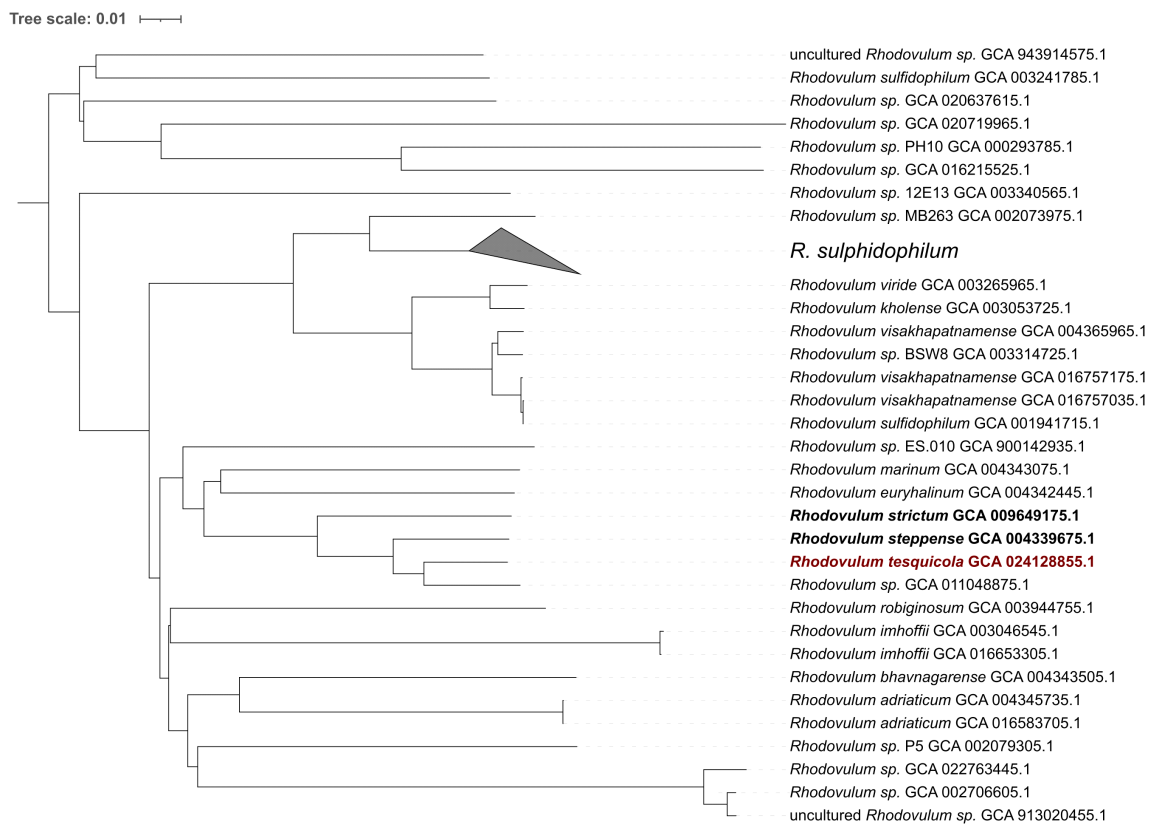
Figure 2. ANI dissimilarity dendrogram of high quality public *Rhodovulum* genomes. The dendrogram is rooted at the midpoint and scaled by genomic distance between strains.

KEGG module	Module name	Subcategory	Module completeness		
			R. tesquicola	*R. steppense*	*R. strictum*
M00135	GABA biosynthesis, eukaryotes, putrescine => GABA	Polyamine biosynthesis	0.25	0.5	0.5
M00033	Ectoine biosynthesis, aspartate => ectoine	Serine and threonine metabolism	1	1	0.4
M00555	Betaine biosynthesis, choline => betaine	Serine and threonine metabolism	0	0.5	0.4
M00793	dTDP-L-rhamnose biosynthesis	Polyketide sugar unit biosynthesis	0.06	1	0.06
M00552	D-galactonate degradation, De Ley-Doudoroff pathway, D-galactonate => glycerate-3P	Other carbohydrate metabolism	0.4	0.4	0.67
M00565	Trehalose biosynthesis, D-glucose 1P => trehalose	Other carbohydrate metabolism	0.67	0.67	0.04
M00173	Reductive citrate cycle (Arnon-Buchanan cycle)	Carbon fixation	0.8	0.7	0.8
M00579	Phosphate acetyltransferase-acetate kinase pathway, acetyl-CoA => acetate	Carbon fixation	1	0.5	1
M00115	NAD biosynthesis, aspartate => quinolinate => NAD	Cofactor and vitamin metabolism	0.8	0.4	1
M00896	Thiamine biosynthesis, archaea, AIR (+ NAD+) => TMP/TPP	Cofactor and vitamin metabolism	0.6	0.4	0.4
M00912	NAD biosynthesis, tryptophan => quinolinate => NAD	Cofactor and vitamin metabolism	0.63	0.5	0.63
M00546	Purine degradation, xanthine => urea	Purine metabolism	0.67	0.67	0.17
M00545	Trans-cinnamate degradation, trans-cinnamate => acetyl-CoA	Aromatics degradation	0.17	0.17	1
M00568	Catechol ortho-cleavage, catechol => 3-oxoadipate	Aromatics degradation	0.25	0.25	0.67
M00649	Multidrug resistance, efflux pump AdeABC	Drug resistance	0.83	0.83	0
M00714	Multidrug resistance, efflux pump QacA	Drug resistance	0.5	0.5	0
M00564	Helicobacter pylori pathogenicity signature, cagA pathogenicity island	Pathogenicity	0.04	0.04	0.5
M00616	Sulfate-sulfur assimilation	Metabolic capacity	0.33	0.33	0.71
M00618	Acetogen	Metabolic capacity	0.71	0.46	0.71

Figure 3. Completeness of the most different KEGG modules for three *Rhodovulum* genomes. Colors indicate the completeness of the module: red for the least complete, green for the most complete.
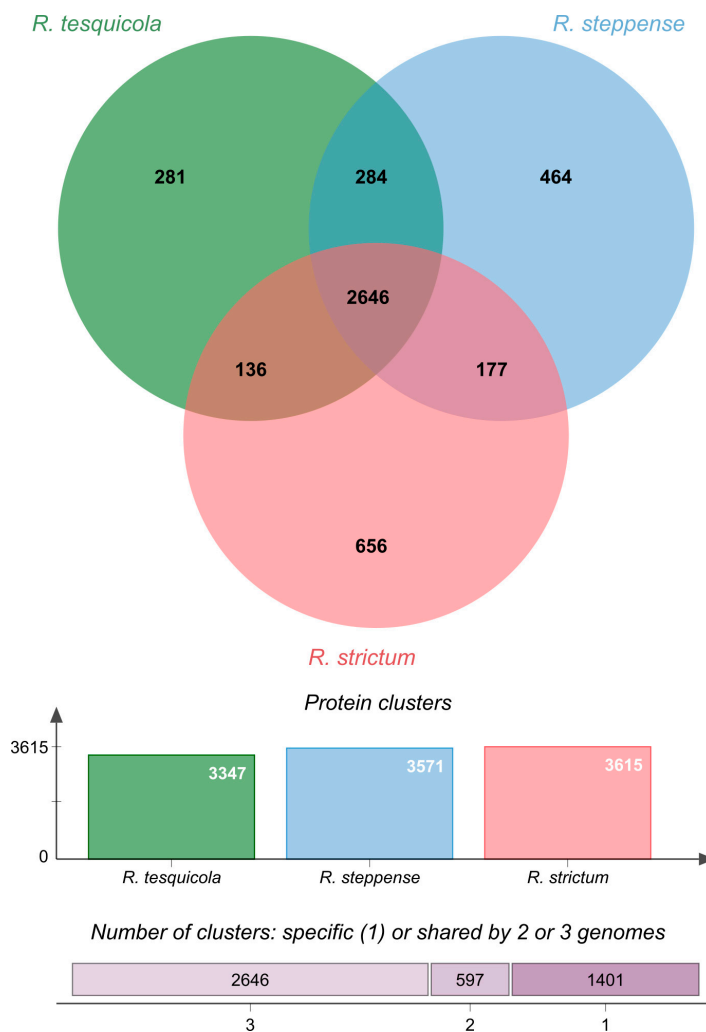
Figure 4. Venn diagram summarizing pangenome analysis of *R. tesquicola*, *R. steppense* and *R. strictum*.

3.2. Carbon Metabolism

Genes responsible for a complete TCA cycle, non-oxidative phase of pentose phosphate pathway, Entner-Doudoroff pathway, glycolysis and gluconeogenesis are present in all the genomes studied (Supplementary Table S1). The anoxygenic phototrophic bacteria can assimilate acetate as a sole organic substrate in a number of anaplerotic pathways to replenish the oxaloacetate pool [42,43]. Among them, the genes for the full ethylmalonyl pathway were found in the genome of *R. tesquicola*. The methylaspartate and glyoxylate pathways are incomplete. Although there is evidence for the isocitrate lyase presence in *R. steppense* [44], we have found no corresponding or homologous genes in any of the three genomes studied. Thus, the detection of the potential presence and functioning of glyoxylate cycle is a subject of further research. The absence of gene for acetyl-CoA synthetase in *R. tesquicola* shows that only the two-step mechanism of acetate assimilation via acetyl phosphate may function in *R. tesquicola*, while *R. steppense* and *R. strictum* have genes for both ways [45]. On the opposite, the gene for malonyl-CoA decarboxylase potentially involved in malonyl/methylmalonyl utilization [46,47] was detected only in *R. tesquicola*, thus expanding the variety of organic substrates for utilization.

3.3. Nitrogen Metabolism

During photoheterotrophic growth, the nitrogenase activity of *R. strictum* was significantly higher than that of *R. steppense* and *R. tesquicola* (Tables 2 and 3).

Table 2. Physiological features of *R. tesquicola*, *R. steppense* and *R. strictum* ([22], this study). b, Biotin; paba, p-aminobenzoate; t, thiamine; n, niacin; +, high nitrogenase activity rate; +−, moderate nitrogenase activity rate.

	R. tesquicola A-36s[T]	*R. steppense* A-20s[T]	*R. strictum* DSM 11289[T]
pH range (optimum)	7.5–10.0 (8.5–9.0)	7.5–10.0 (8.5)	7.5–9.0 (8.0)
NaCl range (optimum) (%)	0.3–10.0 (1–3)	0.5–10.0 (1–5)	0.25–3.00 (0.8)
Vitamin requirement	b, paba, t, n	b, n, t	b, paba, t
Nitrogenase activity	+−	+−	+

Table 3. Nitrogenase activity in the pure cultures of anoxygenic phototrophic *Alphaproteobacteria* (strain designation is indicated in case if only one strain was tested).

Species	Nitrogenase Activity (nmoles C_2H_2/mg Dry Weight per Hour)	Reference
Rhodobacter capsulatus	1430–4900	[19]
Rhodobacter capsulatus B10	3650	[19]
Cereibacter sphaeroides	1110–2755	[19]
Rhodovulum sulfidophilum W4	640	[19]
Rhodothalassium salexigens DSM 2172	630	[19]
Rhodomicrobium vannielii	660–1570	[19]
Rhodopseudomonas palustris	480–700	[19]
Rhodoblastus acidophilus	500–900	[19]
Rhodovulum tesquicola A-36s[T]	278	This work
Rhodovulum steppense A-20s[T]	523	This work
Rhodovulum strictum DSM 11289[T]	4066	This work

The genomes of all three investigated bacteria contain the genes necessary for the biosynthesis of Mo-nitrogenase (*nifHDK*) and iron–molybdenum cofactor FeMoco (*nifB*, *nifEN*, *nifV*); no alternative nitrogenases were found.

All three genomes share a common nitrogenase operon ("slow" operon) (Figure 5). These genes (GH815_10410-GH815_10390 in case of *R. strictum*) have close homologues in several representatives of the genera *Rhodobacter* and *Cereibacter*, as well as in a small number of *Alphaprotobacteria* mainly from *Hyphomicrobiales*. Relatively close homologues

are also found in the representatives of *Firmicutes* and *Chloroflexi* (Figure S1), suggesting a lateral gene transfer event in case of the "slow" operon. In addition, we have found the second nitrogenase operon in *R. strictum* genome (Figure 5). Only a limited number of *Alphaproteobacteria* including *R. strictum* and *Rhodobacter capsulatus* possesses second operon with putative higher nitrogen fixation activity ("fast" operon) (Figure 6).

Figure 5. Nitrogen fixation operons. Top: common nitrogen fixation operon of three studied *Rhodovulum* species ("slow" operon): 1—4Fe-4S binding protein, 2—leucine rich repeat (LRR) protein, 3—*nifZ*, 4—*nifT*, 5—SIR2-like protein, 6—*nifX*, 7—*nifX*-associated nitrogen fixation protein, 8—hypothetical protein, 9—*nif*-specific ferredoxin III, 10—iron-sulfur cluster assembly accessory protein, 11—*nifU*, 12—*nifW*. Bottom: additional operon for nitrogen fixation present in *R. strictum* ("fast" operon): 1—oxidoreductase, 2—nitrogenase, 3—molybdopterin-dependent oxidoreductase (internal stop codon).

Figure 6. Phylogeny of *nifH* gene across *Rhodobacterales*. The tree includes *nifH* homologues missing in *R. steppense* and *R. tesquicola* but found in *R. strictum* and *Rhodobacter capsulatus* strain SB1003 (blue background) and *nifH* homologues appearing in all three *Rhodovulum* strains (red background). Ultrafast bootstrap values are displayed as branch width. The tree is scaled by evolutionary distance (count of substitutions per site).

The *nif* operon in *R. tesquicola* is broken into two parts. 16 genes from *nifA* (NHN26_05215) to *fdxB* (NHN26_05140) are separated from *nifU* (NHN26_05060), *nifV* (NHN26_05055) and *nifW* (NHN26_05050) by a cluster of 15 genes coding mainly for *RnfABCDGE* type electron transport complex. This break may lead to a decreased nitrogenase activity in comparison with *R. steppense*.

Nitrogen fixing phototrophic microorganisms play an important role in the nitrogen uptake by the phototrophic communities of the saline and soda lakes. It has been previously demonstrated that in soda lakes, nitrogen is fixed mainly at low salinities with the nitrogen fixation rates in the light much higher than those in the dark at the salinities of less than 100 g/L [17,48]. The investigated communities consisted mostly of the photoautotrophic heterocystous cyanobacteria with *nifH*-containing *Ectothiorhodospira* sp. present to a lesser extent. In the saline and soda lakes, *Ectothiorhodospiraceae* usually occur at the salinities up to the saturation, while *Rhodobacteraceae* reach high numbers at the mineralization below 100 g/L [21], and thus may be responsible for the nitrogen fixation activity at these conditions. However, in spite of wide distribution of *nif*-related genes among the phototrophic *Alphaproteobacteria* inhabiting saline and soda environments [49], the actual rates of nitrogen fixation in pure cultures as well as *in situ* may vary strongly (Table 3).

The product of the *nifH* gene is dinitrogenase reductase, also known as the Fe protein or component II. The Fe protein has an obligate redox role and is involved in the process of maturation of the P-clusters of the apo-MoFe protein to their catalytically active forms [50–52]. Probably the "fast" and "slow" operon function differs in the expression rate of *nifH* genes, leading to the different rates of nitrogen fixation. As *R. strictum* and *Rba. capsulatus* contain 2 operons with different *nifH* variants (Figure 5) while *R. steppense* and *R. tesquicola* have a single operon, we hypothesize that either the "fast" operon or both of them are responsible for high nitrogenase activity. Phylogenetically, the "fast" operon cluster (Figure S1) includes not only *Rba. capsulatus* and *R. strictum* performing high nitrogenase activity but also *Rhodoblastus sphagnicola*, *Rhodomicrobium vannielii* and *Rhodopseudomonas palustris* with moderate nitrogenase activity, so it is likely that both "fast" and "slow" operons may be required to achieve high activities.

The structure and organization of genes for molybdenum-only and alternative nitrogenases in *Rba. capsulatus* and other organisms are well studied [50,52–55]. The presence of multiple copies of *nif* genes in *Rba. capsulatus* was demonstrated [56], but little is known about the nitrogenase activity depending on the number and structure of nitrogenase operon. The phylogenetic studies of nitrogenase on the phyla level show that most probably nitrogenase first appeared in the methanogenic archaea and later *nif* genes entered bacteria via horizontal gene transfer [57,58]. Nevertheless, the taxa within phylum can have different nitrogenases that can function in a different way depending on the environmental conditions and physiological features [58]. For example, it was demonstrated that *Anabaena variabilis* possesses two types of nitrogenase that function under different environmental conditions and express in different types of cells [59]. In our study, we show that two types of nitrogenase operons are present in the *Rhodovulum* species that probably results in different nitrogen fixation rates, so it is necessary to further investigate this phenomenon.

Unlike *R. tesquicola* and *R. steppense*, both isolated from saline soda lakes, *R. strictum* was isolated from a coastal colored bloom [60]. In spite of the fact that purple nonsulfur bacteria do not predominate in the coastal colored blooms, the members of *Rhodovulum* may represent a significant proportion of the phototrophic bacterial population in such environments [3]. It is probably due to the high nitrogenase activity that gives them the competitive advantage in case of nitrogen removal during the bloom.

3.4. Vitamin Requirements

The vitamin requirements of the investigated bacteria according to the physiological tests are given in Table 2. Biotin requirement corresponds with the fact that biotin biosynthesis pathways are incomplete in the genome of all three bacteria. In contrast to *R. steppense* and *R. strictum*, *R. tesquicola* contains a gene for phosphomethylpyrimidine

synthase which may allow it to start the biosynthesis of thiamine monophosphate. The physiological tests, however, demonstrate that *R. tesquicola* requires additional thiamine. Some of the genes involved into thyamine synthesis are probably not expressed for reasons yet to be discovered. The niacin requirement in *R. steppense* and *R. tesquicola* may be due to the absence of *nadAB* genes which in turn are present in *R. strictum*.

3.5. Osmotic Adaptation

R. tesquicola and *R. steppense* are more halotolerant than *R. strictum* (Table 2). Apart from the complete pathway of ectoin biosynthesis and a partial trehalose synthesis pathway in genomes of the former two bacteria (Figure 3), the ectoine/hydroxyectoine TRAP transporter genes *TeaABC* [61] and glycine betaine uptake system *proVWX* [62] were found. The distribution of the compatible solutes biosynthesis and uptake genes in the investigated moderately halophilic bacteria correlates with the previously described pattern [63].

4. Conclusions

The genomes of *R. tesquicola*, *R. steppense* and *R. strictum* demonstrate significant similarity, except for some features including the vitamin synthesis pathways and the repertoire of the nitrogen fixation genes. The analysis of genome and nitrogen fixation rates shows that the studied *Rhodovulum* members are capable of fixing nitrogen with different rates, and thus may play an important role in nitrogen fixation in high-salt and alkaline conditions. Nevertheless, it becomes evident that the presence of representatives of this group does not necessarily indicate the effective nitrogen fixation in these ecosystems. Moreover, the presence of the "fast" operon does not predict high rates of nitrogen fixation in an organism. Literature data and our results show that the relationship between nitrogen fixation rate and the nitrogenase operon structure is still poorly studied, and thus further studies in this direction appear promising.

References

1. Imhoff, J.F. Genus XIV Rhodovulum. In *Bergey's Manual of Systematic Bacteriology, Volume 2: The Proteobacteria, Part C: The Alpha-, Beta-, Delta-, and Epsilonproteobacteria*; Garrity, G., Brenner, D.J., Krieg, N.R., Staley, J.T., Eds.; Springer: New York, NY, USA, 2005.
2. Hiraishi, A.; Ueda, Y. Intrageneric structure of the genus Rhodobacter: Transfer of Rhodobacter sulfidophilus and related marine species to the genus *Rhodovulum* gen. nov. *Int. J. Syst. Evol. Microbiol.* **1994**, *44*, 15–23. [CrossRef]

3. Hiraishi, A.; Nagao, N.; Yonekawa, C.; Umekage, S.; Kikuchi, Y.; Eki, T.; Hirose, Y. Distribution of phototrophic purple nonsulfur bacteria in massive blooms in coastal and wastewater ditch environments. *Microorganisms* **2020**, *8*, 150. [CrossRef] [PubMed]

4. Imhoff, J.F. True marine and halophilic anoxygenic phototrophic bacteria. *Arch. Microbiol.* **2001**, *176*, 243–254. [CrossRef]

5. Kompantseva, E.I.; Komova, A.V.; Krauzova, V.I.; Kolganova, T.V.; Panteleeva, A.N. Purple nonsulfur bacteria in weakly and moderately mineralized soda lakes of the southern Transbaikal Region and northeastern Mongolia. *Mikrobiologia* **2009**, *78*, 246–253. [CrossRef]

6. Fujimoto, H.; Wakabayashi, M.; Yamashiro, H.; Maeda, I.; Isoda, K.; Kondoh, M.; Kawase, M.; Miyasaka, H.; Yagi, K. Whole-cell arsenite biosensor using photosynthetic bacterium *Rhodovulum sulfidophilum. Appl. Microbiol. Biot.* **2006**, *73*, 332–338. [CrossRef]

7. Cai, J.; Wang, G. Hydrogen production by a marine photosynthetic bacterium, Rhodovulum sulfidophilum P5, isolated from a shrimp pond. *Int. J. Hydrogen. Energy* **2012**, *37*, 15070–15080. [CrossRef]

8. Higuchi-Takeuchi, M.; Morisaki, K.; Toyooka, K.; Numata, K. Synthesis of high-molecular-weight polyhydroxyalkanoates by marine photosynthetic purple bacteria. *PLoS ONE* **2016**, *11*, e0160981. [CrossRef] [PubMed]

9. Gupta, D.; Guzman, M.S.; Rengasamy, K.; Stoica, A.; Singh, R.; Ranaivoarisoa, T.O.; Davenport, E.J.; Bai, W.; McGinley, B.; Meacham, J.M.; et al. Photoferrotrophy and phototrophic extracellular electron uptake is common in the marine anoxygenic phototroph *Rhodovulum sulfidophilum. ISME* **2021**, *15*, 3384–3398. [CrossRef] [PubMed]

10. Melack, J.M.; Kilham, P. Photosynthetic rates of phytoplankton in East African alkaline, saline lakes 1. *Limnol. Oceanogr.* **1974**, *19*, 743–755. [CrossRef]

11. Cloern, J.E.; Cole, B.E.; Oremland, R.S. Autotrophic processes in meromictic Big Soda Lake, Nevada. *Limnol. Oceanogr.* **1983**, *28*, 1049–1061. [CrossRef]

12. Jellison, R.; Miller, L.G.; Melack, J.M.; Dana, G.L. Meromixis in hypersaline Mono Lake, California. 2. Nitrogen fluxes. *Limnol. Oceanogr.* **1993**, *38*, 1020–1039. [CrossRef]

13. Oremland, R.S. Nitrogen fixation dynamics of two diazotrophic communities in Mono Lake, California. *Appl. Environ. Microbiol.* **1990**, *56*, 614–622. [CrossRef]

14. Dubinin, A.V.; Gerasimenko, L.M.; Zavarzin, G.A. Nitrogen-fixation by cyanobacterium Microcoleus chthonoplastes from hypersaline lagoons od Lake Sivash. *Mikrobiologia* **1992**, *61*, 593–597.

15. Pinckney, J.L.; Paerl, H.W. Anoxygenic photosynthesis and nitrogen fixation by a microbial mat community in a Bahamian hypersaline lagoon. *Appl. Environ. Microbiol.* **1997**, *63*, 420–426. [CrossRef]

16. Herbst, D.B. Potential salinity limitations on nitrogen fixation in sediments from Mono Lake, California. *Int. J. Salt Lake Res.* **1998**, *7*, 261–274. [CrossRef]

17. Namsaraev, Z.; Samylina, O.; Sukhacheva, M.; Borisenko, G.; Sorokin, D.Y.; Tourova, T. Effect of salinity on diazotrophic activity and microbial composition of phototrophic communities from Bitter-1 soda lake (Kulunda Steppe, Russia). *Extremophiles* **2018**, *22*, 651–663. [CrossRef]

18. Munson, T.O.; Burris, R.H. Nitrogen fixation by Rhodospirillum rubrum grown in nitrogen-limited continuous culture. *J. Bacteriol.* **1969**, *97*, 1093–1098. [CrossRef]

19. Madigan, M.; Cox, S.S.; Stegeman, R.A. Nitrogen fixation and nitrogenase activities in members of the family Rhodospirillaceae. *J. Bacteriol.* **1984**, *157*, 73–78. [CrossRef]

20. Oda, Y.; Samanta, S.K.; Rey, F.E.; Wu, L.; Liu, X.; Yan, T.; Zhou, J.; Harwood, C.S. Functional genomic analysis of three nitrogenase isozymes in the photosynthetic bacterium *Rhodopseudomonas palustris. J. Bacteriol.* **2005**, *187*, 7784–7794. [CrossRef]

21. Kompantseva, E.I.; Komova, A.V.; Sorokin, D.Y. Communities of anoxygenic phototrophic bacteria in the saline soda lakes of the Kulunda Steppe (Altai Krai). *Mikrobiologia* **2010**, *79*, 89–95. [CrossRef]

22. Kompantseva, E.I.; Komova, A.V.; Novikov, A.A.; Kostrikina, N.A. Rhodovulum tesquicola sp. nov., a haloalkaliphilic purple non-sulfur bacterium from brackish steppe soda lakes. *Int. J. Syst. Evol. Microbiol.* **2012**, *62 Pt 12*, 2962–2966. [CrossRef]

23. Hardy, R.W.; Holsten, R.D.; Jackson, E.K.; Burns, R. The acetylene-ethylene assay for N2 fixation: Laboratory and field evaluation. *Plant Physiol.* **1968**, *43*, 1185–1207. [CrossRef]

24. Korzhenkov, A. ZGA: A flexible pipeline for read processing, de novo assembly and annotation of prokaryotic genomes. *bioRxiv* **2021**. [CrossRef]

25. BBMap—Bushnell B. Available online: Sourceforge.net/projects/bbmap/ (accessed on 12 April 2022).

26. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [CrossRef] [PubMed]

27. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, A.K.; et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **2014**, *9*, e112963. [CrossRef]

28. Tanizawa, Y.; Fujisawa, T.; Nakamura, Y. DFAST: A flexible prokaryotic genome annotation pipeline for faster genome publication. *Bioinformatics* **2018**, *34*, 1037–1039. [CrossRef] [PubMed]

29. Tatusova, T.; DiCuccio, M.; Badretdin, A.; Chetvernin, V.; Nawrocki, E.P.; Zaslavsky, L.; Lomsadze, A.; Pruitt, K.D.; Borodovsky, M.; Ostell, J. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **2016**, *44*, 6614–6624. [CrossRef] [PubMed]

30. Lomsadze, A.; Gemayel, K.; Tang, S.; Borodovsky, M. Modeling leaderless transcription and atypical genes results in more accurate gene prediction in prokaryotes. *Genome Res.* **2018**, *29*, 1079–1089. [CrossRef]

31. Aramaki, T.; Blanc-Mathieu, R.; Endo, H.; Ohkubo, K.; Kanehisa, M.; Goto, S.; Ogata, H. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **2019**, *36*, 2251–2252. [CrossRef]
32. Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y.; Morishima, K. KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **2017**, *45*, D353–D361. [CrossRef]
33. Eren, A.M.; Kiefl, E.; Shaiber, A.; Veseli, I.; Miller, S.E.; Schechter, M.S.; Fink, I.; Pan, J.N.; Yousef, M.; Fogarty, E.C.; et al. Community-led, integrated, reproducible multi-omics with anvi'o. *Nat. Microbiol.* **2021**, *6*, 3–6. [CrossRef] [PubMed]
34. Parks, D.H.; Imelfort, M.; Skennerton, C.T.; Hugenholtz, P.; Tyson, G.W. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **2015**, *25*, 1043–1055. [CrossRef]
35. Jain, C.; Rodriguez-R, L.M.; Phillippy, A.M.; Konstantinidis, K.T.; Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **2018**, *9*, 5114. [CrossRef]
36. Lechner, M.; Findeiß, S.; Steiner, L.; Marz, M.; Stadler, P.F.; Prohaska, S.J. Proteinortho: Detection of (Co-)orthologs in large-scale analysis. *BMC Bioinform.* **2011**, *12*, 124. [CrossRef] [PubMed]
37. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 1–9. [CrossRef]
38. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **2012**, *28*, 3150–3152. [CrossRef] [PubMed]
39. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef]
40. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS ONE* **2010**, *5*, e9490. [CrossRef]
41. Letunic, I.; Bork, P. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **2019**, *47*, W256–W259. [CrossRef]
42. Tang, K.H.; Tang, Y.J.; Blankenship, R.E. Carbon metabolic pathways in phototrophic bacteria and their broader evolutionary implications. *Front. Microbiol.* **2011**, *2*, 165. [CrossRef] [PubMed]
43. Petushkova, E.; Mayorova, E.; Tsygankov, A. TCA Cycle Replenishing Pathways in Photosynthetic Purple Non-Sulfur Bacteria Growing with Acetate. *Life* **2021**, *11*, 711. [CrossRef]
44. Eprintsev, A.T.; Falaleeva, M.I.; Lyashchenko, M.S.; Toropygin, I.Y.; Igamberdiev, A.U. Oligomeric forms of bacterial malate dehydrogenase: A study of the enzyme from the phototrophic non-sulfur bacterium *Rhodovulum steppense* A-20s. *Biosci. Biotechnol. Biochem.* **2018**, *82*, 81–89. [CrossRef]
45. Priefert, H.O.R.S.T.; Steinbüchel, A. Identification and molecular characterization of the acetyl coenzyme a synthetase gene (acoE) of Alcaligenes eutrophus. *J. Bacteriol.* **1992**, *174*, 6590–6599. [CrossRef]
46. Crosby, H.A.; Rank, K.C.; Rayment, I.; Escalante-Semerena, J.C. Structure-guided expansion of the substrate range of methyl-malonyl coenzyme a synthetase (MatB) of *Rhodopseudomonas palustris*. *Appl. Environ. Microbiol.* **2012**, *78*, 6619–6629. [CrossRef]
47. Wang, Z.; Wen, Q.; Harwood, C.S.; Liang, B.; Yang, J. A disjointed pathway for malonate degradation by *Rhodopseudomonas palustris*. *Appl. Environ. Microbiol.* **2020**, *86*, e00631-20. [CrossRef]
48. Samylina, O.S.; Namsaraev, Z.B.; Grouzdev, D.S.; Slobodova, N.V.; Zelenev, V.V.; Borisenko, G.V.; Sorokin, D.Y. The patterns of nitrogen fixation in haloalkaliphilic phototrophic communities of Kulunda Steppe soda lakes (Altai, Russia). *FEMS Microbiol. Ecol.* **2019**, *95*, fiz174. [CrossRef]
49. Tourova, T.P.; Slobodova, N.V.; Bumazhkin, B.K.; Sukhacheva, M.V.; Sorokin, D.Y. Diversity of diazotrophs in the sediments of saline and soda lakes analyzed with the use of the nifH gene as a molecular marker. *Mikrobiologia* **2014**, *83*, 634–647. [CrossRef]
50. Roberts, G.P.; MacNeil, T.; MacNeil, D.; Brill, W.J. Regulation and characterization of protein products coded by the nif (nitrogen fixation) genes of Klebsiella pneumoniae. *J. Bacteriol.* **1978**, *136*, 267–279. [CrossRef]
51. Rangaraj, P.; Ryle, M.J.; Lanzilotta, W.N.; Ludden, P.W.; Shah, V.K. In vitro biosynthesis of iron-molybdenum cofactor and maturation of the nif-encoded apodinitrogenase: Effect of substitution for NifH with site-specifically altered forms of NifH. *J. Biol. Chem.* **1999**, *274*, 19778–19784. [CrossRef]
52. Ribbe, M.W.; Bursey, E.H.; Burgess, B.K. Identification of an Fe protein residue (Glu146) of Azotobacter vinelandii nitrogenase that is specifically involved in FeMo cofactor insertion. *J. Biol. Chem.* **2000**, *275*, 17631–17638. [CrossRef]
53. Avtges, P.; Kranz, R.G.; Haselkorn, R. Isolation and organization of genes for nitrogen fixation in *Rhodopseudomonas capsulata*. *Mol. Gen. Genet.* **1985**, *201*, 363–369. [CrossRef] [PubMed]
54. Schüddekopf, K.; Hennecke, S.; Liese, U.; Kutsche, M.; Klipp, W. Characterization of anf genes specific for the alternative nitrogenase and identification of nif genes required for both nitrogenases in *Rhodobacter capsulatus*. *Mol. Microbiol.* **1993**, *8*, 673–684. [CrossRef]
55. Mus, F.; Alleman, A.B.; Pence, N.; Seefeldt, L.C.; Peters, J.W. Exploring the alternatives of biological nitrogen fixation. *Metallomics* **2018**, *10*, 523–538. [CrossRef]
56. Scolnik, P.A.; Haselkorn, R. Activation of extra copies of genes coding for nitrogenase in *Rhodopseudomonas capsulata*. *Nature* **1984**, *307*, 289–292. [CrossRef] [PubMed]
57. Raymond, J.; Siefert, J.L.; Staples, C.R.; Blankenship, R.E. The natural history of nitrogen fixation. *Mol. Biol. Evol.* **2004**, *21*, 541–554. [CrossRef] [PubMed]

58. Koirala, A.; Brözel, V.S. Phylogeny of Nitrogenase Structural and Assembly Components Reveals New Insights into the Origin and Distribution of Nitrogen Fixation across Bacteria and Archaea. *Microorganisms* **2021**, *9*, 1662. [CrossRef] [PubMed]

59. Thiel, T.; Pratte, B.S. Regulation of three nitrogenase gene clusters in the cyanobacterium *Anabaena variabilis* ATCC 29413. *Life* **2014**, *4*, 944–967. [CrossRef] [PubMed]

60. Hiraishi, A.; Ueda, Y. Isolation and characterization of Rhodovulum strictum sp. nov. and some other purple nonsulfur bacteria from colored blooms in tidal and seawater pools. *Int. J. Syst. Evol. Microbiol.* **1995**, *45*, 319–326.

61. Grammann, K.; Volke, A.; Kunte, H.J. New type of osmoregulated solute transporter identified in halophilic members of the bacteria domain: TRAP transporter TeaABC mediates uptake of ectoine and hydroxyectoine in *Halomonas elongata* DSM 2581T. *J. Bacteriol.* **2002**, *184*, 3078–3085. [CrossRef] [PubMed]

62. Tsuzuki, M.; Moskvin, O.V.; Kuribayashi, M.; Sato, K.; Retamal, S.; Abo, M.; Zeilstra-Ryalls, J.; Gomelsky, M. Salt stress-induced changes in the transcriptome, compatible solutes, and membrane lipids in the facultatively phototrophic bacterium *Rhodobacter sphaeroides*. *Appl. Environ. Microbiol.* **2011**, *77*, 7551–7559. [CrossRef] [PubMed]

63. Imhoff, J.F.; Rahn, T.; Künzel, S.; Keller, A.; Neulinger, S.C. Osmotic adaptation and compatible solute biosynthesis of phototrophic bacteria as revealed from genome analyses. *Microorganisms* **2020**, *9*, 46. [CrossRef] [PubMed]

Article

Salt- and pH-Dependent Thermal Stability of Photocomplexes from Extremophilic Bacteriochlorophyll *b*-Containing *Halorhodospira* Species

Yukihiro Kimura [1,*]**, Kazuna Nakata** [1]**, Shingo Nojima** [1]**, Shinji Takenaka** [1]**, Michael T. Madigan** [2]
and Zheng-Yu Wang-Otomo [3,*]

1   Department of Agrobioscience, Graduate School of Agriculture, Kobe University, Nada, Kobe 657-8501, Japan;
    kn.tea0818@gmail.com (K.N.); nojimashingo@gmail.com (S.N.); stakenaka@people.kobe-u.ac.jp (S.T.)
2   Department of Microbiology, Southern Illinois University, Carbondale, IL 62901, USA;
    madigan@micro.siu.edu
3   Faculty of Science, Ibaraki University, Mito 310-8512, Japan
*   Correspondence: ykimura@people.kobe-u.ac.jp (Y.K.); wang@ml.ibaraki.ac.jp (Z.-Y.W.-O.);
    Tel.: +81-78-803-5819: (Y.K.); +81-29-228-8352 (Z.-Y.W.-O.)

Abstract: *Halorhodospira* (*Hlr.*) species are the most halophilic and alkaliphilic of all purple bacteria. *Hlr. halochloris* exhibits the lowest LH1 Q_y transition energy among phototrophic organisms and is the only known triply extremophilic anoxygenic phototroph, displaying a thermophilic, halophilic, and alkaliphilic phenotype. Recently, we reported that electrostatic charges are responsible for the unusual spectroscopic properties of the *Hlr. halochloris* LH1 complex. In the present work, we examined the effects of salt and pH on the spectroscopic properties and thermal stability of LH1-RCs from *Hlr. halochloris* compared with its mesophilic counterpart, *Hlr. abdelmalekii*. Experiments in which the photocomplexes were subjected to different levels of salt or variable pH revealed that the thermal stability of LH1-RCs from both species was largely retained in the presence of high salt concentrations and/or at alkaline pH but was markedly reduced by lowering the salt concentration and/or pH. Based on the amino acid sequences of LH1 polypeptides and their composition of acidic/basic residues and the Hofmeister series for cation/anion species, we discuss the importance of electrostatic charge in stabilizing the *Hlr. halochloris* LH1-RC complex to allow it to perform photosynthesis in its warm, hypersaline, and alkaline habitat.

Keywords: purple phototrophic bacteria; extremophile; *Halorhodospira halochloris*; *Halorhodospira abdelmalekii*; light-harvesting 1 reaction center; bacteriochlorophyll *b*; thermal stability; salt- and pH-dependence

1. Introduction

Purple phototrophic bacteria have evolved to adapt to various conditions, and some species thrive at extremes of temperature, salinity, and pH [1,2]. *Halorhodospira* (*Hlr.*) *halochloris* (formerly *Ectothiorhodospira halochloris*) is a purple sulfur bacterium (*Gammaproteobacteria*) that was isolated from warm and extremely saline and alkaline Egyptian soda lakes [3]. *Hlr. halochloris* is the only known triply extremophilic anoxygenic phototroph, displaying a thermophilic, halophilic, and alkaliphilic phenotype; the organism grows optimally at 47–50 °C, pH 8–9, and 14–27% (*w/v*) salinity [4]. *Hlr. abdelmalekii* (formerly *Ectothiorhodospira abdelmalekii*), isolated from alkaline soda lakes of the Wadi Natrun in Egypt, is a mesophilic counterpart of *Hlr. halochloris*, growing optimally at 30–40 °C, pH 8.5, and 14–16% (*w/v*) salinity [5]. Both of these phototrophs assemble their light-harvesting machinery with bacteriochlorophyll (BChl) *b* [2,3,5–12]; because of this, they are part of a distinct minority (< 5%) of known purple bacteria. The light-harvesting 1 (LH1) complexes of BChl *b*-containing purple bacteria exhibit Q_y absorption bands at wavelengths beyond

1000 nm whose energy is much lower than that absorbed by BChl *a*-containing purple bacteria, whose LH1 complexes typically absorb near 880 nm [13].

Among purple phototrophic bacteria, several species are thermophilic. For Bchl *a*-containing species, these include *Thermochromatium* (*Tch.*) *tepidum* from the Mammoth Hot Springs of Yellowstone National Park, USA [14], *Allochromatium* (*Alc.*) *tepidum* from a sulfidic hot spring in New Zealand [15,16], *Caldichromatium* (*Ccm.*) *japonicum* from the Nakabusa hot springs in the Northern Alps of Japan [17], and *Hlr. halophila* from Summer Lake, a seasonally dry hypersaline lake in Oregon (USA) [18]. For BChl *b*-containing species, these include *Blastochloris* (*Blc.*) *tepida*, isolated from a hot spring in New Mexico (USA) [9,11], and *Hlr. halochloris* [3]. Several factors have been linked to the thermal stability of the LH1-RC complexes of these unusual phototrophs. The first of these is Ca^{2+}. Thermal stability of the *Tch. tepidum* LH1-RC is achieved by the binding of Ca^{2+} to LH1 polypeptides in a stoichiometric ratio of 1:1 (16 Ca^{2+} to 16 αβ-polypeptides) [19,20], as revealed by a high-resolution X-ray crystallographic structural analysis of the complex [21]. Ca^{2+}-dependent thermal stabilization was also reported for the *Alc. tepidum* LH1-RC [22]; however, in this case, the stoichiometric ratio of Ca^{2+} to αβ-polypeptide was lower (0.375) than that of *Tch. tepidum* [22] and has been confirmed to be 6 Ca^{2+} to 16 αβ-polypeptides in the cryo-EM structure of the *Alc. tepidum* LH1-RC [23]. In this species, each LH1 ring contains multiple pairs of α- and β-polypeptides, but only six α1-polypeptides bind Ca^{2+} with β1- or β3-polypeptides to form the Ca^{2+}-binding sites that reinforce the complex. As for the other thermophiles that contain Bchl *a* (*Hlr. halophila* and *Ccm. japonicum*), the factors responsible for LH1-RC thermal stability remain to be determined.

An alternative factor conferring LH1-RC thermal stability is the presence of carotenoids with extended conjugation. The purple nonsulfur bacterium *Blc. tepida* grows optimally at 42 °C and up to 47 °C [9,11]. However, in contrast to *Tch. tepidum*, Ca^{2+} is neither present in, nor has any effect on, the thermostability of *Blc. tepida* LH1-RC [24]. In comparative studies of *Blc. tepida* with its mesophilic counterpart, *Blc. viridis*, it was discovered that lycopene family carotenoids with elongated conjugations are produced by the thermophile and that these strengthen hydrophobic interactions with LH1-RC proteins; as a result, the *Blc. tepida* complexes remain functional at higher temperatures than is possible in *Blc. viridis*.

In a previous study, we reported salt- and pH-sensitive interconvertible LH1 Q_y band shifts of the purified LH1-RC complex from *Hlr. halochloris* [25] and demonstrated that the native spectral properties of the LH1-RC were maintained in the presence of high concentrations of salts and/or at alkaline pH (>7). During these experiments, it was observed that lowering the salt concentration and/or the pH induced Q_y blue-shifts and decreased the band intensities, suggesting that the thermal stability of the *Hlr. halochloris* LH1-RC complex might also be controlled by salt or pH. Therefore, in the present study, we systematically examined the effects of salts and pH on the thermal stability of the LH1-RC complex from *Hlr. halochloris* and compared them with those from its mesophilic counterpart, *Hlr. abdelmalekii*. Based on our experimental results, along with comparative sequence analyses of relevant LH1-RC proteins, the molecular strategies employed by *Hlr. halochloris* to maintain functional LH1-RC complexes are more fully revealed and are discussed in relation to strategies employed by other purple bacteria.

2. Materials and Methods

2.1. Growth Conditions and Preparation of LH1-RC Complexes

Hlr. halochloris strain DSM 1059 and *Hlr. abdelmalekii* strain DSM 2110 were grown photosynthetically (anoxic/light) at 43 °C and 28 °C, respectively, under illumination of a 60 W incandescent lamp for 7–10 days in the medium of Imhoff and Trüper [3] with slight modifications [25]. The harvested cells suspended in 20 mM Tris-HCl buffer (pH 8.5) were disrupted by sonication (Sonopuls HD3200, Bandelin, Berlin, Germany) and ultracentrifuged at 195,000× *g* for 60 min (L60, Beckman) to obtain photosynthetic membranes (chromatophores). The latter were solubilized with 1.0% (*w/v*) *n*-dodecyl β-ᴅ-maltopyranoside (DDM) at ambient temperature and ultracentrifuged at 195,000× *g*

for 60 min to extract crude LH1-RCs. The supernatant was loaded onto an anion exchange column (Toyopearl DEAE-650s, TOSOH) equilibrated with 20 mM Tris-HCl (pH 8.5) and 0.08% (*w/v*) DDM at 4 °C. LH1-RC complexes were eluted with a linear gradient of NaCl concentration (100–400 mM) and the fractions with $A_{LH1\ Qy}/A_{280}$ ratios > 0.8 were collected. Salt concentrations were adjusted by extensive washing of the purified *Hlr. halochloris* and *Hlr. abdelmalekii* LH1-RCs with a buffer containing 20 mM Tris-HCl (pH 8), 0.08% DDM and an indicated concentration of salts (0–1 M) by ultrafiltration (Amicon Ultra-15, Millipore).

2.2. Evaluation of Thermal Stability

Absorption spectra were recorded on a UV-mini 1200 (Shimadzu) and V-730Bio spectrophotometer (JASCO). Thermal degradation of LH1-RC complexes was monitored as reductions in LH1 Q_y band intensities after incubation at 45 °C for 0–60 min. For all experiments, sample concentrations were normalized with respect to the Q_x band at 600 nm because this band was shown to be unaffected by salt concentration or pH.

3. Results

3.1. NaCl and pH Effects on Spectroscopic Properties of LH1-RC Complexes from Halorhodospira Species

Figure 1A,B show the absorption spectra of desalted LH1-RCs from *Hlr. halochloris* and *Hlr. abdelmalekii* at pH 8 after supplementation with 0–1 M NaCl. For *Hlr. halochloris* (Figure 1A), the desalted LH1 Q_y band appeared at 956 nm in the absence of NaCl and was red-shifted as the NaCl concentration increased, reaching 1014 nm at 1 M NaCl. The salt-dependent and interconvertible spectral changes of the *Hlr. halochloris* LH1-RCs were largely comparable to those reported previously [25] but were significantly different from those observed for the mesophilic *Hlr. abdelmalekii* LH1-RCs, as shown in Figure 1B. At pH 8, the desalted LH1 Q_y band of this species was not completely blue-shifted to ~955 nm. However, a salt-dependent and interconvertible spectral change was also confirmed for *Hlr. abdelmalekii* LH1-RCs when the buffer pH was decreased to 7 (Figure 1C). The LH1 Q_y band appeared at 955 nm upon desalting and was red-shifted as the NaCl concentration increased, reaching 1002 nm at 1 M NaCl. These results suggest that although there may be a common mechanism responsible for the salt-dependent LH1 spectral shifts of *Hlr. halochloris* and *Hlr. abdelmalekii*, the salt requirements for maintaining the native spectral properties of the LH1-RCs differ in the two *Halorhodospira* species at a given pH.

Figure 1D,E exhibit the absorption spectra of the purified LH1-RCs from *Hlr. halochloris* and *Hlr. abdelmalekii*, respectively, at pH 5–8 in the presence of 200 mM NaCl. At pH 8, the LH1 Q_y band appeared at 1012 nm for *Hlr. halochloris* and 1002 nm for *Hlr. abdelmalekii*, and both Q_y bands were blue-shifted to 952 nm at lower pH. The pH-dependent spectral changes were interconvertible, not only for *Hlr. halochloris* (Figure 1D) [25,26] but also for *Hlr. abdelmalekii* (Figure 1E). In addition, their pH requirements were nearly identical, as can be seen in plots of the ratio of two LH1 Q_y band intensities (A_{1012}/A_{952} for *Hlr. halochloris* and A_{1002}/A_{952} for *Hlr. abdelmalekii*) at various pHs (Figure 1F). The salt- and pH-dependent spectral changes of these *Halorhodospira* species were nonspecific to coexisting ion species and their valences [25], and distinct from the Ca^{2+}-specific spectral changes of several BChl a-containing purple sulfur bacteria, including *Tch. tepidum* [27], *Alc. tepidum* [22], and *Thiorhodovibrio* strain 970 [28]. These results also support our previous conclusion that electrostatic charges control the LH1 Q_y transition energies of LH1-RC complexes from *Hlr. halochloris* [25], as they likely do in *Hlr. abdelmalekii* as well.

3.2. NaCl and pH Effects on the Thermal Stability of LH1-RC Complexes from Halorhodospira Species

To better understand the relation between electrostatic charge and protein stability, the effects of salt and pH on the thermal stability of LH1-RCs from *Hlr. halochloris* and *Hlr. abdelmalekii* were examined. Figure 2A shows the rate of thermal degradation of the *Hlr. halochloris* LH1-RC upon incubation at 45 °C and pH 8 for 0–60 min in the presence of 1 M NaCl. Over time, the LH1 Q_y band intensity at 1014 nm decreased slightly and a

faint band increased at 680 nm, assignable to oxidized BChl *b* decomposed from the LH1 complexes [26]. However, in the absence of NaCl, the LH1 Q_y band (blue-shifted to 952 nm, Figure 1B) steadily decreased to near zero after 1 h while concomitantly, the oxidized BChl b band at 680 nm steadily increased (Figure 2B). Similar experiments were performed at pH 8 with different NaCl concentrations, and their relative Q_y band intensities were plotted against incubation time in Figure 2C. As can be seen, the thermal stability of *Hlr. halochloris* LH1-RCs was largely retained in the presence of 1 M NaCl but markedly reduced as NaCl concentrations were decreased, indicating that the photocomplexes of this species require salt for thermal stability.
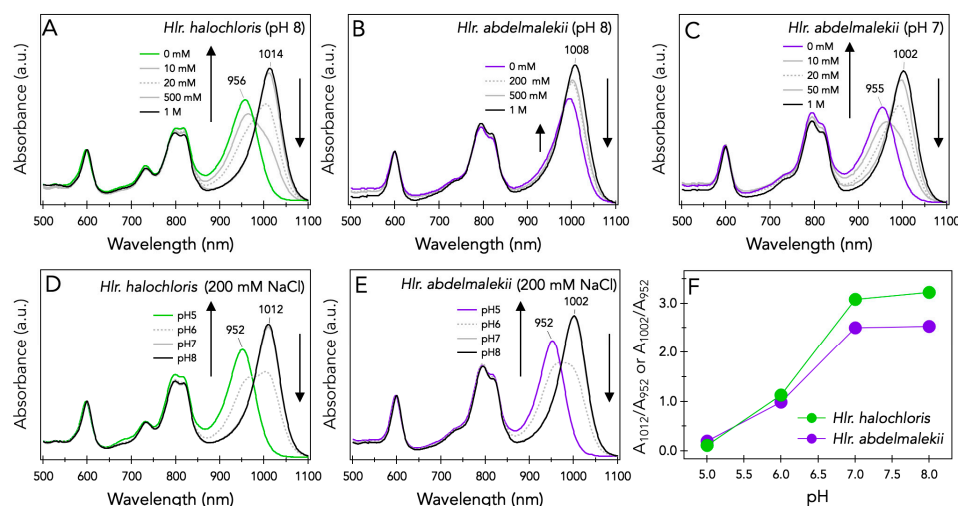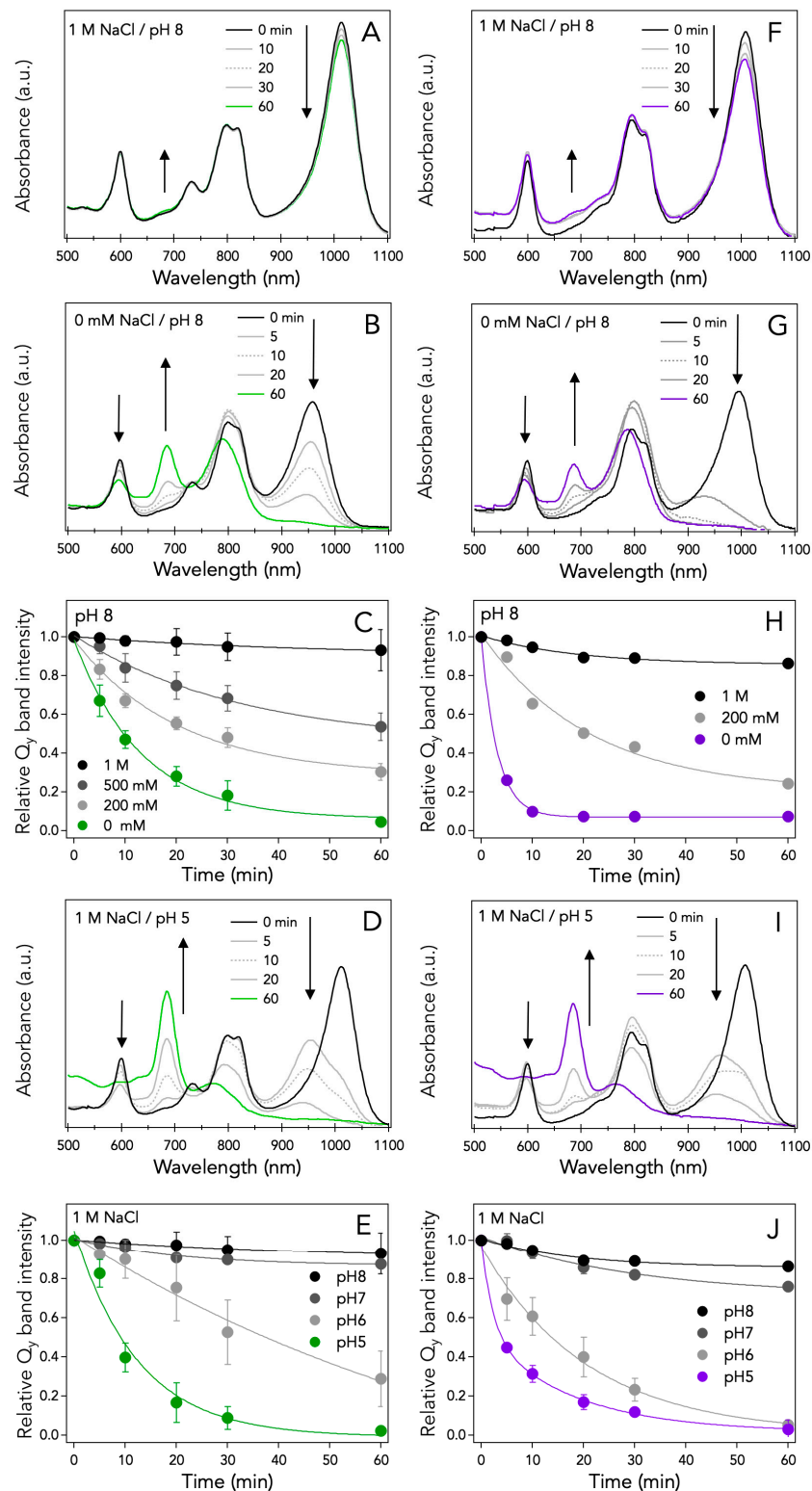

Figure 1. Absorption spectra of the LH1-RC complexes from *Hlr. halochloris* (**A**,**D**) and *Hlr. abdelmalekii* (**B**,**C**,**E**) at pH 8.5 in the presence of indicated concentration of NaCl at pH 8 (**A**,**B**) or pH 7 (**C**), and those at different pHs in the presence of 200 mM of NaCl (**D**,**E**). Panel (**F**) shows plots of the ratio of two LH1 Q_y band intensities (A_{1012}/A_{952} for *Hlr. halochloris* and A_{1002}/A_{952} for *Hlr. abdelmalekii*) at various pHs.

In additional experiments, the effects of pH on the thermal stability of *Hlr. halochloris* LH1-RCs suspended in 1 M NaCl were examined. Thermal stability was largely retained at pH 8 (Figure 2A) but was much reduced at pH 5 (Figure 2D); in the latter, the LH1 Q_y band was virtually eliminated and the oxidized BChl *b* band at 680 nm reached its maximum. Moreover, thermal decomposition of *Hlr. halochloris* LH1-RCs at acidic pH (Figure 2D) was more pronounced than that at reduced NaCl concentrations (Figure 2B), as deduced from the 680 nm band intensity of oxidized BChl *b*. Similar experiments were performed in the presence of 1 M NaCl at different pH values, and relative Q_y band intensities are plotted against incubation time in Figure 2E. The data show that thermal stability is largely retained at pH 7 but markedly reduced as the pH drops below neutrality, demonstrating a pH-dependent thermal stability of *Hlr. halochloris* LH1-RCs. Thus, as one might expect considering the geochemical conditions in the habitat of *Hlr. halochloris* [3], the thermal stability of this organism's photocomplexes is greatest at both elevated pH and salt concentration.

These salt- and pH-dependent effects on the thermal stabilization of purified *Hlr. halochloris* LH1-RCs were confirmed using photosynthetic membranes (chromatophores) in which LH1-RCs remain embedded in the lipid bilayer (Figure S1). The relative Q_y band intensities of *Hlr. halochloris* chromatophores were evaluated after incubation at the indicated pH and temperature for 60 min in the presence (Figure S1A) or absence (Figure S1B) of 4 M NaCl. The relative Q_y band intensities were almost completely diminished in the absence of NaCl (Figure S1B) or at low pH in the presence of 4 M NaCl (Figure S1A). These results suggest that the salt- and pH-dependent thermal stability of photocomplexes in this species is not closely associated with the transmembrane regions, but instead, with the surfaces of the LH1s and/or RCs where solvent molecules are readily accessible.

Figure 2. Absorption spectra of the LH1-RCs from *Hlr. halochloris* (**A**,**B**) and *Hlr. abdelmalekii* (**F**,**G**) at pH 8 in the presence of 1 M NaCl (**A**,**F**) or in the absence of NaCl (**B**,**G**) after incubation at 45 °C for indicated time. The time profiles of the relative Q_y band intensities for the *Hlr. halochloris* (**C**) and *Hlr. abdelmalekii* (**H**) LH1-RCs, respectively, incubated at 45 °C and pH 8 in the presence of indicated concentration of NaCl. Absorption spectra of the LH1-RCs from *Hlr. halochloris* (**D**) and *Hlr. abdelmalekii* (**I**) at pH 5 in the presence of 1 M NaCl. The relative Q_y band intensities of the *Hlr. halochloris* (**E**) and *Hlr. abdelmalekii* (**J**) LH1-RCs after incubation at 45 °C for 60 min at indicated pH in the presence of 1 M NaCl.

Similar studies of LH1-RC thermal degradation were performed on the mesophilic *Hlr. abdelmalekii* to compare it to that of *Hlr. halochloris*. Figure 2F shows the thermal degradation of the *Hlr. abdelmalekii* LH1-RC upon incubation at 45 °C and pH 8 in the presence of 1 M NaCl. Similar to *Hlr. halochloris*, the LH1 Q_y band intensity at 1008 nm was slightly decreased along with a faint increase of the oxidized BChl *b* band at 680 nm. By contrast to *Hlr. halochloris*, thermal decomposition of *Hlr. abdelmalekii* LH1-RCs was rapid in the absence of NaCl at pH 8 (Figure 2G); even with salt present at pH 5, degradation was rapid, as in *Hlr. halochloris* (Figure 2D,I). The degrees of thermal stability of *Hlr. halochloris* and *Hlr. abdelmalekii* LH1-RCs seemed comparable at pH 8 if salt was present (Figure 2A,F). However, significant differences in thermal stability were apparent as salt levels decreased at pH 8 (Figure 2H) or when pH decreased in the presence of salt (Figure 2J); these differences are clearly seen in the life times obtained with a first-order kinetics analysis of each decay (Tables S1 and S2). These results indicate an overall greater thermal stability for *Hlr. halochloris* LH1-RCs compared with *Hlr. abdelmalekii* LH1-RCs. These data therefore indicate that not only are the spectral properties of LH1 complexes in both *Halorhodospira* species controlled by electrostatic charges, but also their thermal stability.

3.3. Effect of Other Salts on Thermal Stability of LH1-RCs from Hlr. halochloris

To further probe how electrostatic charge influences the thermal stability of *Hlr. halochloris* LH1-RC complexes, we tested the capacity of different sodium (Figure 3A) or chloride (Figure 3B) salts to substitute for NaCl. Upon incubation at 45 °C for 60 min in the presence of 500 mM of each sodium salt (Figure 3A), the LH1 Q_y band intensity of *Hlr. halochloris* was largely retained by Na_2SO_4, only moderately affected by NaCl and CH_3COONa, and was reduced by approximately half with $NaHCO_3$ and $NaNO_3$. Since sodium ions are present in each of these salts, the effect of each on thermal stability is attributed to anions in the order of $SO_4^{2-} > Cl^- \approx CH_3COO^- > HCO_3^- > NO_3^-$. This order is in general agreement with the Hofmeister series—a classification of ions in order of their ability to solubilize proteins or stabilize secondary and tertiary structures of proteins [29,30]. A standard Hofmeister anion series is $SO_4^{2-} > HPO_4^{2-} > CH_3COO^- > Cl^- > NO_3^- > ClO_3^- > I^- > ClO_4^- > SCN^-$ [31], in which the most potent stabilizing anions are located on the left side of each pair. The Hofmeister series also orders a standard cation series as $NH_4^+ > K^+ > Na^+ > Li^+ > Mg^{2+} > Ca^{2+}$ [31], in which the most effective stabilizing cations are also positioned on the left side of each pair. Curiously, however, the Hofmeister cation series is in opposition to the results obtained with different chloride salts; the ability of cations to stabilize the *Hlr. halochloris* LH1-RCs was in the order of $Ca^{2+} > Na^+ > Mg^{2+} > Sr^{2+} > K^+ >> NH_4^+$ (Figure 3B).
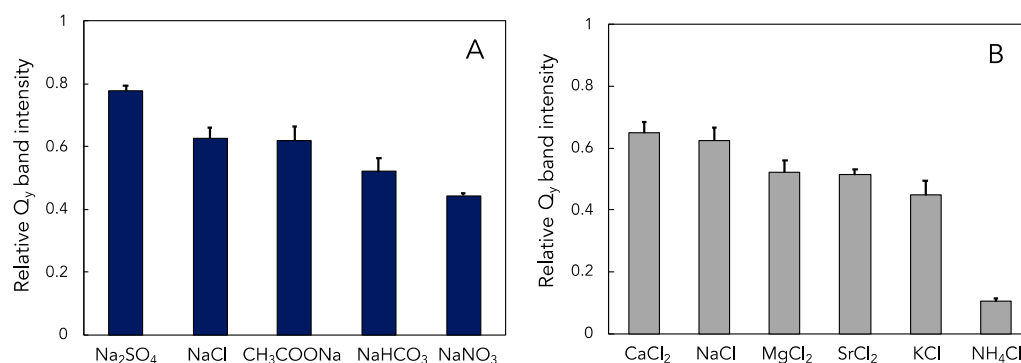
Figure 3. The relative Q_y band intensities of the *Hlr. halochloris* LH1-RCs in the presence of 500 mM of various sodium salts (**A**) or chlorides (**B**) after incubation at 45 °C and pH 8 for 60 min.

4. Discussion

4.1. Effects of Salt Concentration and pH on the Spectral Properties of the LH1-RC Complexes from Halorhodospira Species

In a previous study, we proposed that two factors may responsible for the salt- and pH-dependent spectral changes of the *Hlr. halochloris* LH1-RCs [25]. The first is the axial ligand for Mg^{2+} of BChl *b* bound to LH1 β-polypeptide. Figure 4 compares the amino acid sequences of the LH1 α- or β-polypeptides from several purple bacteria including *Hlr. halochloris* and *Hlr. abdelmalekii.* The conserved His^0 residues in LH1 α-polypeptides are a plausible binding site for α-BChl molecules, and this is also true for most LH1 β-polypeptides other than PufB1 and PufB2 of *Hlr. halochloris*, in which Asn is substituted for the β-BChl binding His^0 [32–34]. Considering that PufB1/PufB2 and PufB3 were comparably expressed, it was hypothesized that the amide oxygen of the Asn side chain modulates its LH1 Q_y transition energy by serving as an axial ligand for the *Hlr. halochloris* LH1 β-BChl *b* [25]. However, this hypothesis does not hold for *Hlr. abdelmalekii*, because such Asn residues are absent from its LH1 β-polypeptide (NCBI Reference Sequence: WP_200191743.1), a polypeptide almost identical to PufB3 of *Hlr. halochloris* (Figure 4).

```
LH1 α-polypeptide
                                            0       +10      +20
                                            |        |        |            %(acidic, basic)
 Hlr. halochloris pufA1              ···HFACISSERFNWLEGAPAAEYYMDEDPGIWKRTSYDG   (12.3,  9.2)
 Hlr. halochloris pufA2              ···HFACISSERFNWLEGAPAAEYYMDEDPGIWKRTSYDG   (12.3,  9.2)
 Hlr. halochloris pufA3              ···HVILMTTERFNWLQGAPAAEYYSEVVEDGAALSPRLV   ( 9.2,  9.2)
 Hlr. abdelmalekii (WP_200194663.1)  ···HFACISSERFNWLEGAPAAEYYMDEDPGIWARTSYDG   (12.3,  7.7)
 Hlr. abdelmalekii (WP_200191741.1)  ···HVILMHTERFNWLQGAPAAEYYSEVIEDSTALSPRTV   ( 9.2,  9.2)
 Blastochloris tepida               ···HFGLLSTNRLNWWEFQRGLP-AA                  ( 5.2, 10.3)
 Blastochloris viridis              ···HFGLLSTDRLNWWEFQRGLPKAA                  ( 6.8, 10.2)
 Rhodobacter sphaeroides            ···HLILLSTPSYNWLEISAAKYNRVAVAE             ( 5.2, 10.3)
 Roseospirillum paruvum 930I         ···HFLCLGHDRFNWLEGNPAATKAAAAAVTMPVNPVA     ( 4.5,  6.0)
 Thermochromatium tepidum            ···HMIVLST-DLNWLDDNIPVSYQALGKK             ( 6.6,  8.2)
 Thirhodovibrio strain 970 pufA1     ···HMVVLSS-DLNWHDDQIPVGYQLQGETLPVNLEMKAAQ  ( 8.8,  1.8)


LH1 β-polypeptide
                                            0       +10      +20
                                            |        |        |            %(acidic, basic)
 Hlr. halochloris pufB1              ···NALAYAWRPFHQGRAGNRLEDHAPDYVRS           (14.3,  7.9)
 Hlr. halochloris pufB2              ···NALAYAWRPFHQGRAGNRLEDHAPDYVRS           (14.3,  7.9)
 Hlr. halochloris pufB3              ···HLLVWAWRPFWPPEQGYRLEDFAPEEIRT           (16.7,  7.6)
 Hlr. abdelmalekii (WP_200191743.1)  ···HLLVWAWRPFWPPEQGYRLEDFAPEEVRT···       (16.7,  7.6)
 Blastochloris tepida               ···HYLVWTARPWIAPIPKGWV                      ( 9.1,  7.3)
 Blastochloris viridis              ···HYLVWTARPWIAPIPKGWV                      ( 9.1,  7.3)
 Rhodobacter sphaeroides            ···HLAVYIWRPWF                              (10.2,  4.1)
 Roseospirillum paruvum 930I         ···HVLAWMWRPFWPGAEGWVMDTAQNLTFLA           ( 8.8,  2.9)
 Thermochromatium tepidum            ···HLLAWLYRPWL                             (10.6,  6.4)
 Thirhodovibrio strain 970pufB1      ···HILSWLYRPWL                             ( 8.3,  5.6)
```

Figure 4. Amino acid sequences of the LH1 α- and β-polypeptides from several purple phototrophic bacteria. The sequences are aligned relative to the His residues (His^0, green) coordinating BChl molecules. Acidic (Asp, Glu) and basic (Arg, Lys) residues are highlighted with magenta and cyan, respectively. Unique Cys residues in the vicinity of His^0 are colored with orange. The numbers of right side represent percentages of acidic and basic residues in each polypeptide.

A second factor that may affect salt- and pH-dependent spectral changes in *Hlr. halochloris* LH1-RCs is LH1 α-Cys^{+3} [25]. Such a residue is rarely present near an α-BChl *b*-binding His^0 and is replaced with Leu or Val in most purple bacteria other than *Roseospirillum parvum* [33,34]. Based on the pKa values of the Cys side chain (ranging from 2.9 to 9.8 inside proteins) [35] and its close association with LH1 BChl molecules [36], we proposed the side chain thiol/thiolate of α-Cys^{+3} as a candidate for modulating the point-charge around molecules of BChl *b* that trigger the pH-dependent interconvertible LH1 Q_y band shifts of *Hlr. halochloris* [25]. The present study demonstrated that the *Hlr. abdelmalekii* LH1-RCs exhibited salt- and pH-dependent interconvertible spectral changes as well (Figure 1). In this respect, one of the LH1 α-polypeptides from *Hlr. abdelmalekii* (NCBI Reference Sequence: WP_200194663.1) is almost identical to PufA1/A2 of *Hlr. halochloris* and has the α-Cys^{+3} residue (Figure 4), although the expression of this (or any *Hlr. abdelmalekii* LH1 polypeptides) is unknown [37]. However, considering these results, we suggest that the α-Cys^{+3} thiolate anions present at alkaline pH contribute to the ultra-

red-shift of the LH1 Q_y band while the protonated form at acidic pH results in a LH1 Q_y band blue-shift. Under high salt conditions, the thiolate anions may be surrounded by cations, thus disturbing the protonation of the thiolate anions and maintaining the red-shifted LH1 Q_y bands. By contrast, protonation of α-Cys^{+3} thiolate anions would be favored when cations were limiting, leading to blue-shifts of the LH1 Q_y band.

4.2. Effects of Amino Acid Compositions on the Thermal Stability of the LH1-RC Complexes from Halorhodospira Species

It is known that halophilic *Archaea* such as *Halobacterium* store high intracellular concentrations of inorganic salts (primarily K$^+$) for osmoregulation purposes under hypersaline conditions by balancing the excess positive charges present from molar concentrations of K$^+$ [38]. In contrast, most halophilic *Bacteria* and eukaryotes accumulate organic compatible solutes [39] instead of inorganic salts for osmoregulation. As for *Hlr. halochloris*, it accumulates glycine betaine, with minor amounts of ectoine and trehalose as compatible solutes [40–42] but does not accumulate KCl [43]. Furthermore, phototrophic growth of *Hlr. halochloris* cells was almost completely suppressed when NaCl (>3 M) in the medium was depleted or replaced with KCl (Figure S2). These findings suggest that a high concentration of NaCl is indispensable for physiological regulation in *Hlr. halochloris* cells and that NaCl is not functionally replaceable with KCl.

The salt- and pH-dependent stabilization of the LH1-RC from *Halorhodospira species* was confirmed not only in purified LH1-RCs but also in photosynthetic membranes (Figure S1). This indicates that surfaces of the LH1-RCs in contact with solvents (rather than transmembrane domains) are the regions of the proteins that sense and respond to changes in salt concentration and pH. In addition, LH1 BChl molecules that exhibit a salt- and pH-dependent spectral change are located on the C-terminal side (periplasmic side) of the photosynthetic membranes. Therefore, we focused our analyses on the amino acid composition of LH1 polypeptides (Figure 4) and the RC C-subunits (Figure S3) from various purple bacteria. It has been observed that halophilic proteins contain increased levels of negative charges due to the presence of acidic amino acids on their surfaces [38,44–47]. In agreement with this, the contents of acidic residues in both *Hlr. halochloris* and *Hlr. abdelmalekii* LH1 α- and β-polypeptides are significantly greater than those in non-halophilic purple bacteria, and the excess acidic residues are predominantly present in their C-terminal regions (Figure 4). In addition to LH1 polypeptides, the C-subunits of *Halorhodospira* RCs also exhibit high contents of acidic residues compared with those of nonhalophilic purple bacteria (Figure S3). Collectively, these acidic residues present in the LH1-RC surfaces exposed to solvents may be necessary for the *Hlr. halochloris* and *Hlr. abdelmalekii* LH1-RC complexes to function at the high salinities experienced in their habitats.

The contents of acidic residues in the LH1-RC complexes of both halophiles are also much greater than those of basic residues (Figures 4 and S3). These views are compatible with the observation that halophilic proteins show a biased amino acid composition on the protein surface, with a large excess of acidic residues and a reduced number of basic residues compared to corresponding proteins from nonhalophiles [46,48]. The highly negative charges of the abundant acidic residues on the halophilic protein surfaces improve their solubility and flexibility at high salt concentrations and allow them to function normally. By contrast, their electrostatic repulsions are thought to be a major factor responsible for the pronounced instability of proteins at lower salt concentrations [45], consistent with the present salt-dependent destabilization of the *Hlr. halochloris* and *Hlr. abdelmalekii* LH1-RC complexes (Figure 2C,H). Therefore, cations of salts may help stabilize the excess negative charges on the surfaces of halophilic LH1-RCs.

However, it is also known that highly negative charges in halophilic proteins can be compensated with tightly bound water dipoles, rather than excess cations, by forming water shells and hydrogen bonds to help maintain protein structure [38,49,50]. Water shells may therefore protect proteins from attack by exogenous molecules at higher temperature and contribute to their thermal stability. Thus, it is also possible that water shells exist

in the *Hlr. halochloris* and *Hlr. abdelmalekii* LH1-RCs. If true, the much-reduced thermal stability of the halophilic LH1-RCs observed at acidic pH could be due to the loss of water shells upon protonation of surface acidic residues. Furthermore, the significant difference observed in the thermal stability of the thermophilic and mesophilic LH1-RCs (Figure 2) could be due to the slightly higher ratio of acidic residues in the RC C-subunit and/or LH1 polypeptides of *Hlr. halochloris* compared with those of *Hlr. abdelmalekii*. Such water shells were confirmed in a high-resolution X-ray crystallographic LH1-RC structural analysis of *Tch. tepidum*—a thermophile, as opposed to a halophile—in which a number of water molecules cover the periplasmic and cytoplasmic surfaces of the LH1-RC complex as water shells [21]. To prove this hypothesis, high-resolution 3D structures of the *Hlr. halochloris* and *Hlr. abdelmalekii* LH1-RCs would be needed.

4.3. Effects of Anions and Cations on the Thermal Stability of the LH1-RC Complexes from Halorhodospira Species

Halophilic proteins maintain functionally active conformations in the presence of high concentrations of kosmotropic ions (small size, high surface charge density, strong hydration and forming water structures) but become inactive in the presence of high concentrations of chaotropic salts (large size, low surface charge density, weak hydration and breaking water structures) [30,51]. These high concentrations of coexisting anions and cations may affect the water structure on halophilic proteins, as discussed in Section 4.2. In the present study, the stabilizing effect of anions was in the order of $SO_4^{2-} > Cl^- \approx CH_3COO^- > HCO_3^- > NO_3^-$ (Figure 3), consistent with the Hofmeister series [29–31]. It can thus be concluded that kosmotropic anions function as stabilizing factors in the *Hlr. halochloris* LH1-RCs. By contrast, the Hofmeister cation series was in opposition to our results obtained with different chloride salts (Figure 3). Given that anions appear to have a larger effect on protein stability than do cations [30], the thermal stability of the *Hlr. halochloris* LH1-RCs is generally consistent with that predicted from the Hofmeister anion series. However, our results with different cations could signal that yet another mechanism contributes to the stability of halophilic proteins. Regarding the Hofmeister series, it was proposed that specific interactions between ions and proteins and ions and water molecules directly contacting the proteins may be as important as the changes that occur in the general water structure [30,31,51]. Therefore, the specific interactions of cations and the LH1-RC and/or possible water shells may be crucial for understanding the stability of LH1-RCs in both *Halorhodospira* species. For example, a possible explanation for the enhanced photocomplex stability observed with Ca^{2+} (Figure 3B) is that a part of acidic residues largely present in the C-terminal side of the *Hlr. halochloris* LH1-RCs may form specific binding sites for Ca^{2+} that reinforce protein structure in a manner similar to that of the Ca^{2+}-binding LH1-RC proteins present in several purple sulfur bacterial complexes [21,23,52]. To confirm this hypothesis, highly resolved structural information is required for the LH1-RC complexes from these *Halorhodospira* species.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/microorganisms10050959/s1, Table S1: Lifetimes for the thermal degradation of the LH1-RC complexes from *Hlr. halochloris* and *Hlr. abdelmalekii* at pH 8 in the presence of indicated concentration of NaCl; Table S2: Lifetimes for the thermal degradation of the LH1-RC complexes from *Hlr. halochloris* and *Hlr. abdelmalekii* in the presence of 1M NaCl at various pHs; Figure S1: The relative Q_y band intensities of *Hlr. halochloris* chromatophores in the presence of 4 M NaCl (A) or in the absence of NaCl (B) after incubation at the indicated pH and temperature for 60 min; Figure S2: Phototrophic growth curves for *Hlr. halochloris* in the presence of 3 M NaCl (black) or 3M KCl (purple) or in the absence of salts (gray) monitored as LH1 Q_y peak intensity at 1020 nm. Photos of *Hlr. halochloris* cultures grown under the indicated condition for 10 days are shown in the inset; Figure S3. Amino acid sequences of the RC C-subunits from several extremophilic purple phototrophic bacteria. Acidic (Asp, Glu) and basic (Arg, Lys) residues are highlighted with magenta and cyan, respectively. The numbers on the right side represent percentages of acidic and basic residues in each subunit.

References

1. Madigan, M.T. Anoxygenic phototrophic bacteria from extreme environments. *Photosynth. Res.* **2003**, *76*, 157–171. [CrossRef] [PubMed]
2. Imhoff, J.F.; Rahn, T.; Kunzel, S.; Neulinger, S.C. Photosynthesis Is widely distributed among proteobacteria as demonstrated by the phylogeny of pufLM reaction center proteins. *Front. Microbiol.* **2018**, *8*, 2679. [CrossRef]
3. Imhoff, J.F.; Trüper, H.G. *Ectothiorhodospira halochloris* sp. nov., s new extremely halophilic phototropic bacterium containing bacteriochlorophyll *b*. *Arch. Microbiol.* **1977**, *114*, 115–121. [CrossRef]
4. Imhoff, J.F.; Suling, J. The phylogenetic relationship among *Ectothiorhodospiraceae*: A reevaluation of their taxonomy on the basis of 16S rDNA analyses. *Arch. Microbiol.* **1996**, *165*, 106–113. [CrossRef] [PubMed]
5. Imhoff, J.F.; Trüper, H.G. *Ectothiorhodospira abdelmalekii* sp. nov. a new halophilic and alkaliphilic phototropic bacterium. *Zent. Bakteriol. I Abt. Orig. C Allg. Angew. Okol. Mikrobiol.* **1981**, *2*, 228–234. [CrossRef]
6. Drews, G.; Giesbrecht, P. *Rhodopseudomonas viridis,* nov. spec., ein neu isoliertes, obligat phototrophes bakterium. *Arch. Mikrobiol.* **1966**, *53*, 255–262. [CrossRef]
7. Eimhjellen, K.E. *Thiocapsa pfennigii* sp. nov. a new species of the phototrophic sulfur bacteria. *Arch. Mikrobiol.* **1970**, *73*, 193–194. [CrossRef]
8. Keppen, O.I.; Gorlenko, V.M. New species of purple budding bacteria containing bacteriochlorophyll-*b*. *Microbiol.* **1975**, *44*, 224–229.
9. Resnick, S.M.; Madigan, M.T. Isolation and characterization of a mildly thermophilic nonsulfur purple bacterium containing bacteriochlorophyll-*b*. *FEMS Microbiol. Lett.* **1989**, *65*, 165–170. [CrossRef]
10. Hiraishi, A. Transfer of the bacteriochlorophyll *b*-containing phototrophic bacteria *Rhodopseudomonas viridis* and *Rhodopseudomonas sulfoviridis* to the genus *Blastochloris* gen. nov. *Int. J. Syst. Bacteriol.* **1997**, *47*, 217–219. [CrossRef]
11. Madigan, M.T.; Resnick, S.M.; Kempher, M.L.; Dohnalkova, A.C.; Takaichi, S.; Wang-Otomo, Z.Y.; Toyoda, A.; Kurokawa, K.; Mori, H.; Tsukatani, Y. *Blastochloris tepida*, sp. nov., a thermophilic species of the bacteriochlorophyll *b*-containing genus *Blastochloris*. *Arch. Microbiol.* **2019**, *201*, 1351–1359. [CrossRef] [PubMed]
12. Bryantseva, I.A.; Gorlenko, V.M.; Kompantseva, E.I.; Imhoff, J.F. *Thioalkalicoccus limnaeus* gen. nov., sp nov., a new alkaliphilic purple sulfur bacterium with bacteriochlorophyll *b*. *Int. J. Syst. Evol. Micr.* **2000**, *50*, 2157–2163. [CrossRef] [PubMed]
13. Cogdell, R.J.; Gall, A.; Köhler, J. The architecture and function of the light-harvesting apparatus of purple bacteria: From single molecules to in vivo membranes. *Q. Rev. Biophys.* **2006**, *39*, 227–324. [CrossRef] [PubMed]
14. Madigan, M.T. A novel photosynthetic purple bacterium isolated from a Yellowstone hot spring. *Science* **1984**, *225*, 313–315. [CrossRef]
15. Madigan, M.T.; Absher, J.N.; Mayers, J.E.; Asao, M.; Jung, D.O.; Bender, K.S.; Kempher, M.L.; Hayward, M.K.; Sanguedolce, S.A.; Brown, A.C.; et al. *Allochromatium tepidum*, sp. nov., a hot spring species of purple sulfur bacteria. *Arch. Microbiol.* **2022**, *204*, 115. [CrossRef] [PubMed]
16. Castenholz, R.W.; Bauld, J.; Jorgenson, B.B. Anoxygenic microbial mats of hot springs—Thermophilic *Chlorobium* sp. *Fems. Microbiol. Ecol.* **1990**, *74*, 325–336. [CrossRef]
17. Saini, M.K.; ChihChe, W.; Soulier, N.; Sebastian, A.; Albert, I.; Thiel, V.; Bryant, D.A.; Hanada, S.; Tank, M. *Caldichromatium japonicum* gen. nov., sp. nov., a novel thermophilic phototrophic purple sulphur bacterium of the *Chromatiaceae* isolated from Nakabusa hot springs, Japan. *Int. J. Syst. Evol. Micr.* **2020**, *70*, 5701–5710. [CrossRef]
18. Raymond, J.C.; Sistrom, W.R. *Ectothiorhodospira Halophila*—A New Species of Genus *Ectothiorhodospira*. *Arch. Mikrobiol.* **1969**, *69*, 121–126. [CrossRef]

19. Kimura, Y.; Yu, L.J.; Hirano, Y.; Suzuki, H.; Wang, Z.Y. Calcium ions are required for the enhanced thermal stability of the light-harvesting-reaction center core complex from thermophilic purple sulfur bacterium *Thermochromatium tepidum*. *J. Biol. Chem.* **2009**, *284*, 93–99. [CrossRef]

20. Kimura, Y.; Yura, Y.; Hayashi, Y.; Li, Y.; Onoda, M.; Yu, L.-J.; Wang-Otomo, Z.-Y.; Ohno, T. Spectroscopic and thermodynamic characterization of the metal-binding sites in the LH1–RC complex from thermophilic photosynthetic bacterium *Thermochromatium tepidum*. *J. Phys. Chem. B* **2016**, *120*, 12466–12473. [CrossRef]

21. Yu, L.J.; Suga, M.; Wang-Otomo, Z.Y.; Shen, J.R. Structure of photosynthetic LH1-RC supercomplex at 1.9 Å resolution. *Nature* **2018**, *556*, 209–213. [CrossRef] [PubMed]

22. Kimura, Y.; Lyu, S.; Okoshi, A.; Okazaki, K.; Nakamura, N.; Ohashi, A.; Ohno, T.; Kobayashi, M.; Imanishi, M.; Takaichi, S.; et al. Effects of calcium ions on the thermostability and spectroscopic properties of the LH1-RC complex from a new thermophilic purple bacterium *Allochromatium tepidum*. *J. Phys. Chem. B* **2017**, *121*, 5025–5032. [CrossRef] [PubMed]

23. Tani, K.; Kobayashi, K.; Hosogi, N.; Ji, X.-C.; Nagashima, S.; Nagashima, K.V.P.; Tsukatani, Y.; Kanno, R.; Hall, M.; Yu, L.-J.; et al. A Ca^{2+}-binding motif underlies the unusual properties of certain photosynthetic bacterial core light-harvesting complexes. *J. Biol. Chem.* **2022**. [CrossRef] [PubMed]

24. Seto, R.; Takaichi, S.; Kurihara, T.; Kishi, R.; Honda, M.; Takenaka, S.; Tsukatani, Y.; Madigan, M.T.; Wang-Otomo, Z.Y.; Kimura, Y. Lycopene-family carotenoids confer thermostability on photocomplexes from a new thermophilic purple bacterium. *Biochemistry* **2020**, *59*, 2351–2358. [CrossRef] [PubMed]

25. Kimura, Y.; Nojima, S.; Nakata, K.; Yamashita, T.; Wang, X.P.; Takenaka, S.; Akimoto, S.; Kobayashi, M.; Madigan, M.T.; Wang-Otomo, Z.Y.; et al. Electrostatic charge controls the lowest LH1 Q_y transition energy in the triply extremophilic purple phototrophic bacterium, *Halorhodospira halochloris*. *Biochim. Biophys. Acta* **2021**, *1862*, 148473. [CrossRef]

26. Steiner, R.; Scheer, H. Characterization of a B800-1020 antenna from the photosynthetic bacteria *Ectothiorhodospira halochloris* and *Ectothiorhodospira abdelmalekii*. *Biochim. Biophys. Acta* **1985**, *807*, 278–284. [CrossRef]

27. Kimura, Y.; Hirano, Y.; Yu, L.J.; Suzuki, H.; Kobayashi, M.; Wang, Z.Y. Calcium ions are involved in the unusual red shift of the light-harvesting 1 Q_y transition of the core complex in thermophilic purple sulfur bacterium *Thermochromatium tepidum*. *J. Biol. Chem.* **2008**, *283*, 13867–13873. [CrossRef]

28. Imanishi, M.; Takenouchi, M.; Takaichi, S.; Nakagawa, S.; Saga, Y.; Takenaka, S.; Madigan, M.T.; Overmann, J.; Wang-Otomo, Z.Y.; Kimura, Y. A dual role for Ca^{2+} in expanding the spectral diversity and stability of light-harvesting 1 reaction center photocomplexes of purple phototrophic bacteria. *Biochemistry* **2019**, *58*, 2844–2852. [CrossRef]

29. Hofmeister, F. Zur Lehre von der Wirkung der Salze. *Arch. Exp. Pathol. Pharmacol.* **1888**, *24*, 247–260. [CrossRef]

30. Yang, Z. Hofmeister effects: An explanation for the impact of ionic liquids on biocatalysis. *J. Biotechnol.* **2009**, *144*, 12–22. [CrossRef]

31. Zhang, Y.J.; Cremer, P.S. Interactions between macromolecules and ions: The Hofmeister series. *Curr. Opin. Chem. Biol.* **2006**, *10*, 658–663. [CrossRef] [PubMed]

32. Wagnerhuber, R.; Brunisholz, R.A.; Bissig, I.; Frank, G.; Zuber, H. A new possible binding-site for bacteriochlorophyll-*b* in a light-harvesting polypeptide of the bacterium *Ectothiorhodospira halochloris*. *FEBS Lett.* **1988**, *233*, 7–11. [CrossRef]

33. Tuschak, C.; Beatty, J.T.; Overmann, J. Photosynthesis genes and LH1 proteins of *Roseospirillum parvum* 930I, a purple non-sulfur bacterium with unusual spectral properties. *Photosynth. Res.* **2004**, *81*, 181–199. [CrossRef]

34. Tsukatani, Y.; Hirose, Y.; Harada, J.; Yonekawa, C.; Tamiaki, H. Unusual features in the photosynthetic machinery of *Halorhodospira halochloris* DSM 1059 revealed by complete genome sequencing. *Photosynth. Res.* **2019**, *140*, 311–319. [CrossRef] [PubMed]

35. Awoonor-Williams, E.; Rowley, C.N. Evaluation of methods for the calculation of the pka of cysteine residues in proteins. *J. Chem. Theory. Comput.* **2016**, *12*, 4662–4673. [CrossRef] [PubMed]

36. Qian, P.; Siebert, C.A.; Wang, P.Y.; Canniffe, D.P.; Hunter, C.N. Cryo-EM structure of the *Blastochloris viridis* LH1-RC complex at 2.9 Å. *Nature* **2018**, *556*, 203–208. [CrossRef] [PubMed]

37. Wagnerhuber, R.; Brunisholz, R.A.; Bissig, I.; Frank, G.; Suter, F.; Zuber, H. The primary structure of the antenna polypeptides of *Ectothiorhodospira halochloris* and *Ectothiorhodospira halophila*—Four core-type antenna polypeptides in *E. halochloris* and *E. halophila*. *Eur. J. Biochem.* **1992**, *205*, 917–925. [CrossRef]

38. Mevarech, M.; Frolow, F.; Gloss, L.M. Halophilic enzymes: Proteins with a grain of salt. *Biophys. Chem.* **2000**, *86*, 155–164. [CrossRef]

39. Roberts, M.F. Organic compatible solutes of halotolerant and halophilic microorganisms. *Saline Syst.* **2005**, *1*, 5. [CrossRef]

40. Galinski, E.A.; Trüper, H.G. Betaine, a Compatible Solute in the Extremely Halophilic Phototropic Bacterium *Ectothiorhodospira halochloris*. *FEMS Microbiol. Lett.* **1982**, *13*, 357–360. [CrossRef]

41. Galinski, E.A.; Pfeiffer, H.P.; Trüper, H.G. 1,4,5,6-Tetrahydro-2-Methyl-4-Pyrimidinecarboxylic Acid—A Novel Cyclic Amino-Acid from Halophilic Phototrophic Bacteria of the Genus *Ectothiorhodospira*. *Eur. J. Biochem.* **1985**, *149*, 135–139. [CrossRef] [PubMed]

42. Oren, A. Life at high salt concentrations, intracellular KCl concentrations, and acidic proteomes. *Front. Microbiol.* **2013**, *4*, 315. [CrossRef] [PubMed]

43. Deole, R.; Challacombe, J.; Raiford, D.W.; Hoff, W.D. An Extremely Halophilic Proteobacterium Combines a Highly Acidic Proteome with a Low Cytoplasmic Potassium Content. *J. Biol. Chem.* **2013**, *288*, 581–588. [CrossRef] [PubMed]

44. Lanyi, J.K. Salt-Dependent Properties of Proteins from Extremely Halophilic Bacteria. *Bacteriol. Rev.* **1974**, *38*, 272–290. [CrossRef] [PubMed]

45. Elcock, A.H.; McCammon, J.A. Electrostatic contributions to the stability of halophilic proteins. *J. Mol. Biol.* **1998**, *280*, 731–748. [CrossRef] [PubMed]

46. Fukuchi, S.; Yoshimune, K.; Wakayama, M.; Moriguchi, M.; Nishikawa, K. Unique amino acid composition of proteins in halophilic bacteria. *J. Mol. Biol.* **2003**, *327*, 347–357. [CrossRef]

47. DasSarma, S.; DasSarma, P. Halophiles and their enzymes: Negativity put to good use. *Curr. Opin. Microbiol.* **2015**, *25*, 120–126. [CrossRef]

48. Fukuchi, S.; Nishikawa, K. Protein surface amino acid compositions distinctively differ between thermophilic and mesophilic bacteria. *J. Mol. Biol.* **2001**, *309*, 835–843. [CrossRef]

49. Richard, S.B.; Madern, D.; Garcin, E.; Zaccai, G. Halophilic adaptation: Novel solvent protein interactions observed in the 2.9 and 2.6 a resolution structures of the wild type and a mutant of malate dehydrogenase from *Haloarcula marismortui. Biochemistry* **2000**, *39*, 992–1000. [CrossRef]

50. Britton, K.L.; Baker, P.J.; Fisher, M.; Ruzheinikov, S.; Gilmour, D.J.; Bonete, M.J.; Ferrer, J.; Pire, C.; Esclapez, J.; Rice, D.W. Analysis of protein solvent interactions in glucose dehydrogenase from the extreme halophile Haloferax mediterranei. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 4846–4851. [CrossRef]

51. Okur, H.I.; Hladilkova, J.; Rembert, K.B.; Cho, Y.; Heyda, J.; Dzubiella, J.; Cremer, P.S.; Jungwirth, P. Beyond the Hofmeister Series: Ion-Specific Effects on Proteins and Their Biological Functions. *J. Phys. Chem. B* **2017**, *121*, 1997–2014. [CrossRef] [PubMed]

52. Tani, K.; Kanno, R.; Makino, Y.; Hall, M.; Takenouchi, M.; Imanishi, M.; Yu, L.J.; Overmann, J.; Madigan, M.T.; Kimura, Y.; et al. Cryo-EM structure of a Ca^{2+}-bound photosynthetic LH1-RC complex containing multiple αβ-polypeptides. *Nat. Commun.* **2020**, *11*, 4955. [CrossRef] [PubMed]

Article

Examination of Genetic Control Elements in the Phototrophic Firmicute *Heliomicrobium modesticaldum*

Alexandria M. Layton [ID] and Kevin E. Redding *[ID]

Center for Bioenergy and Photosynthesis, School of Molecular Sciences, Arizona State University, Tempe, AZ 85287, USA; amlayto1@asu.edu
* Correspondence: kevin.redding@asu.edu; Tel.: +1-480-282-0013

Abstract: *Heliomicrobium modesticaldum* has been used as a model organism for the Heliobacteria, the only phototrophic family in the Firmicutes. It is a moderately thermophilic anoxygenic phototrophic bacterium that is capable of fermentative growth in the dark. The genetic manipulation of *H. modesticaldum* is still in its infancy. Methods to introduce genes through the use of exogenous plasmids and to delete genes from the chromosome through the use of the native CRISPR/Cas system have been developed in the last several years. To expand our genetic toolkit, it was necessary to control gene expression. In this study, we analyzed constitutive and inducible promoters developed for clostridia for their use in *H. modesticaldum* and further tested two reporters, *adhB* and *lacZ*, as indicators of promoter strength. Alcohol dehydrogenase (AdhB) was unsuitable as a reporter in this species due to high endogenous activity and/or low activity of the reporter, but a thermostable LacZ worked well as a reporter. A set of constitutive promoters previously reported to work in *Clostridium thermocellum* was found to be reliable for controlling the expression of the *lacZ* reporter gene in *H. modesticaldum* at a range of activities spanning an order of magnitude. An anhydrotetracycline-inducible promoter was created by inserting *tetO* operators into a strong constitutive promoter, but it was not fully repressible. The implementation of a xylose-inducible promoter resulted in complete repression of β-gal in the absence of xylose, and reliable expression tunable through the concentration of xylose added to the culture.

Keywords: Heliobacteria; promoters; reporters; gene expression; phototrophic bacteria

1. Introduction

The Heliobacteria are the only phototrophic members of the phylum Firmicutes. Although unable to fix inorganic carbon, they are capable of growing phototrophically (using light as an energy source) via electron transport driven by their photochemical reaction center (RC) and chemiosmotic ATP production, or chemotrophically through fermentation of a suitable carbon source. In the case of *Heliomicrobium* (previously *Heliobacterium*) *modesticaldum*, this is primarily pyruvate or lactate, but other species can use a greater range of organic molecules [1–3]. This metabolic flexibility renders phototrophy optional and allows for genetic manipulation of components of the phototrophic apparatus without ill consequences. This, combined with the simplicity of their phototrophic apparatus, make Heliobacteria ideal candidates not only for studying the fundamental mechanisms and evolution of phototrophy but also for studying other aspects of metabolism [4–6].

H. modesticaldum is a moderate thermophile of the family Heliobacteriaceae and has been used as a model system for this group of organisms. Since its discovery and isolation from Icelandic volcanic soil in 1995 [2], its genome has been fully sequenced [7], transcriptomes have been reported [8], and methods have been developed for the introduction of plasmids via conjugation from *Escherichia coli* [9]. A method to leverage its endogenous CRISPR/Cas system to perform precise chromosome editing [9,10] now allows for reverse genetics to be applied to this organism [11–14]. However, one tool that is lacking is the

ability to control and fine tune gene expression. The purpose of this report is to remedy this deficiency.

H. modesticaldum has displayed stringency when it comes to genetic elements and manipulation. To introduce non-native plasmids, the plasmids must first be pre-methylated at presumed restriction endonuclease sites and then conjugated using an engineered *E. coli* host; without the methylation, the DNA will be cleaved and destroyed upon insertion into the cells [9]. In addition, while the Cas9 system from *Streptococcus pyogenes* could be adapted for use in other organisms, *H. modesticaldum* could not tolerate it. Similar to other clostridia [15], the organism's endogenous CRISPR/Cas system was employed to perform gene deletions and chromosome editing [10]. Both of these genetic manipulation techniques were developed by adapting techniques used for other members of the Clostridiales. We took the same route to find methods of controlling gene expression in *H. modesticaldum*.

In this study, we analyzed a series of promoters used in Clostridia for their activity in *H. modesticaldum*. A set of constitutive *C. thermocellum* promoters previously analyzed by Olson et al., 2015 [16] was chosen for several reasons: *C. thermocellum* is a member of the Clostridiales family much like *H. modesticaldum*, *C. thermocellum* grows at a similar temperature to *H. modesticaldum*, the promoters are orthogonal and therefore should not affect the activities of native genes, and these promoters were already previously characterized in *C. thermocellum* and *E. coli*. We then utilized several of the constitutive promoters to attempt to construct an anhydrotetracycline inducible promoter. We have also tested a xylose-inducible promoter. In addition, as a result of this study, two reporters, alcohol dehydrogenase and β-galactosidase, were tested in *H. modesticaldum* for the first time.

2. Materials and Methods

2.1. Strains and Culture Conditions

E. coli strains (TOP10 for activity assays, s17-1 for conjugations) were grown in Luria-Bertani (LB) broth (BD Difco, Franklin Lakes, NJ, USA). When needed, the antibiotics erythromycin (Sigma-Aldrich, St. Louis, MO, USA) and chloramphenicol (Acros, Geel, Belgium) were added to a final concentration of 100 µg mL^{-1} and 15 µg mL^{-1}, respectively. Cultures were grown overnight at 37 °C while shaking at 250 rpm.

All work utilizing *H. modesticaldum* was performed in an anaerobic chamber (Coy Laboratory Products, Grass Lake, MI, USA), with an atmosphere of 20% CO_2 and 2–4% H_2, and balance N_2. *H. modesticaldum* strains were grown in pyruvate yeast extract (PYE) medium following the recipe and procedure outlined previously [9]. When necessary, erythromycin was added to a final concentration of 10 µg mL^{-1}. Cultures were grown at 50 °C under 790-nm light-emitting diodes (LEDs) (Marubeni America, New York, NY, USA) at a flux of 30 µmol photons m^{-2} s^{-1}.

2.2. Plasmid Construction

All constructs were made using the vector pMTL86251 [17]. Plasmids containing promoter regions originating from *C. thermocellum* and thermostable *lacZ* and *adhB* reporter genes were gifts from Daniel Olson (Dartmouth College) [16]. Promoter sequences are listed in Table S1. All oligonucleotide primers were purchased from Integrated DNA Technologies (Coralville, IA, USA) and are listed in Table S2. Restriction enzymes and Q5 DNA polymerase used during plasmid construction were purchased from New England Biolabs (Ipswich, MA, USA).

The seven promoters (gapDH, eno, cbp_2, 2926, 0815, 0966, and 2638) and two reporters (*adhB*, *lacZ*) were PCR amplified individually from the plasmids provided by Daniel Olson (pDGO89, pDGO95, pDGO102, pDGO105, pDGO106, pDGO108, and pDGO117, listed in Table S3) using primers containing restriction sites (PspOMI upstream and BamHI downstream for promoters, BglII upstream and SalI downstream for reporters). The PCR products were digested with their respective enzymes then ligated into pMTL86251 at the NotI and XhoI sites using a modified Golden Gate assembly protocol. Modifications to the protocol include that the ligation mixture contained NotI and XhoI restriction enzymes in

place of a Type IIS enzyme, and heat cycles were set up as follows: 10 min at 37 °C then 10 min at 16 °C, repeated 10 times, followed by 85 °C for 20 min, then held at 4 °C. The module containing the gene coding for *xylR* in the reverse orientation from the promoter region containing the *xylO* sequence (P_{xyl}) was synthesized by GenScript (Piscataway, NJ, USA) with a NotI site downstream of the *xylR* and a BamHI site downstream of P_{xyl} and cloned into pMTL86251 at those sites. Once received from Genscript, the *lacZ* reporter (amplified using the same primers as before, adding BglII and SalI sites) was cloned into this plasmid at the BamHI and XhoI sites, resulting in pALX3. Constructs were confirmed by Sanger sequencing.

To add the *tetO1* sequences, the P_{gapDH}—*lacZ* module was amplified from plasmid pAL66 and inserted into pUC19 using BamHI and SalI sites. The *tetO1* sequences were inserted using primers that amplified around pUC19 to generate a single linear product and ligated (recircularized) using Genscript's GenBuilder Cloning Kit. After confirming the *tetO* sequences through the use of Sanger sequencing, the $P_{gapDH/tetO}$—*lacZ* module was amplified from the pUC19 using primers to add BglII (upstream of $P_{gapDH/tetO}$) and SalI (downstream of *lacZ*) sites. The *tetR* gene was amplified from genomic preps of XL1-Blue *E. coli* cells using primers to add PspOMI (downstream) and EcoRI (upstream) sites, and the eno promoter was amplified using primers to add EcoRI (downstream) and BglII (upstream) sites. All three products ($P_{gapDH/tetO}$—*lacZ*, *tetR*, and eno) were digested with their respective enzymes and inserted into pMTL86251 at the NotI and XhoI sites using the modified Golden Gate assembly process described earlier.

Plasmids were transformed into *E. coli* TOP10 and s17-1 using standard techniques and were further conjugated into *H. modesticaldum* using the procedure described previously [9].

2.3. Induction Tests

Anhydrotetracycline (aTc) (Sigma-Aldrich) was prepared in ethanol (10 µg mL^{-1}) and added to the appropriate final concentration (ranging from 0 to 200 µg L^{-1}) when the cells entered the log phase ($OD_{735} \approx 0.5$). Cells were assayed after induction for 3 h. For some growth tests, aTc was added at the time of inoculation. For all xylose induction tests, cells were grown in PYE media, with xylose (Sigma-Aldrich) added at the time of inoculation.

2.4. Cell Free Extracts

Cell free extracts were obtained following the protocol outlined previously [4], with a few alterations. *E. coli* and *H. modesticaldum* cells were harvested by centrifugation at 5000× *g* for 15 min. Cells were washed and resuspended in 20 mM Tris-HCl buffer pH 8.0 (Santa Cruz Biochemicals, Dallas, TX, USA) containing 100 µL of 3 mg mL^{-1} lysozyme (Goldbio, St. Louis, MO, USA), then incubated at 37 °C for 20 min. Phenyl-methanesulfonyl fluoride (PMSF) (Sigma-Aldrich) was added to 1 mM, and the suspension was sonicated in an Aquasonic sonication bath (Model 75T, VWR, Scientific Products, Radnor, PA, USA) for 20 min with minute-long pauses every 5 min. The lysate was centrifuged at 20,000× *g* for 30 min, and the pellet was discarded. Protein concentration in the extracts was determined using the Pierce BCA protein assay kit (Thermo Scientific, Waltham, MA, USA).

2.5. Reporter Activity Assays

Activity assays were modeled after those described previously [16]. All assays were performed using a 96-well plate and read using an Epoch Microplate spectrophotometer (BioTek Instruments, Inc., Winooski, VT, USA). All reagents used for the assays were purchased from Sigma-Aldrich (St. Louis, MO, USA).

2.5.1. Alcohol Dehydrogenase Assays

In this assay, 4 mM acetaldehyde and 0.4 mM NADPH were added to extract (50 µg of protein) in 100 mM Tris-HCl buffer (pH 7.6), and the absorbance at 340 nm was read every 7 s over a period of 10 min to detect the rapid loss of NADPH.

2.5.2. β-Galactosidase Assays

To measure β-galactosidase (β-gal) activity, 50 mM β-mercaptoethanol, 1 mM MgCl$_2$, and 0.66 mg mL^{-1} ONPG were added to 50 μg of protein in 100 mM sodium phosphate buffer (pH 7.5). The absorbance at 420 nm was immediately read, and then the plate was incubated at 37 °C. Absorbance was read after 30, 60, and 90 min of incubation to calculate a linear rate of activity corresponding to the cleavage of ONPG. The increase in A$_{420}$ was converted into units (U) by using the extinction coefficient of ONP (ε = 4500 M^{-1} s^{-1}) and further to calculate the nmols of ONP produced per second of the reaction.

2.6. Calculations—Curve Fittings and Error Calculations

Growth data were fit to the logistic equation as described [13]. β-gal activity data (product vs. time) were fit to a linear equation. The pro Fit software package (QuantumSoft; http://www.quansoft.com/, accessed on 20 March 2022) was used to produce all fittings.

3. Results

To develop a gene expression system, two enzymes were tested for use as reporters in *H. modesticaldum*. Alcohol dehydrogenase and β-galactosidase have been widely used as reporters and have been tested in many bacterial species, including thermophilic Clostridia [16]. The constitutive and inducible promoters chosen for use in *H. modesticaldum* have been tested in other clostridia, including *C. thermocellum* [16], *Clostridium perfringens* [18], and *Clostridium acetobutilicum* [19].

3.1. Testing of Reporters in H. modesticaldum

The reporter gene *adhB* (encoding an alcohol dehydrogenase from *Thermoanaerobacter pseudethanolicus*) was cloned behind several clostridial promoters that had been previously tested in *C. thermocellum* [16]. The expression of AdhB was tested in both *H. modesticaldum* and *E. coli* by measuring alcohol dehydrogenase activity in cellular extracts after the addition of acetaldehyde and NADPH. The reduction of acetaldehyde to ethanol was monitored by the decrease of absorption of NADPH at 340 nm (Figure 1A). Wild-type (WT) *E. coli* cells displayed a higher level of basal alcohol dehydrogenase activity than WT *H. modesticaldum* cells (−0.55 mAU$_{340}$ s^{-1} vs. −0.25 mAU$_{340}$ s^{-1}, respectively). *E. coli* transformants harboring the plasmid containing the *adhB* reporter gene under the control of promoter 0815 had significantly more alcohol dehydrogenase activity (−0.85 mAU$_{340}$ s^{-1}), an almost two-fold increase from WT expression levels. In contrast, cellular extracts of *H. modesticaldum* transformants containing plasmids using several different promoters to drive *adhB* expression exhibited activities that were barely distinguishable from WT (−0.26 or −0.28 mAU$_{340}$ s^{-1} for gapDH and 2926 promoters, respectively (Figure 1A)).

The *lacZ* gene from *Geobacillus stearothermophilus* [20], which encodes a thermostable beta-galactosidase (β-gal), was cloned behind several clostridial promoters. The resulting plasmids were introduced into *H. modesticaldum*, and cell-free extracts were prepared from them. β-gal activity was assayed by the addition of extract to an assay mixture containing *o*-nitrophenylgalactoside (ONPG); cleavage of ONPG into *o*-nitrophenol (ONP) and galactose was monitored by an increase in the absorbance of ONP at 420 nm. Cellular extracts from WT *H. modesticaldum* cells displayed no detectable β-gal activity (~0.05 μAU$_{420}$ s^{-1}; R^2 = 0.1). In contrast, extracts from two *H. modesticaldum* transconjugants with promoters driving *lacZ* expression displayed much higher levels of β-gal activity: 0.6 mAU$_{420}$ s^{-1} (gapDH promoter) and 0.3 mAU$_{420}$ s^{-1} (2926 promoter) (Figure 1B). In all subsequent studies, the *G. stearothermophilus lacZ* gene was used as the reporter, and β-gal activities were measured over a time of 90 min to generate a linear regression curve and estimate the rate of ONP generation. Throughout this report, β-gal activity is described as units (U, nmol of ONP formed per second of reaction) per mg of protein.
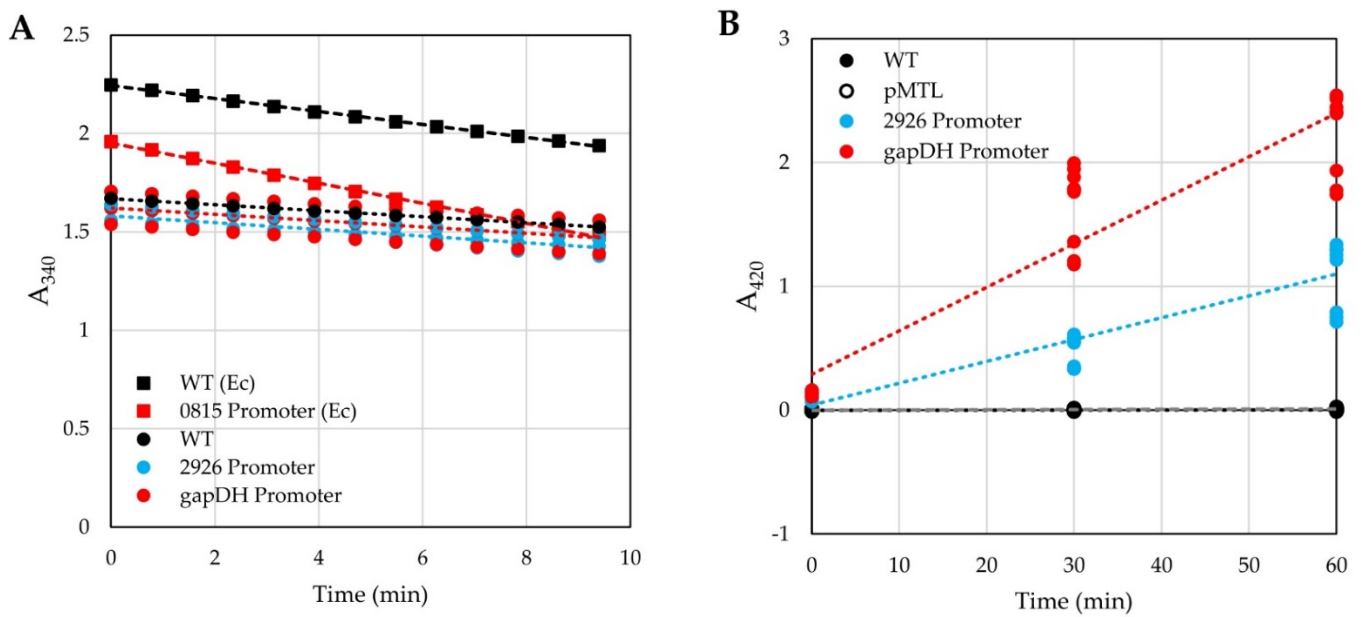
Figure 1. A reporter expression study utilizing two different reporters in *H. modesticaldum* and *E. coli*. Alcohol dehydrogenase (**A**) and β-gal (**B**) activities in extracts from unmodified (WT) and *H. modesticaldum* transformed with empty vector (pMTL) or expression plasmids using the 2926 or gapDH promoters. As a positive control, assays using extracts from *E. coli* (Ec) cells that were either unmodified (WT) or transformed with a vector using the 0815 promoter are included for comparison in panel (**A**). Each point represents a single technical or biological replicate. *E. coli* and WT *H. modesticaldum* controls in (**A**) are displayed with a single replicate (*n* = 1), while 2926 and gapDH samples display each individual replicate (*n* = 9); all samples in (**B**) are displayed with each individual replicate (*n* = 9). Dotted or dashed lines represent the trendline generated for each dataset.

3.2. Testing of Constitutive Promoters from C. thermocellum in H. modesticaldum

Seven promoters from *C. thermocellum* [16] were assayed for their ability to drive expression of *lacZ* in *H. modesticaldum*. Measured activities ranged from 0.067 to 0.79 U/mg protein, with the tested promoters forming two distinct groups (Figure 2). Use of promoters 2926, eno, cbp_2, and gapDH resulted in higher expression of β-gal in *H. modesticaldum*, with cellular activities ranging from 0.39 to 0.79 U/mg protein. Use of promoters 2638, 0815, and 0966 resulted in lower expression (0.067 to 0.11 U/mg protein).
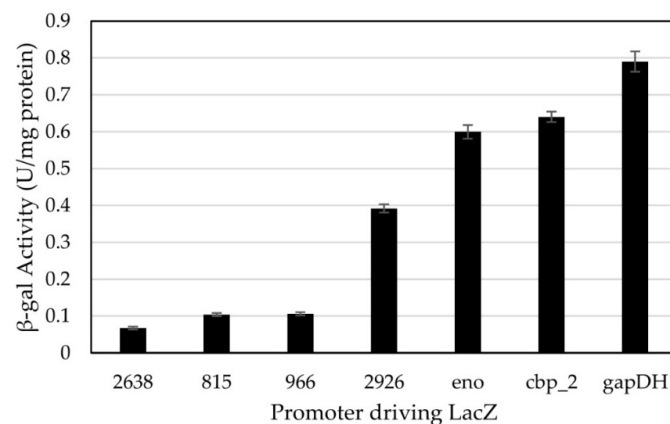
Figure 2. β-gal activities resulting from the expression of *lacZ* driven by different promoters in *H. modesticaldum* transformants. Units represent the nmoles of ONP formed per second of reaction. Data are represented as the average rate of three biological samples, with error bars representing the standard error.

Because introducing plasmids into *H. modesticaldum* requires a conjugation step from *E. coli*, it is helpful to know the activities of the promoters in both bacterial species. Therefore, activities of *E. coli* Top10 transformants harboring the same plasmids were also measured. Overall, activities resulting from the promoters in the two species displayed a positive correlation (R^2 = 0.486), indicating that if expression of *lacZ* driven by a given promoter is low or high in one species, it will tend to be similarly low or high in the other (Figure 3).
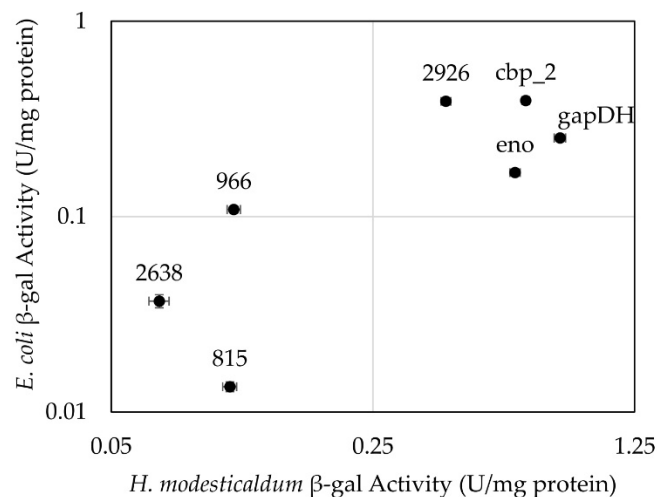
Figure 3. β-gal activities in *E. coli* and *H. modesticaldum* transformants plotted on a logarithmic scale. Data points are representative of the average rate of β-gal activity for each promoter. Error bars represent the standard error. (Values listed in Table S4).

3.3. Conversion of a Constitutive Promoter into an Anhydrotetracycline-Inducible Promoter

At the start of these studies, there was no prior knowledge of which inducible systems would work in *H. modesticaldum*. We decided to construct such a system based off a system previously reported for *C. acetobutylicum* [19], involving the insertion of an orthogonal operator into one of the clostridial promoters previously tested. This system is based on the TetR/*tetO* tetracycline resistance operon. In this system, the TetR repressor binds to the *tetO* operator, inhibiting the transcription of the downstream gene (originally coding for a protein conferring resistance) [21,22]. The binding of tetracycline (Tc) to the TetR homodimer results in a conformational change, removing it from the operator and alleviating repression. As we know that *H. modesticaldum* is sensitive to Tc [9], we expect that it (and related derivatives) are able to enter the cell. Anhydrotetracycline (aTc) has a lower affinity for the ribosome than Tc, making it a weaker antibiotic, but has a higher affinity for the TetR protein, making it a better inducer [23].

We characterized the effects of aTc on the growth of untransformed *H. modesticaldum* cells in liquid cultures and found that *H. modesticaldum* can tolerate up to 100 μg L^{-1} aTc without noticeable effects on growth. In addition, *H. modesticaldum* cells can tolerate higher levels of aTc (up to 150 ng mL^{-1}) if the aTc is added after the cells have grown to an O.D. of about 0.5 (Figure 4). These levels are slightly lower than those tolerated by *C. acetobutylicum* [19]; however, they are also only marginally higher than the levels at which *H. modesticaldum* can tolerate Tc, which has a minimal inhibitory concentration (MIC) of 200 μg L^{-1} on agar plates [1].

We chose the constitutive promoter associated with the highest β-gal expression in *H. modesticaldum*—gapDH—as the starting point for an aTc-inducible system. The gapDH promoter has two annotated −35/−10 regions [16]; therefore, the *tetO* sequence *tetO1* [22] was added to both sites (Figure 5). The replacement of the 15- or 17-bp spacer region between the −35 and −10 boxes with the 18-bp *tet* operator resulted in only a ~7% drop in β-gal expression (Figure 6).
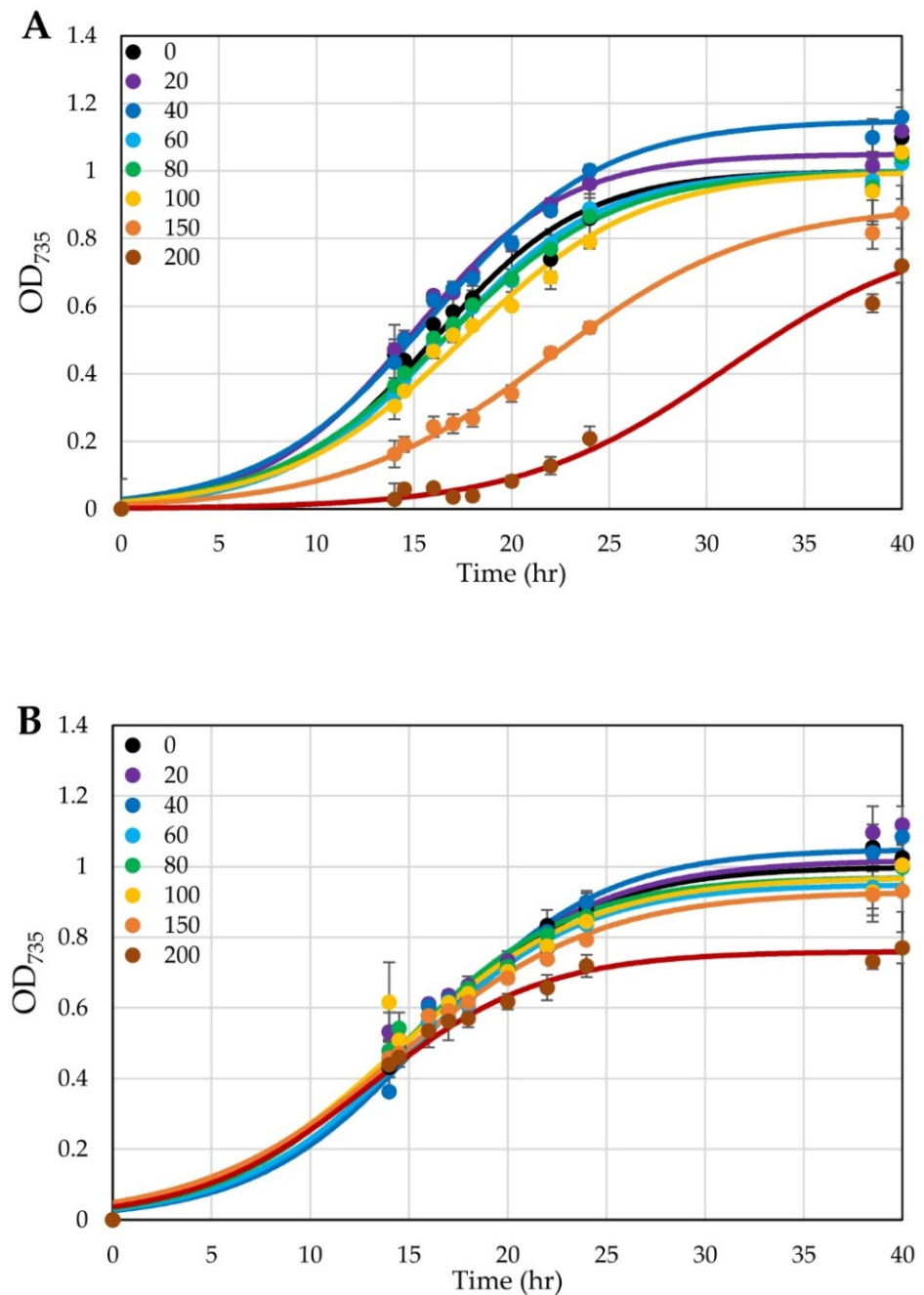
Figure 4. The effect of aTc on the growth of WT *H. modesticaldum*. Tested aTc concentrations ranged from 0 to 200 μg L^{-1}, as indicated in the legends; aTc was added either at the time of inoculation (**A**) or at 14.5 h (**B**). Each point represents the average of three replicates; error bars represent a standard deviation. Curves represent fittings to a logistic function [13].

The *tetR* gene was amplified from *E. coli* and placed behind the next strongest clostridial promoter (eno) directly upstream and in the reverse orientation of the P$_{gapDH/tetO}$—*lacZ* module. The expression of TetR driven by the eno promoter resulted in a ~6-fold reduction of β-gal activity, indicating that it was functioning as a repressor (Figure 6). We then tested the ability of aTc to alleviate the repression due to TetR. A concentration of 150 μg L^{-1} aTc was sufficient to fully alleviate TetR-dependent repression (Figure 6).
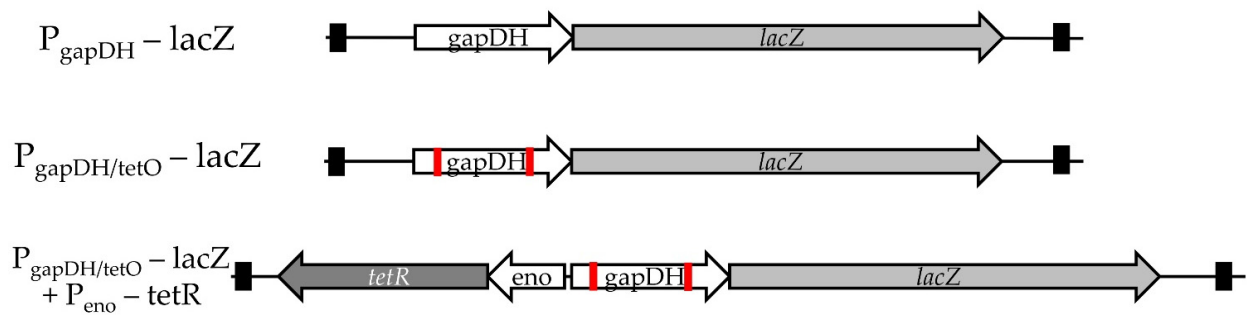
Figure 5. The schematics of each construct made to test the aTc inducible system. P_{gapDH}—*lacZ* is the starting point: the P_{gapDH} promoter driving the expression of *lacZ* (plasmid pAL66). $P_{gapDH/tetO}$—*lacZ* represents the insertion of a *tetO1* box into each $-35/-10$ region of the gapDH promoter (pAL88). $P_{gapDH/tetO}$—*lacZ* + P_{eno}—*tetR* represents the addition of the *tetR* gene driven by P_{eno} in the opposite orientation to the $P_{gapDH/tetO}$—*lacZ* gene (pAL111). White arrow boxes represent promoter sequences, while gray arrow boxes are open reading frames. Solid red boxes represent the *tetO1* boxes. Solid black boxes represent the terminators present at either end of the MCS in plasmid pMTL86251.
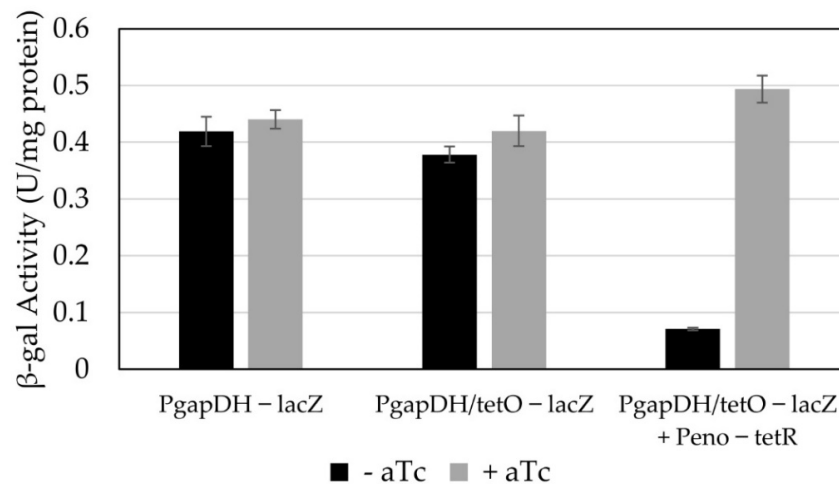
Figure 6. Anhydrotetracycline-inducible promoter construct plasmids and their associated β-gal activity in cultures with and without the addition of aTc (150 μg L^{-1}). The values represent the average specific activity in technical triplicates for P_{gapDH}—*lacZ* and $P_{gapDH/tetO}$—*lacZ* controls (*n* = 3), and the average rate of ONP production of biological and technical triplicates of the $P_{gapDH/tetO}$—*lacZ* + P_{eno}—*tetR* constructs (*n* = 9). The error bars represent the standard deviation of the sets.

3.4. Testing Use of a Xylose-Inducible Promoter System in H. modesticaldum

We also tested the efficacy of a previously developed xylose-inducible promoter [18] in *H. modesticaldum.* It was unclear if this could work as this species is unable to use any carbohydrates as a carbon or energy source, so we did not know if the cells could take up xylose [2,4]. Similar to the aTc system, the xylose-induced system involves an operator (*xylO*) within the promoter region, which will be bound by the repressor XylR in the absence of xylose. Once xylose is introduced, it will bind to the XylR, which will dissociate from the operator, allowing for the transcription of the downstream gene of interest [18]. We first tested the effect of xylose upon the growth of untransformed *H. modesticaldum*. At xylose concentrations below 0.35%, there was no discernible inhibition of growth, but at concentrations of 0.5% and higher, growth was noticeably inhibited (Figure 7). The origin of this effect is unclear at this time, but similar effects have been seen when growing *H. modesticaldum* with other carbohydrates. If pyruvate was removed from the medium, the cells were unable to grow on 0.35% xylose, indicating that xylose does not serve as a carbon source, as expected.
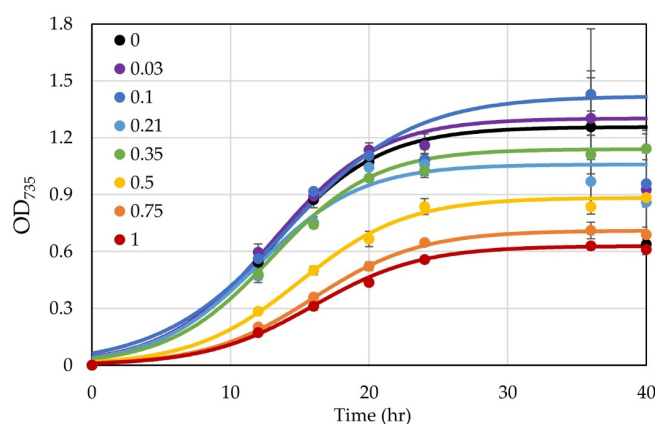
Figure 7. The growth of *H. modesticaldum* with varying concentrations of xylose from 0 to 1 %. Xylose concentrations (%) are displayed in the legend. Data points represent the average, and error bars represent the standard deviation of the biological triplicate (*n* = 3). Curves represent fittings of points to the logistic function [13].

The P_{xyl} promoter with the *xylR* gene (based on genes from the operon from *Clostridioides difficile*) was cloned in front of *lacZ* to make plasmid pALX3. *H. modesticaldum* transconjugants harboring pALX3 were grown with a range of xylose concentrations, and cellular extracts were prepared from log-phase cells. We found a strikingly linear correlation (R^2 = 0.944) between xylose concentration and β-gal activity between 0.03 and 0.3% xylose (Figure 8). The activities ranged from 0.02 to 0.16 U/mg protein. These activities place this promoter at the lower end of the strength scale. Concentrations of xylose above 0.3% yielded similar or lower promoter activity, with a peak around 0.5% xylose (Figure S1). This is most likely due to the negative effects of xylose upon growth.
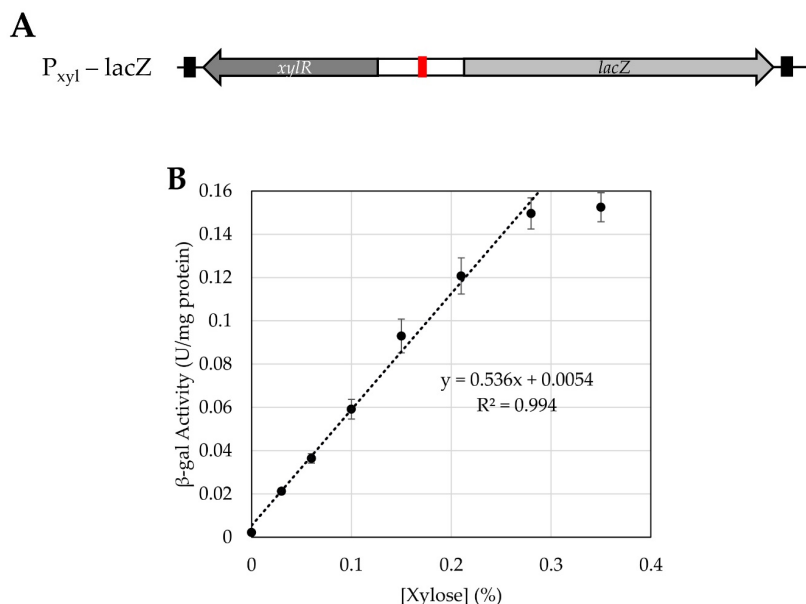
Figure 8. Xylose inducible promoter gene set-up (**A**) and promoter activity tuned by concentration of xylose present in media (**B**). In the schematic (**A**), the red box represents the operator, *xylO*, present in the promoter for the *lacZ* reporter. The *xylR* gene encoding the repressor is upstream and in the opposite orientation of this site; it is expressed constitutively. Terminators are present at either end of this sequence, represented as black boxes. In (**B**), specific activity is reported as β-gal units per mg protein versus xylose concentration (*w/v*). Each point represents the average of a biological and technical triplicate (*n* = 9), and error is represented as standard error.

4. Discussion

4.1. Reporters

From the alcohol dehydrogenase activities assays, we see that *H. modesticaldum* displays some endogenous alcohol dehydrogenase activity, which could have masked the exogenous alcohol dehydrogenase activity as a result of the introduced promoters. However, the alcohol dehydrogenase assays performed with *E. coli* did show a marked difference in activity as a result of the exogenous AdhB, despite also having an active endogenous alcohol dehydrogenase (AdhE). Therefore, the lack of detectable inducible alcohol dehydrogenase activity in *H. modesticaldum* suggests that the exogenous AdhB is inactive or very weakly active in this context. For either of these reasons—the presence of endogenous activity or the lack of detectable exogenous activity—*adhB* is useless as a reporter gene for *H. modesticaldum*.

In contrast, the thermostable β-galactosidase reporter was easily measured and could be used to indicate differences in promoter strength in *H. modesticaldum*. There is negligible endogenous β-gal activity displayed by WT or empty-vector-transformed *H. modesticaldum* samples, indicating that all measured β-gal activity is a result of the promoter driving gene expression. However, we did find evidence that β-galactosidase expression had a negative effect on growth when expressed at high levels (e.g., the under control of the gapDH promoter, which had the highest activity). While β-gal is not a perfect reporter, since it does trigger these negative growth effects, it is the one reporter we were able to find during the duration of this study that provided clear, measurable differences in *H. modesticaldum* conjugants. Unfortunately, we are unable to use other more common fluorescent reporters, such as GFP (which also requires oxygen for fluorophore synthesis), due to quenching by the endogenous pigments. We will continue to analyze other reporters for their use in *H. modesticaldum*, focusing on those that could be used for real-time in vivo monitoring.

4.2. Constitutive Promoters

Using β-galactosidase as a reporter, we were able to determine the relative activity of a set of clostridial promoters in *H. modesticaldum*. The promoters covered a range of strengths, making up two distinct groups—high and low strength—while also displaying many differences amongst key sequence elements. If one focuses on the ribosome binding sites, several strong promoters (cbp_2, gapDH, and 2926) share a similar sequence (AGGAGG). However, a weak promoter (2638) shares the same sequence, while a strong promoter (eno) has a truncated sequence (GGAG). Looking at the promoter sequences themselves does not provide insight into predictable markers of strength. The predicted ribosome binding sites, −35 regions, and −10 boxes vary amongst all of the promoters and do not exhibit discernible patterns amongst the weak or the strong promoters.

In comparison to *E. coli*, the promoters resulted in higher β-gal specific activity in *H. modesticaldum*. While *H. modesticaldum* and *E. coli* both have very similar ribosome binding sites, sharing AGGAGG, there must be more elements in the actual sequence of the promoter that dictate strength. Studies have shown that some clostridia are sensitive to the full sequence of the promoter, not just the −35 and −10 boxes, displaying drastic decreases in promoter strength when these sequences are altered [24]. Because *H. modesticaldum* is more closely related to *C. thermocellum*, the source of these promoters, than *E. coli* is, this may explain the higher activities in *H. modesticaldum* compared to *E. coli*. In addition, this sensitivity to the full promoter sequence may also explain why the activities of the promoters in *H. modesticaldum* are still different from those reported in *C. thermocellum* [16].

4.3. Inducible Promoters

We attempted to construct an inducible promoter for *H. modesticaldum* based on the anhydrotetracycline-inducible system developed and used for *C. acetobutylicum* [19]. To achieve this, we used several of the constitutive clostridial promoters to build various aspects of the system. The strongest promoter (P_{gapDH}) was used to construct the inducible promoter, assuming that insertion of *tetO1* and any additional alterations would

not strongly affect the promoter's strength. The insertion of the *tetO* sequences into both -35/-10 boxes of gapDH only resulted in a 7% decrease in the correlated β-gal activity.

It was unknown how active the promoter in front of the *tetR* gene needed to be to guarantee repression. Our initial study using a truncated version of the native *Clostridium perfringens* thiolase promoter ($P_{minipthl}$) [4] to control the expression of *tetR* suggested that it was too strong for the system, as the $P_{gapDH/tetO}$ promoter was strongly repressed but repression could not be relieved by aTc concentrations under which cells could grow (Figure S2). Therefore, we chose the next highest active promoter from our library of constitutive promoters (P_{eno}). However, even when driving TetR expression from P_{eno}, there was still residual $P_{gapDH/tetO}$ promoter activity in the absence of aTc (at a similar level to one of the weakest constitutive promoters, e.g., 2638). While alleviation of the repression was strong, resulting in almost complete restoration of the activity of the promoter to that of the unmodified P_{gapDH}, the breakthrough activity of this promoter renders it suboptimal for applications in which repression must be very tight (e.g., the expression of a potentially toxic enzyme). This system would be good, however, for high-level expression of proteins whose constitutive presence can be tolerated by *H. modesticaldum*.

Compared to the aTc system, the xylose-inducible promoter adapted from *C. perfringens* is simple and was easily introduced into *H. modesticaldum*. The *P_{xyl}* promoter was inducible under increasing xylose concentrations from 0.03% to about 0.3%, and *H. modesticaldum* was shown to tolerate these concentrations of xylose without affecting their growth. The fact that this species does not metabolize xylose should ensure a constant inducer concentration. Thus, the *P_{xyl}* promoter was inducible and highly tunable in the heliobacterial system. While the activity of the fully induced *P_{xyl}* promoter is not the strongest, it should be useful for precise control of long-term expression of a gene of interest. The lack of detectable background activity in the absence of xylose makes this a promising system for the introduction of potentially toxic proteins, or for the complementation of a null mutation in an essential gene.

References

1. Kimble, L.K.; Stevenson, A.K.; Madigan, M.T. Chemotrophic Growth of Heliobacteria in Darkness. *FEMS Microbiol. Lett.* **1994**, *115*, 51–56. [CrossRef] [PubMed]
2. Kimble, L.K.; Mandelco, L.; Woese, C.R.; Madigan, M.T. *Heliobacterium Modesticaldum*, Sp. Nov., a Thermophilic Heliobacterium of Hot Springs and Volcanic Soils. *Arch. Microbiol.* **1995**, *163*, 259–267. [CrossRef]
3. Orf, G.S.; Redding, K.E. The Heliobacteria. *Encycl. Biochem.* **2021**, 1–13.
4. Tang, K.-H.; Yue, H.; Blankenship, R.E. Energy Metabolism of *Heliobacterium Modesticaldum* during Phototrophic and Chemotrophic Growth. *BMC Microbiol.* **2010**, *10*, 150. [CrossRef] [PubMed]
5. Tang, K.H.; Feng, X.; Zhuang, W.Q.; Alvarez-Cohen, L.; Blankenship, R.E.; Tang, Y.J. Carbon Flow of Heliobacteria Is Related More to Clostridia than to the Green Sulfur Bacteria. *J. Biol. Chem.* **2010**, *285*, 35104–35112. [CrossRef]
6. Grégoire, D.S.; Lavoie, N.C.; Poulain, A.J. Heliobacteria Reveal Fermentation as a Key Pathway for Mercury Reduction in Anoxic Environments. *Environ. Sci. Technol.* **2018**, *52*, 4145–4153. [CrossRef]
7. Sattley, W.M.; Madigan, M.T.; Swingley, W.D.; Cheung, P.C.; Clocksin, K.M.; Conrad, A.L.; Dejesa, L.C.; Honchak, B.M.; Jung, D.O.; Karbach, L.E.; et al. The Genome of *Heliobacterium Modesticaldum*, a Phototrophic Representative of the Firmicutes Containing the Simplest Photosynthetic Apparatus. *J. Bacteriol.* **2008**, *190*, 4687–4696. [CrossRef]
8. Sheehy, D.; Kuang Lu, Y.; Osman, F.; Alattar, Z.; Flores, C.; Sussman, H.; Zaare, S.; Dooling, M.; Meraban, A.; Baker, P.; et al. Genome-Wide Transcriptional Response during the Shift to N_2-Fixing Conditions in *Heliobacterium modesticaldum*. *J. Proteom. Bioinform.* **2018**, *11*, 143–160. [CrossRef]
9. Baker, P.L.; Orf, G.S.; Khan, Z.; Espinoza, L.; Leung, S.; Kevershan, K.; Redding, K.E. A Molecular Biology Tool Kit for the Phototrophic Firmicute, *Heliobacterium modesticaldum*. *Appl. Environ. Microbiol.* **2019**, *85*, e01287-19. [CrossRef]
10. Baker, P.L.; Orf, G.S.; Kevershan, K.; Pyne, M.E.; Bicer, T.; Redding, K.E. Using the Endogenous CRISPR-Cas System of *Heliobacterium Modesticaldum* To Delete the Photochemical Reaction Center Core Subunit Gene. *Appl. Environ. Microbiol.* **2019**, *85*, e01644-19. [CrossRef]
11. Gisriel, C.; Sarrou, I.; Ferlez, B.; Golbeck, J.H.; Redding, K.E.; Fromme, R. Structure of a Symmetric Photosynthetic Reaction Center–Photosystem. *Science* **2017**, *357*, 1021–1025. [CrossRef] [PubMed]
12. Heinnickel, M.; Golbeck, J.H. Heliobacterial Photosynthesis. *Photosynth. Res.* **2007**, *92*, 35–53. [CrossRef] [PubMed]
13. Leung, S.W.; Baker, P.L.; Redding, K.E. Deletion of the Cytochrome Bc Complex from *Heliobacterium Modesticaldum* Results in Viable but Non-Phototrophic Cells. *Photosynth. Res.* **2021**, *148*, 137–152. [CrossRef] [PubMed]
14. Orf, G.S.; Redding, K.E. Expression and Purification of Affinity-Tagged Variants of the Photosynthetic Reaction Center from *Heliobacterium modesticaldum*. *Photosynth. Res.* **2019**, *142*, 335–348. [CrossRef]
15. Pyne, M.E.; Bruder, M.R.; Moo-Young, M.; Chung, D.A.; Chou, C.P. Harnessing Heterologous and Endogenous CRISPR-Cas Machineries for Efficient Markerless Genome Editing in Clostridium. *Sci. Rep.* **2016**, *6*, 25666. [CrossRef]
16. Olson, D.G.; Maloney, M.; Lanahan, A.A.; Hon, S.; Hauser, L.J.; Lynd, L.R. Identifying Promoters for Gene Expression in *Clostridium thermocellum*. *Metab. Eng. Commun.* **2015**, *2*, 23–29. [CrossRef]
17. Heap, J.T.; Pennington, O.J.; Cartman, S.T.; Minton, N.P. A Modular System for Clostridium Shuttle Plasmids. *J. Microbiol. Methods* **2009**, *78*, 79–85. [CrossRef]
18. Nariya, H.; Miyata, S.; Kuwahara, T.; Okabe, A. Development and Characterization of a Xylose-Inducible Gene Expression System for *Clostridium Perfringens*. *Appl. Environ. Microbiol.* **2011**, *77*, 8439–8441. [CrossRef]
19. Dong, H.; Tao, W.; Zhang, Y.; Li, Y. Development of an Anhydrotetracycline-Inducible Gene Expression System for Solvent-Producing *Clostridium acetobutylicum*: A Useful Tool for Strain Engineering. *Metab. Eng.* **2012**, *14*, 59–67. [CrossRef]
20. Lin, P.P.; Rabe, K.S.; Takasumi, J.L.; Kadisch, M.; Arnold, F.H.; Liao, J.C. Isobutanol Production at Elevated Temperatures in Thermophilic *Geobacillus thermoglucosidasius*. *Metab. Eng.* **2014**, *24*, 1–8. [CrossRef]
21. Ramos, J.; Martinez-Bueno, M.; Molina-Henares, A.; Teran, W.; Watanabe, K.; Zhang, X.; Gallegos, M.T.; Brennan, R.; Tobes, R. The TetR Family of Transcriptional Repressors. *Microbiol. Mol. Biol. Rev.* **2005**, *69*, 326–356. [CrossRef] [PubMed]
22. Bertram, R.; Hillen, W. The Application of Tet Repressor in Prokaryotic Gene Regulation and Expression. *Microb. Biotechnol.* **2008**, *1*, 2–16. [CrossRef] [PubMed]
23. Sprengel, R.; Hasan, M.T. Tetracycline-Controlled Genetic Switches. *Handb. Exp. Pharmacol.* **2007**, *178*, 49–72. [CrossRef]
24. Mordaka, P.M.; Heap, J.T. Stringency of Synthetic Promoter Sequences in Clostridium Revealed and Circumvented by Tuning Promoter Library Mutation Rates. *ACS Synth. Biol.* **2018**, *7*, 672–681. [CrossRef] [PubMed]

Article

Genomic Features of the Bundle-Forming Heliobacterium *Heliophilum fasciatum*

Kelly S. Bender [1], Michael T. Madigan [1], Kyleigh L. Williamson [1], Marisa H. Mayer [2], Mary N. Parenteau [2], Linda L. Jahnke [2], Paula V. Welander [3], Sophia A. Sanguedolce [4], Abigail C. Brown [4] and W. Matthew Sattley [4,*]

[1] Microbiology Program, School of Biological Sciences, Southern Illinois University, Carbondale, IL 62901, USA; benderk@siu.edu (K.S.B.); madigan@siu.edu (M.T.M.); kyleigh.williamson@siu.edu (K.L.W.)
[2] Exobiology Branch, NASA Ames Research Center, Moffett Field, CA 94035, USA; marisa.h.mayer@nasa.gov (M.H.M.); mary.n.parenteau@nasa.gov (M.N.P.); linda.l.jahnke@nasa.gov (L.L.J.)
[3] Department of Earth System Science, Stanford University, Stanford, CA 94305, USA; welander@stanford.edu
[4] Division of Natural Sciences, Indiana Wesleyan University, Marion, IN 46953, USA; sophia.sanguedolce@myemail.indwes.edu (S.A.S.); abbey.brown2@myemail.indwes.edu (A.C.B.)
* Correspondence: matthew.sattley@indwes.edu; Tel.: +1-765-677-2128

Abstract: Eight species of heliobacteria have had their genomes sequenced. However, only two of these genomes have been analyzed in detail, those from the thermophilic *Heliomicrobium (Hmi.) modesticaldum* and the alkaliphilic *Heliorestis (Hrs.) convoluta*. Here we present analyses of the draft genome sequence of a species of heliobacterium that grows optimally at a moderate temperature and neutral pH. The organism, *Heliophilum (Hph.) fasciatum*, is phylogenetically unique among cultured heliobacteria and was isolated from rice soil, a common habitat for heliobacteria. The *Hph. fasciatum* genome contains 3.14 Mbp—similar to that of other reported heliobacteria—but has a G+C base ratio that lies between that of *Hmi. modesticaldum* and *Hrs. convoluta*. Many of the genomic features of *Hmi. modesticaldum* and *Hrs. convoluta*, such as the absence of genes encoding autotrophic pathways, the presence of a superoperonal cluster of photosynthesis-related genes, and genes encoding endospore-specific proteins, are also characteristic of the *Hph. fasciatum* genome. However, despite the fact that *Hph. fasciatum* is diazotrophic, classical *nif* genes encoding the alpha and beta subunits of dinitrogenase (*nifDK*) present in other heliobacteria could not be identified. Instead, genes encoding several highly divergent NifDK homologs were present, at least one of which likely encodes a functional dinitrogenase and another a methylthio-alkane reductase (MarDK) for sulfur assimilation. A classical NifH (dinitrogenase reductase) homolog was also absent in *Hph. fasciatum*, but a related protein was identified that likely carries out this function as well as electron delivery to MarDK. The N_2-fixing system of *Hph. fasciatum* is therefore distinct from that of other heliobacteria and may have unusual properties.

Keywords: anoxygenic phototrophs; heliobacteria; bacteriochlorophyll *g*; *Heliophilum fasciatum*; genome sequence

1. Introduction

Heliobacteriaceae comprise a family of anoxygenic phototrophic bacteria phylogenetically and physiologically distinct from purple and green bacteria [1,2]. Unlike the latter two groups, heliobacteria do not contain bacteriochlorophyll (Bchl) *a*, *b*, *c*, *d*, or *e*, but instead, produce Bchl *g* as their main pigment, a bacteriochlorophyll structurally most closely related to green plant chlorophyll *a* [3]. In addition, heliobacteria also differ from purple and green bacteria in that heliobacteria are of gram-positive lineage, are obligate anaerobes, lack all known biochemical pathways for autotrophic growth, and produce heat-resistant endospores [4–6].

Several species of heliobacteria have been isolated since the discovery of *Heliobacterium (Hbt.) chlorum* by Gest and Favinger in the early 1980s [1,7] reviewed in [2,4,8], and described species form two broad groups on the basis of their pH optima for growth:

neutrophilic species, with pH optima near 7, and alkaliphilic species, with pH optima near 9. This physiological dichotomy of heliobacteria is mirrored in their phylogeny (Figure 1). Although several species of heliobacteria have had their genomes sequenced, thus far only two species—both extremophiles—have had their genomes analyzed in detail. These include the thermophilic (and neutrophilic) species *Heliomicrobium* (previously *Heliobacterium*) [8] *modesticaldum* [9–11] and the alkaliphilic species *Heliorestis convoluta* [12,13]. These studies have provided the genomic support for the aforementioned unusual physiological and biochemical features of the heliobacteria and confirmed that all characterized species (with the exception of *Candidatus* "Heliomonas lunata" [14]) are robust diazotrophs [15,16].
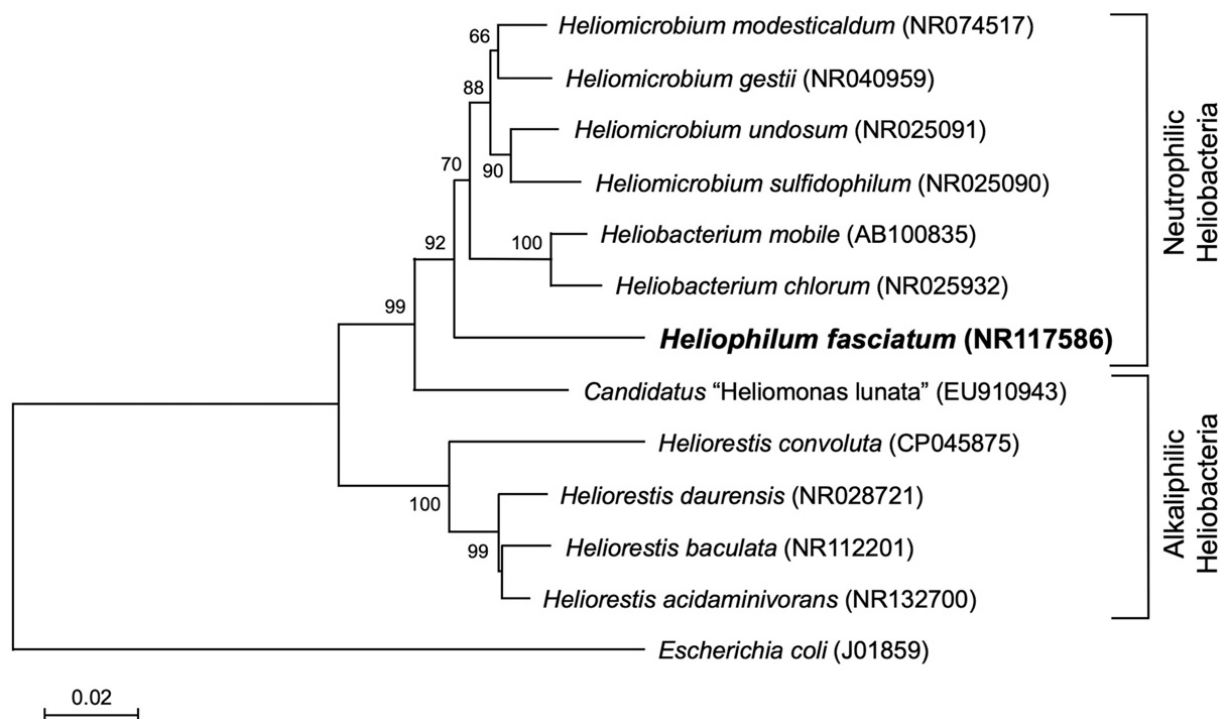
Figure 1. Phylogeny of *Heliophilum fasciatum* based on 16S ribosomal RNA gene sequence analysis. The multiple alignments included 16S rRNA gene sequences from all cultured type strains of the *Heliobacteriaceae*, as well as *Candidatus* "Heliomonas lunata" and the outgroup species *Escherichia coli*. The neighbor-joining tree was constructed from 1344 nucleotide positions. The scale bar indicates the number of base substitutions per site. Bootstrap values > 50 (500 replicates) are indicated at each node, and GenBank accession numbers are shown in parentheses.

In 1996, Ormerod et al. [16] described a new genus of bundle-forming heliobacteria in which a raft of apparently loosely attached rod-shaped cells displayed swimming motility as a unit (Figure 2). Besides exhibiting the usual assortment of heliobacterial physiological properties [5], this organism, named *Heliophilum fasciatum* ("sun-loving bundle-forming heliobacterium"), differed in several respects from *Hmi. modesticaldum* and *Hrs. convoluta*, most notably in its habitat and phylogenetic status (Table 1). By 16S rRNA gene sequencing criteria, *Hph. fasciatum* lies basal to all known neutrophilic heliobacteria and can be considered a "bridge" species between neutrophilic and alkaliphilic clades (Figure 1). All three of these heliobacteria assimilate acetate and pyruvate during photoheterotrophic growth and only *Hrs. convoluta* is unable to grow fermentatively in darkness (Table 1). In addition, each of these three species is morphologically distinct, shows distinct pH and temperature optima for growth, and was isolated from geographically well separated and geochemically distinct habitats (Tables 1 and 2).
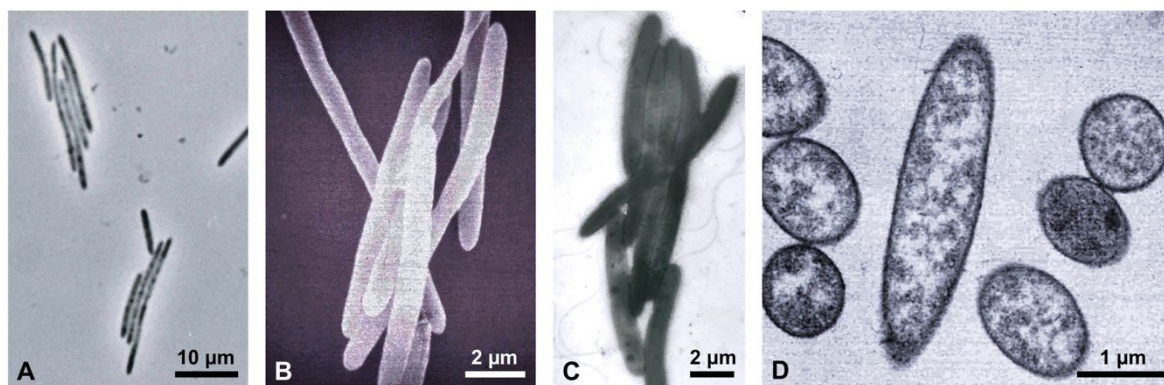
Figure 2. *Heliophilum fasciatum*. (**A**) Phase-contrast micrograph of cells formed into motile bundles. (**B**) Scanning electron micrograph of a cell bundle. (**C**) Negatively stained transmission electron micrograph with visible flagella. (**D**) Thin sectioned transmission electron micrograph showing absence of internal photosynthetic membranes, typical of the heliobacteria.

Table 1. Properties of *Heliophilum fasciatum* strain Tanzania[T] compared with those of *Heliorestis convoluta* strain HH[T] and *Heliomicrobium modesticaldum* strain Ice1[T].

Property [a]	Tanzania[T]	HH[T]	Ice1[T]
Cell morphology	Rods in bundles	Ring-shaped coils	Rods
Flagellar motility	Yes	Yes	Yes
Absorption maxima [b]	792 nm	786 nm	788 nm
Characteristic carotenoid	4,4′-diaponeu-rosporene	OH-diaponeurosporene glucoside ester	4,4′diaponeu-rosporene
Growth temperature optima (°C)/pH optima	37/7	33/8.5−9	50−52/6.5
Photoassimilation of acetate or pyruvate	Yes	Yes	Yes
Other C sources photoassimilated	Lactate, Butyrate [c], ethanol [c]	Propionate, butyrate	Lactate, glucose, fructose, ribose [d]
Chemotrophic (dark, fermentative) growth	Yes	No	Yes
16S rRNA gene sequence identity to Tanzania[T] (%)	100	90	93

[a] data from [9,12,16]; [b] absorption maxima for bacteriochlorophyll *g* in intact cells suspended in anoxic 30% bovine serum albumin; [c] in the presence of bicarbonate; [d] weak growth on sugars compared to growth on pyruvate or lactate [17].

Table 2. Heliobacteria with sequenced genomes.

Genus/Species	Habitat/pH Optimum	Protein-Encoding Genes [a]
Heliobacterium chlorum	Garden soil/7	3876 (NZ_JACVHF000000000.1)
Heliobacterium mobile	Paddy soil, Thailand/7	3699 (NZ_WNKU00000000.1)
Heliomicrobium modesticaldum	Hot springs, Reykjanes, Iceland/6-7 [b]	2662 (NC_010337.2)
Heliomicrobium gestii	Paddy soil, Dar es Salaam, Tanzania/7	3266 (NZ_WXEX00000000.1)
Heliomicrobium undosum	Microbial mat, Garga Hot Springs, Siberia/7.5	3356 (NZ_WXEY00000000.1)
Heliorestis acidaminivorans	Lake Hamra, Wadi Natroun Egypt/9	2765 (NZ_WBXO00000000.1)
Heliorestis convoluta	Lake Hamra, Wadi, Natroun Egypt/8.5	2909 (NZ_CP045875.1)
Heliophilum fasciatum	Paddy soil, Dar es Salaam, Tanzania/7	[c] 2834 (NZ_SLXT00000000.1)/2951

[a] known from complete genomes of *Hmi. modesticaldum* [10] and *Hrs. convoluta* [13] and predicted from draft genome sequences of the other species. Genbank reference sequence numbers listed in parentheses; [b] strain dependent [9]; [c] number preceding the slash is from Genbank annotation; number following the slash is from this study.

Despite the fact that *Hph. fasciatum* is phylogenetically and phenotypically distinct from *Hmi. modesticaldum* and *Hrs. convoluta* (Figure 1 and Table 1), this species has remained in the background of heliobacterial research until its genome was recently sequenced as part of a systematic study of the heliobacteria to support taxonomic revisions in the family [8].

Genomic studies of heliobacteria have lagged behind those of purple and green bacteria, and to date, a total of only eight species of heliobacteria have had their genomes sequenced, six of which only to the draft stage. Sequence data have shown the size of heliobacterial genomes to be quite variable, as the largest heliobacterial genome encodes more than a thousand additional proteins than does the smallest (Table 2). In a comparative study of the genomes of the eight species of heliobacteria listed in Table 2 [8], the authors showed that a phylogenetic tree constructed from the genomes of these species closely mirrored the tree constructed from 16S rRNA gene sequences (Figure 1). In addition, the authors used genomic data along with the comparative sequences of a few key genes and operons to propose that three of the species in the genus *Heliobacterium* should be grouped into their own genus, *Heliomicrobium* [8]; these include *Hmi. modesticaldum*, *Hmi. gestii*, and *Hmi. undosum* (Table 2). Analyses of genes encoding particular photocomplex and endospore proteins supported the taxonomic proposals, as did genes encoding carbon monoxide dehydrogenase, an enzyme previously unsuspected in the heliobacteria [8].

Thus far, detailed analyses of the genomes of only *Hmi. modesticaldum* and *Hrs. convoluta* have been published [10,11,13]. Here we augment these studies with an analysis of the major gene sets of *Hph. fasciatum* using data generated from our own sequencing project. Our results focus on genes encoding proteins of major pathways of carbon, nitrogen, and sulfur metabolism, and genes encoding the formation of endospores and flagellar components, comparing them where appropriate with those from other heliobacteria, in particular, the two species with complete genome sequences (Tables 1 and 2). Our analyses have uncovered several similarities and differences in the genetic capabilities of *Hph. fasciatum* compared with those of other heliobacteria, including in particular the fact that despite its diazotrophic phenotype, *Hph. fasciatum* lacks genes encoding the canonical nitrogenase universally present in all other cultured heliobacteria and in most other anoxygenic phototrophic bacteria.

2. Materials and Methods

The *Heliophilum fasciatum* strain Tanzania[T] was obtained from the Leibniz-Institut DSMZ, Braunschweig, Germany, as DSM 11170[T]. Cells of *Hph. fasciatum* were grown phototrophically in medium PYE [15], and genomic DNA was isolated using the JetFlex™ Genomic DNA Purification Kit (ThermoFisher Cat No. A30700). The genome was sequenced through the U.S. Department of Energy Joint Genome Institute (JGI) Community Science Program. Whole-genome shotgun sequencing was done with Illumina NovaSeq sequencing (library 300 bp), and the resulting fragments were assembled and annotated by the JGI Integrated Microbial Genomes (IMG) annotation pipeline (IMGAP v.5.0.23 with gene calling program CRT 1.8.2). The genome sequence of *Hph. fasciatum* Tanzania[T] is publically available (listed as *Heliophilum fasciatum* MTM in the JGI/IMG database) under the genome ID 2929297113. The genome of the exact same strain of *Hph. fasciatum* was independently sequenced at JGI by another group [8] and is publicly available as genome ID 2795386140; the latter sequence was also accessioned into Genbank as reference sequence NZ_SLXT00000000.1.

Phylogenetic analyses of 16S ribosomal RNA genes and protein sequences were performed using MEGA version X [18,19]. The *Hph. fasciatum* strain Tanzania[T] 16S rRNA gene sequence was aligned with corresponding sequences from other heliobacterial type species and *Escherichia coli* (J01859), used as the outgroup to root the phylogenetic tree. The tree was drawn using MEGA X according to the parameters described in the legend in Figure 1. For trees of *nif*-related proteins, *Hph. fasciatum* NflHDK amino acid sequences were used as queries in blastp [20] searches to identify other heliobacteria homologs. The genetic organization of the *Hmi. gestii* strain DSM 11169 *nif* and *anf* loci were obtained from Genbank genome accession PRJNA599378. Protein phylogenetic analyses were performed on a subset of NifHDK homologs previously designated as belonging to nitrogenase-like protein groups I–IV [21,22]. Following ClustalW [23] analysis, the resulting alignments were assembled into maximum-likelihood trees using the LG + G substitution model [24]

with 100 replicates. Note that the files for NifD and NifK homologs were concatenated prior to tree assembly.

3. Results and Discussion

3.1. General Genomic Properties

Table 3 compares genomic statistics of *Hph. fasciatum* strain Tanzania[T] with those of *Hmi. modesticaldum* and *Hrs. convoluta*. The genome of *Hph. fasciatum* consists of a single chromosome and was sequenced to 477X coverage and assembled to yield a total of 75 contigs. The estimated size of the *Hph. fasciatum* genome is 3,141,306 base pairs, and its G+C base ratio is approximately equidistant from that of *Hrs. convoluta*, whose genome is near the lowest of genomic G+C ratios in heliobacteria, and *Hmi. modesticaldum*, whose genome is near the highest (Table 3). The average nucleotide identity between the three species' genomes clearly reflected their phylogenetic positions (Figure 1), and the *Hph. fasciatum* genome is clearly the most gene-dense of the three species. Moreover, the percentage of hypothetical proteins encoded by the *Hph. fasciatum* genome is similar to that of *Hmi. modesticaldum*, and both of these were significantly lower than that of *Hrs. convoluta* (Table 3).

Table 3. Comparative genome statistics for *Heliophilum fasciatum* Tanzania[T], *Heliorestis convoluta*. HH[T], and *Heliomicrobium modesticaldum* Ice1[T a].

Characteristic	Tanzania[T]	HH[T]	Ice1[T]
Accession number	Ga043916 [b]	CP045875 [b]	CP000930 [b]
Genome size (bp) [c]	3,141,306	3,218,981	3,075,407
Contigs	75	1	1
Genome G+C (%)	50.9	43.1	56
Coding DNA (%)	95.7	90.1	90.6
Total ORFs	2951	3263	3138
Hypothetical proteins (%)	22.9	27.5	23.8
rRNAs	15	9	24
tRNAs	89	105	104
Average nucleotide identity (ANI) to genome of Tanzania[T] (%) [d]	100	66.7	68.8

[a] data for strain HH[T] taken from [13] and for strain Ice1[T] from [10]; [b] for strain Tanzania[T], this is the Gold Analysis Project ID, Joint Genome Institute Integrated. Microbial Genomes and Microbiomes (listed in the IMG as *Heliophilum fasciatum* MTM). For strain HH[T] and strain Ice1[T], these are accession numbers from Genbank; [c] each genome consists of a single chromosome with no plasmids. The genomes of strains HH[T] and Ice1[T] were each closed into a single contig and thus the genome size is precise. The strain Tanzania[T] genome is the draft sequence from a total of 75 contigs; [d] data from [8].

3.2. Phototrophy

Genes encoding phototrophy in heliobacteria are organized into an assemblage called the photosynthesis gene cluster (PGC) [10,11]. The arrangement of photosynthesis genes in the PGC of all heliobacteria sequenced thus far is nearly identical [8], and this trend extends to *Hph. fasciatum*. As in other heliobacteria, the *Hph. fasciatum bch* genes encoding biosynthetic proteins of the major heliobacterial pigment Bchl *g* were split into two sub-clusters consisting of *bchJGMELNBIDH* and *bchXZY*, with the gene encoding the homodimeric type I (Fe–S type) reaction center protein PshA positioned between *bchG* and *bchM*. Other genes, including those encoding proteins for carotenoid and cofactor biosynthesis and major aspects of cell division and electron transport, were arranged identically to their complements in other sequenced heliobacteria, including *Hmi. modesticaldum* and *Hrs. convoluta* [10,13].

In heliobacteria, the *bchJGMELNBIDH* and *bchXZY* gene clusters are separated by about 20 genes encoding a variety of proteins, most prominently those necessary for cofactor biosynthesis. However, an unusual feature of the PGC of *Hph. fasciatum* was the insertion between the two *bch* gene clusters of 11 additional genes encoding proteins unrelated to photosynthesis. This assortment of genes encodes a variety of proteins including the spore-coat protein CotF (discussed later), carbon monoxide dehydrogenase, a nucleotide-

binding protein for DNA uptake, a transcription regulator, and a transposase. Based on the heliobacterial genomes currently available and the phylogenetic divergence of *Hph. fasciatum* from other heliobacteria (Figure 1), this multiple-gene insertion may be unique to the PGC of this particular species.

3.3. Central Carbon Metabolism

Although all heliobacteria presumably exhibit a common mechanism for light-harvesting photochemistry (for a recent summary, see Orf and Redding [5]), carbon source utilization varies somewhat among species of *Heliobacteriaceae*. Although a few heliobacteria (e.g., *Hmi. gestii* and *Hmi. modesticaldum*) are able to use certain carbohydrates as a carbon source [16, 17], all heliobacteria catabolize pyruvate as a preferred carbon source, and most others can use yeast extract, lactate, or certain fatty acids for photoheterotrophic growth [4]. In addition to the best photoheterotrophic growth on pyruvate or lactate, good growth of *Hph. fasciatum* was demonstrated on acetate or butyrate plus CO_2, and weaker but measurable growth on ethanol plus CO_2 [16]. As is true for all other heliobacteria thus far studied, photoautotrophic growth (light/$H_2 + CO_2$) of *Hph. fasciatum* has not been demonstrated, and therefore, as a group, heliobacteria are phototrophic but not photosynthetic in the usual sense of this word since they lack a complete autotrophic pathway. Besides H_2, photoautotrophic growth of *Hph. fasciatum* using sulfide as an electron donor also was not observed [16]. For *Hph. fasciatum*, sulfide levels above 0.1 mM were growth inhibitory [16], and thus tests for photoautotrophic growth at concentrations above this were not performed.

The use of lactate as a carbon source by *Hph. fasciatum* is afforded by the presence of genes encoding both lactate permease and lactate dehydrogenase. In addition, genes encoding alcohol dehydrogenase, aldehyde dehydrogenase, and acetyl-CoA synthetase were all identified in the *Hph. fasciatum* genome and presumably account for the utilization of ethanol as a growth substrate, an ability shared with *Hmi. gestii*. Based on the presence of several key genes (e.g., those genes encoding coenzyme A transferase, acyl-(butyryl)-CoA dehydrogenase, enoyl-CoA hydratase, 3-hydroxybutyryl-CoA dehydrogenase, and acetyl-CoA C-acetyltransferase), butyrate metabolism in *Hph. fasciatum* appears to be identical to that described in *Hrs. convoluta* [13].

Although photoassimilation of propionate was not tested in the species description of *Hph. fasciatum*, the ability of this species to use propionate as a carbon source seems likely because genes encoding propionyl-CoA carboxylase, methylmalonyl-CoA epimerase, and methylmalonyl-CoA mutase were all identified in the *Hph. fasciatum* genome. This genetic profile matches that of *Hrs. convoluta* and *Hrs. acidaminivorans*, both of which photoassimilate and grow photoheterotrophically on propionate (notably, this synteny was also observed in the genomes of *Hmi. modesticaldum* and *Hmi. gestii*, species that have not been shown to utilize propionate but may also be able to do so). It is likely that this same gene complement also occurs in *Hrs. daurensis*, a species in which growth on propionate was also observed [25] but for which no genome sequence is available. Although propionate did not support photoheterotrophic growth of either *Hrs. baculata* [26] or *Hmi. sulfidophilum* [27], it did serve as a suitable carbon source for most species of *Heliorestis*, as well as *Hmi. undosum* [27]. Considering the prevalence of these genes in heliobacteria, propionate assimilation may be more widespread among heliobacteria than previously thought.

According to culture-based studies, all neutrophilic heliobacteria can grow by pyruvate fermentation in darkness [4,5]. By contrast, alkaliphilic heliobacteria, including all species of *Heliorestis* and *Candidatus* "Heliomonas lunata" (Figure 1), cannot grow in darkness and thus appear to be obligate photoheterotrophs. Depending on the species of heliobacterium, pyruvate is fermented with or without the production of H_2 [28], suggesting that pyruvate is metabolized by either pyruvate:ferredoxin oxidoreductase (H_2 produced) [29] or pyruvate:formate lyase (no H_2 produced). Genes encoding pyruvate:ferredoxin oxidoreductase are present in the *Hph. fasciatum* genome, although the biochemistry of pyruvate fermentation in this species has not been studied. However, the absence of key genes encoding a

H$_2$-evolving [Fe-Fe] hydrogenase in all genome-sequenced *Heliomicrobium* species (Table 2) and in *Hph. fasciatum* suggests that these species do not evolve H$_2$ while fermenting pyruvate. Curiously, however, as in species of *Heliomicrobium* [5], the genome of *Hph. fasciatum* contained a homolog of *hydA*, which encodes the H$_2$-evolving HydA polypeptide of the [Fe-Fe] hydrogenase, but lacked the *hydEFG* maturase genes required to synthesize a functional enzyme. By contrast, both *hydA* (two copies) and *hydEFG* are present in the genomes of *Hbt. chlorum* and *Hbt. mobile* and the latter species evolves H$_2$ during pyruvate fermentation [28]. Surprisingly though, *Hrs. convoluta* and *Hrs. acidaminivorans*—alkaliphiles that do not ferment pyruvate—also possess two copies of *hydA* and one copy of *hydEFG*. Thus, the distribution (or indeed, the very existence) of functional [Fe-Fe] hydrogenases in heliobacteria requires further study. In contrast to an [Fe-Fe] H$_2$-evolving hydrogenase, genes encoding a [Ni-Fe] uptake-type hydrogenase are present in the *Hph. fasciatum* genome and in all heliobacteria with sequenced genomes (Table 2).

Although the genome of *Hph. fasciatum* encodes no obvious mechanism for carbohydrate uptake, such as the ribose ABC transporter identified in *Hmi. modesticaldum* and *Hrs. convoluta* [10,13], genes encoding complete glycolytic and nonoxidative pentose phosphate pathways were identified. As is true for all other sequenced heliobacteria, genes encoding glucose 6-phosphate dehydrogenase and 6-phosphogluconolactonase were absent in *Hph. fasciatum*, suggesting that the possession of incomplete Entner–Doudoroff and oxidative pentose phosphate pathways is universal among the *Heliobacteriaceae*.

Genes encoding enzymes of a complete citric acid cycle (CAC) were identified in *Hph. fasciatum*. The unusual citrate (*re*)-synthase identified in the genomes of both *Hmi. modesticaldum* [10] and *Hrs. convoluta* [13] was also encoded in the *Hph. fasciatum* genome. Described in *Hmi. modesticaldum* by Tang et al. [17], this unorthodox form of citrate synthase may have ancestral lineage within nonphototrophic clostridia, in which the enzyme is also found [5], and is presumably a common feature of heliobacterial central carbon metabolism.

Like other heliobacteria, *Hph. fasciatum* lacks a gene encoding citrate lyase and is, therefore, incapable of autotrophic growth using the reverse CAC present in green sulfur bacteria. Likewise, key genes of the Calvin–Benson cycle are absent in the *Hph. fasciatum* genome. Although a gene homologous to the C-terminal domain of the ribulose-bisphosphate carboxylase (RuBisCO) large subunit (*cbbL*) was identified in the genomes of *Hph. fasciatum*, *Hbt. mobile*, *Hbt. chlorum*, and *Hmi. undosum*, (but not in the genomes of *Hmi. modesticaldum*, *Hmi. gestii*, *Hrs. convoluta*, and *Hrs. acidaminivorans*), no gene encoding the corresponding small subunit of RuBisCO (*cbbS*) was identified. Moreover, as is the case with other heliobacteria, no gene encoding phosphoribulokinase was present in the genome of *Hph. fasciatum*.

Presumably, anaplerotic CO$_2$ assimilation through the activity of phosphoenolpyruvate (PEP) carboxykinase (encoded by *pckA*) is an important mechanism of non-autotrophic CO$_2$ fixation for all heliobacteria [5,17]. Interestingly, however, although *pckA* is present in the genomes of all other heliobacteria for which a sequence is available, no homolog encoding PEP carboxykinase could be identified in the genome of *Hph. fasciatum*. By contrast, as previously mentioned, the *Hph. fasciatum* genome does encode pyruvate:ferredoxin oxidoreductase, and in addition to evolving CO$_2$ during pyruvate fermentation, this enzyme can assimilate CO$_2$ by carboxylating acetyl-CoA to pyruvate. To replenish CAC intermediates drawn off for biosynthetic purposes in *Hph. fasciatum*, oxaloacetate can be synthesized from pyruvate using an encoded ATP-dependent pyruvate carboxylase.

Nearly all heliobacteria require biotin as a growth factor, with the only documented exceptions being *Hrs. convoluta* and *Hrs. acidaminivorans* [12,14]. The genomes of both of these species contain the full complement of biotin biosynthesis genes (*bioABCDF*). By contrast, the genomes of most other sequenced species of heliobacteria contain *bioABD* but lack genes encoding BioC (malonyl-[acyl-carrier protein] O-methyltransferase) and BioF (8-amino-7-oxononanoate synthase), enzymes required for biotin biosynthesis. Among sequenced heliobacteria, *Hph. fasciatum* is uniquely deficient in regard to biotin biosynthesis,

as its genome contains *bioB*, which encodes biotin synthase but lacks all other genes of the biotin (*bio*) operon.

3.4. Nitrogen Metabolism: Utilization of Fixed Nitrogen

Heliophilum fasciatum grows using ammonia, glutamine, yeast extract, or dinitrogen as nitrogen sources; glutamate or aspartate are not used [16]. Two separate and divergent genes encoding ammonium transporter (Amt) proteins are present to import ammonia. Genes encoding the common glutamine synthetase–glutamate synthase pathway are also present, allowing for the incorporation of ammonia into key organic nitrogen compounds. In addition, a gene encoding the substrate-binding domain of an ABC glutamine transporter is present and may account for the organism's ability to use externally supplied glutamine. While *Hph. fasciatum* has not been shown to use nitrite as a nitrogen source, and nitrite transporters have not been identified in any heliobacterium [5], a NrfAH-type nitrite reductase is encoded in the *Hph. fasciatum* genome. By contrast, genes encoding an assimilative nitrate reduction system could not be identified.

3.5. Nitrogen Metabolism: Nitrogen Fixation

Although the genomics supporting the utilization of fixed nitrogen compounds by *Hph. fasciatum* is not unusual, such is not true of the genomics supporting its diazotrophic growth. A nitrogenase system has been detected by acetylene reduction in N_2-grown cells of *Hph. fasciatum*, albeit activities are expressed at lower levels than in cells of *Heliomicrobium gestii* or *Heliomicrobium mobile* [15,16]. Six-genes encoding NifHDK (dinitrogenase reductase and the α- and β-subunits of dinitrogenase, respectively) and NifENB (nitrogenase assembly and maturase proteins) are present in virtually all diazotrophs that produce a FeMo-cofactor-containing nitrogenase and are considered the minimal gene set required for a diazotrophic phenotype [30]. Curiously, however, experiments to detect genes encoding dinitrogenase reductase (*nifH*) and the α subunit of dinitrogenase (*nifD*) in *Hph. fasciatum* by Southern blot and PCR analyses were unsuccessful, suggesting that a unique enzyme or pathway for nitrogen fixation might exist in this phototroph [31]. The *Hph. fasciatum* genome supports this hypothesis by revealing genes encoding five distinct NifDK-like proteins that contain the pfam 00148 oxidoreductase domain present in the α- and β-subunits of all known FeMo-cofactor-containing dinitrogenases [30]. All genes encoding these *n*itrogen *f*ixation-*l*ike proteins in *Hph. fasciatium* (genes and proteins abbreviated *nfl* and Nfl, respectively, and numbered 1–5) are clustered, with genes encoding NflDK$_{1-4}$ located at the end of the contig Ga0493917_07 and those encoding NflDK$_5$ positioned near the beginning of contig Ga0493917_01 (Figure 3A). In contrast to this arrangement, a "traditional *nif* cluster" containing *nifI*$_1$, *nifI*$_2$, *nifH*, *nifD*, *nifK*, *nifE*, *nifN*, *nifX*, *fdxB*, *nifB*, and *nifV* present in the genomes of *Hmi. modesticaldum* [10], *Hrs. convoluta* [13], and *Hmi. gestii* (Figure 3B), is absent from the *Hph. fasciatum* genome. Instead, single copies of *nfl* genes similar to *nifB*, *nifH*, *nifX*, and *fdxB* are present in the same chromosomal locus in *Hph. fasciatum* as genes encoding NflDK$_1$ (Figure 3A). A NifV (homocitrate synthase) homolog is encoded elsewhere in the *Hph. fasciatum* chromosome; however, homologs of NifE and NifN could not be identified. Additionally, genes encoding an alternative nitrogenase, such as the FeFe-cofactor nitrogenase of *Hmi. gestii* (Figure 3C), were also absent from *Hph. fasciatum*.
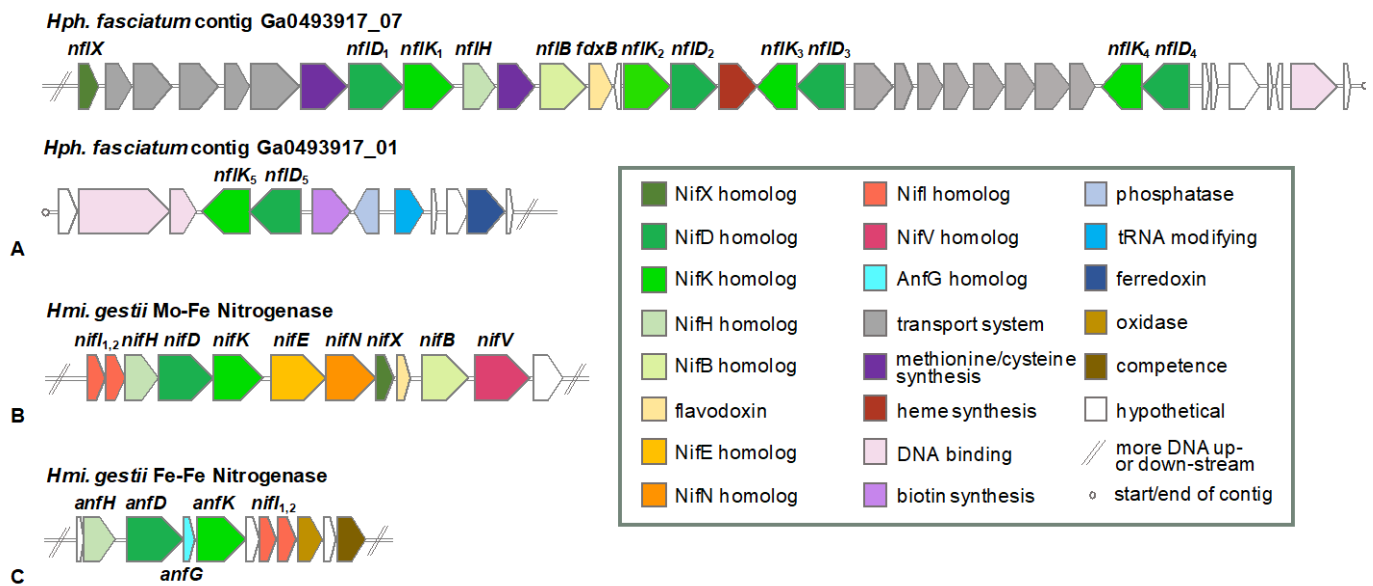
Figure 3. Genetic organization of nitrogen fixation-related genes and surrounding genes. (**A**) *Hph. fasciatum* loci containing genes encoding the nitrogen fixation-like homologs NflDK$_{1-5}$. The following JGI/IMG coordinates are illustrated: Ga0493917_07: 106,630 to 145,155 and Ga0493917_01: 75 to 11,134; (**B**) *Hmi. gestii* locus containing genes encoding the Mo-Fe nitrogenase (NifHDK). The following Genbank coordinates are illustrated: NZ_WXEX01000020.1: 13,816 to 26,039; (**C**) *Hmi. gestii* locus containing genes encoding the alternative Fe-Fe nitrogenase (AnfHDGK). The following Genbank coordinates are illustrated: NZ_WXEX01000005.1: 245,244 to 257,234. Predicted functions are described in the key.

The apparent absence of genes encoding NifEN in *Hph. fasciatum* is of particular interest but is not without precedence. FeMo nitrogenases lacking NifE or NifN (or both) are thought to be "ancestral nitrogenases", enzymes that employ a truncated cofactor assembly pathway and catalyze an activity in addition to N_2 reduction [30,32]. For example, the diazotrophic methanogen *Methanocaldococcus* sp. strain FS406-22 and bacterium *Caldicellulosiruptor* sp. strain YA01 synthesize nitrogenases that lack NifN but still contain NifE [33,34]. However, the genome of *Endomicrobium proavitum* strain Rsa215, a diazotrophic free-living termite gut bacterium, is missing genes encoding both NifE and NifN [35] indicating that a functional nitrogenase can be synthesized that lacks these proteins.

Because the NifEN proteins are hypothesized to be encoded by genes originating from a *nifDK* duplication [36], the *Hph. fasciatum* NflDK$_{1-5}$ homologs (Figure 3A) were compared to NifDK and NifEN homologs from other heliobacteria as well as from the model diazotrophs *Azotobacter vinelandii* and *Rhodospirillum rubrum*. The results showed that none of the five *Hph. fasciatum* NflDK proteins formed a clade with NifEN homologs (data not shown) or with NifDK (Figure 4A). Thus, despite possessing genes encoding molybdate transport (ModA) and molybdenum cofactor synthesis (Moa/Moe), *Hph. fasciatum* is the only known species of heliobacteria to lack both NifEN and an ortholog of the classical NifDK nitrogenase.

A search for NflDK$_{1-5}$ homologs in other species of heliobacteria identified two such proteins in *Hbt. chlorum*, four in *Hbt. mobile*, and five in *Hmi. gestii* (Figure 4B). In addition to the five *Hmi. gestii* NflDK homologs, the *Hmi. gestii* genome encodes a classical NifDK as well as a homolog of the Fe-Fe alternative dinitrogenase (AnfDK) of *Azotobacter vinelandii* and *Rsp. rubrum* (Figures 3B,C and 4A,B) (alternative nitrogenases lack Mo but contain Fe-Fe or V-Fe cofactors and are thought to be redundant enzymes that support nitrogen fixation when Mo is limiting [37]). Our discovery of *anfDK* in *Hmi. gestii* (Figure 3C) complements a previous finding of an *anfH*-like gene in this species [32] and physiological evidence for an Fe-Fe nitrogenase system [38]. Moreover, an

examination of the chromosomal locus encoding AnfHDK in *Hmi. gestii* also indicated the presence of a gene encoding the δ-subunit (*anfG*) of its Fe-Fe nitrogenase (Figure 3C).
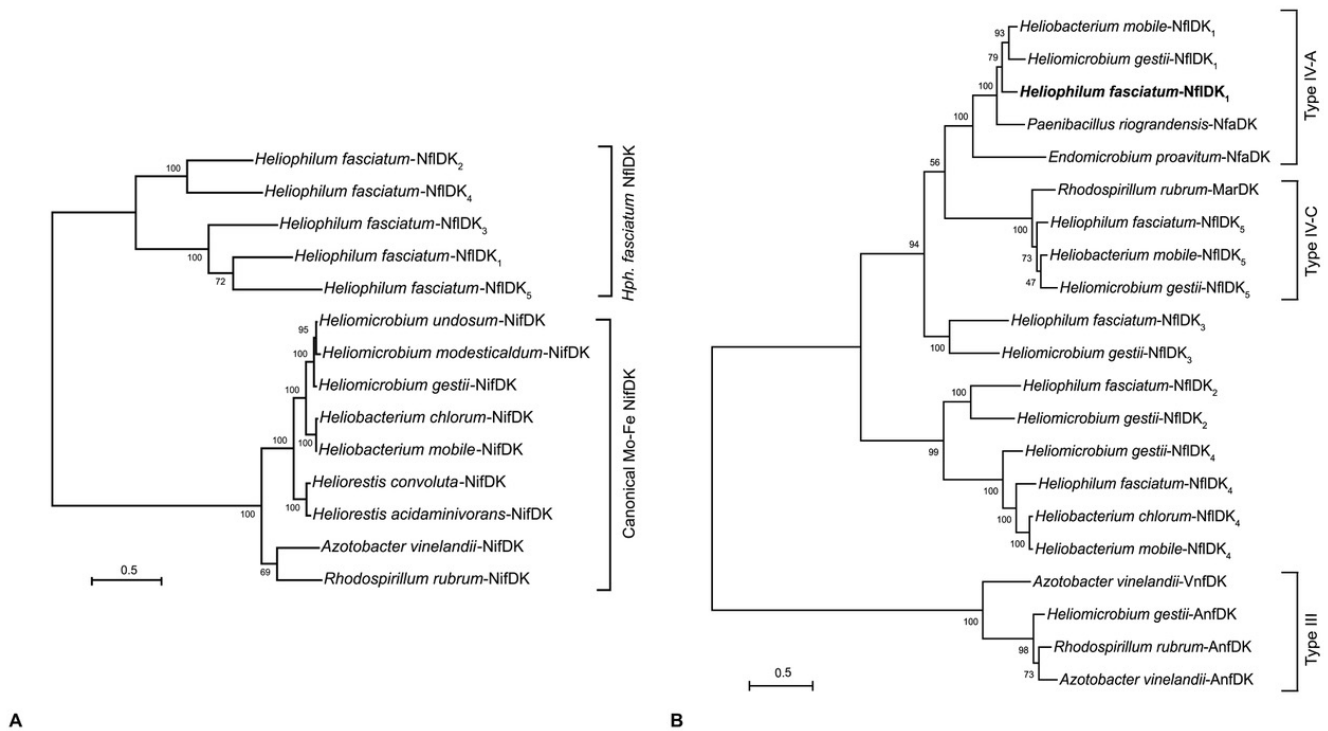
Figure 4. Maximum-likelihood phylogenetic trees of *Hph. fasciatum* concatenated NflDK sequences. (**A**) Comparison with canonical Mo-Fe nitrogenase NifDK sequences. (**B**) Comparison with Type III and IV nitrogen fixation-like sequences. The nodes represent bootstrap values based on 100 replicates, and scale bars indicate 0.5 changes per position. Taxa accession numbers are as follows with those beginning with a number corresponding to the JGI/IMG database and those beginning with a letter from Genbank: *Hph. fasciatum*—NflDK$_1$ (2929298242, 2929298243), NflDK$_2$ (2929298250, 2929298249), NflDK$_3$ (2929298253, 2929298252), NflDK$_4$ (2929298263, 2929298262), NflDK$_5$ (2929297118, 2929297117); *Hmi. gestii*—NifDK (WP_161263156, WP_161263155), AnfDK (WP_161261510, WP_161261560), NflDK$_1$ (WP_161260453, WP_161260452), NflDK$_2$ (WP_161260457, WP_161260458), NflDK$_3$ (WP_161260450, WP_161260449), NflDK$_4$ (WP_161260444, WP_161260443), NflDK$_5$ (WP_161260462, WP_161260461); *Hmi. modesticaldum*—NifDK (641558455, 641558456); *Hmi. undosum*—NifDK (WP_161255976, WP_161255978); *Hrs. acidaminivorans*—NifDK (2914075535, 2914075536); *Hrs. convoluta*—NifDK (QGG48552, QGG48553); *Hbt. chlorum*—NifDK (WP_188040647, WP_188040648), NflDK$_4$ (WP_188039480, WP_188039481); *Hbt. mobile*—NifDK (WP_155476316, WP_155476317), NflDK$_1$ (WP_155475363, WP_155475362), NflDK$_4$ (WP_155475372, WP_155475371), NflDK$_5$ (WP_155475353, WP_155475354); *A. vinelandii*—NifDK (643803063, 643803064), AnfDK (643807739, 643807737), VnfDK (643803186, 643803184); *Rsp. rubrum*—NifDK (637825100, 637825101), AnfDK (637825484, 637825482), MarDK (637824885, 637824884); *E. proavitum*—NfaDK (WP_052570612, WP_052570613); *P. riograndensis*—NfaDK (WP_020429361, WP_046504163).

Phylogenetic studies of nitrogenases have revealed at least six groups (designated I–VI) of "*nif*-like" genes distributed throughout *Bacteria* and *Archaea*, with members of groups IV–VI predicted to encode alternative functions such as roles in (bacterio)chlorophyll biosynthesis or archaeal methyl coenzyme-M reductase cofactor F_{430} biosynthesis, as well as various aspects of sulfur metabolism [21,22,31,35,39–43]. Phylogenetic analysis of the *Hph. fasciatum* NflDK$_{1–5}$ homologs indicates that they do not form a clade with Nif, Anf, or Vnf nitrogenases (groups I–III). Instead, the NflDK$_{1–5}$ proteins are more similar to the group IV MarDK, NflDK, and NfaDK (*nitrogen fixation IV—subgroup A*) proteins of *Rhodospirillum rubrum*, *Paenibacillus riograndensis*, and *Endomicrobium proavitum*,

respectively [22,35] (Figure 4B). Whereas the MarDK proteins of *Rsp. rubrum* do not function as a nitrogenase [22], and the role of the NfaDK proteins in *P. riograndensis* is unknown (*P. riograndensis* also possesses canonical Nif and alternative Anf nitrogenases) [44], the *E. proavitum* NfaDK homolog falls within subgroup IV-A and is predicted to possess all the ligands required for N_2 reduction [35]. Moreover, despite lacking homologs to NifEN, cells of *E. proavitum* incorporate $^{15}N_2$ into biomass, presumably through the activity of NfaDK, the only *nif*-like genes present in the genome [35]. Therefore, based on its placement in a clade with *E. proavitum* NfaDK (Figure 4B) and the absence of genes encoding a classical nitrogenase system (Figure 4A), we predict that diazotrophy in *Hph. fasciatum* is supported by its NfaDK homolog $NflDK_1$. If true, and if such an enzyme is a less robust nitrogenase than are classical nitrogenases, this could explain the lower nitrogenase activities observed in cells of *Hph. fasciatum* compared with those of other heliobacterial species [15,16].

While the roles of *Hph. fasciatum* $NflDK_{2-4}$ are unknown, it is notable that the *Hph. fasciatum*, *Hmi. gestii*, and *Hbt. mobile* $NflDK_5$ homologs form a clade with the *Rhodospirillum rubrum* MarDK proteins of the MarBHDK system that synthesizes methionine from volatile organic sulfur compounds (VOSCs) such as dimethyl sulfide and (2-methylthio)ethanol [22] (Figure 4B). MarDK are group IV-C nitrogenase-like proteins encoding the methylthio-alkane reductase that cleaves the C–S bond of VOSCs. Utilization of the MarBHDK pathway to assimilate sulfur from organic sources may compensate for the apparent lack of a sulfite reduction system in *Hph. fasciatum* (see Section 3.5). *Hmi. gestii*, *Hbt. mobile*, and *Hbt. chlorum* also possess an additional *nifH*-like (*nflH*) gene linked to the genes encoding these MarDK homologs. These NifH homologs form a clade with the MarH protein of *Rsp. rubrum* (Figure 5) and likely encode the electron-delivering subunit of a methylthio-alkane reductase. *Hph. fasciatum* possesses only one *nifH*-like gene, and it lies adjacent to the genes encoding $NflDK_1$ in the genome (Figure 3A). Surprisingly, however, this NifH-like protein along with NfaH proteins of *P. riograndensis* and the *E. proavitum* nitrogenase form a clade with MarH instead of with NifH from other heliobacteria or AnfH or VnfH of Fe-Fe and V-Fe nitrogenases (Figure 5). All of these proteins possess the same critical residues necessary for reductive activity, which may allow for their cross-reactivity.

How the *Hph. fasciatum* and *E. proavitum* genomes have come to share these divergent NifHDK homologs is puzzling. Whereas both microbes are obligate anaerobes and fermentative, one is phototrophic while the other is symbiotic, and thus their ecology is unrelated. Why *Hph. fasciatum* is the only known heliobacterium to lack a conventional nitrogenase is also an intriguing question. Although we acknowledge that our analysis is not based on a single-contig closed genome, we nevertheless feel, for three reasons, that genes encoding a canonical nitrogenase system in *Hph. fasciatum* are truly absent. First, sequencing coverage of the *Hph. fasciatum* genome was high (477X), thus reducing the chance of losing or overlooking multi-gene operons. Second, experiments by others using PCR and Southern blots to detect *nifH* and *nifD* in *Hph. fasciatum* were negative, although both genes were easily identified in other heliobacterial species [31]. Finally, using degenerate primers targeting *nifD*, *anfD*, and *vnfD* [45], no PCR amplicons were obtained from *Hph. fasciatum* genomic DNA, although a canonical *nifD* was obtained from two other heliobacterial species used as controls (K.S.B., unpublished results).

Another interesting aspect of the *Hph. fasciatum* nitrogenase story is the lack of genes encoding obvious nitrogen metabolism regulators. A gene encoding a P-II nitrogen regulatory protein is present upstream of a gene encoding an ammonia transporter (2929297543), but this is not close to the loci encoding the $NflDK_{1-5}$ proteins. Further studies are thus needed to determine the expression patterns of the *Hph. fasciatum* $NflDK_{1-5}$ homologs, as well as mutational and $^{15}N_2$ incorporation experiments to confirm the nitrogenase activity of *Hph. fasciatum* $NflDK_1$ and test the N_2-reducing ability of mutant derivatives lacking one or more homologs. The significance of encoding multiple NifDK-type proteins is also unknown, but it is not unusual for genomes encoding group IV nitrogenase-like enzymes

to contain multiple copies of *nifDK*-like genes [43]. It is therefore possible that these genes encode important biochemical functions that are yet to be recognized.
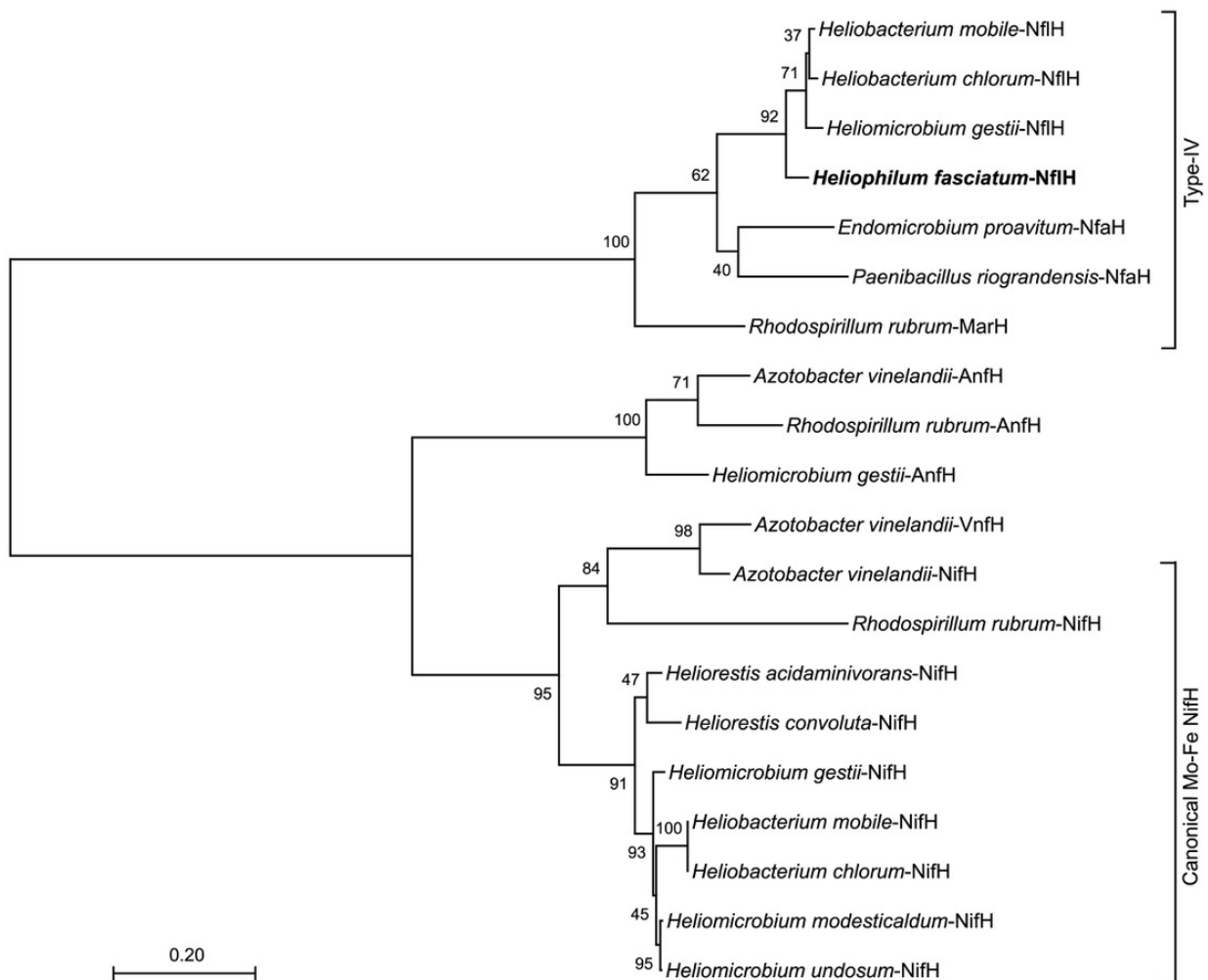
Figure 5. Maximum-likelihood phylogenetic tree of the *Heliophilum fasciatum* NflH sequence. The nodes represent bootstrap values based on 100 replicates and the scale bar indicates 0.2 changes per position. Taxa accession numbers are as follows with those beginning with a number corresponding to the JGI/IMG database (those beginning with a letter are from Genbank): *Hph. fasciatum*—NflH (2929298244); *Hmi. gestii*—NifH (WP_161263157.1), AnfH (WP_170294419), NflH (WP_161260455); *Hmi. modesticaldum*—NifH (641558454); *Hmi. undosum*—NifH (WP_161255974.1); *Hrs. acidaminivorans*—NifH (2914075534); *Hrs. convoluta*—NifH (QGG48551.1); *Hbt. chlorum*—NifH (WP_155476315.1), NflH (WP_188039485.1); *Hbt. mobile*—NifH (WP_155476315.1), NflH (WP_155475351.1); *A. vinelandii*—NifH (643803062), AnfH (643807740), VnfH (643803191); *Rsp. rubrum*—NifH (637825099), AnfH (637825485), MarH (637824886); *E. proavitum*—NfaH (WP_052570620.1); *P. riograndensis*—NfaH (WP_020429374.1).

3.6. Sulfur Metabolism

Unlike many heliobacteria that require a reduced source of sulfur for biosynthetic purposes [4,5,9], *Hph. fasciatum* grows in defined media with sulfate as the only sulfur source [16]. In this regard, the *Hph. fasciatum* genome encodes the widely distributed CysPTWA-mediated sulfate uptake system: CysNDC proteins that make adenylphosphosulfate (APS) and phosphoadenylphosphosulfate (PAPS) from sulfate, and AprAB and CysH that make sulfite from APS and PAPS, respectively. However, the *Hph. fasciatum* genome lacks homologs of CysJI, the classical alpha-beta sulfite reductase that reduces sulfite to sulfide in a wide variety of bacteria. Two genes were present that encoded

a protein weakly similar to CysI (beta subunit of sulfite reductase). These genes (IDs 292929318 and 2929297321), were nearly identical and were annotated as "sulfite reductase beta subunit-like" but were only 20% identical to authentic CysI from *Escherichia coli* or *Bacillus subtilis*. Although it is possible that the *Hph. fasciatum* CysI-like protein participates in sulfite reduction, the absence of a homolog encoding the alpha subunit (CysJ) of sulfite reductase leaves open the question of exactly how sulfite reduction occurs in *Hph. fasciatum*.

Most heliobacteria will grow in media containing sulfide, and some species are remarkably sulfide tolerant [9,14]. *Hph. fasciatum* is an exception and is completely growth inhibited by as little as 0.1 mM sulfide [16]. Consistent with this, *Hph. fasciatum* lacks the common sulfide oxidation system—sulfide:quinone oxidoreductase—of purple and green sulfur bacteria [46]. In addition, *Hph. fasciatum* lacks *sox* genes (as do all known heliobacteria), which encode enzymes necessary to oxidize thiosulfate, the other reduced inorganic sulfur compound commonly oxidized as an electron donor by anoxygenic phototrophs. Thus, once sulfide is available in the *Hph. fasciatum* cell, from either the reduction of sulfate or from organic sources, it is likely to be quickly incorporated into sulfur-containing amino acids. In this regard, *Hph. fasciatum* is well equipped. Its genome encodes cysteine synthase that makes cysteine from serine and sulfide, as well as all genes necessary to convert aspartate to homocysteine, from which methionine is made. In addition, the genome encodes an ABC transporter for methionine and an ABC system for the uptake of branched-chain amino acids. However, as for methionine, all *Hph. fasciatum* genes necessary for the biosynthesis of valine, leucine, and isoleucine could be identified.

3.7. Motility

Cells of *Hph. fasciatum* are large rods (Figure 2) that display an unusual form of motility, unique among heliobacteria. *Hph. fasciatum* cells form bundles of parallel cells resembling the rafts of *Bacillus* or *Proteus* that form in biofilms [47,48]; the *Hph. fasciatum* bundles are motile as a unit and move with a rolling motion [16]. Bacterial flagella consist of three major components: the basal body, which traverses the cytoplasmic membrane (and if present, the outer membrane), the hook complex, and the filament, the rotating component that imparts cell motility [49]. Although both gram-positive and gram-negative bacterial flagella contain a basal body and hook that connects with the filament, gram-positive species lack the L and R rings that imbed the flagellum into the outer membrane of gram-negative bacteria [50]. In agreement with the *Bacillus subtilis* genome, only two of the three Flh proteins (FlhAB), highly conserved proteins required for exporting flagellar components across the cytoplasmic membrane [51], were identified in the *Hph. fasciatum* genome. A gene encoding FlhE, a protein found primarily in gram-negative enteric bacteria that functions to prevent proton leakage during the secretion of flagellar components across the cytoplasmic membrane, is missing, but it has been shown to be unessential for swimming motility [52].

Genes encoding Flg proteins (responsible for forming or regulating the formation of the flagellar rod, P ring, and hook) [53,54] in the *Hph. fasciatum* genome also followed the *Bacillus* pattern; genes encoding all eight of the Flg proteins of *Bacillus* (FlgBCDEFGKLMN) were identified. In addition, an unidentified protein with significant homology to FlgJ, a protein missing in *Bacillus*, was identified in the *Hph. fasciatum* genome and shared 50–60% identity with FlgJ from several gram-positive bacteria. FlgJ is a peptidoglycan hydrolase [54] and is found in the genomes of all heliobacteria listed in Table 2. Flg proteins not encoded in the *Hph. fasciatum* genome include FlgAHI and FlgOPQT, outer membrane proteins or proteins otherwise unessential for motility in gram-positive bacteria [53,54].

Genes encoding most Fli proteins responsible for forming the basal body, hook, filament, and cytoplasmic flagellar components (FliACDEFGHIJKLMNOPQRS) [53,54] were identified in the *Hph. fasciatum* genome, but the filament chaperone protein FliT, unessential in many motile bacteria, and regulatory proteins FliZY showed less than 50% identity to homologs from *Bacillus*. Mot proteins form the stator complex in the flagellar basal body [53], and the *Hph. fasciatum* genome encodes MotA and MotB, as does *Bacillus*. As

expected, MotCD, proteins that form a second stator in *Pseudomonas aeruginosa* and some other gram-negative bacteria, and MotXY, proteins that form an alternative stator in bacteria that use a Na$^+$-motive force instead of a H$^+$-motive force to power the flagellum, such as *Vibrio alginolyticus*, were not encoded in the *Heliophilum* genome.

The *Heliophilum* genome encoded several chemotaxis proteins, suggesting that the bundled cells can respond to chemical attractants and repellants. Chemotaxis-specific proteins included virtually all "Che" proteins (CheABCDRWXY) plus at least five chemical sensors, the methyl-accepting chemotaxis proteins [55]. The only major *Escherichia coli* Che gene not identified in *Hph. fasciatum* was CheZ. This protein has phosphatase activity and functions to dephosphorylate CheY, the protein that controls the direction of flagellar rotation. CheZ is distributed only in certain *Proteobacteria*, and its function is replaced in gram-positive bacteria by CheC and CheX [56], both of which were encoded in the *Hph. fasciatum* genome.

Although no obvious answer emerged to explain the bundle-forming phenotype of *Hph. fasciatum*, this behavior could be a manifestation of biofilm formation. In this connection, cultures of *Hph. fasciatum* streaked on agar plates form spreading, diffuse colonies [16]. Rafts of cells are common in biofilm-forming bacterial species, and the dynamic nature that exists between free-swimming individual cells and rafts of cells indicates that specific "glue-like" substances are not required to form rafts [47]. Instead, it is likely that rafting is the product of intercellular bundling of flagella to yield cells with their flagella interwoven in phase [47,57]; notably, such flagellar aggregates were observed in the original description of *Hph. fasciatum* [16]. From an ecological perspective, rafts might help *Hph. fasciatum* become established on a moist surface, such as a rice plant root, where individual cells could be washed away.

3.8. Sporulation

True to its phylogenetic roots with the endospore-forming *Firmicutes*, *Hph. fasciatum* forms endospores—dormant cells highly resistant to physiochemical extremes [58]—complete with the signature molecule of these structures, dipicolinic acid–Ca^{2+} complexes [16]. Moreover, experiments to hybridize *Hph. fasciatum* DNA with endospore-specific gene probes from *Bacillus subtilis* confirmed the genetic capacity for this phototroph (and several other species of heliobacteria) to produce endospores [6].

Over 500 gene products have been linked to endospores in *Bacillus subtilis*, several of which are unessential for endospore production [59]. Sporulation-specific genes in *Hmi. modesticaldum* were found to be but a small subset of those found in *B. subtilis*, although several key sporulation genes were identified. As in *Hmi. modesticaldum* and *Hrs. convoluta*, all five sigma factors absolutely essential for the formation of endospores (sigma E, F, G, H, and K) were identified in the *Hph. fasciatum* genome. An ortholog of the master transcription factor Spo0A, active during the early stages of sporulation [60], is present in all heliobacterial genomes, including that of *Heliophilum*. However, Spo0M, a key protein that regulates several genes essential for sporulation, is absent from the *Hmi. modesticaldum* and *Hrs. convoluta* genomes [10,13]. By contrast, the *Hph. fasciatum* genome contained a putative *spo0M* gene (30–40% identity to *spo0M* from several sporulating bacteria). *Hbt. chlorum*, *Hmi. gestii*, and *Hmi. undosum* but not *Hbt. mobile* also contained a *spo0M* homolog, although all of these species encoded proteins less than 40% identical to the *Hph. fasciatum* protein, suggesting considerable divergence in this protein in those heliobacteria that encode it.

Also present in the *Hph. fasciatum* genome were genes encoding a variety of small acid-soluble spore proteins (SASPs). These proteins function to protect DNA in endospores from damage by desiccation, ultraviolet radiation, and to some degree, from heat [60–62]. Genes encoding the major alpha/beta-type SASPs (*sspA* and *sspB*) were present and showed high homology to genes from all other heliobacteria and chemotrophic endospore-forming bacteria. Three different α/β-type proteins were found in the *Hph. fasciatum* genome compared with just two in the genomes of the two alkaliphiles, *Hrs. convoluta* and *Hrs. acidaminovorans*, and

four in each of the other five genome-sequenced heliobacterial species that grow optimally at neutral pH (Figure 1). In addition, the *Hph. fasciatum* genome encoded a minor SASP found in *B. subtilis* but not encoded in any other heliobacterial genome. This SASP was of the "thioredoxin type" (*B. subtilis sspT*), a protein encoded only in the genomes of a subset of sporulating bacteria [61]. Although the function of this minor SASP is unknown, it is clearly an endospore-associated protein because its transcription is dependent on the spore-specific sigma factor, σF. The genomes of all heliobacteria other than *Hph. fasciatum* encoded a different minor SASP, *sspH*, a gene under the control of σK in the developing forespore and whose product is also of unknown function.

The endospore is covered by a coat consisting of both structural and enzymatic proteins, and in *B. subtilis*, these proteins are encoded by a variety of genes, including over 20 *cot* genes [63]. Although Cot proteins are widespread in *Bacillus* species, in *Clostridium* species many *cot* genes are missing [63]. The *Hph. fasciatum* genome contained a handful of *cot* homologs. CotJC was encoded in the genomes of all heliobacteria listed in Table 2 and is one of the few Cot proteins universally present in species of *Bacillus* and *Clostridium*. The inner spore core protein GerQ is also present in all heliobacterial genomes. Genes encoding the spore-coat proteins CotJA and CotJB were annotated in *Hph. fasciatum* but the proteins bore no close identity to *B. subtilis* CotJA and CotJB. Notably, however, the *Hph. fasciatum* CotJA and CotJB homologs bore significant similarities to CotJA/JB from *Clostridium perfringens*, suggesting that there may be different forms of these endospore-specific proteins in aerobes compared with anaerobes.

Finally, in contrast to all other heliobacteria, a homolog of CotSA, a protein universally distributed in species of *Bacillus* and *Clostridium* [63], was autoannotated in the *Hph. fasciatum* genome as a cell wall glycosyltransferase with ~30% identity to *B. subtilis* CotSA. A protein similar to the *B. subtilis* spore-coat protein CotF was also identified in the *Hph. fasciatum* genome but is absent from the genomes of other sequenced heliobacteria; CotF is a spore-coat protein widely distributed among endospore-forming bacteria [63]. Genes encoding homologs of the *B. subtilis* spore-coat proteins CotABCDEGHIKMNOPQR-SUVWXYZ or CotIC, NE, NH, NW, and OO could not be identified in the *Hph. fasciatum* genome. Thus, as regards the genetics of sporulation, *Hph. fasciatum* has the essential genes necessary to form endospores but, like other heliobacteria with analyzed genomes, is missing many of the unessential genes present in the genome of the well-studied *Bacillus subtilis*.

4. Conclusions

The genome of *Heliophilum fasciatum* represents the first detailed analysis of a genome from a species of heliobacteria that grows optimally at moderate temperatures and neutral pH. Although no firm explanation emerged for the mechanism that allows this organism to move about its habitat in loosely attached bundles, the genomics supporting the physiology of this species revealed many things in common with other heliobacteria (e.g., lack of autotrophy, ability to grow in darkness by fermentation, production of endospores, to mention a few), as well as an important feature possibly unique to this species: diazotrophy in the absence of a canonical nitrogenase. Whether nitrogen fixation in the Tanzanian rice soil habitat of *Hph. fasciatum* [16] is best served by its unusual nitrogenase system remains to be seen. However, the fact that all other heliobacteria with sequenced genomes (some species of which originate from rice soils) contain a canonical nitrogenase (Figures 3–5), suggests that the ancestor of *Hph. fasciatum* may have been incapable of diazotrophy but that eventually, this species acquired an enzyme with nitrogenase-like activity from either duplications or lateral transfers of genes encoding functionally related proteins. Nitrogen fixation by phototrophic purple bacteria in paddy soil environments is well known [64,65]. *Hph. fasciatum* and other paddy-dwelling heliobacterial species could also be contributing fixed nitrogen to rice plants, perhaps in exchange for organic compounds to fuel their photoheterotrophic metabolism.

References

1. Gest, H.; Favinger, J.L. *Heliobacterium chlorum*, an anoxygenic brownish-green photosynthetic bacterium containing a "new" form of bacteriochlorophyll. *Arch. Microbiol.* **1983**, *136*, 11–16. [CrossRef]
2. Madigan, M.T.; Ormerod, J.G. Taxonomy, physiology, and ecology of heliobacteria. In *Anoxygenic Photosynthetic Bacteria*; Blankenship, R.E., Madigan, M.T., Bauer, C.E., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1995; Volume 2, pp. 17–30. ISBN 0-306-47954-0.
3. Brockmann, H.; Lipinski, A. Bacteriochlorophyll *g*. A new bacteriochlorophyll from *Heliobacterium chlorum*. *Arch. Microbiol.* **1983**, *136*, 17–19. [CrossRef]
4. Sattley, W.M.; Madigan, M.T. The Family *Heliobacteriaceae*. In *The Prokaryotes—Firmicutes and Tenericutes*; Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., Thompson, F., Eds.; Springer: Berlin/Heidelberg, Germany, 2014; pp. 185–196. ISBN 978-3-642-30120-9.
5. Orf, G.S.; Redding, K.E. The Heliobacteria. In *Encyclopedia of Biological Chemistry*; Joseph, J., Ed.; Elsevier: Oxford, UK, 2021; Volume 2, pp. 352–364.
6. Kimble-Long, L.K.; Madigan, M.T. Molecular evidence that the capacity for endosporulation is universal among phototrophic heliobacteria. *FEMS Microbiol. Lett.* **2001**, *199*, 191–195. [CrossRef]
7. Stevenson, A.K.; Kimble, L.K.; Woese, C.R.; Madigan, M.T. Characterization of new phototrophic heliobacteria and their habitats. *Photosynth. Res.* **1997**, *53*, 1–12. [CrossRef]
8. Kyndt, J.A.; Salama, D.M.; Meyer, T.E.; Imhoff, J.F. Phylogenetic relationship of phototrophic heliobacteria and systematic reconsideration of species and genus assignments based on genome sequences of eight species. *Int. J. Syst. Evol. Microbiol.* **2021**, *71*, 4729. [CrossRef]
9. Kimble, L.K.; Mandelco, L.; Woese, C.R.; Madigan, M.T. *Heliobacterium modesticaldum*, sp. nov., a thermophilic heliobacterium of hot springs and volcanic soils. *Arch. Microbiol.* **1995**, *163*, 259–267. [CrossRef]
10. Sattley, W.M.; Madigan, M.T.; Swingley, W.D.; Cheung, P.C.; Clocksin, K.M.; Conrad, A.L.; Dejesa, L.C.; Honchak, B.M.; Jung, D.; Karbach, L.E.; et al. The Genome of *Heliobacterium modesticaldum*, a Phototrophic Representative of the Firmicutes Containing the Simplest Photosynthetic Apparatus. *J. Bacteriol.* **2008**, *190*, 4687–4696. [CrossRef]
11. Sattley, W.M.; Swingley, W.D. Properties and evolutionary implications of the heliobacterial genome. In *Genome Evolution of Photosynthetic Bacteria*; Beatty, T.J., Ed.; Academic Press: San Diego, CA, USA, 2013; Volume 66, pp. 67–98. ISBN 9780123979230.
12. Asao, M.; Jung, D.; Achenbach, L.A.; Madigan, M.T. *Heliorestis convoluta* sp. nov., a coiled, alkaliphilic heliobacterium from the Wadi El Natroun, Egypt. *Extremophiles* **2006**, *10*, 403–410. [CrossRef]
13. Dewey, E.D.; Stokes, L.M.; Burchell, B.M.; Shaffer, K.N.; Huntington, A.M.; Baker, J.M.; Nadendla, S.; Giglio, M.G.; Bender, K.S.; Touchman, J.W.; et al. Analysis of the Complete Genome of the Alkaliphilic and Phototrophic Firmicute *Heliorestis convoluta* Strain HHT. *Microorganisms* **2020**, *8*, 313. [CrossRef]
14. Asao, M.; Takaichi, S.; Madigan, M.T. Amino acid-assimilating phototrophic heliobacteria from soda lake environments: *Heliorestis acidaminivorans* sp. nov. and 'Candidatus Heliomonas lunata'. *Extremophiles* **2012**, *16*, 585–595. [CrossRef]
15. Kimble, L.K.; Madigan, M.T. Nitrogen fixation and nitrogen metabolism in heliobacteria. *Arch. Microbiol.* **1992**, *158*, 155–161. [CrossRef]
16. Ormerod, J.G.; Kimble, L.K.; Nesbakken, T.; Torgersen, Y.A.; Woese, C.R.; Madigan, M.T. *Heliophilum fasciatum* gen. nov. sp. nov. and *Heliobacterium gestii* sp. nov.: Endospore-forming heliobacteria from rice field soils. *Arch. Microbiol.* **1996**, *165*, 226–234. [CrossRef] [PubMed]
17. Tang, K.-H.; Yue, H.; Blankenship, R.E. Energy metabolism of *Heliobacterium modesticaldum* during phototrophic and chemotrophic growth. *BMC Microbiol.* **2010**, *10*, 150. [CrossRef] [PubMed]

18. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef] [PubMed]

19. Stecher, G.; Tamura, K.; Kumar, S. Molecular Evolutionary Genetics Analysis (MEGA) for macOS. *Mol. Biol. Evol.* **2020**, *37*, 1237–1239. [CrossRef]

20. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [CrossRef]

21. Raymond, J.; Siefert, J.L.; Staples, C.R.; Blankenship, R.E. The Natural History of Nitrogen Fixation. *Mol. Biol. Evol.* **2004**, *21*, 541–554. [CrossRef]

22. North, J.A.; Narrowe, A.B.; Xiong, W.; Byerly, K.M.; Zhao, G.; Young, S.J.; Murali, S.; Wildenthal, J.A.; Cannon, W.R.; Wrighton, K.C.; et al. A nitrogenase-like enzyme system catalyzes methionine, ethylene, and methane biogenesis. *Science* **2020**, *369*, 1094–1098. [CrossRef]

23. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [CrossRef]

24. Le, S.Q.; Gascuel, O. An Improved General Amino Acid Replacement Matrix. *Mol. Biol. Evol.* **2008**, *25*, 1307–1320. [CrossRef]

25. Bryantseva, I.A.; Gorlenko, V.M.; Kompantseva, E.I.; Achenbach, L.A.; Madigan, M.T. *Heliorestis daurensis*, gen. nov. sp. nov., an alkaliphilic rod-to-coiled-shaped phototrophic heliobacterium from a Siberian soda lake. *Arch. Microbiol.* **1999**, *172*, 167–174. [CrossRef] [PubMed]

26. Bryantseva, I.A.; Gorlenko, V.M.; Kompantseva, E.I.; Tourova, T.P.; Kuznetsov, B.B.; Osipov, G.A. Alkaliphilic heliobacterium *Heliorestis baculata* sp. nov. and emended description of the genus *Heliorestis*. *Arch. Microbiol.* **2000**, *174*, 283–291. [CrossRef] [PubMed]

27. Briantseva, I.A.; Gorlenko, V.M.; Turova, T.P.; Kuznetsov, B.B.; Lysenko, A.M.; Bykova, S.A.; Gal'Chenko, V.F.; Mitiushina, L.L.; Osipov, G.A. *Heliobacterium sulfidophilum* sp. nov. and *Heliobacterum undosum* sp. nov.: Sulfide-oxidizing heliobacteria from thermal sulfidic springs. *Mikrobiologia* **2000**, *69*, 396–406.

28. Kimble, L.K.; Stevenson, A.K.; Madigan, M.T. Chemotrophic growth of heliobacteria in darkness. *FEMS Microbiol. Lett.* **1994**, *115*, 51–55. [CrossRef]

29. Pickett, M.W.; Williamson, M.P.; Kelly, D.J. An enzyme and 13C-NMR study of carbon metabolism in heliobacteria. *Photosynth. Res.* **1994**, *41*, 75–88. [CrossRef]

30. Dos Santos, P.C.; Fang, Z.; Mason, S.W.; Setubal, J.C.; Dixon, R. Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. *BMC Genom.* **2012**, *13*, 162. [CrossRef]

31. Enkh-Amgalan, J.; Kawasaki, H.; Seki, T. NifH and NifD sequences of heliobacteria: A new lineage in the nitrogenase phylogeny. *FEMS Microbiol. Lett.* **2005**, *243*, 73–79. [CrossRef]

32. Garcia, A.K.; McShea, H.; Kolaczkowski, B.; Kaçar, B. Reconstructing the evolutionary history of nitrogenases: Evidence for ancestral molybdenum-cofactor utilization. *Geobiology* **2020**, *18*, 394–411. [CrossRef]

33. Chen, Y.; Nishihara, A.; Haruta, S. Nitrogen-fixing Ability and Nitrogen Fixation-related Genes of Thermophilic Fermentative Bacteria in the Genus *Caldicellulosiruptor*. *Microbes Environ.* **2021**, *36*, ME21018. [CrossRef]

34. Lie, T.J.; Kuo, Y.P.; Leite, M.; Costa, K.C.; Harwood, C.S.; Leigh, J.A. A genetic study of *Nif*-associated genes in a hyperthermophilic methanogen. *Microbiol. Spectr.* **2022**, *10*, e02093-21. [CrossRef]

35. Zheng, H.; Dietrich, C.; Radek, R.; Brune, A. *Endomicrobium proavitum*, the first isolate of *Endomicrobia* class. nov. (phylum *Elusimicrobia*)—An ultramicrobacterium with an unusual cell cycle that fixes nitrogen with a Group IV nitrogenase. *Environ. Microbiol.* **2015**, *18*, 191–204. [CrossRef] [PubMed]

36. Boyd, E.S.; Anbar, A.D.; Miller, S.; Hamilton, T.L.; Lavin, M.; Peters, J.W. A late methanogen origin for molybdenum-dependent nitrogenase. *Geobiology* **2011**, *9*, 221–232. [CrossRef] [PubMed]

37. Harwood, C.S. Iron-only and vanadium nitrogenases: Fail-safe enzymes or something more? *Ann. Rev. Microbiol.* **2020**, *74*, 247–266. [CrossRef] [PubMed]

38. Kimble, L.K.; Madigan, M.T. Evidence for an alternative nitrogenase in *Heliobacterium gestii*. *FEMS Microbiol. Lett.* **1992**, *100*, 255–260. [CrossRef]

39. Nomata, J.; Mizoguchi, T.; Tamiaki, H.; Fujita, Y. A second nitrogenase-like enzyme for bacteriochlorophyll biosynthesis: Reconstitution of chlorophyllide a reductase with purified X-protein (BchX) and YZ-protein (BchY-BchZ) from *Rhodobacter capsulatus*. *J. Biol. Chem.* **2006**, *281*, 15021–15028. [CrossRef]

40. Muraki, N.; Nomata, J.; Ebata, K.; Mizoguchi, T.; Shiba, T.; Tamiaki, H.; Kurisu, G.; Fujita, Y. X-ray crystal structure of the light-independent protochlorophyllide reductase. *Nature* **2010**, *465*, 110–114. [CrossRef]

41. Moore, S.; Sowa, S.T.; Schuchardt, C.; Deery, E.; Lawrence, A.D.; Ramos, J.V.; Billig, S.; Birkemeyer, C.; Chivers, P.; Howard, M.; et al. Elucidation of the biosynthesis of the methane catalyst coenzyme F430. *Nature* **2017**, *543*, 78–82. [CrossRef]

42. Zheng, K.; Ngo, P.D.; Owens, V.L.; Yang, X.-P.; Mansoorabadi, S.O. The biosynthetic pathway of coenzyme F430 in methanogenic and methanotrophic archaea. *Science* **2016**, *354*, 339–342. [CrossRef]

43. Méheust, R.; Castelle, C.; Carnevali, P.B.M.; Farag, I.F.; He, C.; Chen, L.-X.; Amano, Y.; Hug, L.; Banfield, J.F. Groundwater *Elusimicrobia* are metabolically diverse compared to gut microbiome *Elusimicrobia* and some have a novel nitrogenase paralog. *ISME J.* **2020**, *14*, 2907–2922. [CrossRef]

44. Fernandes, G.D.C.; Trarbach, L.J.; de Campos, S.B.; Beneduzi, A.; Passaglia, L.M. Alternative nitrogenase and pseudogenes: Unique features of the *Paenibacillus riograndensis* nitrogen fixation system. *Res. Microbiol.* **2014**, *165*, 571–580. [CrossRef]

45. McRose, D.L.; Zhang, X.; Kraepiel, A.M.L.; Morel, F.M.M. Diversity and Activity of Alternative Nitrogenases in Sequenced Genomes and Coastal Environments. *Front. Microbiol.* **2017**, *8*, 267. [CrossRef] [PubMed]

46. Shahak, Y.; Schütz, M.; Bronstein, M.; Greisbeck, C.; Hauska, G.; Padan, E. Sulfide-dependent anoxygenic photosynthesis in prokaryotes—Sulfide-Quinone Reductase (SQR), the initial step. In *The Phototrophic Prokaryotes*; Peschek, G.A., Loffelhardt, W.L., Schmetterer, G., Eds.; Kluwer Academic/Plenum: New York, NY, USA, 1999; pp. 217–228.

47. Kearns, D.; Losick, R. Swarming motility in undomesticated *Bacillus subtilis*. *Mol. Microbiol.* **2004**, *49*, 581–590. [CrossRef] [PubMed]

48. Kearns, D.B. A field guide to bacterial swarming motility. *Nat. Rev. Genet.* **2010**, *8*, 634–644. [CrossRef] [PubMed]

49. Apel, D.; Surette, M.G. Bringing order to a complex molecular machine: The assembly of the bacterial flagella. *Biochim. Biophys. Acta (BBA)-Biomembr.* **2008**, *1778*, 1851–1858. [CrossRef] [PubMed]

50. Albanna, A.; Alomari, A. Comparison 2D flagellar proteins structure in gram positive vs. gram-negative motile bacteria. *EurAsian J. Biosci.* **2020**, *17*, 6275–6281.

51. Nakamura, S.; Minamino, T. Flagella-Driven Motility of Bacteria. *Biomolecules* **2019**, *9*, 279. [CrossRef]

52. Stafford, G.P.; Hughes, C. *Salmonella typhimurium flhE*, a conserved flagellar regulon gene required for swarming. *Microbiology* **2007**, *153*, 541–547. [CrossRef]

53. Girgis, H.S.; Liu, Y.; Ryu, W.S.; Tavazoie, S. A Comprehensive Genetic Characterization of Bacterial Motility. *PLoS Genet.* **2007**, *3*, e154–e160. [CrossRef]

54. Herlihey, F.A.; Moynihan, P.; Clarke, A.J. The Essential Protein for Bacterial Flagella Formation FlgJ Functions as a β-*N*-Acetylglucosaminidase. *J. Biol. Chem.* **2014**, *289*, 31029–31042. [CrossRef]

55. Wadhams, G.H.; Armitage, J.P. Making sense of it all: Bacterial chemotaxis. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 1024–1037. [CrossRef] [PubMed]

56. Szurmant, H.; Ordal, G.W. Diversity in Chemotaxis Mechanisms among the Bacteria and Archaea. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 301–319. [CrossRef] [PubMed]

57. Jones, B.V.; Young, R.; Mahenthiralingam, E.; Stickler, D.J. Ultrastructure of *Proteus mirabilis* Swarmer Cell Rafts and Role of Swarming in Catheter-Associated Urinary Tract Infection. *Infect. Immun.* **2004**, *72*, 3941–3950. [CrossRef] [PubMed]

58. Nicholson, W.L.; Munakata, N.; Horneck, G.; Melosh, H.J.; Setlow, P. Resistance of *Bacillus* Endospores to Extreme Terrestrial and Extraterrestrial Environments. *Microbiol. Mol. Biol. Rev.* **2000**, *64*, 548–572. [CrossRef] [PubMed]

59. Traag, B.A.; Pugliese, A.; Eisen, J.A.; Losick, R. Gene Conservation among Endospore-Forming Bacteria Reveals Additional Sporulation Genes in *Bacillus subtilis*. *J. Bacteriol.* **2013**, *195*, 253–260. [CrossRef] [PubMed]

60. Piggot, P.J.; Hilbert, D.W. Sporulation of *Bacillus subtilis*. *Curr. Opin. Microbiol.* **2004**, *7*, 579–586. [CrossRef] [PubMed]

61. Setlow, P. Mechanisms for the prevention of damage to DNA in spores of *bacillus* species. *Annu. Rev. Microbiol.* **1995**, *49*, 29–54. [CrossRef] [PubMed]

62. Bagyan, I.; Setlow, B.; Setlow, P. New Small, Acid-Soluble Proteins Unique to Spores of *Bacillus subtilis*: Identification of the Coding Genes and Regulation and Function of Two of These Genes. *J. Bacteriol.* **1998**, *180*, 6704–6712. [CrossRef]

63. Henriques, A.O.; Moran, C.P., Jr. Structure, Assembly, and Function of the Spore Surface Layers. *Annu. Rev. Microbiol.* **2007**, *61*, 555–588. [CrossRef]

64. Habte, M.; Alexander, M. Nitrogen Fixation by Photosynthetic Bacteria in Lowland Rice Culture. *Appl. Environ. Microbiol.* **1980**, *39*, 342–347. [CrossRef]

65. Maeda, I. Potential of Phototrophic Purple Nonsulfur Bacteria to Fix Nitrogen in Rice Fields. *Microorganisms* **2021**, *10*, 28. [CrossRef]

Article

Taxonomic Re-Evaluation and Genomic Comparison of Novel Extracellular Electron Uptake-Capable *Rhodovulum visakhapatnamense* and *Rhodovulum sulfidophilum* Isolates

Emily J. Davenport and Arpita Bose *

Department of Biology, Washington University in St. Louis, St. Louis, MO 63130, USA; demily@wustl.edu
* Correspondence: abose@wustl.edu

Abstract: *Rhodovulum* spp. are anoxygenic phototrophic purple bacteria with versatile metabolisms, including the ability to obtain electrons from minerals in their environment to drive photosynthesis, a relatively novel process called phototrophic extracellular electron uptake (pEEU). A total of 15 strains of *Rhodovulum sulfidophilum* were isolated from a marine estuary to observe these metabolisms in marine phototrophs. One representative strain, *Rhodovulum sulfidophilum* strain AB26, can perform phototrophic iron oxidation (photoferrotrophy) and couples carbon dioxide fixation to pEEU. Here, we reclassify two *R. sulfidophilum* isolates, strain AB26 and strain AB19, as *Rhodovulum visakhapatnamense* using taxonomic re-evaluation based on 16S and *pufM* phylogenetic analyses. The strain AB26 genome consists of 4,380,746 base-pairs, including two plasmids, and encodes 4296 predicted protein-coding genes. Strain AB26 contains 22 histidine kinases, 20 response regulators, and dedicates ~16% of its genome to transport. Transcriptomic data under aerobic, photoheterotrophy, photoautotrophy, and pEEU reveals how gene expression varies between metabolisms in a novel *R. visakhapatnamense* strain. Genome comparison led by transcriptomic data under pEEU reveals potential pEEU-relevant genes both unique to *R. visakhapatnamense* strains and shared within the *R. sulfidophilum* genomes. With these data we identify potential pEEU-important transcripts and how speciation may affect molecular mechanisms of pEEU in *Rhodovulum* species from the same environment.

Keywords: phototrophic bacteria; phototrophic extracellular electron uptake; comparative genomics; transcriptomics; *Rhodovulum sulfidophilum*; *Rhodovulum visakhapatnamense*

1. Introduction

Bacterial metabolisms are diverse, allowing for growth in nearly any environment [1]. Some bacteria are capable of obtaining electrons from insoluble extracellular sources within their environments via a process called extracellular electron uptake (EEU) [2]. Various forms of this metabolism exist with the common theme being that microbes interact with solid-phase conductive substances as sources of electrons [3]. Through EEU, microbes drive biogeochemical cycles in all corners of the earth at undetermined rates [4]. This metabolism also has potential to improve bioremediation of contaminated sites and the biosynthesis of sustainable fuels [5–7].

EEU was first observed in species such as mineral-reducing, *Shewanella oneidensis* MR1 and the photoautotroph, *Rhodopseudomonas palustris* TIE-1 (hereafter referred to as TIE-1) [8–10]. One well-characterized electron uptake system is known; the PioABC operon studied in TIE-1 [11]. Studies of pEEU in TIE-1 revealed the Calvin–Benson–Bassham cycle as an electron sink during pEEU [12,13]. This link to carbon dioxide fixation connects pEEU to the carbon cycle. Furthermore, because TIE-1 can use various soluble and insoluble forms of iron, pEEU connects to the cycling of iron and other elements that are bound to iron minerals [3,14]. Until recently pEEU has been studied solely in freshwater phototrophs, leaving a knowledge gap regarding the prevalence of pEEU in other ecosystems.

In an effort to address this, we isolated bacteria from a microbial mat in a marine estuary in Woods Hole, Massachusetts, USA (henceforth referred to as Woods Hole isolates) [15,16]. In total, 15 of the Woods Hole isolates were initially identified as *Rhodovulum sulfidophilum*, a purple non-sulfur bacterium capable of many diverse metabolisms including anoxygenic photosynthesis (non-oxygen-evolving). We sequenced and assembled the genomes of 12 of the 15 isolates based on these data (3 isolates, namely *Rhodovulum sulfidophilum* strain AB14, strain AB26, AB30, were assembled previously [15]). We have previously reported that all 15 strains were capable of heterotrophic, photoautotrophic, and photoheterotrophic growth as well as photoferrotrophy [12]. A representative strain, *R. sulfidophilum* AB26, was selected for further growth studies and found to be capable of pEEU in bioelectrochemical systems (BES) using a novel uptake pathway consisting of a di-heme *c*-type cytochrome (EeuP), with carbon dioxide fixation as a cellular sink for acquired electrons [12]. EeuP is found in diverse bacterial lineages without other characterized pEEU uptake systems in their genomes, which may suggest its role as a novel uptake system [12].

In assembling and releasing the remaining 12 genomes of the Woods Hole isolates, we discovered 2 isolates, strain AB19 and the representative pEEU strain AB26, that no longer fit into *R. sulfidophilum* but rather the closely related, novel *Rhodovulum visakhapatnamense* species, originally isolated from a tidal seawater enriched under phototrophic conditions [17]. Prior to this study, *R. visakhapatnamense* was represented by a single-strain *R. visakhapatnamense* strain JA181. Like other *Rhodovulum* species, strain JA181 is capable of chemoorganotrophy in the dark and photoheterotrophy under illumination [17]. Currently, three *R. visakhapatnamense* genomes are publicly available: strain AB26, strain AB19, and strain JA181 [17]. Given this new taxonomic classification and the previously described transcriptomic data our group possesses [11], we aim to describe the genome content and expression profiles under these growth conditions to further characterize members of *R. visakhapatnamense*. Comparative genomic analysis based on transcript data of putative pEEU-important transcripts within *R. visakhapatnamense* strain AB26 reveals the distribution of homologous genes within the genomes of the 15 Woods Hole isolates.

2. Materials and Methods

2.1. Microbial Isolation and Cultivtion Conditions

As previously described [11], Rhodovulum sulfidophilum strains were isolated in July 2014 from independent microbial mat samples from the Trunk River estuary in Woods Hole, MA. Enrichments were cultivated photoheterotrophically in anoxic artificial seawater (SW) medium supplemented with 20 mM acetate or 10 mM sodium thiosulfate. Enrichments were cultivated with ~850-nm light at 30 °C, and passaged six times, followed by streaking aerobically six times on Bacto agar with Difco marine broth (MB) 2216 (BD Diagnostic Systems, Hunt Valley, MD USA) to isolate single colonies. All growth experiments were carried out at 30 °C unless otherwise noted. AB26 strains were grown in MB broth under dark with shaking at 200 rpm for aerobic heterotrophic growth. All phototrophic (photoautotrophic and photoheterotrophic) growth experiments were carried out anaerobically in SW media supplemented with 70 mM sodium bicarbonate and 1 mM sodium sulfate in sealed sterile serum bottles. The phototrophic cultures were grown without shaking under light (with a single 60 W incandescent light bulb at 25 cm). For anaerobic photoautotrophic growth, AB26 strains were grown with H_2 (80% H_2: 20% CO_2 at 50 kPa) or 10 mM sodium thiosulfate or 5 mM Fe(II) chloride in SW medium. To prevent iron precipitation, the photoautotrophic culture with Fe(II) contained 10 mM nitrilotriacetic acid (NTA). For anaerobic photoheterotrophic growth, AB26 strains were grown in SW medium with 10 mM acetate. The headspace atmosphere of the phototrophic cultures consisted of 80% N_2 and 20% CO_2, except for hydrogen. Where a change in culture medium was required, cells were washed three times in basal SW medium post-centrifugation at $5000\times g$.

2.2. Genome Sequencing

Genome analysis was performed as previously described [15]. Genomic DNA was isolated from mid-log phase cell culture in marine broth using the DNeasy Blood and Tissue Kit (Qiagen, Germantown, MD, USA). Samples were prepped for Illumina 250-bp paired-end sequencing using Nextera sample prep kit (Illumina, Inc., San Diego, CA, USA), and sequenced on Illumina MiSeq platform with V2 chemistry. Reads were quality checked and adapter trimmed using Trimmomatic version 0.33 using default parameters for paired-end reads [18]. Reads were de novo assembled using CLC Genomics Workbench (CLC Bio-Qiagen, Aarhaus, Denmark), and scaffolds generated using MeDuSa [19] and *R. sulfidophilum* DSM 2351 as an alignment guide. Alignment of reads to *R. sulfidophilum* DSM 2351 was performed using Bowtie2 version 2.2.29 [20] short read mapper, and annotated with National Center for Biotechnology Information Prokaryotic Genome Annotation Pipeline [21].

2.3. Phylogenetic Analysis, Taxonomic Re-Evaluation, and Comparative Genomics

ANI between strains was performed using JSpeciesWS [22] and visualized with Morpheus (Morpheus, https://software.broadinstitute.org/morpheus (accessed on 3 February 2022)). Phylogenetic trees were built using MEGA11 [23]. For strain-level phylogenetic analysis, 16S rRNA sequences and the photosynthetic reaction center subunit M (*pufM*) protein sequences were used to build trees [24,25]. The 16S tree used Kimura 2-parameter model [26] and a gamma distribution of 5. The *pufM* tree used Jones-Taylor-Thornton (JTT) model [27] and a gamma distribution of 5. For phylogenetic analysis of EeuP protein sequences among *R. sulfidophilum* strains, the JTT model was used with a gamma distribution of 5. All sequences were aligned using the ClustalW algorithm. BLASTp was performed using a local protein database (Supplemental Materials File S2).

2.4. RNA Isolation, RNA Sequencing, and Differential Expression Analysis

As previously described [12], Cell cultures were sampled anaerobically and immediately mixed with 1:1 with RNAlater® (Qiagen, Germantown, MD, USA). RNA was extracted from cells using the RNeasy Mini Kit (Qiagen, USA) and DNA removal performed using Turbo DNA-*free* Kit (Ambion, Austin, TX, USA). RNA samples were tested for purity using PCR. Illumina single-end 50-bp libraries were prepared and sequenced at Washington University's Genome Technology Access Center on an Illumina HiSeq3000 (Illumina Inc., Madison, WI, USA). Reads were mapped to the strain AB26 genome using TopHat2 version 2.1.1 and the gff3 annotation file as a guide for sequence alignment. Bowtie 2 version 2.3.3.1 was used to index the reference genome FASTA file. The number of reads mapping to each feature were counted by HTSeq version 0.9.1. Differentially expressed genes were determined in DESEQ2 version 1.16.1 using the HTSeq read counts. Significantly differentially expressed genes were identified using an adjusted *p*-value cutoff of 0.05.

2.5. Supplemental Materials & Methods

Supplemental figures, tables, and database files can be found in the Supplemental materials. Accession numbers: *Rhodovulum visakhapatnamense* strains AB26 GCA_001941715.1; AB19 GCA_016757185.1; JA181 GCA_004365965.1; *Rhodovulum sulfidophilum* strains: AB14 GCA_001941695.1; AB15 GCA_021409265.1; AB16 GCA_016757145.1; AB17 GCA_021409205.1; AB18 GCA_021409185.1; AB20 GCA_016757195.1; AB21 GCA_021409225.1; AB22 GCA_016757115.1; AB23 GCA_016757055.1; AB28 GCA_021409165.1; AB30 GCA_001941745.1; AB33 GCA_021568715.1; AB35 GCA_016757045.1; DSM1374 GCA_001633165.1; DSM2351 GCA_001548075.1; IM796 GCA_016653205.1; SE1 GCA_010119435.1; SNK001 GCA_001633145.1.

3. Results and Discussion

3.1. Taxonomic Re-Evaluation of Strain AB26 and Strain AB19

A total of 15 isolates were cultured from a marine microbial mat in Woods Hole, MA, USA and identified as belonging to *R. sulfidophilum* through 16S sequencing [12,15].

Prior to sequencing the Woods Hole isolates, few *R. sulfidophilum* genomes were publicly available. Whole genome alignments of strain AB26 showed only 57.8% of the genome sequence aligned to the available reference genome *R. sulfidophilum* DSM2351 [15]. Growth experiments of select isolates revealed differences in doubling times in the same growth conditions [12], and therefore we sought to further investigate the phylogenetic relatedness of the isolated strains. 16S and photosynthetic reaction center subunit M (*pufM*) phylogenetic trees were produced to determine the relatedness between isolates from Woods Hole (Figure 1).
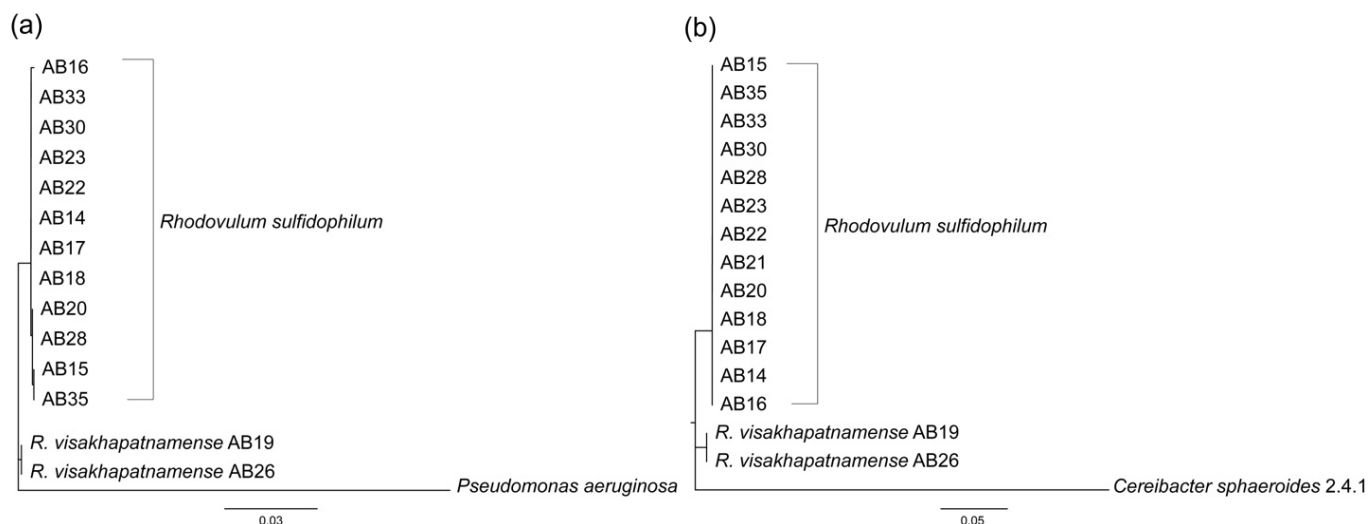
Figure 1. 16S rRNA (**a**) and *pufM* (**b**) phylogenetic trees between Woods Hole *R. sulfidophilum* isolates. (**a**) 16S phylogenetic tree constructed by maximum-likelihood method using the Kimura-2 parameter model [26], tested by bootstrapping (500 re-samplings) in MEGA11 [22] with *Pseudomonas aeruginosa* as outgroup. (**b**) *pufM* tree constructed by maximum-likelihood using the Jones–Taylor–Thornton parameter model [27], tested by bootstrapping (500 re-samplings) in MEGA11 [22], using *Cereibacter sphaeroides* 2.4.1 as outgroup.

16S phylogenetic analysis shows strain AB26 and *R. sulfidophilum* strain AB19 reside in a separate branch from all other Woods Hole isolates, with a second clade consisting of strain AB33, strain AB16, and strain AB30. The remaining Woods Hole isolates generally nest within a group together. The pattern of the 16S phylogenetic tree is mirrored in the photosynthetic reaction center subunit M (*pufM*) tree in that strain AB26 and strain AB19 again reside in a separate branch from the remaining Woods Hole isolates. Further investigation using average nucleotide identities (ANI) revealed strain AB26 and strain AB19 ANI values are far below the cut-off value for same-species relatedness (>~94%, Figure 2) [27]. The remaining Woods Hole isolates all show ANI values indicative of same-species relatedness. Two groups of strains, AB18 and AB33 and strains AB35 and AB15, show highest similarity to each other. One other isolate, strain AB28, showed lower ANI values to all other strains (94–95% ANI) but were still within same species-values (Figure 2).
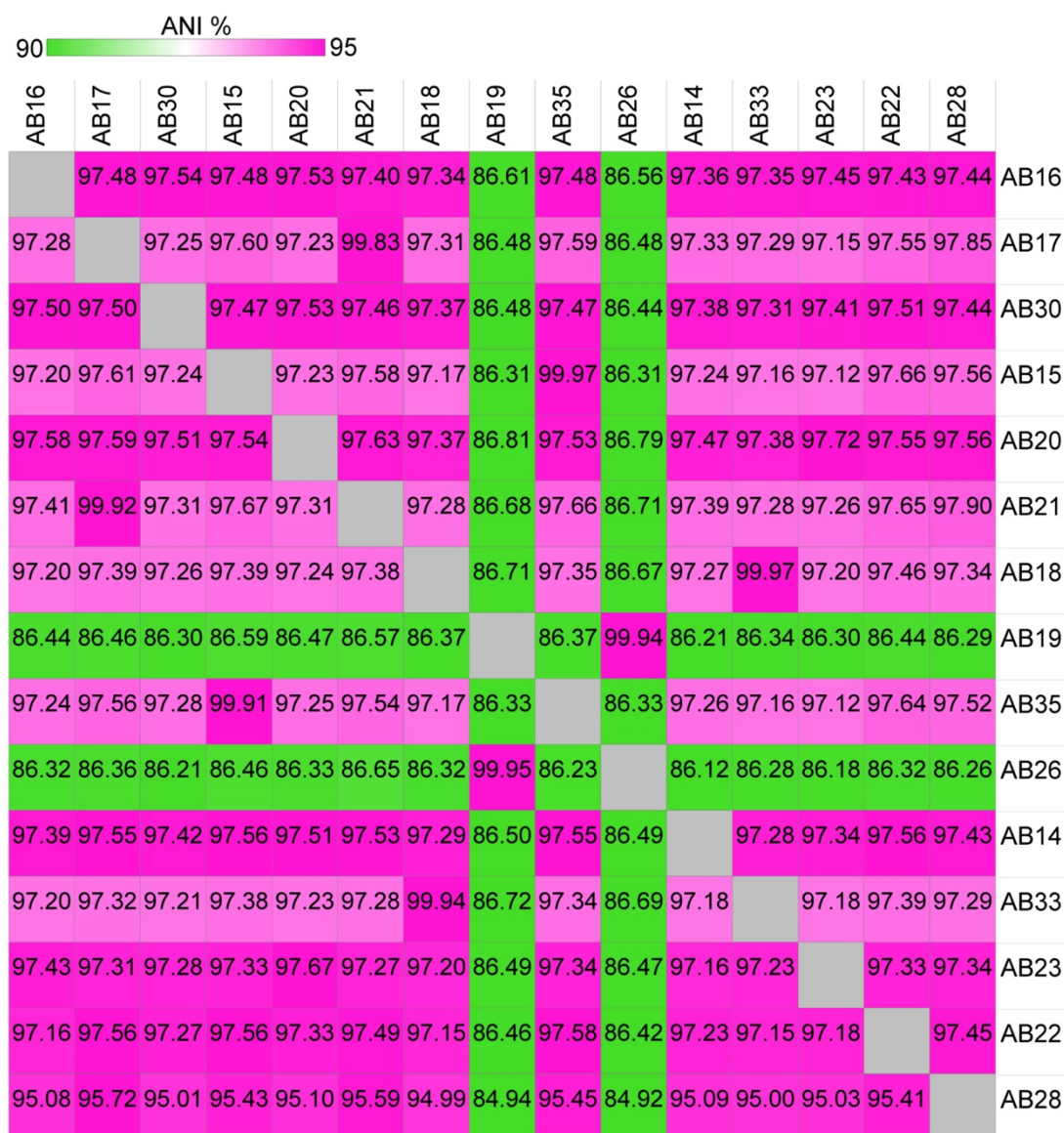
Figure 2. Heatmap showing average nucleotide identities (ANI) values from pairwise comparisons of the 15 Woods Hole isolates. ANI values were calculated using JSpeciesWSd [22] and visualized as a heatmap using Morpheus (Morpheus, https://software.broadinstitute.org/morpheus (accessed on 3 February 2022).

To determine the proper species designation of strain AB26 and strain AB19, ANI comparisons between strain AB26 and strain AB28 were performed to include the closely related species *Rhodovulum viride* and *Rhodovulum visakhapatnamese* (Supplemental Figure S1) [17,28]. Strain AB28 did not show any increased similarities to other *Rhodovulum* genomes, suggesting the current taxonomy is correct, but may change as more *Rhodovulum* genomes are assembled. ANI analysis shows strain AB26 and strain AB19 share ANI values of 97.95% with *R. visakhapatnamense*, placing strain AB26 and strain AB19 in a new species (Figure 3).
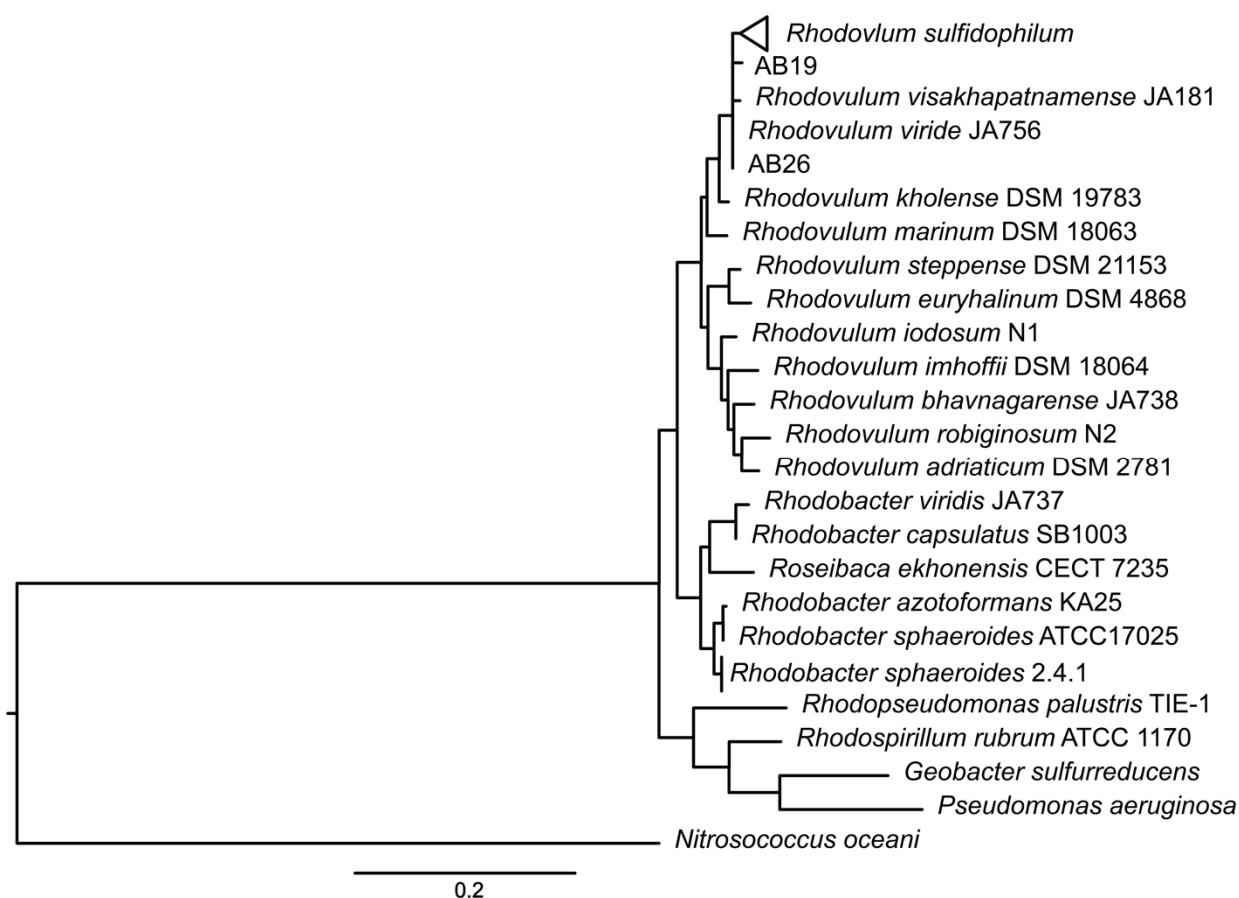
Figure 3. 16S phylogenetic tree showing the new taxonomic assignment of strain AB19 and strain AB26 as *Rhodovulum visakhapatnamense*. Tree constructed by maximum-likelihood method using the Kimura-2 parameter model [26], tested by bootstrapping (500 resamplings) in MEGA11 [23].

3.2. Strain AB26 Genome Features and Expression Analysis

Strain AB26 was chosen as a representative strain for this study due to its preferable growth characteristics [12]. The genome of strain AB26 consists of 4,380,746 bp of DNA and a GC content of 67.9%, assembled into 3 scaffolds. strain AB26 has 4375 genes, with 4296 protein-encoding genes (Table 1). The strain AB26 genome contains genes allowing for diverse metabolic capabilities that purple non-sulfur bacteria are known for. This is in line with previous observations that strain AB26 is capable of aerobic heterotrophic growth in the dark, photoheterotrophy using succinate and acetate for carbon and electrons, and photoautotrophy, using inorganic substances, including thiosulfate and hydrogen (H_2).

Table 1. Genome features of strain AB26. Data for table derived from NCBI and JGI IMG. Strain AB26 is also capable of photosynthesis via pEEU, using electrons obtained from a poised electrode (pEEU) or photoferrotrophy, obtaining electrons by oxidizing Fe(II) [12].

		% Total
Genome Size (bp)	4,380,746	
GC%	67.9	
Scaffolds	3	
Genes total number	4375	
Protein coding genes with predicted function	3212	73.42
rRNA genes	3	

Table 1. *Cont.*

Genome Size (bp)	4,380,746	
tRNA genes	49	
Other RNA genes	7	
Protein coding genes with function prediction	3212	73.4
Without function prediction	1084	24.8
Biosynthetic gene clusters	5	
Genes in biosynthetic clusters	115	2.6
Protein coding genes coding signal peptides	370	8.5
Protein coding genes coding transmembrane proteins	919	21.0
COG clusters	1869	56.4
Pfam clusters	2235	65.0
TIGRfam clusters	1083	83.6

3.2.1. Phototrophy

R. visakhapatnamense is a member of the purple nonsulfur bacteria capable of anoxygenic photosynthesis [29]. The genes necessary for growth via photosynthesis occur in a ~50-kb region of the genome spanning four separate operons. The first operon spans a 13-kb region and encodes the reaction center, bacteriochlorophyll synthesis, carotenoid synthesis, and light-harvesting proteins. Reaction center subunits M (BV509_00330) and L (BV509_00335) share 70% identity with the reaction center genes in the model anoxygenic phototroph *Cereibacter sphaeroides* [30]. The second photosynthetic gene cluster (PGC) consists of tetrapyrrole biosynthesis genes *bchEJ*. The third PGC is homologous to the *puc* operon, which encodes the structural polypeptides of the light-harvesting-II peripheral antenna complex, and also contains a PucC family protein [29]. PucC encoded in model *puc* operons regulates response to oxygen tension and light intensity [30]. The fourth PGC spans 32-kb and contains further pigment synthesis genes, the photosynthetic reaction center subunit H, and various response regulators.

Transcriptomic data shows photosynthetic reaction center subunit M is significantly expressed during pEEU compared to aerobic growth (log_2fold change ~1; $p < 0.0001$), but not when grown with H_2 compared to aerobic growth (log_2fold change ~0.5; $p > 0.05$). The same is true for photosynthetic reaction center subunit L during pEEU compared to aerobic (log_2fold change ~1; $p < 0.05$), but not growth with H_2 compared to aerobic growth (log_2fold change ~0.5; $p > 0.05$).

3.2.2. Carbon Dioxide Fixation

The strain AB26 genome contains form I (BV509_05525) and form II (BV509_05520) of RuBisCO, the enzyme necessary for carbon dioxide fixation via the Calvin–Benson–Bassham (CBB) cycle [29]. The two forms reside in an operon together with the CbbQ RuBisCO-activating protein (BV509_05515), and a LsyR family transcriptional regulator (BV509_15210), which shares similarity with the RuBisCO regulator CbbR. The genes required for photorespiration via the Calvin–Benson–Bassham cycle are housed 2-Mb upstream from the RuBisCO enzyme gene cluster. As previously described [12], transcriptomic data reveals upregulation of form I RuBisCO (log_2fold change ~2, $p < 0.05$) but downregulation of form II RuBisCO (log_2fold change ~−0.35, $p > 0.1$) during pEEU, and upregulation of a formate dehydrogenase operon (log_2fold change ~4; $p < 0.0001$) [12]. The high upregulation of formate dehydrogenase may suggest an involvement of this enzyme in catalyzing CO_2 fixation [31].

3.2.3. Reducing Power

Strain AB26 is metabolically diverse, capable of aerobic heterotrophy, anaerobic photoheterotrophy, or anaerobic photoautotrophy using various inorganic compounds for electron donors (Figure 4). Strain AB26 contains a 16-kb operon coding for the synthesis and assembly of a nickel-dependent hydrogenase (BV509_06105-BV509_06195, Figure 4a). Expression of this nickel-dependent hydrogenase was highest during growth with thiosulfate compared to aerobic growth, with expression of genes within the operon ranging from ~1–3.5-fold ($p < 0.001$, Figure 4a). Strain AB26 is also capable of growth using thiosulfate via *sox*-like genes organized into 3 clusters. The first cluster contains two SoxXY-like carrier proteins (BV509_04530- BV509_04535) and a single metallohydrolase SoxH (BV509_04575). The second cluster contains the canonical *soxABCXYZ* genes (BV509_09605-BV509_09655). *soxR* (BV509_15000) is 1-Mb downstream, and is a putative transcriptional regulator of the *soxABCXYZ* genes [32]. *soxR* is slightly upregulated during growth with thiosulfate compared to aerobic (log$_2$fold change ~1.3; $p < 0.0001$, Figure 4b), which corresponds with the increased expression of other genes in the *sox* operon. However, *soxR* is not upregulated during growth with acetate as compared to aerobic growth (log$_2$fold change ~-0.3; $p < 0.0001$), despite increased expression patterns of the other *sox* genes (1–2-fold; $p < 0.0001$, Figure 4b). A carbon monoxide dehydrogenase (BV509_12715-BV509_12730, Figure 4a) and a formate dehydrogenase operon (BV509_15145-BV509_15165, Figure 4a) are also present in the strain AB26 genome.
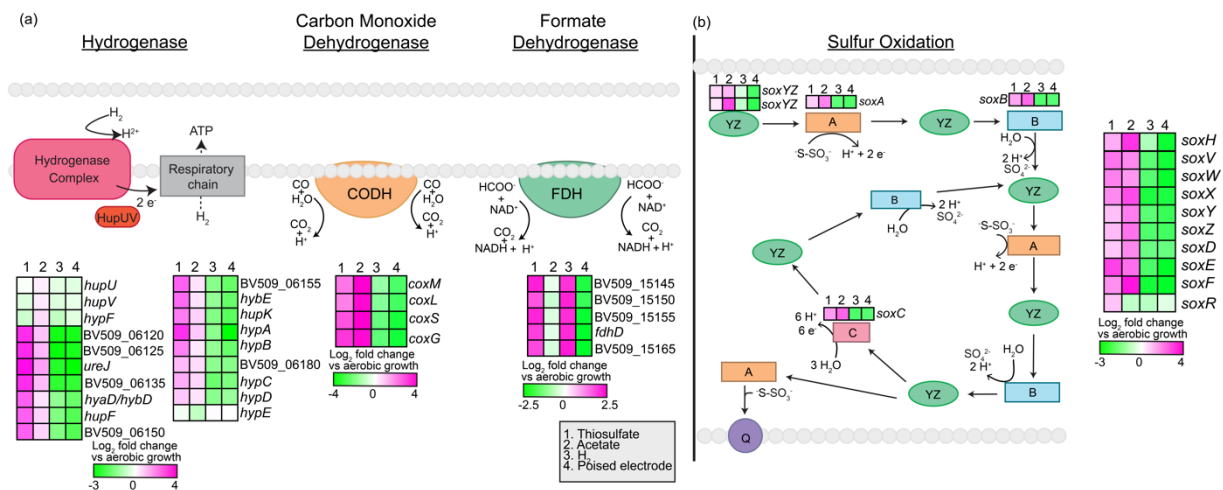
Figure 4. Genes involved in oxidation of inorganic compounds as a source of reducing power and their expression analysis via RNASeq. (**a**) nickel-dependent hydrogenase, carbon monoxide dehydrogenase, and formate dehydrogenase and their expression based on RNASeq. (**b**) sulfur oxidation pathway [33] and expression of the putative genes based on RNASeq. Heatmaps describe expression of four modes of metabolism (thiosulfate, acetate, H$_2$, and pEEU) compared to aerobic growth. (**a**) CODH (carbon monoxide dehydrogenase), FDH (formate dehydrogenase), CO (carbon monoxide), HCOO$^-$ (formate) (**b**) Q (quinone pool).

3.2.4. Regulation and Signaling

To be successful in transient environments requires genomic resources to sense and respond. Strain AB26 dedicates ~5% of its genome to signal transduction and regulation. Typical bacterial genomes dedicate 5–6% to regulation and sensing. Some purple nonsulfur bacteria such as *R. palustris* TIE-1 allot ~10% of their genome to environmental sensing, similar to other soil bacteria [34]. Strain AB26 contains 22 histidine kinases, 4 of which contain an HAMP domain (domain present in Histidine kinases, Adenylyl cyclases, Methyl-accepting proteins and Phosphatases) involved with chemotaxis [35]. Thirteen PAS domain-containing proteins are likely involved in light and oxygen sensing [36], and 2 GAF domains are potentially associated with photoreceptors [37]. Strain AB26

contains 19 RNA polymerase sigma factors and 121 transcriptional regulators belonging to 30 different families. During pEEU in strain AB26, transcripts encoding histidine kinases responding to envelope stress (BV509_04670, $\log_2$fold change ~2, $p < 0.0001$, Figure 5a), nitrate assimilation (BV509_11520, $\log_2$fold change ~3, $p < 0.0001$, Figure 5b), a sigma-70 factor involved in exocytoplasmic stress (BV509_00135, $\log_2$fold change ~2, $p < 0.01$, Figure 5e),and an anti-sigma antagonist (BV509_19770 $\log_2$fold change ~3, $p < 0.0001$, Figure 5d) were upregulated. This suggests that strain AB26 is potentially under stress during pEEU compared to growth under H_2. Also upregulated during pEEU is a flagellar switching response regulator (BV509_14575, ($\log_2$fold change ~2, $p < 0.0001$, Figure 5c).
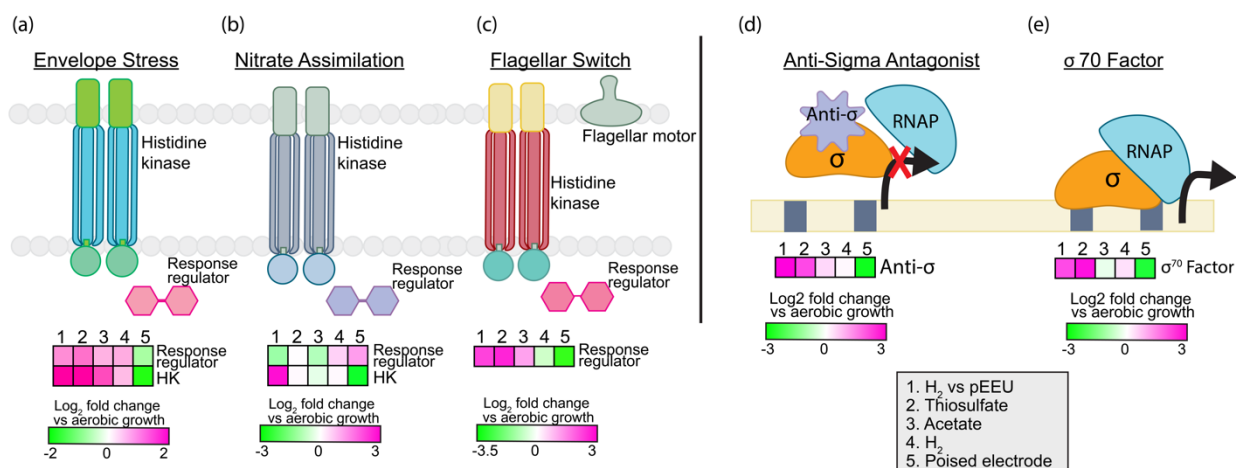
Figure 5. Expression analysis of significantly upregulated environmental responses such as envelope stress (**a**) nitrate assimilation (**b**) and flagellar switch (**c**) under pEEU compared to photoautotrophy with H_2 (growth condition 1). Heatmaps also describe expression of transcription regulators anti-sigma antagonist (**d**) and sigma 70 factor (**e**) under four modes of metabolism (thiosulfate, acetate, H_2, and pEEU) compared to aerobic growth.

3.2.5. Biodegradation and Carbon Storage

The diverse metabolisms of purple non-sulfur phototrophs allow for proliferation in industrial waste environments [38]. *Rhodopseudomonas palustris* and *Cereibacter sphaeroides* are two purple phototrophs known for their ability to metabolize diverse compounds including fatty acids, amino acids, aromatic compounds, and lignin. They are, therefore, important species in the bioremediation of wastewater and polluted environments [39–42]. Some purple phototrophs are also capable of synthesizing carbon storage polymers such as polyhydroxalkanoates (PHA), which they degrade in carbon-starved environments. The genome of strain AB26 encodes a variety of genes for the metabolism of aromatic compounds: a benzoate transporter (BV509_16875) and a 4-hydroxybenzoate octaprenyl-transferase (BV509_17190) for the first step of benzoate degradation, a salicylate esterase (BV509_RS17105), a biphenyl-2,3-diol 1,2-dioxygenase III-related protein (BV509_11920) for the degradation of biphenyl, a type II 3-dehydroquinate dehydratase (BV509_13850) for the degradation of quinate, and for the degradation of gentisate, a fumarylacetoacetate hydrolase family protein (BV509_03940) and a maleylacetoacetate isomerase (BV509_16060). Strain AB26 contains two operons for the degradation of phenylacetate (BV509_03065-BV509_03095, BV509_05275-BV509_05285) [43]. Similar to *R. palustris* TIE-1 [42], strain AB26 encodes genes for the synthesis and catabolism of polyhydroxybutyrate (PHB). *phaA*, which encodes a beta-ketothiolase, and *phaB*, which encodes an acetoacetyl-CoA reductase, are organized separately from each other and the remaining PHB gene cluster. The remainder of the PHB cluster contains the PHA depolymerase gene *phaZ*, the class 1 poly(R)-hydroxyalkanoic acid synthase *phaC*, which encodes the phasin protein affecting PHB granule accumulation and utilization, and the PHA synthesis repressor gene *phaR* (BV509_06270-BV509_06285). Unlike TIE-1, strain AB26 only has single-gene copies for the

synthesis and utilization of PHB. As previously described [12], the PHB biosynthesis enzymes were downregulated during pEEU as compared to aerobic growth (log_2 fold change ~0, $p > 0.05$), and the repressor *phaR* was slightly upregulated (log_2 fold change ~1, $p < 0.01$).

3.2.6. Transporters

The strain AB26 genome encodes 561 transport system genes, dedicating ~16% of its genome to transport. Of these, 347 transport genes are involved in primary transport, with 45 different ABC transport systems and 23 ATPase genes. Strain AB26 contains type III secretion system genes for flagella biosynthesis, and seven type II secretion system families (Figure 6).
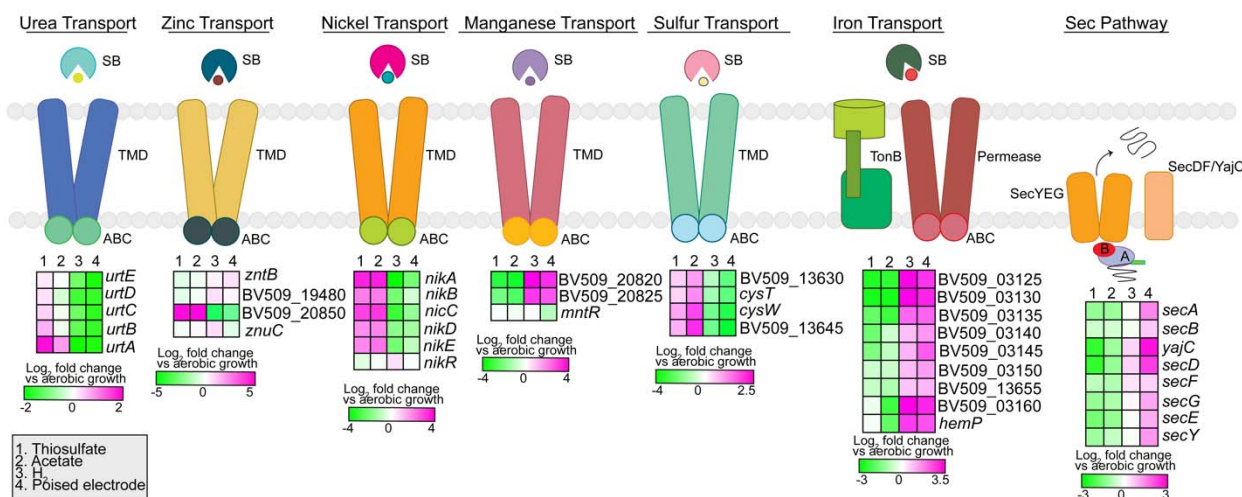
Figure 6. Expression analysis of transport systems in strain AB26. The Sec pathway genes are all upregulated during pEEU, which supports the need for protein export into the periplasmic space to facilitate extracellular electron transport. Heatmaps describe expression of four modes of metabolism (thiosulfate, acetate, H_2, and pEEU) compared to aerobic growth. SB (substrate binding domain), TMD (transmembrane domain), ABC (ATP cassette binding domain).

In total, 60 of the ABC transport genes were identified as various amino acid transport systems, with 22 genes dedicated to the transport of branched chain amino acids. For iron acquisition, strain AB26 contains 11 *tonB* genes and 5 iron transport systems. The lack of siderophore biosynthesis genes suggests that strain AB26 may transport siderophores produced by other bacteria into the cell. Strain AB26 contains two ferrous iron transport genes (BV509_02645-BV509_02650) for aerobic iron acquisition. Nutrient acquisition (Figure 6) for strain AB26 includes urea transport (BV509_09920-BV509_09945) and sulfate transport (BV509_13630-13645). Located on a plasmid are various metal uptake systems [15] including zinc (BV509_08850, BV509_19470, BV509_19480, BV509_20850), nickel (BV509_20695-BV509_20715, BV509_11055), and manganese (BV509_03550, BV509_20820-20825, Figure 6), which may help maintain metal ion homeostasis. Strain AB26 contains four microcin transporters, 14 multidrug transporters, and 9 efflux transporters. The Sec pathway was upregulated compared to aerobic growth only during pEEU, supporting the requirement of protein export in the periplasmic space to facilitate extracellular electron transfer across the membranes (Figure 6)

Strain AB26 contains 176 passive transport genes, with two multidrug and toxic compound extrusion (MATE) genes, 10 resistance-nodulation-cell division (RND) pumps, 23 major facilitator superfamily (MFS) genes, and 59 tripartite ATP-independent periplasmic (TRAP) transport genes. The TRAP family transporters are common in marine bacteria such as strain AB26 due to the natural Na^+ gradient in their environment [44].

3.2.7. Nitrogen Assimilation and Fixation

Strain AB26 uses the glutamine synthetase and glutamate synthase for ammonia assimilation in the cell. Strain AB26 genome contains two ammonia transporters, three glutamine synthetases, and a glutamine–synthetase adenyltransferase to regulate activity of glutamine–synthetases. Strain AB26 also encodes genes to convert urea to ammonia (BV509_09920-BV50-0995, Figure 7). Strain AB26 genome contains structural genes for a single molybdenum nitrogenase (BV509_14955-BV509_14965), assembly and cofactor genes, and the nitrogen fixation-specific sigma factor RpoN to fix nitrogen during anaerobic growth (Figure 7). During pEEU, one glutamine synthetase is highly upregulated (BV509_16830, $\log_2$fold change ~2.5, $p < 0.0001$), a second glutamine synthetase is slightly upregulated (BV509_16820, $\log_2$fold change ~0.75, $p < 0.001$; Figure 7a), and both are downregulated during photoheterotrophic growth compared to aerobic growth. Conversely, during pEEU, the glutamate synthase is downregulated as compared to aerobic growth (Figure 7a).
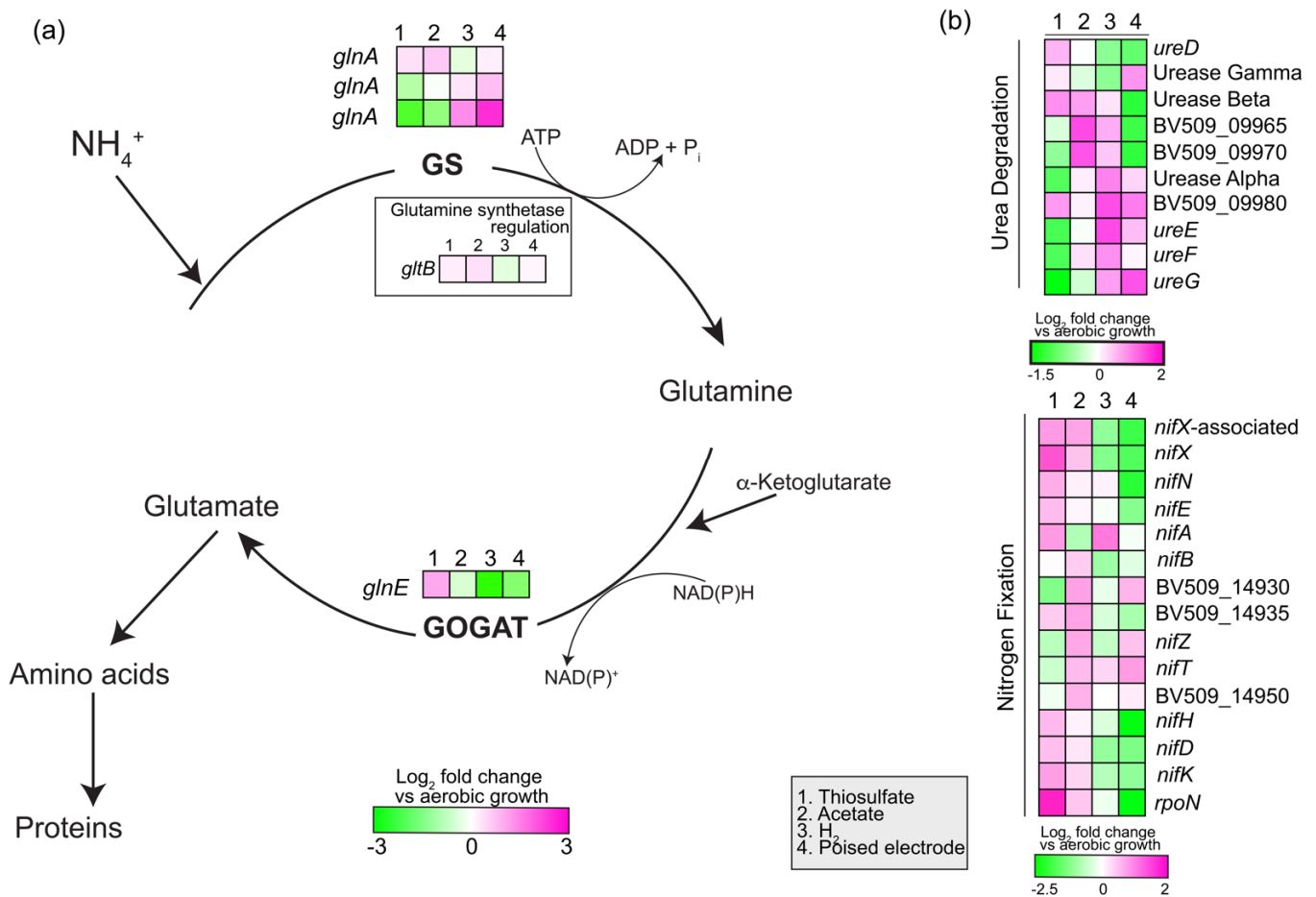
Figure 7. Expression analysis of genes involved in the glutamine synthetase/glutamate synthase (GS/GOGAT) pathway for ammonia assimilation in strain AB26 (**a**). Gene expression in urea degradation and nitrogen fixation (**b**). Heatmaps describe expression of four modes of metabolism (thiosulfate, acetate, H_2, and pEEU) compared to aerobic growth. (**a**) GS (glutamine synthetase), GOGAT (glutamate synthase) (**b**) BV509_09965 (hypothetical protein), BV509_09970 (DUF1127 domain-containing protein), BV509_09980 (DUF1127 domain-containing protein), BV509_14930 (4Fe-4S dicluster domain-containing protein), BV509_14935 (hypothetical protein), BV509_14950 (SIR2 family protein).

3.2.8. Extracellular Electron Uptake (EEU) Genes

To perform pEEU, Gram-negative bacteria require uptake systems consisting of electron transfer proteins spanning the outer membrane, periplasmic space, and inner membrane. These systems may also include porin proteins to span membranes and facilitate

the exchange of electrons from the cell exterior to the interior. While some characterized electron uptake systems exist [11,45], these uptake systems or homologs of them are not present in all bacteria that are hypothesized to be able to perform pEEU. Our group recently identified a novel protein important to phototrophic extracellular electron uptake in strain AB26, which does not encode any known EEU pathways in its genome; the di-heme *c*-type cytochrome EeuP [12]. As previously reported, EeuP is a novel pEEU-important protein found in A26 but also in species spanning the Proteobacteria and Acidobacteria. So far, no other components of the novel EEU pathway have been identified in strain AB26 other than EeuP. To identify other potential proteins that might be important in pEEU, a RNASeq guided comparative genomic approach with specific targets, such as cytochrome c-containing proteins, would be useful. This is because these proteins are often associated with extracellular electron transfer [3].

3.3. RNA-Seq Led Genomic Comparison between Rhodovulum spp.

Our group previously performed RNA-Seq in strain AB26 between photoautotrophy with H_2/CO_2 as electron donor/carbon source compared to pEEU via poised electrodes [12] in an effort to characterize the molecular mechanisms of pEEU in this novel isolate and identify the cellular sinks of electrons derived from extracellular sources. With these data, and the full genome sequences of the 14 other isolates comprising two *Rhodovulum* species, we can identify homologs of pEEU-important transcripts and infer ability to perform pEEU based on their presence or absence in each strain. Supplementary Table S1 shows the general genome features of all 15 Woods Hole isolates and publicly available genomes for both *R. visakhapatnamense* [46] and *R. sulfidophilum* DSM 2351 [47] and DSM1374 [48].

3.3.1. C-Type Cytochromes (Cyt_c)

Electron transfer with solid phase donors is commonly carried out by *c*-type cytochromes. *c*-type cytochromes mediate single electron transfers via a covalently bonded heme group [49,50], making them ideal candidates for pEEU processes. Therefore, our analysis of the pEEU vs. hydrogen transcript data focuses on identifying upregulated transcripts encoding *c*-type cytochromes, or upregulated hypothetical proteins and searching for the canonical CXXCH (where X is any amino acid residue) heme-binding motif within the protein sequence.

Diheme Cyt_c EeuP

RNA-Seq identified a significantly upregulated transcript ($\log_2$fold change ~2, $p < 0.0001$) in strain AB26 encoding the hypothetical protein later identified to be the di-heme EeuP (BV509_10070). BLASTp analysis identified homologous sequences of EeuP in all isolates except strain AB21. BLASTp results for strain AB21 returned two partial cyt_c sequences and one partial hypothetical protein sequence. These partial sequences separately align with the intact EeuP homologs and may be an artifact of assembly. From this we can hypothesize that all the isolates are pEEU-capable. Maximum likelihood phylogenetic analysis of the EeuP protein sequence (Figure 8) reveals EeuP in the isolates reflects the same relationship shown in the 16S and *pufM* phylogenetic trees (Figure 1), in that strain AB19 and strain AB26 EeuP sequences group outside of the *R. sulfidophilum* EeuP sequences. As previously reported [12], EeuP homologous sequences were also identified in Alpha-, Beta-, and Gammaproteobacteria classes and in one *Acidobacter* species. Further analysis of the species containing EeuP homologs may reveal the relationship between EeuP conservation and metabolic/environmental adaptation.
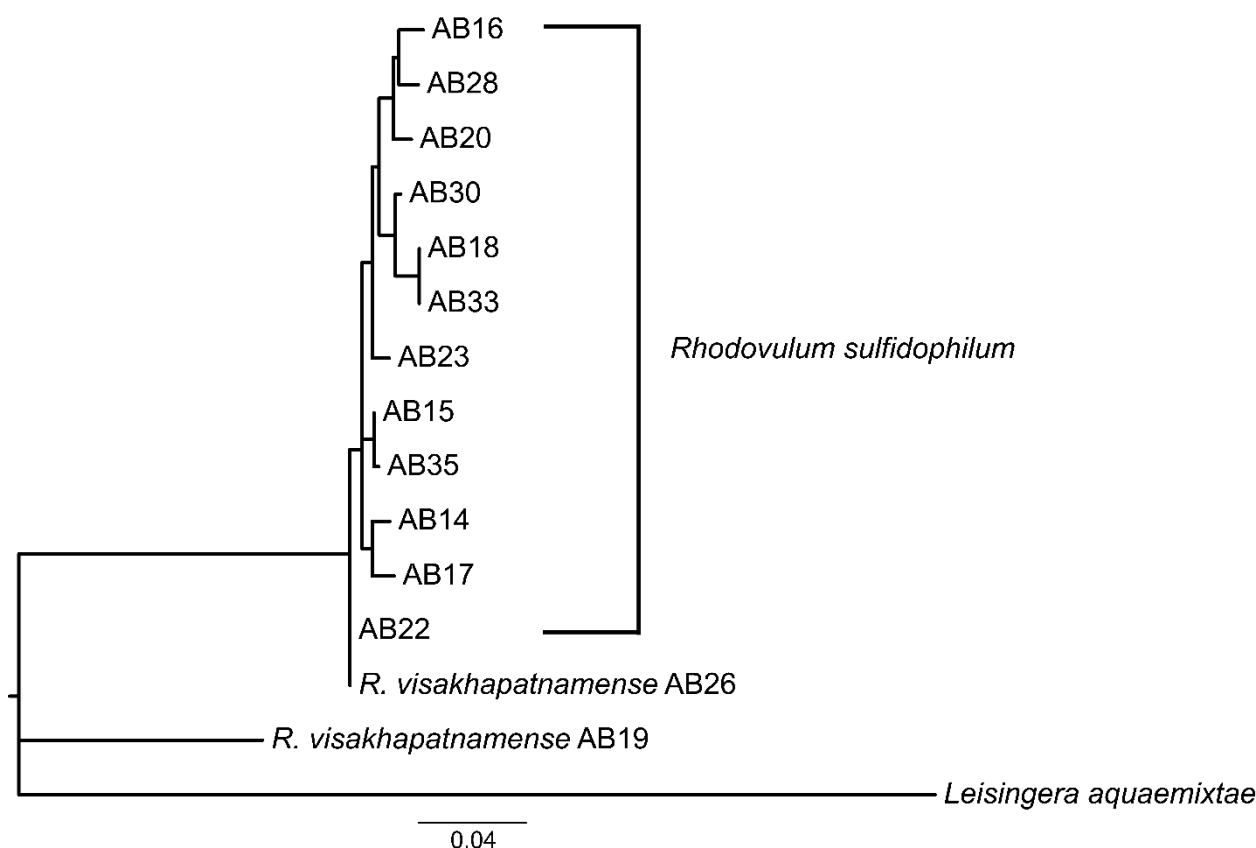
Figure 8. Phylogenetic tree showing protein sequence similarity of the diheme EeuP in Woods Hole isolates. Tree constructed by maximum-likelihood method using the Kimura-2 [25] parameter model, tested by bootstrapping (500 resamplings) in MEGA11 [22].

Hypothetical Protein BV509_18570

Another significantly upregulated transcript ($\log_2$fold change ~6, $p < 0.0001$) in strain AB26 encodes a hypothetical protein (BV509_18570). BV509_18570 is a 413-residue long protein containing a single heme-binding motif (CXXCH). InterProScan [51] predicts the molecular functions as electron transfer and heme binding. PRED-TAT [52] did not predict a secretory signal peptide (Sec), which would be expected for an electron transport protein involved in trans-membrane transport. Surprisingly, BLASTp analysis revealed no homologs of this open reading frame (ORF) within any of the *R. sulfidophilum* isolates. Homologs of BV509_18570 were found in *R. visakhapatnamense* strain AB19 (JMJ92_02255) and *R. visakhapatnamense* strain JA181 (EV657_1244). The homologous sequences are each 625 amino acid residues in length, ~200 amino acid residues longer than the strain AB26 BV509_18570. Protein sequence alignments in MAUVE [53] show JMJ92_02255 and EV657_1244 sequences align 212 residues before aligning to BV509_18570, and these additional residues contain a second heme-binding motif and a Sec signal peptide sequence. Therefore, it is possible the 413-residue long BV509_18570 may be truncated due to an assembly error and, similar to the related homologs, contains an additional heme-binding motif and Sec signal peptide sequence. BV509_18570 could be a another di-heme cyt$_c$ similar to EeuP, which was also annotated as a hypothetical protein, and be important for pEEU specific to *R. visakhapatnamense* spp. This hypothesis is supported by the high $\log_2$fold change of this transcript during pEEU and presence of a Sec signal peptide sequence, suggesting the transport of this protein across the inner membrane to the periplasmic space.

Diheme Cyt$_c$ BV509_14915

This transcript (log$_2$fold change ~6, $p < 0.0001$) encodes a cyt$_c$ with two heme-binding motifs. BV509_14915 is 347 residues long and contains a Sec signal peptide. Because of this, BLASTp analysis shows all 15 marine *Rhodovulum* spp. isolates contain a homolog of this di-heme cyt$_c$. Given the high log$_2$fold change, Sec signal peptide sequence, heme-binding motifs, and the presence of homologs within the other isolates' genomes, this cytochrome could be a conserved component of the pEEU molecular pathway in the *Rhodovulum* spp. isolates.

Cyt$_c$ BV509_15885

The transcript encoding the gene BV509_15885 is annotated as a cyt$_c$. RNA-Seq data did not identify this cyt$_c$ as significantly upregulated during pEEU (log$_2$fold change -0.48), but it was identified through heme staining (membrane fraction) followed by mass spectroscopy compared to hydrogen grown cultures [12]. This cyt$_c$ contains 5 heme-binding motifs and a predicted Sec signal peptide. BLASTp analysis identified homologs of this cyt$_c$ in strain AB19 and *R. visakhapatnamense* strain JA181. Homologs in the *R. sulfidophilum* isolates show 71–75% identity.

3.3.2. Transcriptional Regulators

To understand expression patterns and potential regulators of pEEU genes, the pEEU vs. hydrogen transcript data was analyzed for any upregulated transcriptional regulators.

Mar Family Transcription Regulator BV509_01095

BV509_01095 encodes a hypothetical protein identified as a member of the multi antibiotic resistance family transcriptional regulator (log$_2$fold change ~6, $p < 0.001$). InterProScan predicts phosphorelay signal transduction function. BV509_01095 is present only in strain AB19 and strain AB26 via BLASTp analysis, but not *R. visakhapatnamense* strain JA181 or any of the *R. sulfidophilum* isolates.

Transcriptional Regulator BV509_15825

This transcript is another transcriptional regulator upregulated during pEEU in strain AB26 (log$_2$fold change ~4, $p < 0.0001$). This regulator is also predicted to be involved in signal transduction response regulation via InterProScan, with DNA-binding wing-helix domain and tetratricopeptide repeat domain. BLASTp reveals this transcription regulator is unique to strain AB19, strain AB26, and *R. visakhapatnamense* strain JA181 but not the *R. sulfidophilum* isolates.

4. Conclusions

With the current accessibility of genome sequencing and ability to isolate bacterial species from diverse environments, this paper shows that taxonomy requires continuous addressing. Here, we provide a current taxonomic classification of 15 *Rhodovulum* spp. isolated from a marine estuary, specifically the re-classification of two of the original 15 isolates previously identified as *R. sulfidophilum* now as *R. visakhapatnamense*. These isolates are capable of many modes of metabolism, including photoferrotrophy and pEEU. To expand the characterization of *R. visakhapatnamense*, which at present is represented by 3 genomes assembled at the scaffold level, we describe the genome contents of *R. visakhapatnamense* strain AB26, and pair the genome data with expression data under anaerobic photoheterotrophy (acetate), photoautotrophy (H$_2$ or thiosulfate as electron donor), and photosynthetic extracellular electron uptake (poised electrode as electron donor). Upregulated transcripts under pEEU of interest to the analysis include *c*-type cytochromes and transcriptional regulators. BLASTp shows some transcripts of interest are unique to *R. visakhapatnamense* genomes and could be indicative of adaptation of molecular mechanisms for pEEU explained by speciation.

What will be interesting to observe in the future is continued comparison of expression profiles and proteomes under pEEU and photoferrotrophy between *R. sulfidophilum* and *R. visakhapatnamense* strains from Woods Hole, as this data points to differences between species. Mutational analysis of genes of interest in each species will be necessary to determine their role under these growth conditions. With the data presented here, we can begin to characterize the metabolic pathway of pEEU in *Rhodovolum* species.

References

1. Shu, W.-S.; Huang, L.-N. Microbial diversity in extreme environments. *Nat. Rev. Microbiol.* **2022**, *20*, 219–235. [CrossRef]
2. Rosenbaum, M.; Aulenta, F.; Villano, M.; Angenent, L.T. Cathodes as electron donors for microbial metabolism: Which extracellular electron transfer mechanisms are involved? *Bioresour. Technol.* **2011**, *102*, 324–333. [CrossRef] [PubMed]
3. Gupta, D.; Guzman, M.S.; Bose, A. Extracellular electron uptake by autotrophic microbes: Physiological, ecological, and evolutionary implications. *J. Ind. Microbiol. Biotechnol.* **2020**, *47*, 863–876. [CrossRef] [PubMed]
4. Falkowski, P.G.; Fenchel, T.; Delong, E.F. The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science* **2008**, *320*, 1034–1039. [CrossRef]
5. Lovley, D.R.; Nevin, K.P. Electrobiocommodities: Powering microbial production of fuels and commodity chemicals from carbon dioxide with electricity. *Curr. Opin. Biotechnol.* **2013**, *24*, 385–390. [CrossRef]
6. Rabaey, K.; Rodriguez, J.; Blackall, L.L.; Keller, J.; Gross, P.; Batstone, D.; Verstraete, W.; Nealson, K.H. Microbial ecology meets electrochemistry: Electricity-driven and driving communities. *ISME J.* **2007**, *1*, 9–18. [CrossRef] [PubMed]
7. Rabaey, K.; Rozendal, R.A. Microbial electrosynthesis—Revisiting the electrical route for microbial production. *Nat. Rev. Microbiol.* **2010**, *8*, 706–716. [CrossRef]
8. Bose, A.; Gardel, E.J.; Vidoudez, C.; Parra, E.; Girguis, P.R. Electron uptake by iron-oxidizing phototrophic bacteria. *Nat. Commun.* **2014**, *5*, 3391. [CrossRef] [PubMed]
9. Rowe, A.R.; Rajeev, P.; Jain, A.; Pirbadian, S.; Okamoto, A.; Gralnick, J.A.; El-Naggar, M.Y.; Nealson, K.H. Tracking electron uptake from a cathode into *Shewanella* cells: Implications for energy acquisition from solid-substrate electron donors. *MBio* **2018**, *9*, e02203-17. [CrossRef]

10. Tefft, N.M.; TerAvest, M.A. Reversing an extracellular electron transfer pathway for electrode-driven acetoin reduction. *ACS Synth. Biol.* **2019**, *8*, 1590–1600. [CrossRef]

11. Jiao, Y.; Newman, D.K. The pio operon is essential for phototrophic Fe (II) oxidation in *Rhodopseudomonas palustris* TIE-1. *J. Bacteriol.* **2007**, *189*, 1765–1773. [CrossRef] [PubMed]

12. Gupta, D.; Guzman, M.S.; Rengasamy, K.; Stoica, A.; Singh, R.; Ranaivoarisoa, T.O.; Davenport, E.J.; Bai, W.; McGinley, B.; Meacham, J.M.; et al. Photoferrotrophy and phototrophic extracellular electron uptake is common in the marine anoxygenic phototroph *Rhodovulum sulfidophilum*. *ISME J.* **2021**, *15*, 3384–3398. [CrossRef] [PubMed]

13. Guzman, M.S.; Rengasamy, K.; Binkley, M.M.; Jones, C.; Ranaivoarisoa, T.O.; Singh, R.; Fike, D.A.; Meacham, J.M.; Bose, A. Phototrophic extracellular electron uptake is linked to carbon dioxide fixation in the bacterium *Rhodopseudomonas palustris*. *Nat. Commun.* **2019**, *10*, 1355. [CrossRef] [PubMed]

14. Shi, L.; Dong, H.; Reguera, G.; Beyenal, H.; Lu, A.; Liu, J.; Yu, H.-Q.; Fredrickson, J.K. Extracellular electron transfer mechanisms between microorganisms and minerals. *Nat. Rev. Microbiol.* **2016**, *14*, 651–662. [CrossRef]

15. Guzman, M.S.; McGinley, B.; Santiago-Merced, N.; Gupta, D.; Bose, A.J.G.A. Draft genome sequences of three closely related isolates of the purple nonsulfur bacterium *Rhodovulum sulfidophilum*. *Genome Announc.* **2017**, *5*, e00029-17. [CrossRef]

16. Guzman, M.S.; Bose, A. Draft genome sequences of four Rhodobacter sphaeroides strains isolated from a marine ecosystem. *Microbiol. Resour. Announc.* **2019**, *8*, e01648-18. [CrossRef]

17. Srinivas, T.; Kumar, P.A.; Sasikala, C.; Ramana, C.V.; Imhoff, J.F. *Rhodovulum visakhapatnamense* sp. nov. *Int. J. Syst. Evol. Microbiol.* **2007**, *57*, 1762–1764. [CrossRef]

18. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]

19. Bosi, E.; Donati, B.; Galardini, M.; Brunetti, S.; Sagot, M.-F.; Lió, P.; Crescenzi, P.; Fani, R.; Fondi, M. MeDuSa: A multi-draft based scaffolder. *Bioinformatics.* **2015**, *31*, 2443–2451. [CrossRef]

20. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef]

21. Zhao, Y.; Wu, J.; Yang, J.; Sun, S.; Xiao, J.; Yu, J. PGAP: Pan-genomes analysis pipeline. *Bioinformatics* **2011**, *28*, 416–418. [CrossRef] [PubMed]

22. Richter, M.; Rosselló-Móra, R.; Oliver Glöckner, F.; Peplies, J. JSpeciesWS: A web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* **2015**, *32*, 929–931. [CrossRef] [PubMed]

23. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027. [CrossRef] [PubMed]

24. Béjà, O.; Suzuki, M.T.; Heidelberg, J.F.; Nelson, W.C.; Preston, C.M.; Hamada, T.; Eisen, J.A.; Fraser, C.M.; DeLong, E.F. Unsuspected diversity among marine aerobic anoxygenic phototrophs. *Nature* **2002**, *415*, 630–633. [CrossRef] [PubMed]

25. Yutin, N.; Suzuki, M.T.; Béjà, O. Novel primers reveal wider diversity among marine aerobic anoxygenic phototrophs. *Appl. Environ. Microbiol.* **2005**, *71*, 8958–8962. [CrossRef] [PubMed]

26. Kimura, M.; Ohta, T. On the stochastic model for estimation of mutational distance between homologous proteins. *J. Mol. Evol.* **1972**, *2*, 87–90. [CrossRef] [PubMed]

27. Jones, D.T.; Taylor, W.R.; Thornton, J.M. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* **1992**, *8*, 275–282. [CrossRef]

28. Konstantinidis, K.T.; Tiedje, J.M. Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 2567–2572. [CrossRef]

29. Blankenship, R.E. *Molecular Mechanisms of Photosynthesis*; John Wiley & Sons: Hoboken, NJ, USA, 2021.

30. Nickens, D.G.; Bauer, C.E. Analysis of the puc operon promoter from *Rhodobacter capsulatus*. *J. Bacteriol.* **1998**, *180*, 4270–4277. [CrossRef]

31. Hartmann, T.; Leimkühler, S. The oxygen-tolerant and NAD$^+$-dependent formate dehydrogenase from *Rhodobacter capsulatus* is able to catalyze the reduction of CO_2 to formate. *FEBS J.* **2013**, *280*, 6083–6096. [CrossRef]

32. Lahiri, C.; Mandal, S.; Ghosh, W.; Dam, B.; Roy, P. A novel gene cluster *soxSRT* is essential for the chemolithotrophic oxidation of thiosulfate and tetrathionate by *Pseudaminobacter salicylatoxidans* KCT001. *Curr. Microbiol.* **2006**, *52*, 267–273. [CrossRef] [PubMed]

33. Grabarczyk, D.B.; Berks, B.C. Intermediates in the Sox sulfur oxidation pathway are bound to a sulfane conjugate of the carrier protein SoxYZ. *PLoS ONE* **2017**, *12*, e0173395. [CrossRef] [PubMed]

34. Larimer, F.W.; Chain, P.; Hauser, L.; Lamerdin, J.; Malfatti, S.; Do, L.; Land, M.L.; Pelletier, D.A.; Beatty, J.T.; Lang, A.S.; et al. Complete genome sequence of the metabolically versatile photosynthetic bacterium *Rhodopseudomonas palustris*. *Nat. Biotechnol.* **2004**, *22*, 55–61. [CrossRef] [PubMed]

35. Aravind, L.; Ponting, C.P. The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins. *FEMS Microbiol. Lett.* **1999**, *176*, 111–116. [CrossRef]

36. Kyndt, J.A.; Meyer, T.E.; Cusanovich, M.A. Photoactive yellow protein, bacteriophytochrome, and sensory rhodopsin in purple phototrophic bacteria. *Photochem. Photobiol. Sci.* **2004**, *3*, 519–530. [CrossRef]

37. Martinez, S.E.; Beavo, J.A.; Hol, W.G. GAF domains: Two–billion–year–old molecular switches that bind cyclic nucleotides. *Mol. Interv.* **2002**, *2*, 317–323. [CrossRef]

38. Hiraishi, A.; Kitamura, H. Distribution of phototrophic purple nonsulfur bacteria in activated sludge systems and other aquatic environments. *Nippon Suisan Gakkaishi* **1984**, *50*, 1929–1937. [CrossRef]

39. Wu, P.; Zhang, Y.; Chen, Z.; Wang, Y.; Zhu, F.; Cao, B.; Wu, Y.; Li, N. The organophosphorus pesticides in soil was degradated by *Rhodobacter sphaeroides* after wastewater treatment. *Biochem. Eng. J.* **2019**, *141*, 247–251. [CrossRef]

40. Hassan, M.A.; Shirai, Y.; Kusubayashi, N.; Karim, M.I.A.; Nakanishi, K.; Hasimoto, K. The production of polyhydroxyalkanoate from anaerobically treated palm oil mill effluent by *Rhodobacter sphaeroides*. *J. Biosci. Bioeng.* **1997**, *83*, 485–488. [CrossRef]

41. Brandl, H.; Gross, R.A.; Lenz, R.W.; Lloyd, R.; Fuller, R.C. The accumulation of poly (3-hydroxyalkanoates) in *Rhodobacter sphaeroides*. *Arch. Microbiol.* **1991**, *155*, 337–340. [CrossRef]

42. Ranaivoarisoa, T.O.; Singh, R.; Rengasamy, K.; Guzman, M.S.; Bose, A. Towards sustainable bioplastic production using the photoautotrophic bacterium *Rhodopseudomonas palustris* TIE-1. *J. Ind. Microbiol. Biotechnol.* **2019**, *46*, 1401–1417. [CrossRef] [PubMed]

43. Fuchs, G.; Boll, M.; Heider, J. Microbial degradation of aromatic compounds—From one strategy to four. *Nat. Rev. Microbiol.* **2011**, *9*, 803–816. [CrossRef] [PubMed]

44. Mulligan, C.; Kelly, D.J.; Thomas, G.H. Tripartite ATP-independent periplasmic transporters: Application of a relational database for genome-wide analysis of transporter gene frequency and organization. *J. Mol. Microbiol. Biotechnol.* **2007**, *12*, 218–226. [CrossRef] [PubMed]

45. Shi, L.; Rosso, K.M.; Zachara, J.M.; Fredrickson, J.K. Mtr extracellular electron-transfer pathways in Fe (III)-reducing or Fe (II)-oxidizing bacteria: A genomic perspective. *Biochem. Soc. Trans.* **2012**, *40*, 1261–1267. [CrossRef] [PubMed]

46. Srinivas, A.; Kumar, B.V.; Sree, B.D.; Tushar, L.; Sasikala, C.; Ramana, C.V.J. *Rhodovulum salis* sp. nov. and *Rhodovulum viride* sp. nov., phototrophic Alphaproteobacteria isolated from marine habitats. *Int. J. Syst. Evol. Microbiol.* **2014**, *64*, 957–962. [CrossRef] [PubMed]

47. Nagao, N.; Hirose, Y.; Misawa, N.; Ohtsubo, Y.; Umekage, S.; Kikuchi, Y. Complete genome sequence of *Rhodovulum sulfidophilum* DSM 2351, an extracellular nucleic acid-producing bacterium. *Genome Announc.* **2015**, *3*, e00388-15. [CrossRef] [PubMed]

48. Masuda, S.; Hori, K.; Maruyama, F.; Ren, S.; Sugimoto, S.; Yamamoto, N.; Mori, H.; Yamada, T.; Sato, S.; Tabata, S.; et al. Whole-genome sequence of the purple photosynthetic bacterium *Rhodovulum sulfidophilum* strain W4. *Genome Announc.* **2013**, *1*, e00577-13. [CrossRef]

49. Allen, J.W.; Daltrop, O.; Stevens, J.M.; Ferguson, S.J. C-type cytochromes: Diverse structures and biogenesis systems pose evolutionary problems. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2003**, *358*, 255–266. [CrossRef]

50. Estevez-Canales, M.; Kuzume, A.; Borjas, Z.; Füeg, M.; Lovley, D.; Wandlowski, T.; Esteve-Núñez, A. A severe reduction in the cytochrome C content of *Geobacter sulfurreducens* eliminates its capacity for extracellular electron transfer. *Environ. Microbiol. Rep.* **2015**, *7*, 219–226. [CrossRef]

51. Jones, P.; Binns, D.; Chang, H.-Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al. InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **2014**, *30*, 1236–1240. [CrossRef]

52. Bagos, P.G.; Nikolaou, E.P.; Liakopoulos, T.D.; Tsirigos, K.D. Combined prediction of Tat and Sec signal peptides with hidden Markov models. *Bioinformatics* **2010**, *26*, 2811–2817. [CrossRef] [PubMed]

53. Darling, A.C.; Mau, B.; Blattner, F.R.; Perna, N.T. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **2004**, *14*, 1394–1403. [CrossRef] [PubMed]

Article

Genome Sequence and Characterization of a Xanthorhodopsin-Containing, Aerobic Anoxygenic Phototrophic *Rhodobacter* Species, Isolated from Mesophilic Conditions at Yellowstone National Park

John A. Kyndt [1,*] **, Sydney Robertson** [1] **, Isabella B. Shoffstall** [2] **, Robert F. Ramaley** [3] **and Terrance E. Meyer** [4,†]

1   College of Science and Technology, Bellevue University, Bellevue, NE 68005, USA; srobertson011@yahoo.com
2   Department of Chemistry and Biochemistry, New Mexico State University, Las Cruces, NM 88003, USA; bella4@nmsu.edu
3   Department of Biochemistry and Molecular Biology, University of Nebraska Medical Center, Omaha, NE 68198, USA; rramaley@unmc.edu
4   Department of Chemistry and Biochemistry, The University of Arizona, Tucson, AZ 85721, USA; temeyer@email.arizona.edu
*   Correspondence: jkyndt@bellevue.edu
†   T.E.M. passed away during the preparation of this manuscript.

Abstract: The genus *Rhodobacter* consists of purple nonsulfur photosynthetic alphaproteobacteria known for their diverse metabolic capabilities. Here, we report the genome sequence and initial characterization of a novel *Rhodobacter* species, strain M37P, isolated from Mushroom hot spring runoff in Yellowstone National Park at 37 °C. Genome-based analyses and initial growth characteristics helped to define the differentiating characteristics of this species and identified it as an aerobic anoxygenic phototroph (AAP). This is the first AAP identified in the genus *Rhodobacter*. Strain M37P has a pinkish-red pigmentation that is present under aerobic dark conditions but disappears under light incubation. Whole genome-based analysis and average nucleotide identity (ANI) comparison indicate that strain M37P belongs to a specific clade of recently identified species that are genetically and physiologically unique from other representative *Rhodobacter* species. The genome encodes a unique xanthorhodopsin, not found in any other *Rhodobacter* species, which may be responsible for the pinkish-red pigmentation. These analyses indicates that strain M37P is a unique species that is well-adapted to optimized growth in the Yellowstone hot spring runoff, for which we propose the name *Rhodobacter calidifons* sp. nov.

Keywords: aerobic anoxygenic phototrophs; AAP; *Rhodobacter*; purple nonsulfur bacteria; hot spring; Yellowstone; xanthorhodopsin

1. Introduction

Species of the genus *Rhodobacter* are purple nonsulfur photosynthetic alphaproteobacteria that have unique morphological and physiological properties [1,2]. *Rhodobacter* belongs to the *Alphaproteobacteria* (Pseudomonadota phylum) and representatives are found in a wide range of environments. The two best-studied species in the *Rhodobacter* genus are *Rba. capsulatus* and *Rba. sphaeroides*, which are often used as model systems for bacterial photosynthesis studies. In more recent developments, the use and studies of *Rhodobacter* have been expanded to a wide range of applications, from the consideration of *Rba. sphaeroides* for in vivo imaging of tumor cells with natural advantages over other fluorescent materials [3], to the use of various *Rhodobacter* LPS molecules for TLR-mediated chemokine expression [4–7], and all the way to the use of *Rhodobacter* extracts as a food supplement to regulate immune responses in humans and animals [7].

Members of the genus *Rhodobacter* all perform anoxygenic photosynthesis. They also fix nitrogen and thereby play key roles in biogeochemical cycles. Most *Rhodobacter* species

can grow aerobically, but when doing so repress the synthesis of photosynthetic pigments, and all oxidize sulfide at low concentrations but with differences in the oxidation products. *Rba. blasticus* is an unusual species that produces lamellar membranes and, prior to the recent discovery of *Rba. thermarum*, there were no others sharing this property [8,9].

Currently, there are 21 validly named species in the *Rhodobacter* genus: *Rba. aestuarii* [10], *Rba. alkalitolerans* [11], *Rba. amnigenus* [12], *Rba. azollae* [13], *Rba. azotoformans* [14], *Rba. blasticus* [8], *Rba. capsulatus* [1,15], *Rba. flagellatus* [16], *Rba. johrii* [17], *Rba. lacus* [13], *Rba. maris* [18], *Rba. ovatus* [19], *Rba. sediminis* [20], *Rba. sediminicola* [21], *Rba. sphaeroides* [22], *Rba. tardus* [23], *Rba. thermarum* [9], *Rba. veldkampii* [24], *Rba. vinaykumarii* [25], *Rba. viridis* [26], and *Rba. xinxiangensis* [27]. Eight of these were discovered in the last three years and not all of these have their genomes sequenced; however, many metagenome-derived uncultivated *Rhodobacter* genomes are found in NCBI GenBank. More will undoubtedly be added since metagenomic sequencing is becoming more prevalent in environmental analysis. A recent effort to resolve the diversity in the taxonomy of the *Rhodobacter* genus [28] identified the existence of four monophyletic clusters, with interspersing chemotrophic species, and proposed the reclassification of several members of the genus *Rhodobacter* into three new genera: *Luteovulum*, *Phaeovulum*, and *Fuscovulum*. Although a revision of the taxonomy of the *Rhodobacter* was certainly due, that study was performed using the genomic comparison of only 12 strains, and still leaves many of the species designated as *Rhodobacter* dispersed within the larger *Rhodobacter* phylogeny. For the purpose of the current discussion, the *Rhodobacter* designation is used and the new genus names (*Luteovulum*, *Phaeovulum*, and *Fuscovulum*) are indicated where appropriate.

A novel *Rhodobacter* species, designated strain M37P, was isolated in 2006 by RFR from the Mushroom hot spring runoff within the Lower Geyser Basin of Yellowstone National Park. The isolate was purified using aerobic *Thermus* medium with rifampin [29]. A partial 16S rRNA (470 bp) construct of the pure culture was sequenced (GenBank accession DQ363135), and identified "*Rhodobacter gluconicum*" as the closest relative at that time (97% identity; 451/466 bp), which unfortunately is not available in culture collections and a description has not been published. Since the initial isolation, several new species have come to light that share an even closer relationship. Due to its unique habitat and the fact that very little is known about this new isolate, we decided to sequence its genomic DNA and compare it to the better characterized *Rhodobacter* species.

2. Materials and Methods

2.1. Strain Isolation

The M37P strain was originally isolated from a runoff channel, at a 37 °C temperature sampling point, which flowed from Mushroom Spring, a slightly alkaline (pH 8.0–8.2) hot spring in the lower geyser basin of Yellowstone National Park (US). The strain was labeled M37P, for Mushroom Spring, 37 °C Pink-colored isolate. The original isolate was grown aerobically on Thermus medium [29] containing 0.1% yeast extract and 0.1% tryptone in Castenholz salts (pH 7.6–7.8). Cultures were grown on agar plates incubated at 37 °C and allowed to grow for 2–3 weeks. Pinkish-red colonies were picked and consecutively transferred to new plates three times to obtain pure cultures and subsequently used to inoculate liquid medium.

2.2. Cultivation

Subsequent growth was also conducted in Thermus medium containing 0.3% yeast extract, 0.3% tryptone, and 0.3% casamino acids (final pH 7.6–7.8). Increasing the concentration of yeast extract, tryptone, or casamino acids above the 0.3% level did not increase the growth rate or final cell yield. Metabolism of malate, succinate, glutamate, and glucose for growth was tested in Thermus medium [29] and in two different minimal media, Minimal 22 and RCVB [30], at concentrations of 30 mM. Growth factor requirements were tested for thiamin (1 mg/L), nicotinic acid (1 mg/L), and biotin (15 µg/L) in RCVB medium. Growth was tested in 50–100 mL liquid cultures and agar plates, under aerobic and anaerobic

conditions, both in light and dark conditions for several weeks. The growth temperature range tested in shaking liquid cultures was 30–45 °C. The pH range tested was from 5–10 in liquid Thermus medium containing 0.3% yeast extract, 0.3% tryptone, and 0.3% casamino acids. No growth was observed in any of the anaerobic conditions. Oxidase activity was tested using BBL DrySlides from Becton Dickinson and catalase activity was tested using 3% (w/v) H_2O_2 by assessing bubble production as a positive result. Cells were observed using a MEIJI Techno (Chikumazawa, Japan) MT4200H microscope equipped with brightfield and phase contrast condenser and images were obtained with an attached Motic digital camera (Motic North America, Richmond, BC, Canada).

Pigments were extracted from dark and light incubated cultures using a methanol extraction from 50 mL aerobically grown cultures. Cells were pelleted by centrifugation (5 min at $10,000 \times g$) and resuspended in 5 mL methanol, after which they were centrifuged again and the supernatant was used for spectral analysis. Absorption spectra were obtained using an Evolution 300 UV-vis spectrophotometer (Thermo Scientific, Waltham, MA, USA) with the VisionPro software (version 4.5.0).

2.3. Genome Sequencing

We isolated gDNA of the M37P isolate from living cultures with the GeneJET DNA purification kit (Thermo Scientific). The cells form a visible capsule and cluster together during cultivation, which complicated lysis with the standard protocol. An additional cell lysis step in 5% SDS and proteinase K (1 mg/mL) was needed, and incubation at 65 °C was performed for several hours to assure sufficient lysis of the cells. The quality and quantity of purified DNA were determined using Qubit and NanoDrop and showed a concentration of 21 ng/μL and an A_{260}/A_{280} absorbance ratio of 1.73, which indicated that there was sufficient high-quality DNA for sequencing. The DNA library was prepared following the Nextera DNA Flex library prep kit instructions (Illumina, San Diego, CA, USA) using 300 ng of genomic DNA. The genome was sequenced by an Illumina MiniSeq system with 500 μL of a 1.8-pM library, using paired-end sequencing (2×150 bp). Quality control of the reads was performed using FASTQC within BaseSpace (Illumina; version 1.0.0), using a k-mer size of 5 and contamination filtering. De novo assembly of the data was performed using SPAdes (version 3.10.0) [31] through PATRIC [32]. The assembled genome had a coarse consistency of 98.5% and fine consistency of 97.5% and was considered of good quality. The genome sequence was annotated using the Rapid Annotations using Subsystems Technology (RAST) server (version 2.0) [33]. Completeness and possible contamination were assessed with CheckM [34].

2.4. Whole Genome Comparison

The average percentage nucleotide identity (ANI) was calculated using JSpecies [35], which uses a pairwise genome comparison algorithm to measure probability if two or more genomes belong to the same species. The compared genomes were either selected from the curated reference database GenomesDB (by JSpecies) or uploaded from GenBank for those not present in the JSpecies database. Reciprocal average amino acid identity (AAI) comparison was calculated using the AAI calculator from the Konstantinidis Lab (http://enve-omics.ce.gatech.edu/aai/index; accessed on 1 April 2022) [36]. The following alignment options were used: minimum length, zero aa; minimum identity, 20%; minimum alignments. 50.

A whole genome-based phylogenetic tree was generated using the CodonTree method within PATRIC [32], which used PGFams as homology groups. Representative genomes from different *Rhodobacter* species were selected for the analysis. The number of single-copy genes from PGFams was set at 1000, and 557 PGFams were found among the selected genomes using the CodonTree analysis. Protein sequences were aligned using MUSCLE [37], and the nucleotide coding gene sequences were aligned using the Codon_align function of BioPython. The aligned proteins and coding DNA from single-copy genes

were used for RaxML analysis [38,39] within PATRIC [32]. iTOL was used to draw the phylogenetic trees expressed in the Newick phylogenetic tree format [40].

2.5. Single-Gene Phylogenetic Comparisons

The multiple sequence alignments for the 16S rRNA and xanthorhodopsin comparisons were performed using ClustalW [41]. For the 16S rRNA analysis, the phylogenetic tree was designed by using the maximum likelihood method and general time reversible model [42]. The tree with the highest log likelihood was used. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the maximum composite likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.8636)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 80.81% sites).

For the xanthorhodopsin tree, the evolutionary history was inferred by using the maximum likelihood method and Le_Gascuel_2008 model [43]. The tree with the highest log likelihood (−7279.39) was used. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the JTT model, and then selecting the topology with superior log likelihood value. A discrete gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.1987)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 14.31% sites).

In both cases, 500 bootstrapping rounds were used to construct the tree. The final trees are drawn to scale, with branch lengths measured in the number of substitutions per site. The 16S rDNA analysis involved 29 sequences, derived from whole genomes where available, with the following PATRIC-BRC accession numbers: fig|2715277.3.rna.28 (*Rhodobacter* sp. M37P); fig|2883999.5.rna.51 (*Rhodobacter* sp. Har01); fig|2807096.3.rna.45 (*Rhodobacter* sp. N10); fig|2852097.3.rna.48 (*Rhodobacter* sp. HSP-20); fig|1985673.3.rna.53 (*Rhodobacter* sp. CCP-1); fig|2593022.3.rna.42 (*Rhodobacter* sp. SYSU G02214); fig|2593021.3.rna.42 (*Rhodobacter* sp. SYSU G03088); fig|2593019.3.rna.27 (*Rhodobacter* sp. SYSU G02092); fig|2593020.3.rna.27 (*Rhodobacter* sp. SYSU G02094); fig|2562317.3.rna.48 (*Rhodobacter* sp. SY28-1); fig|2528036.3.rna.32 (*Rhodobacter* sp. YIM 73028); fig|2249421.3.rna.20 (*Rhodobacter* sp. YIM 73036); fig|1063.35.rna.41 (*Rhodobacter sphaeroides* SCJ); fig|1063.34.rna.43 (*Rhodobacter sphaeroides* FY); fig|39723.4.rna.7 (*Rhodobacter sphaeroides* f. sp. denitrificans IL106); fig|272943.71.rna.1 (*Rhodobacter sphaeroides* 2.4.1); fig|1061.9.rna.57 (*Rhodobacter capsulatus* DSM 1710); fig|1054202.4.rna.47 (*Rhodobacter viridis* JA737); fig|1188250.3.rna.48 (*Rhodobacter blasticus* DSM 2131); fig|1185920.4.rna.50 (*Rhodobacter veldkampii* DSM 11550); fig|445629.8.rna.51 (*Rhodobacter johrii* JA192); fig|43057.6.rna.48 (*Rhodobacter azotoformans* KA25); fig|1063.31.rna.46 (*Rhodobacter blasticus* 28/5); fig|1188249.3.rna.49 (*Cereibacter changlensis* JA139); fig|446682.3.rna.35 (*Rhodobacter maris* JA276); fig|439529.3.rna.19 (*Rhodobacter ovatus* JA234); fig|407234.3.rna.49 (*Rhodobacter vinaykumarii* DSM 18714); fig|4535-82.3.rna.47 (*Rhodobacter aestuarii* DSM 19945); fig|1884310.3.rna.15 (*Rhodobacter* sp. 24-YEA-8); fig|272942.6.rna.3 (*Rhodobacter sediminis* SB 1003).

The xanthorhodopsin analysis used 31 protein sequences with the following accession numbers: LGHU01000114 (*Pseudorhodobacter antarcticus* KCTC 23700); 1081.4.con.0047 (*Rhodospirillum sodomense*); ATUR01000008 (*Sphingopyxis baekryungensis* DSM 16222); JASG010-00004 (*Sulfitobacter guttiformis* KCTC 32187); JEMV01000009 (*Sphingobium* sp. Ant17); JPPQ01000083 (*Sphingobium* sp. ba1); 204455.19.con.0021 (*Rhodobacter* sp. M37P); NZ_DS99-0628 (*Octadecabacter arcticus* 238); NC_015581 (*Thioalkalimicrobium cyclicum* ALM1); LAPV01-000235 (*Devosia psychrophila* Cr7-05); 85076.5.con.0174 (*Marichromatium* SE5_Pink); CP006644 (*Sphingomonas sanxanigenens* DSM 19645); MBFX01000002 (*Spiribacter salinus* M7-304); BMGD01000001 (*Blastomonas aquatica* CGMCC 1.12851); CP042288 (*Aquisalinus flavus* D11M-2); PDEQ01000001 (*Longibacter salinarum* KCTC 52045); RAHX01000001 (*Erythrobacter aquimixticola* JSSK-14); NTKP01000020 (*Rhodothermaeota* MED-G12); RZHG01000028 (*Halomonas*

sp. DSM 19434); SJDN01000025 (*Tabrizicola* sp. WMC-M-20); VHHQ01000006 (*Thermus thermophilus* IB-21); NC_007677 (*Salinibacter ruber* DSM 13855); BJXV01000002 (*Halovibrio variabilis* NBRC 102410); CP011494 (*Marinobacter psychrophilus* 20041); LN681225 (*Legionella hackeliae* ATCC35250); MTEJ01000043 (*Thiothrix lacustris* A8); VMKO01000001 (*Halopeptonella vilamensis* DSM 21056); DSHZ01000261 (*Thermus islandicus* SpSt-189); NCEQ01000006 (*Brevundimonas subvibrioides* 32-68-21); RWJI01000002 (*Sphingorhabdus wooponensis* 03SU3-P); AXCW01000007 (*Actinotalea ferrariae* CF5-4)

In both phylogenetic trees, all positions with less than 95% site coverage were eliminated, i.e., fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). Evolutionary analyses were conducted in MEGA X [44], and tree visualization was performed in iTOL [40].

2.6. Structure Modeling

The xanthorhodopsin protein sequence was obtained from the M37P genome and used for 3D modeling using SWISS-MODEL (https://swissmodel.expasy.org; accessed on 9 April 2022) [45]. The SWISS-MODEL template library (SMTL version 12 January 2022, PDB release 7 January 2022) was searched with BLAST [46] and HHblits [47] for evolutionary related structures. The target sequence was searched in the BLAST database against the primary amino acid sequence contained in the SMTL. A total of 21 templates were found. An initial HHblits profile was built using the procedure outlined in [47], followed by 1 iteration of HHblits against Uniclust30 [48]. The obtained profile was then searched against all profiles of the SMTL. A total of 191 templates were found. The best matching target was found to be 6nwd (GII198protein; rhodopsin derived from *Gloeobacter violaceus* PCC 7421) with 45.5% sequence identity, which was used for model construction. Models are built based on the target–template alignment using ProMod3 [49]. Coordinates that are conserved between the target and the template are copied from the template to the model. Insertions and deletions are remodeled using a fragment library. Side chains are then rebuilt. Finally, the geometry of the resulting model is regularized by using a force field. The global and per-residue model quality has been assessed using the QMEAN scoring function [50]; a QSQE score of 0.18 and coverage of 0.98 was obtained.

3. Results and Discussion

3.1. Genome Analysis

Sequencing the *Rhodobacter* sp. M37P genome generated 5,051,274 reads, yielding a total of 420.96 Mbp of sequencing data. The assembled genome consisted of 59 contigs, with the largest contig being 483,911 bp, with a N_{50} of 255,108 bp. The GC content was 67.5% and the length of the M37P genome was 3.63 Mbp. The genome sequence was annotated using the RAST server, which identified 3732 coding sequences (CDS) and 46 tRNAs. There were also 56 short repeat regions. A CheckM analysis showed the genome to have 100% completeness and no contamination. A comparison of genome properties with related *Rhodobacter* species is provided in Table 1. The GC content and genome size are similar to those of *Rhodobacter blasticus* (now *Fuscovulum blasticum*) DSM 2131 (66.5% and 3.58 Mbp), *Rba. capsulatus* DSM 1710 (66.5% and 3.67 Mbp), and slightly lower than *Rba.* (now *Luteovulum*) *sphaeroides* 2.4.1 (68.7% and 4.60 Mbp). The M37P genome sequence contained genes for nitrogen fixation (nitrogenase and nitrogenase reductase) and sulfur oxidation (*Sox*XYZ), which is consistent with *Rhodobacter* metabolism.

Table 1. Genomic and growth properties of *Rhodobacter* species. References where the growth data were obtained are included in the table. Other data were obtained from NCBI GenBank and genome analysis performed in PATRIC [32]. LH II = Light Harvesting complex II; Y = positive; N = negative; n.d. = not determined or not available in literature.

Strain	Genome				Growth						Rhodopsin	Rubisco	LH II	Reference
	size (Mb)	GC %	CDS	Genbank Acc.	malate	succinate	glutamate	glucose	temp. optimum	anaerobic growth				
Rba. sp. M37P	3.63	67.5	3732	JAANHS000000000	Y	Y	Y	Y	38–39	N	Xantho-rhodopsin	none	N	*this paper*
Rba. thermarum YIM73036	3.65	66.8	3695	QMJY00000000	N	Y	N	Y	37–45	N	none	none	N	[9]
Rba. flagellatus SYSU G03088	3.62	67.7	3677	VMDU00000000	N	n.d.	N	N	45	n.d.	none	*rbcS; rbcL*	N	[16]
Rba. blasticus DSM 2131	3.59	66.5	3625	PZKE00000000	Y	n.d.	Y	Y	30–35	Y	none	*rbcII; rbcS; rbcL*	Y	[8]
Rba. capsulatus DSM 1710	3.67	66.5	3601	QKZO00000000	Y	Y	Y	Y	30	Y	none	*rbcII; rbcS; rbcL*	Y	[1,15]
Rba. sediminis SB1003	3.87	66.6	3728	CP001312	Y	n.d.	Y	Y	30–35	Y	none	*rbcII; rbcS; rbcL*	Y	[20]
Rba. viridis JA737	3.87	65.3	3780	QJTK00000000	N	N	N	Y	30	Y	none	*rbcII*	Y	[26]
Rba. ovatus JA234	3.81	66.5	3775	OAOQ00000000	N	Y	N	N	25–30	Y	none	*rbcII; rbcS; rbcL*	Y	[19]
Rba. vinaykumarii JA123	3.49	68.2	3366	OBMN00000000	Y	Y	N	Y	20–30	Y	none	*rbcII; rbcS; rbcL*	Y	[25]
Rba. maris JA276	3.83	65.1	3806	OBMT00000000	Y	N	N	N	25–35	Y	none	*rbcII; rbcS; rbcL*	Y	[18]
Rba. aestuarii JA296	3.84	61.1	3742	QAXT00000000	N	N	N	N	25–30	Y	none	*rbcII; rbcS; rbcL*	Y	[10]
Rba. sphaeroides ATH 2.4.1	4.63	68.7	4657	CP030271	Y	Y	Y	Y	30–34	Y	none	*rbcII; rbcS; rbcL*	Y	[22]
Rba. sphaeroides IL106	4.79	69.1	4848	QRBG00000000	Y	Y	Y	Y	30	Y	none	*rbcII; rbcS; rbcL*	Y	[51]
Rba. azotoformans KA25	4.41	68.4	4505	QAOT00000000	Y	Y	Y	Y	30–35	Y	none	*rbcII; rbcS; rbcL*	Y	[14]
Rba. johrii JA192	4.59	69.3	4438	PZZW00000000	Y	Y	Y	Y	25–35	Y	none	*rbcII; rbcS; rbcL*	Y	[17]

A BLAST (NCBI) comparison using the complete 16S rRNA from our new species indicated *Rba. thermarum* YIM73036 and *Rba. blasticus* DSM 2131 as closest relatives: *Rba. thermarum* YIM73036 97.74% (1429/1462 bp) and *Rba. blasticus* 28/5 97.08% (1430/1473 bp). Using the complete 16S rRNA, the M37P isolate is only 96% (1368/1420 bp) identical to "*Rhodobacter gluconicum*", to which it was previously compared. These values are below the proposed species delineation for 16S rRNA comparisons of 98.7% [52]. A phylogenetic tree based on the 16S rRNA sequences of related *Rhodobacter* species is provided in Figure 1. This shows *Rba. thermarum* YIM73036 and *Rba. flagellatum* SYSU G03088 as the closest relatives, but clearly separated from the new strain M37P on the tree. All three appear to be part of a separate clade on the tree, which also includes some newer strains SY28-1, SYSU G02214, SYSU G02094, and SYSU G02092. These latter strains have not been formally described, but their genomes are available in the GenBank database. Interestingly, all of the SYSU strains were isolated from hot spring sediments in Tibet. The full-length M37P 16S rDNA sequence, derived from the genome, was submitted to GenBank with accession number ON328317.
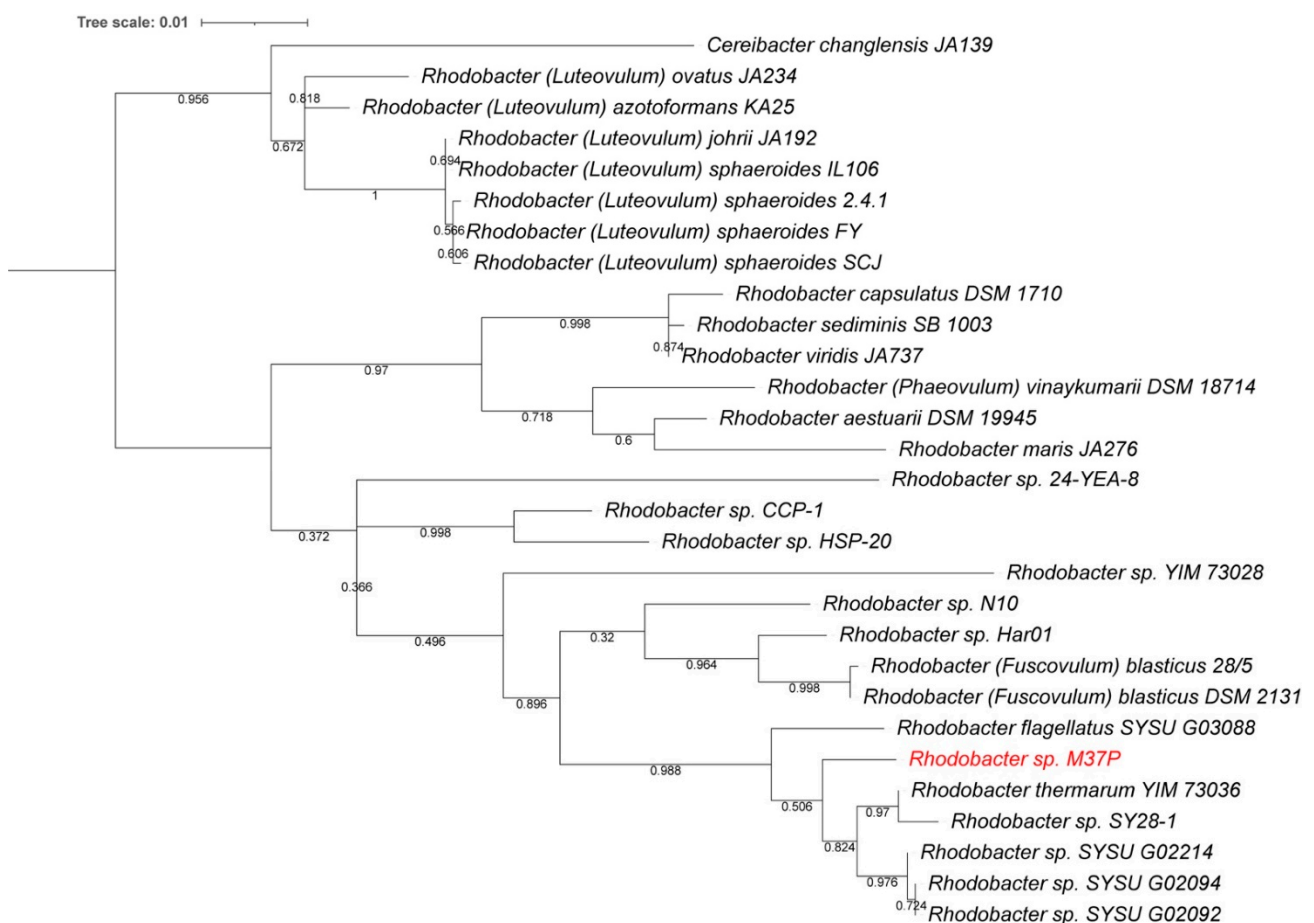
Figure 1. 16S rRNA-derived phylogenetic tree for *Rhodobacteraceae* species. The phylogenetic tree was calculated in MegaX and iTOL was used to draw the phylogenetic trees expressed in the Newick phylogenetic tree format. The length of the branches is proportional to the number of nucleotide substitutions per site, and bootstrap values were generated from 500 bootstrapping rounds. The new isolate, *Rhodobacter* strain M37P, 16 rRNA is indicated in red.

A whole genome-based JSpecies comparison [35] of the average percentage nucleotide identity (ANI) between strain M37P and published *Rhodobacter* genome sequences deposited in GenBank, showed the highest identities with *Rba. thermarum* YIM73036, (87.5%), *Rba. flagellatus* SYSU G03088 (85.2%), *Rba.* strain SY28-1 (82.6%), *Rba.* strain YIM73028 (80.1%), and *Rba. blasticus* DSM 2131 (76.1%). At well below the 95% cutoff for same species,

strain M37P is clearly a distinct species in its own right. All other *Rhodobacter* species showed identities at or lower than 74%.

A reciprocal average amino acid identity (AAI) comparison showed an 84.1% identity (SD 13.7%; from 3441 proteins) with *Rba. thermarum* YIM73036, 83.2% (SD 13.2%; from 4030 proteins) with *Rba.* sp. 28-1, 82.7% (SD 13.4%; from 3419 proteins) with *Rba. flagellates*, and 72.6% (SD 14.9%; from 2498 proteins) with *Rba. blasticus*. These are all below the arbitrary species delineation of 85–90% AAI, and further defines the species individuality of the M37P isolate.

A whole genome-based phylogenetic analysis of the sequenced *Rhodobacter* genomes constructed using RaxML [38,39], is consistent with the ANI and 16S rRNA analysis presented above. As can be seen in Figure 2, the *Rhodobacter* sp. M37P genome is clearly different from any of the other *Rhodobacter* species, and part of a clade of five species (orange in Figure 2) that includes the recently discovered *Rba. thermarum* strain YIM 73036, strain SY28-1, *Rba. flagellatus* SYSU G03088, and strain YIM73028. Except for *Rba. thermarum* and *Rba. flagellatus*, none of these new strains have been fully characterized; however, at least three of the four appear to be isolated from hot spring sediments, similar to M37P. Most closely related is the recently discovered *Rba. thermarum* YIM73036, which can grow at a maximum temperature of 55 °C, although 37–45 °C is optimum [17]. Most *Rhodobacters* grow best at 30–37 °C (Table 1).
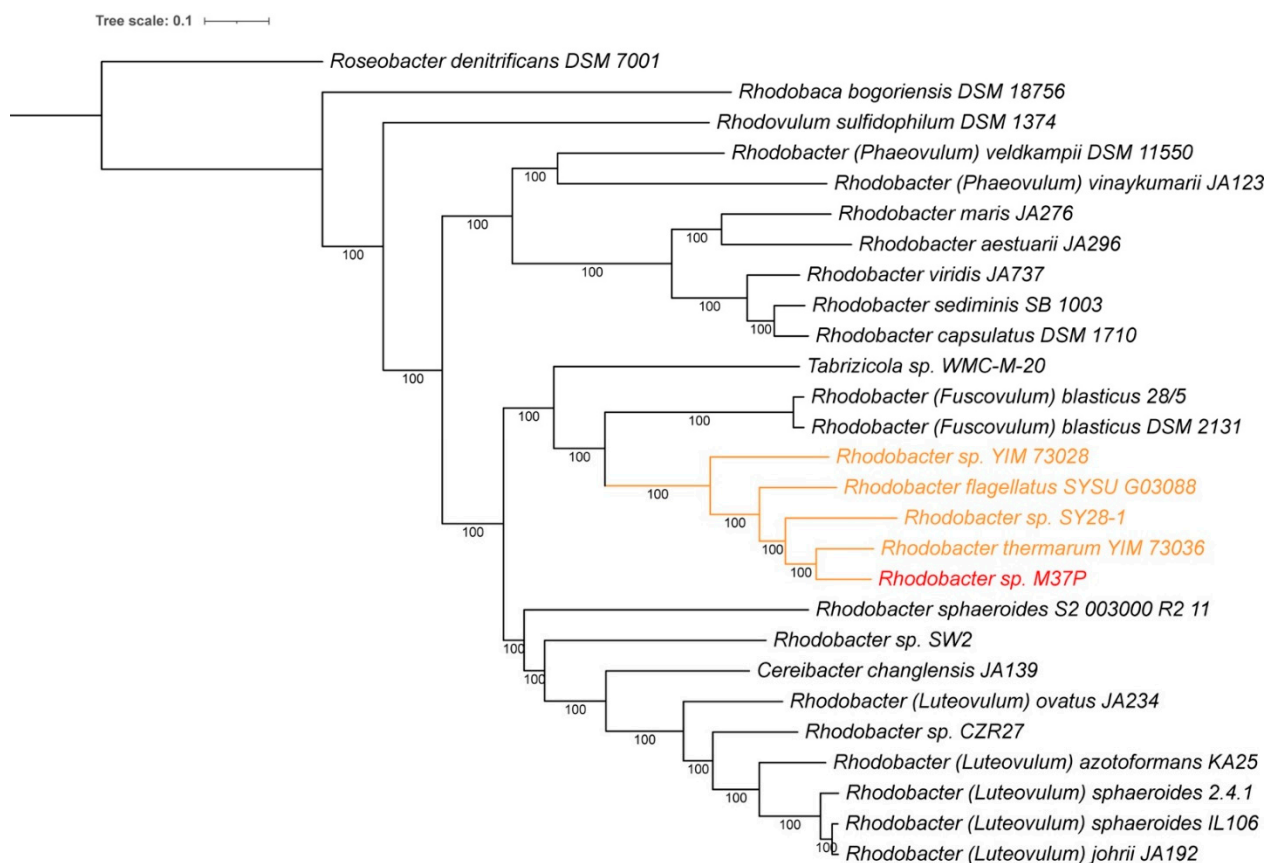
Figure 2. Whole genome-based phylogenetic tree of all sequenced *Rhodobacter* species. *Roseobacter denitrificans* was used as an outgroup and *Rhodovulum sulphidophilum* and *Rhodobaca bogoriensis* were included for comparison. One hundred rounds of the 'rapid bootstrapping' option of RaxML were used to generate the support values for the phylogenetic tree. The branch length tree scale is defined as the mean number of substitutions per site, which is an average across both nucleotide and amino acid changes. Genome accession numbers can be found in Table 1. The *Rhodobacter calidifons* M37P genome is indicated in red and species with ANI > 80% to M37P are colored orange.

3.2. Species and Growth Characteristics

Microscopic examination of the purified strain (Figure 3) showed that cells are rod-shaped, 1 to 2 µm long and 0.5 to 0.7 µm wide. Multiplication appears to occur by binary fission and cells appear to be motile. Interestingly, cells appear to exhibit a directional twitching motility towards the light in glass petri dishes (data not shown). Strain M37P was found to be catalase and oxidase positive. The growth temperature range tested in shaken cultures was 30–45 °C, with an optimum temperature of 38–39 °C and doubling time of 8–9 h. However, the doubling time of M37P on minimal medium is very long (14–20 h under our tested conditions). The temperature optimum of *Rba.* M37P is slightly higher than most *Rhodobacter* species, but is closer to the temperature optimum of the closely related *Rba. thermarum* YIM73036 (Table 1). The optimal pH range was found to be 7.6 to 8.0.
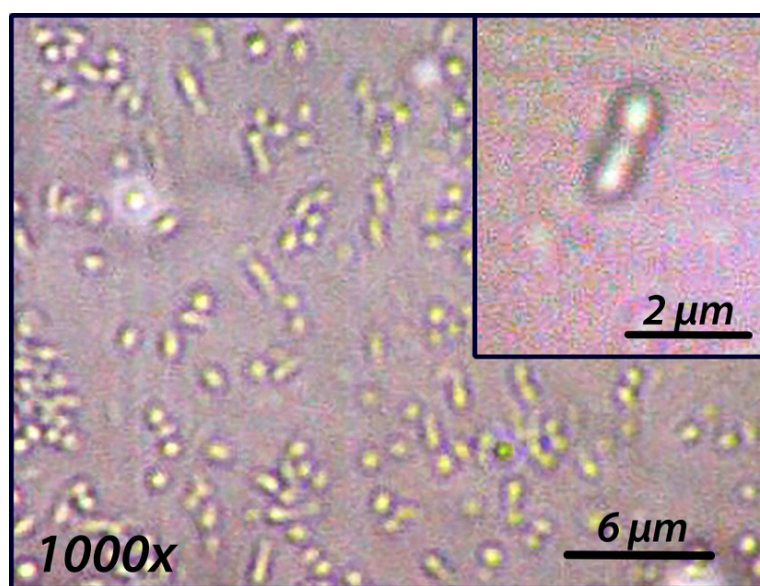
Figure 3. Phase contrast microscopy of *Rhodobacter* strain M37P at 1000× total magnification. Insert is 2.5× digitally magnified and contrast autocorrected.

Like other *Rhodobacter* isolates, we found that M37P will use malate, succinate, glutamate, and glucose for growth supplementation in Thermus medium and in minimal medium (Minimal 22 and RCVB). Many photosynthetic *Rhodobacter* species are able to grow both aerobically and anaerobically. In our growth studies, using both rich nutrient broth or minimal RCVB medium, we were unable to grow *Rhodobacter* sp. M37P under anaerobic conditions in either liquid medium or on plates. This is not typical for *Rhodobacter* species; however, aerobic anoxygenic phototrophs (AAPs) on the other hand are unable to use bacteriochlorophyll for anaerobic growth, but they do perform cyclic photophosphorylation only in the presence of oxygen [53]. AAP lack the RuBisCo enzyme for CO_2 fixation and are therefore not autotrophic [54]. Indeed, when checking the genome of M37P we found that it does contain *puf* genes for the photosynthetic reaction center subunits L, M, and H, but does not contain the RuBisCo gene for fixation of CO_2. RuBisCo is an essential protein in the Calvin–Benson–Bassham cycle that catalyzes the addition of carbon dioxide to ribulose-1,5-bisphosphate. The presence of multiple RuBisCo genes is not uncommon in Proteobacteria [55–58]. In the same species, a single RuBisCo gene can be present (Form II), or alternatively, a two-subunit form of RuBisCo (Form I) consisting of large (*rbc*L) and small (*rbc*S) subunits, or both forms can be found as well [57]. Several of the *Rba. sphaeroides* and *Rba. capsulatus* contain both versions of the rubisco genes (Table 1). However, neither the small or large RuBisCo subunit, nor the single Form II RuBisCo was found annotated in the M37P genome. We also performed a BLAST search within PATRIC using all three of the *Rba. sphaeroides 2.4.1.* RuBisCo genes to see if it may have not been annotated correctly

in the automated annotation, but again no rubisco genes were found in the strain M37P genome. Only a single gene encoding a protein 'similar to a RuBisCo type III' was found annotated. An NCBI-BLAST search of this gene identified it as a RuBisCo-like protein, not involved in CO_2 fixation but playing a role in an uncharacterized sulfur salvage pathway and has been shown to catalyze a novel isomerization reaction of 5-methylthio-D-ribulose 1-phosphate in *Rhodospirillum rubrum* [59].

The lack of anaerobic growth, the presence of *puf* genes, and absence of *rbc* genes indicate that strain M37P is an AAP, which is the first AAP identified in the genus *Rhodobacter*. *Roseobacter denitrificans* Och114 is the closest known AAP relative to *Rhodobacter*, based on 16S rDNA taxonomic analysis [53]. We included *Rsb. denitrificans* Och114 in our whole genome-based phylogenetic tree comparison (Figure 2). This shows the closer relationship of M37P to other *Rhodobacter* species than to *Roseobacter denitrificans*, consistent with the earlier identification of M37P as a *Rhodobacter* species. Several of the other species on the M37P clade are also lacking all of the RuBisCo genes (Figure 2 and Table 1). Unfortunately, not all of them have been characterized; however, only weak anaerobic photo-heterotrophic growth was reported for *Rba. thermarum*, while no photolithoautotrophic growth was observed [9], which is consistent with the lack of RuBisCo. Further comparative studies will be necessary to determine whether other members of this clade are also AAPs.

The pigmentation of M37P is a pinkish, red-centered colony with a lighter edge and the red center becomes more intense upon continuous aerobic growth. Interestingly, the pinkish-red pigment that is present in the culture under aerobic dark conditions, disappears under light conditions, and returns when the cultures are subsequently incubated in the dark. This is different from the aerobic repression of photosynthetic pigment synthesis observed in other *Rhodobacter* species [60,61]. On the other hand, in aerobic anoxygenic phototrophic bacteria, the influence of light intensities was found to inhibit bacteriochlorophyll a (Bchl *a*) and carotenoid synthesis [53,62,63]. A two- to three-fold increase in the concentration of Bchl *a* has been observed in dark-incubated periods, for example, in the AAPs *Erythrobacter hydrolyticum*, *Erythromicrobium sibiricum*, and *Roseobacter denitrificans* [62–64]. The light-dependent regulation of carotenoids appears to be different in various AAPs, where carotenoids are greatly inhibited by blue light in *Roseobacter denitrificans*, but not affected by light of any wavelength in *E. ramosum* [53,63]. Dark-incubated whole cell absorption spectra of *Rba.* strain M37P showed maxima characteristic of BChl *a* (590, 795 and 862 nm). After performing methanol-based pigment extractions of aerobic dark and light cultures of M37P, we observed a small peak at 750 nm, possibly due to bacteriopheophytin (Figure 4, black spectrum). However, in the dark-incubated cultures, there was a high abundance of pigments with peaks at 464 nm, 490 nm, and 524 nm (red spectrum in Figure 4), which were not seen in cultures that were incubated in the light for several hours (black spectrum; Figure 4). A similar light-induced washout of carotenoid pigments was observed in *Rsb. denitrificans* [63]. This indicates that there is a similar regulation of carotenoid synthesis in M37P as in *Rsb. denitrificans*. Compared to anaerobic purple non-sulfur phototrophs, AAP cells typically contain only small amounts of BChl relative to the abundance of carotetoids (about 1:10 ratio [53,62]). This also appears to be the case in M37P, based on the spectra of the extracted pigments (Figure 4), further indicative of the fact that strain M37P is an AAP.

Comparison of the photosynthetic systems of strain M37P and the other *Rhodobacter* genomes also revealed that this strain is lacking the genes for the light harvesting complex II (LHII), which is typically an important photosystem component in *Rhodobacter* (and other photosynthetic bacteria), and augments the collection of solar energy [65,66]. The LHII antenna complexes are peripheral complexes that transfer energy to the LHI complexes, which in turn transfer energy to the reaction centers (RCs) embedded therein [67]. The LHII alpha and beta subunit genes (*pucA* and *pucB*) are typically found in the *pucBACDE* gene cluster. PucC proteins have been implicated in LH1 and LH2 assembly and are part of the bacteriochlorophyll delivery (BCD) family of the major facilitator superfamily [68,69].
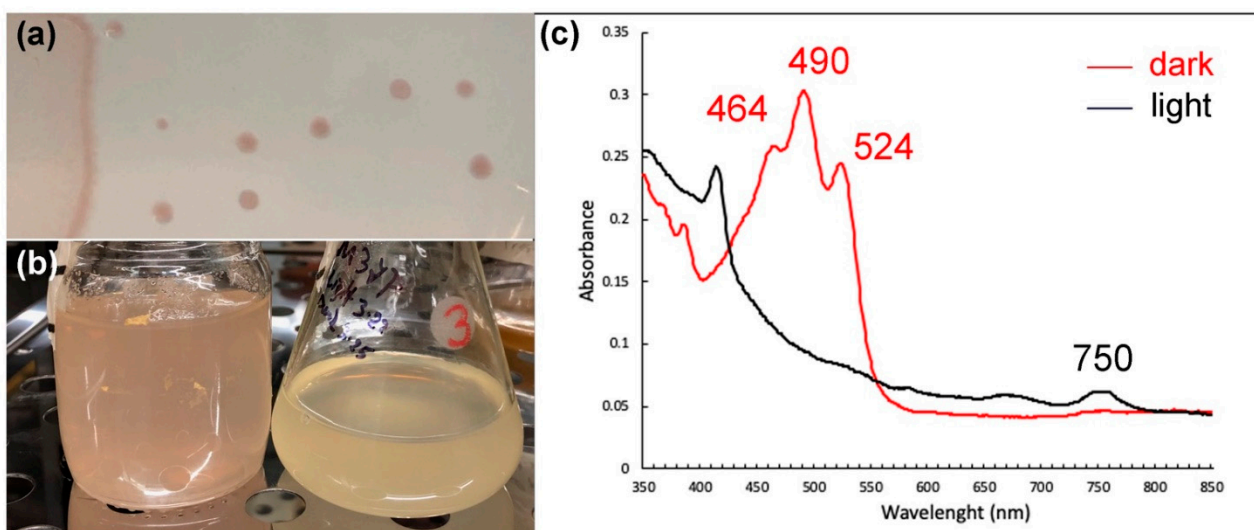
Figure 4. (**a**) Colonies of *Rhodobacter* sp. M37P grown on dark-incubated Thermus medium plates. (**b**) Dark- (left) and light- (right) incubated aerobic liquid cultures of strain M37P. (**c**) Absorption spectra of extracted pigments from dark-incubated (red) and light-incubated (black) aerobic *Rhodobacter* sp. M37P cultures. Spectral peak maxima are indicated.

All of the *Rhodobacter* genomes have a *pucC* gene homolog, including strain M37P, that is clustered with the reaction center H subunit gene in their genome; however, the second *pucC* homolog, and in fact the entire *pucBACDE* gene cluster that encodes the LHII complex genes, is lacking from strain M37P, and from strains YIM73028, SYSU G03088, SY28-1, and YIM73036 (Table 1), all of which form the unique clade with strain M37P (Figures 1 and 2). None of these genes were found annotated in these genomes and a BLAST search in PATRIC using the *puc* genes from the *Rba. sphaeroides*, *Rba. capsulatus*, and *Ceribacter changlensis* LHII gene cluster did not find any *pucBACDE* gene homologs for LHII complexes in any of the members of the M37P clade.

A difference in diversity of LH systems in AAPs as compared to analogous complexes from anaerobic phototrophic bacteria has been observed earlier [53,70]; however, the closest AAP relative of M37P, *Rsb. denitrificans*, does produce a functional peripheral LHII complex. On the other hand, some other AAPs, *Acidiphilium rubrum* and *Porphyrobacter* also appear to lack LHII complexes [71,72]. The lack of LHII or the reduced size of the complex is consistent with a more auxiliary role of phototrophy in AAPs [70].

Although the evolution of phototrophy in Proteobacteria is a complex issue, it has been proposed that AAPs have evolved from anaerobic anoxygenic phototrophs [53,73]. This was mainly hypothesized based on earlier DNA–DNA hybridizations of AAPs and the fact that anoxygenic phototrophy evolved well before the gradual oxygenation of the Earth [74,75]. It is therefore intriguing to speculate that *Rba. calidifons* M37P has evolutionary lost the ability for anaerobic photosynthesis and autotrophic growth, given its close phylogenetic relationship to the anaerobic phototrophic *Rhodobacter* species. However, at this point, it is not certain if the evolution to an AAP would have occurred by the loss of carbon fixation genes only, or what specifically causes the loss of ability for anaerobic photosynthesis. The lack of a LH II complex certainly indicates that there is a more complex evolutionary story involved. A more detailed phylogenetic and biochemical analysis of the phototrophy of *Rhodobacter* M37P and other related AAPs will be necessary to determine whether this occurred by evolutionary gene loss from anaerobic photosynthetic *Rhodobacter* species, or if there is an earlier divergence in evolution that formed the M37P containing clade.

3.3. Xanthorhodopsin

Upon further comparison of the *Rba.* M37P genome to 75 other *Rhodobacter* genomes (Supplementary Table S1), we found that it contains a unique gene for xanthorhodopsin (XR) (PGFam_08288266; locus tag: G8O29_16735). Xanthorhodopsins are light-harvesting proton pumps with a dual chromophore [76–78]. They have one retinal molecule, similar to archaeal bacteriorhodopsin (BR) and the more widespread proteorhodopsin (PR); in addition, XR contains an additional carotenoid antenna chromophore that allows the cells to utilize a wider range of light for energy conversion. Other than strain M37P, none of the analyzed *Rhodobacter* genomes contain any form of rhodopsin gene, nor have any rhodopsin-containing *Rhodobacter* species been described to our knowledge. To identify the closest relatives of the *Rhodobacter* XR (RXR), we searched all PATRIC genomes (derived from NCBI GenBank) and performed a phylogenetic comparison of the translated protein to other annotated xanthorhodopsins.

Figure 5 provides a phylogenetic tree of the closest homologs. M37P xanthorhodopsin (labeled in red) forms a separate clade with closely related XRs from *Aquisalinus flavus* D11M-2 and *Tabrizicola* sp. WMC-M-20. The homology with these proteins was 82% (212/257 aa.) for *Aqs. flavus* and 86% (222/258 aa.) for *Tabrizicola sediminis* WMC-M-20. *Aquisalinus flavus* D11M-2 belongs to the *Parvularculaceae* and was isolated from the saline Lake Dasugan in China and is strictly heterotrophic and aerobic [79]. *Tabrizicola sediminis* WMC-M-20 is an AAP isolated from a saline lake in Tibet [80] and belongs to the larger family of *Rhodobacteriaceae*.
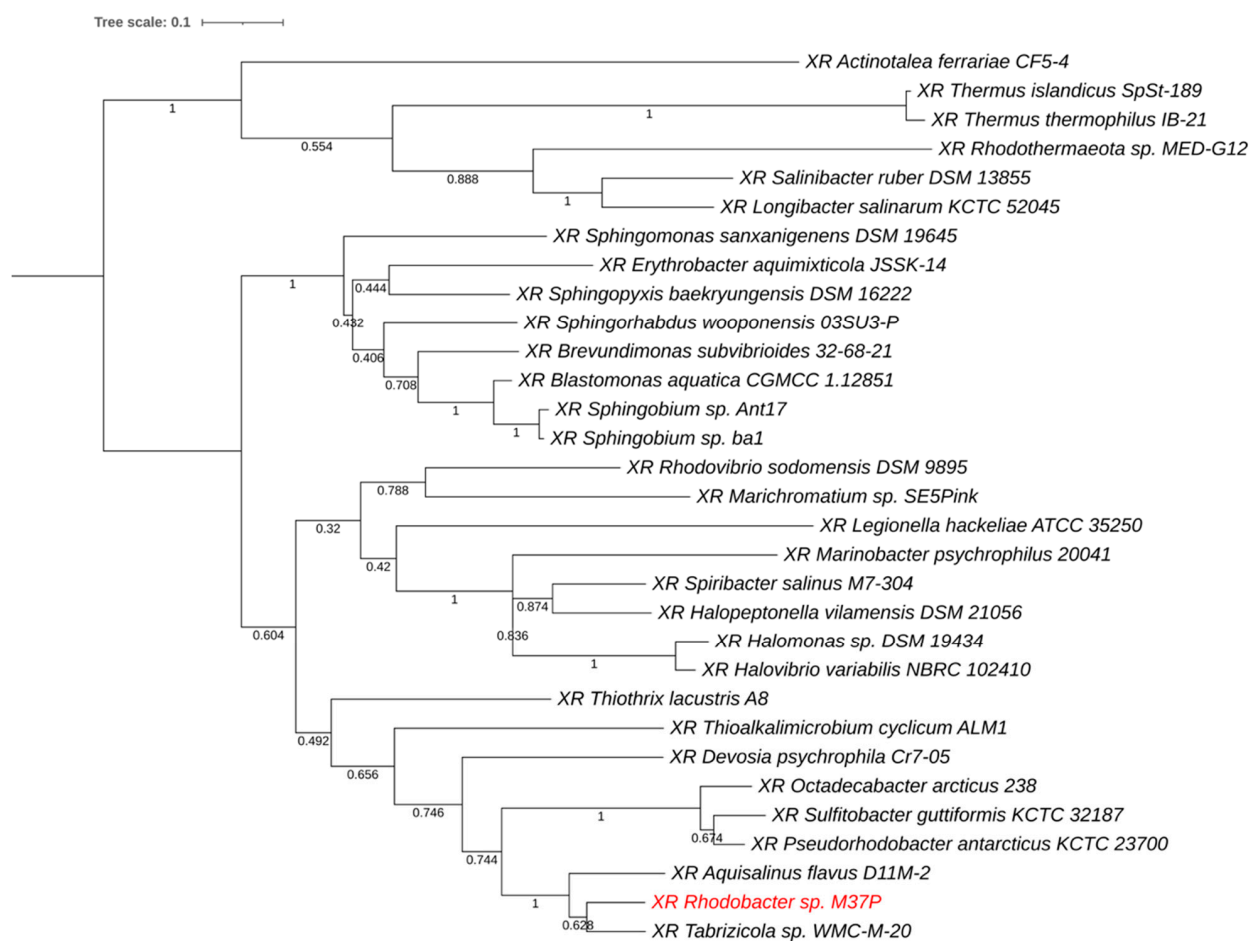
Figure 5. Phylogenetic comparison of xanthorhodopsin from *Rba.* sp. M37P to its closest xanthorhodopsin homologs derived from genome sequences. The M37P XR protein (in red) forms a separate clade with closely related XRs from *Aquisalinus flavus* D11M-2 and *Tabrizicola* sp. WMC-M-20.

The xanthorhodopsin gene in strain M37P is followed by several carotenoid biosynthesis genes (polyprenyl synthetase; phytoene desaturase (neurosporene-forming) (EC 1.3.99.28); phytoene synthase (EC 2.5.1.32); lycopene beta-cyclase (EC 5.5.1.19); and 15,15′-beta-carotene monooxygenase). The rhodopsin retinal chromophore is an oxidative cleavage product of beta-carotene, which is catalyzed by a beta-carotene 15,15′-monooxygenase [81,82]. The presence of the retinal biosynthetic enzyme, beta-carotene 15,15′-monooxygenase, indicates that strain M37P is capable of producing a functional rhodopsin. Three of the other carotenoid biosynthetic enzymes located in between—polyprenyl synthetase, phytoene desaturase, and phytoene synthase—are known to catalyze the formation of neurosporene and lycopene carotenoids all the way from the terpenoid backbone synthesis (Supplementary Figure S1) [83]. Lycopene beta-cyclase is involved in several downstream carotenoid modification pathways, including 7,8-dihydroxy-beta-carotene, beta-carotene, lutein and neurosporaxanthin synthesis [84] (Supplementary Figure S1), and is likely important for the synthesis of carotene or the additional carotenoid that is associated with functional xanthorhodopsins. This indicates that all the genes needed for the biosynthesis of beta-carotene and a functional xanthorhodopsin protein are present in the *Rhodobacter* M37P genome. This synteny is also conserved in the *Tabrizicola* sp. WMC-M-20 and several of the other xanthorhodopsin-containing genomes, which supports this hypothesis.

Proton-pumping rhodopsins consist of seven transmembrane helices that pump protons along the cell membrane and thereby create a proton gradient that is used for ATP production, thereby providing a simple method for light to biological energy conversion.

Bacteriorhodopsin is only found in Archaea, whereas Proteorhodopsin is now found to be widespread amongst marine members of the Bacteria and Archaea [85,86]. On the other hand, only a limited amount of xanthorhodopsins have been described. XR has some unique structural differences with other rhodopsins, such as BR and PR [78,87,88]. XR-type rhodopsins have an additional binding pocket for the carotenoid antenna molecule, and share the structural features of having elongated helices, a flipped B–C loop and a 3-omega motif [78]. To confirm that the *Rba.* sp. M37P gene encodes a true xanthorhodopsin, we created a 3D structural model of RXR using SWISS-MODEL and analyzed it for the potential presence of these distinguishing features.

The RXR structure has seven transmembrane helices, similar to other microbial rhodopsins, and has Lys228 positioned correctly for the Schiff base binding of the retinal chromophore (Figure 6A). A superposition of RXR with other rhodopsin structures shows that it is more similar to XR than BR or PR, as expected from the sequence similarity. The following RMSD (root-mean-square deviation) values were obtained in PyMOL superpositions: 0.090 with XR (1418 atoms) from *Salinibacter ruber* (PDB entry 3DDL); 1.714 with GR (1541 atoms) from (PDB entry 6NWD); 2.010 with BR (954 atoms) from *Halobacterium salinarum* (PDB entry 1C3W); and 3.002 with PR (1187 atoms) from an uncultured gammabacteria (PDB entry 4KLY). The superposition of RXR with *S. ruber* XR (Figure 6) shows that it also has the typical elongated A, B, and G helices and is capable of forming the characteristic β-sheet structure in the BC loop, which is not seen in PR or BR structures. The binding pocket for the carotenoid antenna molecule is also present in the RXR structure, indicating that it is capable of binding a second chromophore (both retinal and salinixanthin were modeled into the structure in Figure 6).

The eubacterial rhodopsin structures also share a 3-omega motif, formed by three non-consecutive aromatic amino acid residues from the extended helices and the B-C loop structure, which are linked through aromatic pi-stacking interactions. The RXR structure also positions Tyr9 (helix A; equivalent to Tyr13 in XR) and Trp66 (helix B; equivalent to Trp70 in XR) in such interactions; however, the third position is substituted by a non-aromatic Ile77 residue (instead of Tyr81 in XR) (Figure 6B). The 3-omega motif is conserved in all different eubacterial rhodopsin pumps, but completely lacking from archaeal BR-pumps, and is thought to be involved in structural interprotomer interactions, rather than being correlated to a specific function [78]. The potential lack of the third pi-stacking aromatic interaction in the RXR B–C-loop, and its possible structural implications will need

further study. Nevertheless, the sequence homology and structural comparison indicate that the rhodopsin gene found in strain M37P indeed encodes a xanthorhodopsin.
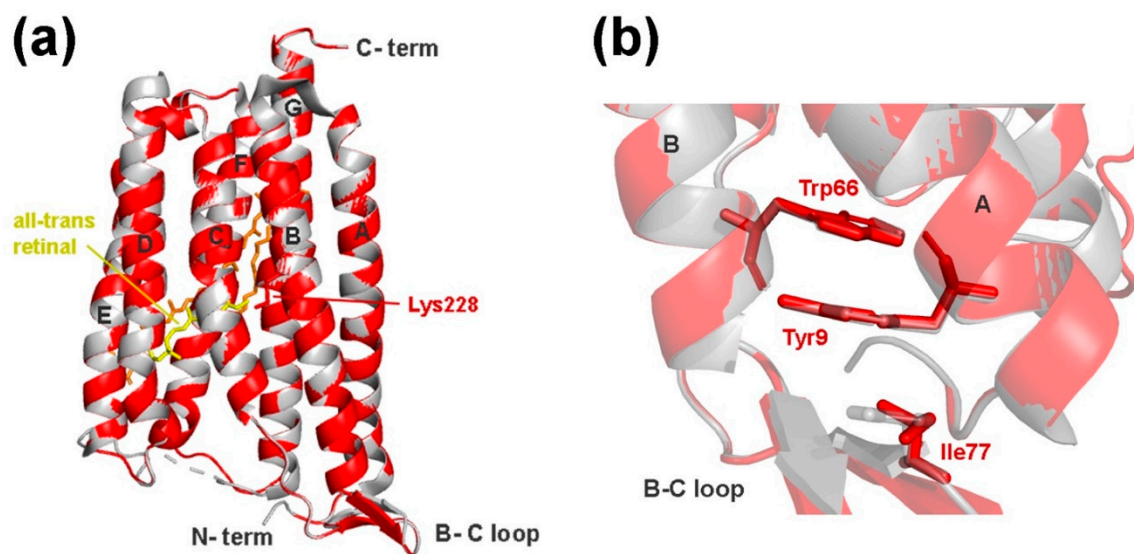
Figure 6. (**a**) Superimposed structures of *Rba.* sp. M37P xanthorhodopsin (RXR; red) and the *Salinibacter ruber* xanthorhodopsin (light gray, PDB entry 3DDL) The retinal and salinixanthin chromophores from XR are labeled in yellow and orange, respectively. (**b**) Detailed structural image of the 3-omega motif formed by pi-stacking of aromatic residues in helix A, B, and the B–C loop. RXR residues are labeled in red and *Salinibacter* XR residues shown in gray.

Interestingly, the absorption spectra of the extracted pigments (Figure 4) correspond well with the major absorption spectra described for the xanthorhodopsin carotenoid component [76–78,89], where absorption spectra at 521 nm, 487 nm and 456 nm are due to structurally bound salinixanthin. Further analysis of the purified holo-protein and HPLC analysis of the M37P carotenoids will be necessary to determine its characteristics; however, based on our current analysis of this unique strain, the pinkish-red pigmentation observed in dark-adapted cultures is due to the production of carotenoids, possibly linked to a functional xanthorhodopsin.

3.4. Xanthorhodopsin in Mushroom Spring Microbial Mats

The potential importance of rhodopsins in the energy metabolism of the Mushroom Spring microbial ecosystem has been mentioned earlier during a metagenomic analysis of the microbial mat communities [90]. No *Rhodobacteriaceae* were identified in that study; however, the assembled undermat metagenome showed sixteen annotated rhodopsin genes, affiliated with seven phylogenetic groups (although six rhodopsins were linked to *Armatiomonadetes* members), representing an unexpected diversity of potential retinal-based phototrophy in the mat [90]. These rhodopsin genes were found as part of partial genomes and it was uncertain whether all of these were producing functional rhodopsins. Nevertheless, the unexpected occurrence of rhodopsin- (mainly xanthorhodopsin) type genes in the Mushroom Spring runoff community indicates an important role for rhodopsin-mediated light harvesting in those microbial mats. The samples for the metagenomics study were taken at a higher temperature point (60 °C) than the M37P sampling, which could explain the lack of *Rhodobacter* species in their analysis.

Although experimental evidence for the exact role of eubacterial rhodopsins is still somewhat scarce, studies with PR from different bacterial organisms have shown an increased survival rate or growth-enhancing effect, particularly in nutrient-poor environments, due to light-induced proton pumping and coupled ATP production via PR [91–94]. Rhodopsin-containing *Vibrio* species appear to have increased long-term survival under starved conditions when exposed to light, while *Flavobacteria* PR-containing species show

a six-fold increase in growth yield under nutrient-poor conditions with light exposure of the appropriate wavelength. Similarly, the presence of XR in *Rba.* M37P is expected to provide either long-term starvation survival or short-term adaptation to low nutrient levels in the Mushroom Spring runoff. Oxygenic cyanobacteria are the predominant primary producers in the Mushroom Spring [90,95]. They are predominantly found in the upper 0–2 mm of the microbial mat where they fix CO_2 and synthesize and excrete metabolites that are then consumed by (photo)heterotrophic members of the community. This will create a light-dependent diurnal cycling of periods of high and low organic nutrient levels, and rhodopsin-containing species would be expected to have an adaptive advantage if they can extend their growth and energy production under the phases of nutrient-poor conditions. Blue–green light (450–550 nm) is not absorbed as well by the Chl *a*, phycocyanin, and carotenoids of cyanobacteria, and the presence of an antenna chromophore in xanthorhodopsin is certainly expected to be advantageous in niches that receive low irradiance. Since *Rhodobacter* M37P is the only *Rhodobacter* species that contains a xanthorhodopsin gene, it is possible that M37P has selectively obtained this xanthorhodopsin gene cluster through horizontal gene transfer from one of the Mushroom Spring species. In doing so, it thereby obtained a selective advantage in the complex microbial community of the Mushroom Spring runoff area.

3.5. Description of Rhodobacter calidifons sp. nov.

Based on the genomic and genetic differences and the unique growth characteristics described above, it is clear that strain M37P is a distinct new species belonging to the genus *Rhodobacter*. We therefore propose the name *Rhodobacter calidifons* sp. nov. (cali'di.fons; L. masc. adj. *calidi* hot; L. masc. noun *fons* spring; *calidifons* hot spring; named after being isolated from a hot spring runoff).

Cells are rod-shaped, 1 to 2 μm long, and 0.5 to 0.7 μm wide. Multiplication is by binary fission. Cells appear to be motile. Catalase and oxidase positive. Aerobically grown cultures are white when grown in the light, and pinkish-red pigmented when grown in dark conditions. Colonies are smooth and glossy and dark-pink in dark-incubated cultures. No anaerobic growth in either dark or light conditions. Aerobic anoxygenic phototroph. Optimal pH range, 7.6 to 8.0. Growth temperature range was 30–45 °C, with an optimum temperature of 38–39 °C. Growth occurs on malate, succinate, glutamate, and glucose. Growth factors required are thiamin and nicotinic acid. Yeast extract and biotin stimulate the growth appreciably. Nitrogen sources: ammonium salts and amino acids. Capable of nitrogen fixation and sulfur oxidation. Photopigments: absorption spectra of whole cells show the maxima characteristic of bacteriochlorophyll a (378, 590, 795, and 862 nm). Methanol-extracted pigment from dark-incubated cultures shows absorption maxima at 464, 490, and 524 nm. Contains xanthorhodopsin. DNA base composition: 67.5 mol% G+C (genome analysis) and a genome size of 3.63 Mb.

GenBank accession number of the genome is JAANHS000000000, and GenBank accession number for the full-length 16S rRNA is ON328317. Strain M37P was isolated from runoff from the Mushroom Hot Spring at Yellowstone National Park.

4. Conclusions

The lack of anaerobic growth and the effects of light on pigment synthesis and degradation, indicate that *Rhodobacter calidifons* M37P is an aerobic anoxygenic phototroph (AAP). The increased pigment synthesis during dark incubation and decreased pigments observed after several hours of light incubation are consistent with the observed effects in other AAPs and is expected to provide supplemental energy production when light is present in the hot spring runoff. Further physiological studies will be required to clarify the molecular nature of the pigment regulation; however, to our knowledge, this is the only *Rhodobacter* species that has been described as an AAP. Although further characterization of the carotenoids and protein purification will be necessary, based on spectral homology and the presence of a unique xanthorhodopsin-encoding gene, we suggest that the increased

pigmentation in *Rba. calidifons* M37P is due to the production of xanthorhodopsin. None of the other *Rhodobacter* genomes contain any form of rhodopsin in their genome, nor has any rhodopsin-producing *Rhodobacter* species ever been described. The *Rhodobacter* xanthorodopsin was possibly obtained through horizontal gene transfer from the microbial mat community of the Mushroom Spring runoff, where the presence of a xanthorhodopsin with an antenna chromophore provides a selective advantage under low irradiation.

AAP bacteria metabolize more efficiently when light is available, suggesting that in oligotrophic waters they can use energy from both light and scarce nutrients simultaneously. In these organisms, light inhibits bacteriochlorophyll a synthesis, so its activity diminishes during the day. On the other hand, rhodopsin-containing cells exploit the light for phototrophy and, due to the light-induced proton pumping, especially in nutrient-poor environments, the cell viability of rhodopsin-containing bacteria increases. The characteristics and adaptation of this unique *Rhodobacter* species indicate that light energy is of major importance in the Mushroom Spring community, even under low light conditions.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/microorganisms10061169/s1, Table S1: *Rhodobacter* genomes used for the comparative study. Figure S1: Beta-carotene biosynthesis pathway overview. Enzymes found in the *Rba. calidifons* M37P xanthorhodopsin gene cluster are marked and labeled in pink.

Author Contributions: Conceptualization, J.A.K., T.E.M. and R.F.R.; methodology, J.A.K., T.E.M. and R.F.R.; software, J.A.K., S.R. and I.B.S.; validation, J.A.K., T.E.M. and R.F.R.; formal analysis, J.A.K., S.R., I.B.S. and R.F.R.; investigation, J.A.K., S.R., I.B.S. and R.F.R.; resources, J.A.K. and R.F.R.; data curation, J.A.K., S.R., I.B.S., R.F.R. and T.E.M.; writing—original draft preparation, J.A.K., T.E.M. and R.F.R.; writing—review and editing, J.A.K., S.R., I.B.S. and R.F.R.; visualization, J.A.K., S.R. and I.B.S.; project administration, J.A.K. and R.F.R.; funding acquisition, J.A.K. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JAANHS000000000. The version described in this paper is version JAANHS010000000.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Imhoff, J.F.; Truper, H.G.; Pfennig, N. Rearrangement of the species and genera of the phototrophic "Purple Nonsulfur Bacteria". *Int. J. Syst. Bacteriol.* **1984**, *34*, 340–343. [CrossRef]
2. Imhoff, J.F. Genus *Rhodobacter.* Imhoff, Trüper and Pfennig 1984, 342^{VP}. In *Bergey's Manual of Systematic Bacteriology*, 2nd ed.; Brenner, D.J., Krieg, N.R., Staley, J.T., Garrity, G.M., Eds.; Springer: New York, NY, USA, 2005; Volume 2, Part C; pp. 161–167.
3. Kwon, S.Y.; Jiang, S.N.; Zheng, J.H.; Choy, H.E.; Min, J.J. *Rhodobacter sphaeroides*, a novel tumor-targeting bacteria that emits natural near-infrared fluorescence. *Microbiol. Immunol.* **2014**, *58*, 172–179. [CrossRef]
4. Christ, W.J.; Asano, O.; Robidoux, A.L.C.; Perez, M.; Wang, Y.; Dubuc, G.R.; Gavin, W.E.; Hawkins, L.D.; McGuinness, P.D.; Mullarkey, M.A.; et al. E5531, a pure endotoxin antagonist of high potency. *Science* **1995**, *268*, 80–83. [CrossRef] [PubMed]
5. A Mullarkey, M.; Rose, J.R.; Bristol, J.R.; Kawata, T.; Kimura, A.; Kobayashi, S.; Przetak, M.; Chow, J.; Gusovsky, F.; Christ, W.J.; et al. Inhibition of endotoxin response by e5564, a novel Toll-Like Receptor 4-directed endotoxin entagonist. *J. Pharmacol. Exp. Ther.* **2003**, *304*, 1093–1102. [CrossRef]
6. Kanie, Y.; Yamaguchi, Y.; Hayashi, A.; Uzawa, J.; Hatakeyama, M.; Hidaka, Y.; Toda, N.; Nakamura, S.; Kanie, O. Structural analysis of a novel lipooligosaccharide (LOS) from *Rhodobacter azotoformans. Carbohydr. Res.* **2019**, *473*, 104–114. [CrossRef] [PubMed]

7. Murakami, K.; Kamimura, D.; Hasebe, R.; Uchida, M.; Abe, N.; Yamamoto, R.; Jiang, J.J.; Hidaka, Y.; Nakanishi, Y.; Fujita, S.; et al. *Rhodobacter azotoformans* LPS (RAP99-LPS) is a TLR4 agonist that inhibits lung metastasis and enhances TLR3-mediated chemokine expression. *Front. Immunol.* **2021**, *12*, 675909. [CrossRef]

8. Eckersley, K.; Dow, C.S. *Rhodopseudomonas blastica* sp. nov: A member of the *Rhodospirillaceae. J. Gen. Microbiol.* **1980**, *119*, 465–473.

9. Khan, I.U.; Habib, N.; Xiao, M.; Li, M.M.; Xian, W.D.; Hejazi, M.S.; Tarhriz, V.; Zhi, X.Y.; Li, W.J. *Rhodobacter thermarum* sp. nov., a novel phototrophic bacterium isolated from sediment of a hot spring. *Antonie Van Leeuwenhoek.* **2019**, *112*, 867–875. [CrossRef]

10. Venkata Ramana, V.; Anil Kumar, P.; Srinivas, T.N.; Sasikala, C.; Ramana, C.V. *Rhodobacter aestuarii* sp. nov., a phototrophic alphaproteo bacterium isolated from an estuarine environment. *Int. J. Syst. Evol. Microbiol.* **2009**, *59 Pt 5*, 1133–1136. [CrossRef] [PubMed]

11. Gandham, S.; Lodha, T.; Chintalapati, S.; Chintalapati, V.R. *Rhodobacter alkalitolerans* sp. nov., isolated from an alkaline brown pond. *Arch. Microbiol.* **2018**, *200*, 1487–1492. [CrossRef] [PubMed]

12. Chen, W.M.; Chang, T.H.; Yang, C.C.; Sheu, D.S.; Jheng, L.C.; Sheu, S.Y. *Rhodobacter amnigenus* sp. nov. and *Rhodobacter ruber* sp. nov., isolated from freshwater habitats. *Int. J. Syst. Evol. Microbiol.* **2021**, *71*, 005150. [CrossRef]

13. Suresh, G.; Sailaja, B.; Ashif, A.; Dave, B.P.; Sasikala, C.; Ramana, C.V. Description of *Rhodobacter azollae* sp. nov. and *Rhodobacter lacus* sp. nov. *Int. J. Syst. Evol. Microbiol.* **2017**, *67*, 3289–3295. [CrossRef] [PubMed]

14. Hiraishi, A.; Muramatsu, K.; Ueda, Y. Molecular genetic analyses of *Rhodobacter azotoformans* sp. nov. and related species of phototrophic bacteria. *Syst. Appl. Microbiol.* **1996**, *19*, 168–177. [CrossRef]

15. Molisch, H. *Die Purpurbakterien Nach Neuen Untersuchungen*; G. Fischer: Jena, Germany, 1907.

16. Xian, W.D.; Liu, Z.T.; Li, M.M.; Liu, L.; Ming, Y.Z.; Xiao, M.; Salam, N.; Li, W.J. *Rhodobacter flagellatus* sp. nov., a thermophilic bacterium isolated from a hot spring. *Int. J. Syst. Evol. Microbiol.* **2020**, *70*, 1541–1546. [CrossRef]

17. Girija, K.R.; Sasikala, C.; Ramana, C.V.; Spröer, C.; Takaichi, S.; Thiel, V.; Imhoff, J.F. *Rhodobacter johrii* sp. nov., an endospore-producing cryptic species isolated from semi-arid tropical soils. *Int. J. Syst. Evol. Microbiol.* **2010**, *60*, 2099–2107. [CrossRef]

18. Venkata Ramana, V.; Sasikala, C.; Ramana, C.V. *Rhodobacter maris* sp. nov., a phototrophic alphaproteobacterium isolated from a marine habitat of India. *Int. J. Syst. Evol. Microbiol.* **2008**, *58*, 1719–1722. [CrossRef]

19. Srinivas, T.N.; Anil Kumar, P.; Sasikala, C.; Spröer, C.; Ramana, C.V. *Rhodobacter ovatus* sp. nov., a phototrophic alphaproteobacterium isolated from a polluted pond. *Int. J. Syst. Evol. Microbiol.* **2008**, *58*, 1379–1383. [CrossRef] [PubMed]

20. Subhash, Y.; Lee, S.S. *Rhodobacter sediminis* sp. nov., isolated from lagoon sediments. *Int. J. Syst. Evol. Microbiol.* **2016**, *66*, 2965–2970. [CrossRef] [PubMed]

21. Suresh, G.; Dhanesh, K.; Krishnaiah, A.; Sasikala, C.; Ramana, C.V. *Rhodobacter sediminicola* sp. nov., isolated from a fresh water pond. *Int. J. Syst. Evol. Microbiol.* **2020**, *70*, 1294–1299. [CrossRef]

22. Van Niel, C.B. The culture, general physiology, morphology and classification of the nonsulfur purple and brown bacteria. *Bacteriol. Rev.* **1944**, *8*, 1–118. [CrossRef]

23. Sheu, C.; Li, Z.H.; Sheu, S.Y.; Yang, C.C.; Chen, W.M. *Tabrizicola oligotrophica* sp. nov. and *Rhodobacter tardus* sp. nov., two new species of bacteria belonging to the family *Rhodobacteraceae. Int. J. Syst. Evol. Microbiol.* **2020**, *70*, 6266–6283. [CrossRef]

24. Hansen, T.A.; Imhoff, J.F. *Rhodobacter veldkampii*, a new species of Phototrophic Purple Nonsulfur Bacteria. *Int. J. Syst. Evol. Microbiol.* **1985**, *35*, 115–116. [CrossRef]

25. Srinivas, T.N.R.; Anil Kumar, P.; Sasikala, C.; Ramana, C.V.; Imhoff, J.F. *Rhodobacter vinaykumarii* sp. nov., a marine phototrophic alphaproteobacterium from tidal waters, and emended description of the genus *Rhodobacter. Int. J. Syst. Evol. Microbiol.* **2007**, *57*, 1984–1987. [CrossRef] [PubMed]

26. Raj, P.S.; Ramaprasad, E.V.V.; Vaseef, S.; Sasikala, C.; Ramana, C.V. *Rhodobacter viridis* sp. nov., a phototrophic bacterium isolated from mud of a stream. *Int. J. Syst. Evol. Microbiol.* **2013**, *63*, 181–186. [CrossRef]

27. Han, H.; Wang, T.; Chen, Z.; Li, Y.; Yao, L. *Rhodobacter xinxiangensis* sp. nov., isolated from pakchoi-cultivated soil contaminated with heavy metal and its potential to reduce Cd and Pb accumulation in pakchoi (*Brassica campestris* L.). *Arch. Microbiol.* **2020**, *202*, 1741–1748. [CrossRef] [PubMed]

28. Suresh, G.; Lodha, T.D.; Indu, B.; Sasikala, C.; Ramana, C.V. Taxogenomics resolves conflict in the genus *Rhodobacter*: A two and half decades pending thought to reclassify the genus *Rhodobacter. Front. Microbiol.* **2019**, *10*, 2480. [CrossRef]

29. Ramaley, R.F.; Hixson, J. Isolation of a nonpigmented, thermophilic bacterium similar to *Thermus aquaticus. J. Bacteriol.* **1970**, *103*, 527–528. [CrossRef]

30. Weaver, P.; Wall, J.D.; Gest, H. Characterization of *Rhodopseudomonas capsulata. Arch. Microbiol.* **1975**, *105*, 207–216. [CrossRef]

31. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comp. Biol.* **2012**, *19*, 455–477. [CrossRef] [PubMed]

32. Wattam, A.R.; Davis, J.J.; Assaf, R.; Boisvert, S.; Brettin, T.; Bun, C.; Conrad, N.; Dietrich, E.M.; Disz, T.; Gabbard, J.L.; et al. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Res.* **2017**, *45*, D535–D542. [CrossRef]

33. Aziz, R.K.; Bartels, D.; Best, A.A.; DeJongh, M.; Disz, T.; Edwards, R.A.; Formsma, K.; Gerdes, S.; Glass, E.M.; Kubal, M.; et al. The RAST server: Rapid annotations using subsystems technology. *BMC Genom.* **2008**, *9*, 75. [CrossRef] [PubMed]

34. Parks, D.H.; Imelfort, M.; Skennerton, C.T.; Hugenholtz, P.; Tyson, G.W. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **2015**, *25*, 1043–1055. [CrossRef] [PubMed]

35. Richter, M.; Rosselló-Móra, R.; Oliver Glöckner, F.O.; Peplies, J. JSpeciesWS: A web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* **2016**, *32*, 929–931. [CrossRef]

36. Rodriguez-R, L.M.; Konstantinidis, K.T. Bypassing cultivation to identify bacterial species. *Microbe* **2014**, *9*, 111–118.

37. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [CrossRef]

38. Stamatakis, A.; Hoover, P.; Rougemont, J.J.S. A rapid bootstrap algorithm for the RAxML web servers. *Syst. Biol.* **2008**, *57*, 758–771. [CrossRef] [PubMed]

39. Stamatakis, A.J.B. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [CrossRef] [PubMed]

40. Letunic, I.; Bork, P. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **2019**, *47*, W256–W259. [CrossRef]

41. Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680. [CrossRef]

42. Nei, M.; Kumar, S. *Molecular Evolution and Phylogenetics*; Oxford University Press: New York, NY, USA, 2000.

43. Le, S.Q.; Gascuel, O. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **2008**, *25*, 1307–1320. [CrossRef]

44. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef] [PubMed]

45. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303. [CrossRef]

46. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 421–430. [CrossRef]

47. Steinegger, M.; Meier, M.; Mirdita, M.; Vöhringer, H.; Haunsberger, S.J.; Söding, J. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinform.* **2019**, *20*, 473. [CrossRef]

48. Mirdita, M.; von den Driesch, L.; Galiez, C.; Martin, M.J.; Söding, J.; Steinegger, M. Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Res.* **2016**, *45*, D170–D176. [CrossRef] [PubMed]

49. Studer, G.; Tauriello, G.; Bienert, S.; Biasini, M.; Johner, N.; Schwede, T. ProMod3—A versatile homology modelling toolbox. *PLoS Comp. Biol.* **2021**, *17*, e1008667. [CrossRef]

50. Studer, G.; Rempfer, C.; Waterhouse, A.M.; Gumienny, G.; Haas, J.; Schwede, T. QMEANDisCo—Distance constraints applied on model quality estimation. *Bioinformatics* **2020**, *36*, 1765–1771. [CrossRef]

51. Satoh, T.; Hoshino, Y.; Kitamura, H. *Rhodopseudomonas sphaeroides* forma sp. denitrificans, a denitrifying strain as a subspecies of *Rhodopseudomonas sphaeroides*. *Arch. Microbiol.* **1976**, *108*, 265–269. [CrossRef]

52. Stackebrandt, E.; Ebers, J. Taxonomic parameters revisited: Tarnished gold standards. *Microbiol. Today* **2006**, *33*, 152–155.

53. Yurkov, V.V.; Beatty, J.T. Aerobic Anoxygenic Phototropic Bacteria. *Microbiol. Mol. Biol. Rev.* **1998**, *62*, 695–724. [CrossRef]

54. Yurkov, V.; Hughes, E. Genes associated with the peculiar phenotypes of the Aerobic Anoxygenic Phototrophs. In *Advances in Botanical Research*; Elsevier: Amsterdam, The Netherlands, 2013; pp. 327–358.

55. Kusian, B.; Bowien, B. Organization and regulation of CBB CO_2 assimilation genes in autotrophic bacteria. *FEMS Microbiol. Rev.* **1997**, *21*, 135–155. [CrossRef] [PubMed]

56. Ashida, H.; Danchin, A.; Yokota, A. Was photosynthetic RuBisCO recruited by acquisitive evolution from RuBisCO-like proteins involved in sulfur metabolism? *Res. Microbiol.* **2005**, *156*, 611–618. [CrossRef]

57. Badger, M.R.; Bek, E.J. Multiple Rubisco forms in proteobacteria: Their functional significance in relation to CO_2 acquisition by the CBB cycle. *J. Exp. Bot.* **2008**, *59*, 1525–1541. [CrossRef]

58. Kyndt, J.A.; Aviles, F.A.; Imhoff, J.F.; Künzel, S.; Neulinger, S.C.; Meyer, T.E. Comparative genome analysis of the photosynthetic betaproteobacteria of the genus *Rhodocyclus*: Heterogeneity within strains assigned to *Rhodocyclus tenuis* and description of *Rhodocyclus gracilis* sp. nov. as a new species. *Microorganisms* **2022**, *10*, 649. [CrossRef]

59. Imker, H.J.; Fedorov, A.A.; Fedorov, E.V.; Almo, S.C.; Gerlt, J.A. Mechanistic diversity in the RuBisCO superfamily: The "enolase" in the methionine salvage pathway in *Geobacillus kaustophilus*. *Biochemistry* **2007**, *46*, 4077–4089. [CrossRef]

60. Pemberton, J.M.; Horne, I.M.; McEwan, A.G. Regulation of photosynthetic gene expression in purple bacteria. *Microbiology* **1998**, *144*, 267–278. [CrossRef]

61. Gregor, J.; Klug, G. Regulation of bacterial photosynthesis genes by oxygen and light. *FEMS Microbiol. Lett.* **1999**, *179*, 1–9. [CrossRef]

62. Yurkov, V.; van Gemerden, H. Impact of light/dark regime on growth rate, biomass formation and bacteriochlorophyll synthesis in *Erythromicrobium hydrolyticum*. *Arch. Microbiol.* **1993**, *159*, 84–89. [CrossRef]

63. Takamiya, K.I.; Shioi, Y.; Shimada, H.; Arata, H. Inhibition of accumulation of bacteriochlorophyll and carotenoids by blue light in an aerobic photosynthetic bacterium *Roseobacter denitrificans*, during anaerobic respiration. *Plant Cell Physiol.* **1992**, *33*, 1171–1174.

64. Yurkov, V.; Krasilnikova, E.N.; Gorlenko, V.M. Effect of light and oxygen on metabolism of the aerobic bacterium *Erythromicrobium sibiricum*. *Microbiology* **1993**, *62*, 35–38.

65. Cartron, M.L.; Olsen, J.D.; Sener, M.; Jackson, P.J.; Brindley, A.A.; Qian, P.; Dickman, M.J.; Leggett, G.J.; Schulten, K.; Neil Hunter, C. Integration of energy and electron transfer processes in the photosynthetic membrane of *Rhodobacter sphaeroides*. *Biochim. Biophys. Acta Bioenerg.* **2014**, *1837*, 1769–1780. [CrossRef]

66. Olsen, J.D.; Tucker, J.D.; Timney, J.A.; Qian, P.; Vassilev, C.; Hunter, C.N. The organization of LH2 complexes in membranes from *Rhodobacter sphaeroides*. *J. Biol. Chem.* **2008**, *283*, 30772–30779. [CrossRef] [PubMed]

67. Gabrielsen, M.; Gardiner, A.T.; Cogdell, R. Peripheral complexes of purple bacteria. In *The Purple Phototrophic Bacteria, Advances in Photosynthesis and Respiration*; Hunter, C.N., Daldal, F., Thurnauer, M.C., Beatty, J.T., Eds.; Springer: Dordrecht, The Netherlands, 2009; Volume 28, pp. 135–153.

68. Jaschke, P.R.; Leblanc, H.N.; Lang, A.S.; Beatty, J.T. The PucC protein of *Rhodobacter capsulatus* mitigates an inhibitory effect of light-harvesting 2 alpha and beta proteins on light-harvesting complex 1. *Photosynth Res.* **2008**, *95*, 279–284. [CrossRef] [PubMed]

69. Mothersole, D.J.; Jackson, P.J.; Vasilev, C.; Tucker, J.D.; Brindley, A.A.; Dickman, M.J.; Hunter, C.N. PucC and LhaA direct efficient assembly of the light-harvesting complexes in *Rhodobacter sphaeroides*. *Mol. Microbiol.* **2016**, *99*, 307–327. [CrossRef]

70. Selyanin, V.; Hauruseu, D.; Koblížek, M. The variability of light-harvesting complexes in aerobic anoxygenic phototrophs. *Photosynth Res.* **2016**, *128*, 35–43. [CrossRef]

71. Wakao, N.; Shiba, T.; Hiraishi, A.; Ito, M.; Sakurai, Y. Distribution of bacteriochlorophyll a in species of the genus *Acidiphilium*. *Curr. Microbiol.* **1993**, *27*, 277–279. [CrossRef]

72. Wakao, N.; Yokoi, N.; Isoyama, N.; Hiraishi, A.; Shimada, K.; Kobayashi, M.; Kise, H.; Iwaki, M.; Itoh, S.; Takaichi, S. Discovery of natural photosynthesis using Zn-containing bacteriochlorophyll in an aerobic bacterium *Acidiphilium rubrum*. *Plant Cell Physiol.* **1996**, *37*, 889–893. [CrossRef]

73. Koblížek, M. Ecology of aerobic anoxygenic phototrophs in aquatic environments. *FEMS Microbiol. Rev.* **2015**, *39*, 854–870. [CrossRef]

74. Bekker, A.; Holland, H.D.; Wang, P.L.; Rumble, D., 3rd; Stein, H.J.; Hannah, J.L.; Coetzee, L.L.; Beukes, N.J. Dating the rise of atmospheric oxygen. *Nature* **2004**, *427*, 117–120. [CrossRef] [PubMed]

75. Koblížek, M.; Zeng, Y.; Horák, A.; Oborník, M. Regressive evolution of photosynthesis in the *Roseobacter* clade. In *Advances in Botanical Research*; Thomas Beatty, J., Ed.; Academic Press: Cambridge, MA, USA; Elsevier: Amsterdam, The Netherlands, 2013; Volume 66, pp. 385–405. [CrossRef]

76. Balashov, S.P.; Imasheva, E.S.; Boichenko, V.A.; Antón, J.; Wang, J.M.; Lanyi, J.K. Xanthorhodopsin: A proton pump with a light-harvesting carotenoid antenna. *Science* **2005**, *309*, 2061–2064. [CrossRef] [PubMed]

77. Lanyi, J.K.; Balashov, S.P. Xanthorhodopsin: A bacteriorhodopsin-like proton pump with a carotenoid antenna. *Biochim. Biophys. Acta* **2008**, *1777*, 684–688. [CrossRef] [PubMed]

78. Morizumi, T.; Ou, W.L.; Van Eps, N.; Inoue, K.; Kandori, H.; Brown, L.S.; Ernst, O.P. X-ray crystallographic structure and oligomerization of *Gloeobacter* rhodopsin. *Sci. Rep.* **2019**, *9*, 11283. [CrossRef] [PubMed]

79. Zhong, Z.-P.; Liu, Y.; Wang, F.; Zhou, Y.-G.; Liu, H.-C.; Liu, Z.-P. *Aquisalinus flavus* gen. nov., sp. nov., a member of the family *Parvularculaceae* isolated from a saline lake. *Int. J. Syst. Evol. Microbiol.* **2016**, *66*, 1813–1817. [CrossRef] [PubMed]

80. Liu, Z.X.; Dorji, P.; Liu, H.C.; Li, A.H.; Zhou, Y.G. *Tabrizicola sediminis* sp. nov., one aerobic anoxygenic photoheterotrophic bacteria from sediment of saline lake. *Int. J. Syst. Evol. Microbiol.* **2019**, *69*, 2565–2570. [CrossRef] [PubMed]

81. Woggon, W.-D. Oxidative cleavage of carotenoids catalyzed by enzyme models and beta-carotene 15,15′-monooxygenase. *Pure Appl. Chem.* **2002**, *74*, 1397–1408. [CrossRef]

82. Kim, J.; Smith, J.J.; Tian, L.; Dellapenna, D. The evolution and function of carotenoid hydroxylases in *Arabidopsis*. *Plant Cell Physiol.* **2009**, *50*, 463–479. [CrossRef]

83. Canniffe, D.P.; Hitchcock, A. Photosynthesis: Carotenoids in Photosynthesis—Structure and Biosynthesis. In *Encyclopedia of Biological Chemistry III*, 3rd ed.; Jez, J., Ed.; Elsevier: Amsterdam, The Netherlands, 2021; pp. 163–185. ISBN 9780128220405.

84. Maresca, J.A.; Graham, J.E.; Wu, M.; Eisen, J.A.; Bryant, D.A. Identification of a fourth family of lycopene cyclases in photosynthetic bacteria. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 11784–11789. [CrossRef]

85. Frigaard, N.U.; Martinez, A.; Mincer, T.J.; DeLong, E.F. Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea. *Nature* **2006**, *439*, 847–850. [CrossRef]

86. Finkel, O.; Béjà, O.; Belkin, S. Global abundance of microbial rhodopsins. *ISME J.* **2013**, *7*, 448–451. [CrossRef]

87. Gushchin, I.; Shevchenko, V.; Polovinkin, V.; Kovalev, K.; Alekseev, A.; Round, E.; Borshchevskiy, V.; Balandin, T.; Popov, A.; Gensch, T.; et al. Crystal structure of a light-driven sodium pump. *Nat. Struct. Mol. Biol.* **2015**, *22*, 390–395. [CrossRef]

88. Kato, H.E.; Inoue, K.; Abe-Yoshizumi, R.; Kato, Y.; Ono, H.; Konno, M.; Hososhima, S.; Ishizuka, T.; Hoque, M.R.; Kunitomo, H.; et al. Structural basis for Na(+) transport mechanism by a light-driven Na(+) pump. *Nature* **2015**, *521*, 48–53. [CrossRef] [PubMed]

89. Boichenko, V.A.; Wang, J.M.; Antón, J.; Lanyi, J.K.; Balashov, S.P. Functions of carotenoids in xanthorhodopsin and archaerhodopsin, from action spectra of photoinhibition of cell respiration. *Biochim. Biophys. Acta–Bioenerg.* **2006**, *1757*, 1649–1656. [CrossRef] [PubMed]

90. Thiel, V.; Hügler, M.; Ward, D.M.; Bryant, D.A. The dark side of the Mushroom Spring microbial mat: Life in the shadow of Chlorophototrophs. II. Metabolic functions of abundant community members predicted from metagenomic snalyses. *Front. Microbiol.* **2017**, *8*, 943. [CrossRef]

91. Gómez-Consarnau, L.; González, J.M.; Coll-Lladó, M.; Gourdon, P.; Pascher, T.; Neutze, R.; Pedrós-Alió, C.; Pinhassi, J. Light stimulates growth of proteorhodopsin-containing marine *Flavobacteria*. *Nature* **2007**, *445*, 210–213. [CrossRef]

92. Gómez-Consarnau, L.; Akram, N.; Lindell, K.; Pedersen, A.; Neutze, R.; Milton, D.L.; González, J.M.; Pinhassi, J. Proteorhodopsin phototrophy promotes survival of marine bacteria during starvation. *PLoS Biol.* **2010**, *8*, e1000358. [CrossRef]

93. Wang, Z.; O'Shaughnessy, T.J.; Soto, C.M.; Rahbar, A.M.; Robertson, K.L.; Lebedev, N.; Vora, G.J. Function and regulation of *Vibrio campbellii* Proteorhodopsin: Acquired phototrophy in a classical organoheterotroph. *PLoS ONE* **2012**, *7*, e38749. [CrossRef] [PubMed]

94. Kumagai, Y.; Yoshizawa, S.; Nakajima, Y.; Watanabe, M.; Fukunaga, T.; Ogura, Y.; Hayashi, T.; Oshima, K.; Hattori, M.; Ikeuchi, M.; et al. Solar-panel and parasol strategies shape the proteorhodopsin distribution pattern in marine *Flavobacteria*. *ISME J.* **2018**, *12*, 1329–1343. [CrossRef]

95. Thiel, V.; Wood, J.M.; Olsen, M.T.; Tank, M.; Klatt, C.G.; Ward, D.M.; Bryant, D.A. The dark side of the Mushroom Spring microbial mat: Life in the shadow of Chlorophototrophs. I. Microbial diversity based on 16S rRNA gene amplicons and metagenomic sequencing. *Front. Microbiol.* **2016**, *7*, 919. [CrossRef]

Comparative Genome Analysis of the Photosynthetic Betaproteobacteria of the Genus *Rhodocyclus*: Heterogeneity within Strains Assigned to *Rhodocyclus tenuis* and Description of *Rhodocyclus gracilis* sp. nov. as a New Species

John A. Kyndt [1],* , Fabiola A. Aviles [1], Johannes F. Imhoff [2] , Sven Künzel [3], Sven C. Neulinger [4]
and Terrance E. Meyer [5],†

[1] College of Science and Technology, Bellevue University, Bellevue, NE 68005, USA;
     avilesacosta7766@gmail.com
[2] GEOMAR Helmholtz Centre for Ocean Research Kiel, RD3 Marine Symbioses, Düsternbrooker Weg 20,
     24105 Kiel, Germany; jimhoff@geomar.de
[3] Max Planck Institute for Evolutionary Biology, 24306 Plön, Germany; kuenzel@evolbio.mpg.de
[4] omics2view.consulting GbR, 24118 Kiel, Germany; sven.neulinger@web.de
[5] Department of Biochemistry, University of Arizona, Tucson, AZ 85721, USA; temeyer@email.arizona.edu
* Correspondence: jkyndt@bellevue.edu
† Terrance E. Meyer passed away while the manuscript was in the final stages of preparation.

Abstract: The genome sequences for *Rhodocyclus purpureus* DSM 168[T] and four strains assigned to *Rhodocyclus tenuis* (DSM 110, DSM 111, DSM 112, and IM 230) have been determined. One of the strains studied (IM 230) has an average nucleotide identity (ANI) of 97% to the recently reported genome of the type strain DSM 109 of *Rcy. tenuis* and is regarded as virtually identical at the species level. The ANI of 80% for three other strains (DSM 110, DSM 111, DSM 112) to the type strain of *Rcy. tenuis* points to a differentiation of these at the species level. *Rcy. purpureus* is equidistant from *Rcy. tenuis* and the new species, based on both ANI (78–80%) and complete proteome comparisons (70% AAI). Strains DSM 110, DSM 111, and DSM 112 are very closely related to each other based on ANI, whole genome, and proteome comparisons but clearly distinct from the *Rcy. tenuis* type strain DSM 109. In addition to the whole genome differentiation, these three strains also contain unique genetic differences in cytochrome genes and contain genes for an anaerobic cobalamin synthesis pathway that is lacking from both *Rcy. tenuis* and *Rcy. purpureus*. Based on genomic and genetic differences, these three strains should be considered to represent a new species, which is distinctly different from both *Rcy. purpureus* and *Rcy. tenuis*, for which the new name *Rhodocyclus gracilis* sp. nov. is proposed.

Keywords: *Rhodocyclus*; *purpureus*; *tenuis*; *gracilis*; HiPIP; cobalamin; whole genome sequencing; taxonomy

1. Introduction

The anaerobic anoxygenic photosynthetic betaproteobacteria are represented by a small group of bacteria currently classified in the *Burkholderiales* (the genera *Rhodoferax* and *Rubrivivax*) and *Rhodocyclales* (the genus *Rhodocyclus*) orders. Genome sequences of several of the *Burkholderiales* species have been previously reported, including strains of *Rubrivivax gelatinosus* [1,2], *Rubrivivax benzoatilyticus* [3], *Rhodoferax fermentans* [4], *Rhodoferax antarcticus* [5], and *Rhodoferax jenense* [6]. There also is a genome sequence for the aerobic anoxygenic photosynthetic bacterium *Roseateles depolymerans* [7], which is related to *Rubrivivax gelatinosus*.

Though the genome sequence of the type strain of *Rhodocyclus tenuis* DSM 109[T] has been reported recently [8], the diversity of this group of bacteria has not been studied in

detail. The first and type species of the genus *Rhodocyclus, Rhodocyclus purpureus,* was discovered by Norbert Pfennig [9]. A second species of this genus was isolated and described based on a single strain as *Rhodospirillum tenue* [10] and later reclassified as *Rhodocyclus tenuis* [11]. Several strains that were assigned to *Rhodocyclus tenuis* have been isolated from various German freshwater lakes (Lake Pluss-See near Plön, a forest ditch near Grünenplan, Garrensee near Ratzeburg, Eutiner See, Nonnenmattweiher near Neuenweg Black Forest, Edebergsee near Plön) and peat bogs (Black Forest) by Hanno Biebl and Norbert Pfennig [12]. A single isolate obtained from a pond ("keyhole pond") in the Botanical Garden in Bonn was obtained by one of us (JFI). Parts of these strains have been studied previously regarding to their physiological properties [12], carotenoids [13], lipopolysaccharide structures [14], sulfate assimilation [15], and lipid and fatty acid composition [16]. Although these studies revealed the heterogeneity of the group, no systematic studies have been performed.

Here, we report on the genome analysis of *Rhodocyclus purpureus* DSM 168[T] and *Rhodocyclus tenuis* DSM 109[T] and four strains previously assigned to *Rhodocyclus tenuis* and provide evidence for the existence of a new species of *Rhodocyclus,* represented by three of the strains studied that have previously been shown to be different from the type strain of *Rcy. tenuis* [12–14].

2. Materials and Methods

2.1. Origin of the Strains of Rhodocyclus Tenuis

The type strain DSM 109[T] (Pfennig 2761) was isolated from a pond near Grünenplan (district Holzminden, Germany); strains DSM 110 (Pfennig 3760) and DSM 111 (Pfennig 3761) from Nonnenmattweiher pond near Neuenweg in the rural district Lörrach in the Black Forest (Germany); and strain DSM 112 (Pfennig 3661) from a kolk in the Hinterzartener Moor in the Black Forest, Germany. Strain IM 230 (Imhoff 230) was isolated by JF Imhoff from a small pond called "Schlüssellochteich" (keyhole pond) in the Botanical Garden of Bonn University.

2.2. Genome Sequencing

Genomic DNA for *Rhodocyclus tenuis* DSM 110, DSM 111, and DSM 112 was obtained from DSMZ. Cells of *Rcy. purpureus* DSM 168[T] and *Rcy. tenuis* IM 230 were grown under the recommended conditions [17] and DNA extracted from well-grown cultures according to Imhoff et al. [18]. The quantity and purity of DNA were determined using Qubit and Nanodrop instruments and showed 260/280 ratios between 1.80 and 1.90. The DNA libraries were prepared with the Nextera® XT DNA Sample Preparation kit from Illumina (San Diego, CA, USA) following the manufacturer's protocol.

Rhodocylus sp. strains DSM 110, DSM 111, and DSM 112 genomes were sequenced using 500 µL of a 1.8 pM library with an Illumina MiniSeq instrument, using paired-end sequencing (2 × 150 bp). Quality control of the reads was performed using FASTQC within BaseSpace (Illumina, San Diego; version 1.0.0), using a k-mer size of 5 and contamination filtering. The data for each were assembled de novo using SPAdes (version 3.10.0; [19,20]) for DSM 110 and DSM 111 or Unicycler within PATRIC [21,22] for DSM 112. Default k-mer lengths were used for both programs. The genome sequences were annotated using RAST (Rapid Annotations using Subsystem Technology; version 2.0; [23]). An EvalG genome quality analysis, using the checkM algorithm [24], ran during PATRIC annotation and showed an estimated 100% completeness and 0% contamination for each of these genomes.

Rcy. purpureus and *Rcy. tenuis* IM 230 genomes were sequenced on an Illumina MiSeq using the MiSeq® Reagent Kit v3 600 cycles sequencing chemistry (Illumina, San Diego, CA, USA), with a cluster density of approximately 1200 K/mm2. Trimmomatic v0.36 [25] was used for read quality filtering. Illumina Nextera XT adapters were removed from the reads. Quality trimming was conducted with a 5-base pair (bp) sliding window, trimming the reads with an average Phred quality score below 30. Read lengths >21 bp after quality trimming were retained. Only single reads (i.e., reads with their mate deleted) were

included into downstream analysis. Reads were further checked for ambiguous base calls, as well as for low complexity, using the DUST algorithm [26]. They were filtered accordingly with an in-house R script in Microsoft R Open v3.3.2 (R Core Team 2016). SPAdes v3.10.0 was used for the pre-assembly of the filtered reads [19,20], using default k-mer lengths. Scaffolds $\geq$ 500 bp of this pre-assembly were subjected to extension and second-round scaffolding with SSPACE standard v3.0 [27]. Scaffolds $\geq$ 2500 bp were assigned to genome bins by MetaBAT v0.32.4 [28], to ensure draft-genome purity of *Rcy. purpureus* DSM 168[T]. From the two resulting genome bins (3.595 and 1.048 Mbp, respectively), the larger one with a G + C content of 66 mol% was selected as the draft genome of *Rcy. purpureus*. Base coverage was determined with BBMap v36.81 (https://sourceforge.net/projects/bbmap (accessed on 29 May 2017)) [29] for filtered reads unambiguously mapped to the scaffolds of the draft genome. Estimated fold-coverage was calculated as the median base coverage over all scaffold positions.

2.3. Whole Genome Comparison

Average percentage nucleotide identity (ANIb) between the whole genomes was calculated using JSpecies [30]. A whole genome-based phylogenetic tree was generated using the CodonTree method within PATRIC [22], which used PGFams (global (cross-genus) protein families) as homology groups. At total of 445 PGFams were found among these selected genomes using the CodonTree analysis, and the aligned proteins and coding DNA from single-copy genes were used for RAxML analysis [31,32]. iTOL was used for tree visualization [33]. A proteome comparison was performed using protein sequence-based genome comparison using bidirectional BLASTP within PATRIC [22]. Average amino acid identities (AAI) values were calculated from the proteome comparison within PATRIC [22], using only bi-directional hits with *Rcy. tenuis* DSM109[T] as the reference strain. Digital DNA–DNA Hybridization (dDDH) data were obtained using the Type (Strain) Genome Server (TYGS) web server (https://tygs.dsmz.de (accessed on 3 April 2021)) [34]. The program used the distance formula d4 to calculate a similarity based on sequence identity.

For synteny analysis, global PATRIC PGFam families were used to generate comparative genome regions to determine a set of genes that match a focus gene [22]. All *Rhodocyclus* genomes were used in the search and were compared to the DSM 110 genome. The gene set is compared to the focus gene using BLAST and sorted by BLAST scores within PATRIC [22]. The *cbi*X (long cobaltochelatase) gene was used as a focus gene to analyze synteny of the cobalamin synthesis gene cluster.

The multiple sequence alignments for the 16S rRNA, HiPIP, and RuBisCo comparisons were performed using Clustal Omega [35]. All of the 16S rRNA sequences were genome derived. The phylogenetic tree was calculated by the neighbor-joining (NJ) method [36] within JALVIEW [37] and a Newick file was generated. iTOL was used to draw the phylogenetic trees expressed in the Newick phylogenetic tree format [33].

3. Results and Discussion

3.1. Whole Genome Analysis

The genomic features of six strains of the genus *Rhodocyclus,* including five new genome sequences and the previously sequenced *Rcy. tenuis* DSM 109[T] genome, were compared (summary in Table 1). Based on the genome size and the G + C content, three groups can be recognized: *Rcy. tenuis* DSM 109[T] and IM 230 represent group 1 and have an identical G + C content (64.7 mol%) and slightly larger genome size compared to those of group 2, with strains DSM 110, DSM 111, and DSM 112. *Rcy. purpureus* DSM 168[T] is the only representative of group 3 and has a similar genome size compared to the group 1 strains but a higher G + C content (66.1 mol%).

Table 1. Overview of genome features of all of the *Rhodocyclus* genome sequences.

Species	Genome Size	GC Content	Contigs	Coverage	CDS	tRNAs	Reference	Genbank Accession #
Rhodocyclus tenuis DSM109	3.85 Mb	64.7	30	100x	3605	51	*Wang et.al., 2020*	SSSP00000000
Rhodocyclus tenuis IM230	3.65 Mb	64.7	28	93x	3359	51	*this study*	NRRZ00000000
Rhodocyclus **sp.** *DSM110*	2.93 Mb	64.5	27	54x	2735	47	*this study*	WIXJ00000000
Rhodocyclus **sp.** *DSM111*	2.93 Mb	64.4	30	55x	2705	47	*this study*	WJED00000000
Rhodocyclus **sp.** *DSM112*	2.98 Mb	64.5	27	73x	2768	47	*this study*	JAATWB000000000
Rhodocyclus purpureus DSM168	3.62 Mb	66.1	69	81x	3600	51	*this study*	NHRX00000000

The Average Nucleotide Identity (ANI) comparison revealed a 97.1% ANI of strains within group 1 and values of 98.8% and higher of strains within group 2 (Table 2). However, the comparison between strains from the two groups shows ANI values below 80%. Applying the arbitrary cutoff value for species differentiation of 95% [30], the two groups should clearly represent different species. *Rcy. purpureus* has ANI values of 80% or less with all of the other strains and is rightfully recognized as a distinct species. For comparison, *Accumulibacter phosphatis* was included (Table 2) as a species from a closely related genus [38], which showed ANI values of 72–73% with all of the *Rhodocyclus* strains.

Table 2. Whole-genome-based average nucleotide identity (ANI) of *Rhodocyclus* species and relatives. ANI values above the species cutoff of 95% are shown in bold.

Rhodocyclus tenuis DSM 109[T]						
97.1	*Rhodocyclus tenuis* IM 230					
80.3	80.3	*Rhodocyclus purpureus* DSM 168[T]				
79.7	79.4	79.9	*Rhodocyclus* sp. DSM 111			
79.8	79.4	77.8	**98.8**	*Rhodocyclus* sp. DSM 110		
79.7	79.5	77.9	**98.9**	**98.9**	*Rhodocyclus* sp. DSM 112	
73.3	72.9	73.1	72.5	72.5	72.5	*Ca. Accumulibacter phosphatis* UW1

The ANI data imply that all of the studied strains are indeed members of the genus *Rhodocyclus*, and *Rhodocyclus purpureus* and *Rhodocylus tenuis* are distinct species of this genus. However, the strains of group 2 (DSM 110, DSM 111, and DSM 112) belong to a new species of the genus *Rhodocyclus*. A whole-genome-based phylogenetic tree (Figure 1) supports these findings and shows the group of strains DSM 110, DSM 111, and DSM 112 as very close relatives to each other but apart from the type strains of *Rcy. tenuis* DSM 109[T] and *Rcy. purpureus* DSM 168[T]. This indicates that this group indeed forms a separate species of the *Rhodocyclus* genus.
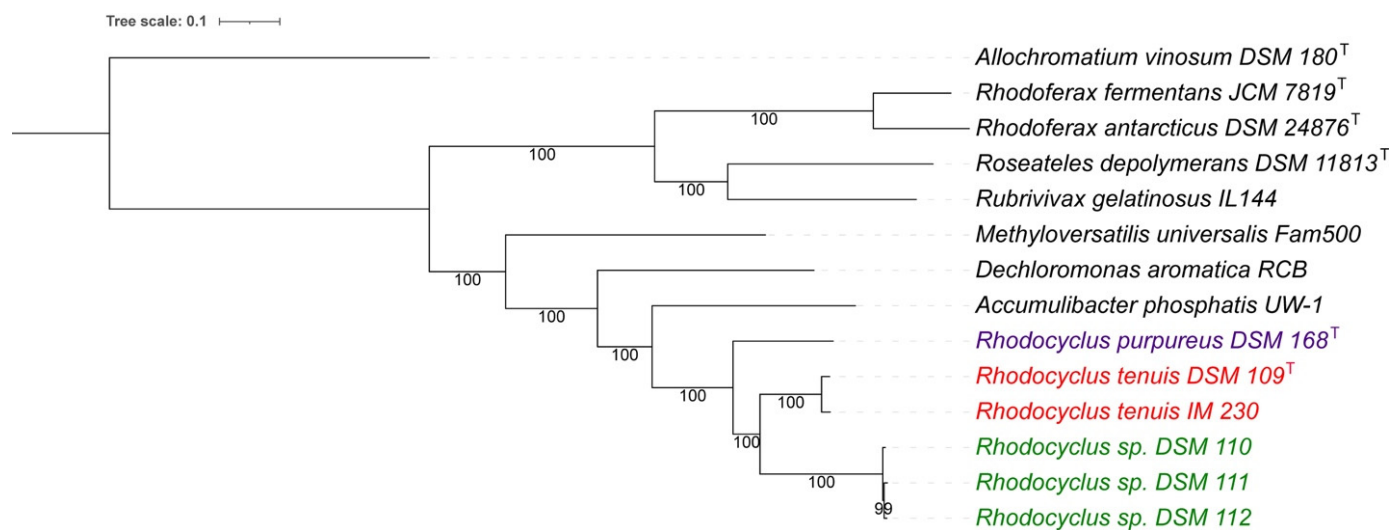
Figure 1. Whole-genome-based phylogenetic tree of all known *Rhodocyclus*, compared to representative genomes of close relatives. One hundred rounds of the 'Rapid bootstrapping' option of RaxML were used to generate the support values for the phylogenetic tree. The branch length tree scale is defined as the mean number of substitutions per site, which is an average across both nucleotide and amino acid changes. The *Rhodocyclus* genomes are colored differently based on their ANI values (with a species cutoff of 95%). *Allochromatium vinosum* DSM 180^T was added as an outgroup.

A protein-sequence-based genome comparison, with the type strain of *Rcy. tenuis* DSM 109^T as a reference, provided a complete proteome comparison (Figure 2). This provided amino acid sequence identity for both bi- and unidirectional hits between each genome and the reference genome (color coded in Figure 2). The average amino acid sequence identity (AAI), using only reciprocal (bi-directional) hits, was calculated for each genome, which showed a 78.4% identity with DSM 110 (from 2203 proteins), 77.8% with DSM 111 (from 2184 proteins), and 78.0% with DSM 112 (from 2185 proteins). Similar values were obtained for *Rcy. purpureus* DSM 168^T with 77.0% identity (from 1938 proteins). However, a much higher value of 97.6% was found when comparing the *Rcy. tenuis* IM 230 proteome (from 3058 proteins) with the DSM 109 reference proteome. These data correspond well with the other analyses and further support the distinction of the strains DSM 110, DSM 111, and DSM 112 at the species level.

Digital DNA–DNA hybridization analyses (dDDH) showed only a distant relationship between the type strain of *Rcy. tenuis* DSM 109^T and the strains DSM 110 (25.0%), DSM 111 (24.9%), and DSM 112 (25.1%). These values are similar to what was obtained between *Rcy. tenuis* DSM 109^T and *Rcy. purpureus* DSM 168^T (24.8%). The dDDH values between *Rcy. purpureus* and the three strains also places them distantly related, with 22.9% (DSM 110), 23.0% (DSM 111), and 23.0% (DSM 112). On the other hand, the dDDH values amongst the three strains, DSM 110, DSM 111, and DSM 112, showed very high DNA–DNA hybridization values (91–92%). Consistent with the analyses provided above, this places these three strains in a closer relationship with each other than any of the other species, equidistant from *Rcy. tenuis* DSM 109^T and *Rcy. purpureus*, DSM 168^T, supporting the placement of those into a separate species group.
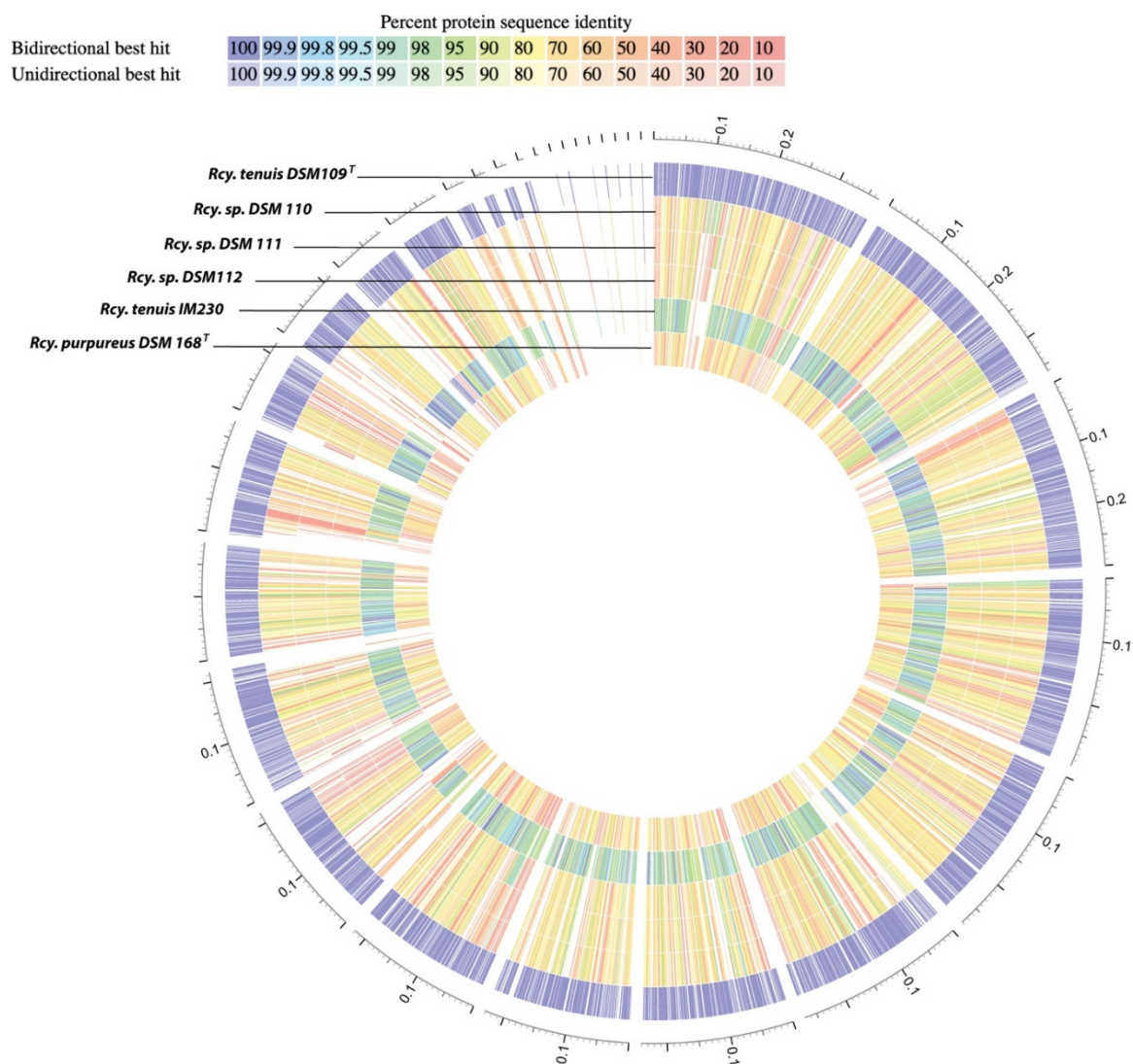
Figure 2. Proteome comparison of *Rhodocyclus* species and relatives based on protein-sequence-based genome comparison using bidirectional BLASTP. *Rcy. tenuis* DSM 109[T] was used as the reference proteome. The percent protein identity is color-coded for each proteome as compared to the reference proteome. *Rcy. tenuis* IM 230 showed average amino acid identities (AAI) of 97.6% (blue-green), while the *Rcy.* DSM 110, *Rcy*. DSM 111, *Rcy.* DSM 112, and *Rcy. purpureus* proteomes are equidistant from the reference proteome (77–78% identity; yellow-orange).

In addition to the whole-genome-based analyses, we also compared the 16S rRNA sequences of all *Rhodocyclus* species and related species. A 16S rRNA-based phylogenetic tree is provided in Figure 3. *Rcy. tenuis* DSM 109[T] and IM 230 have a 99.4% 16S rRNA identity (1541 nt. overlap), while the three strains DSM 110, DSM 111, and DSM 112 only have 97.4% identity with *Rcy. tenuis* DSM 109[T] (1541 nt. overlap). They do have 99.9% identity amongst themselves. The *Rcy. purpureus* 16S rRNA is equidistant from *Rcy tenuis* DSM 109[T] and DSM 110, with 96.6% and 96.0% identity, respectively (1551 nt. overlap). These values are below the proposed species delineation for 16S rRNA comparisons of 98.7% [39] and place *Rcy. purpureus* and the three strains, DSM 110, DSM 111, and DSM 112, on separate clades in the 16S rRNA phylogenetic tree (Figure 3), which is consistent with the whole-genome-based analyses described above.
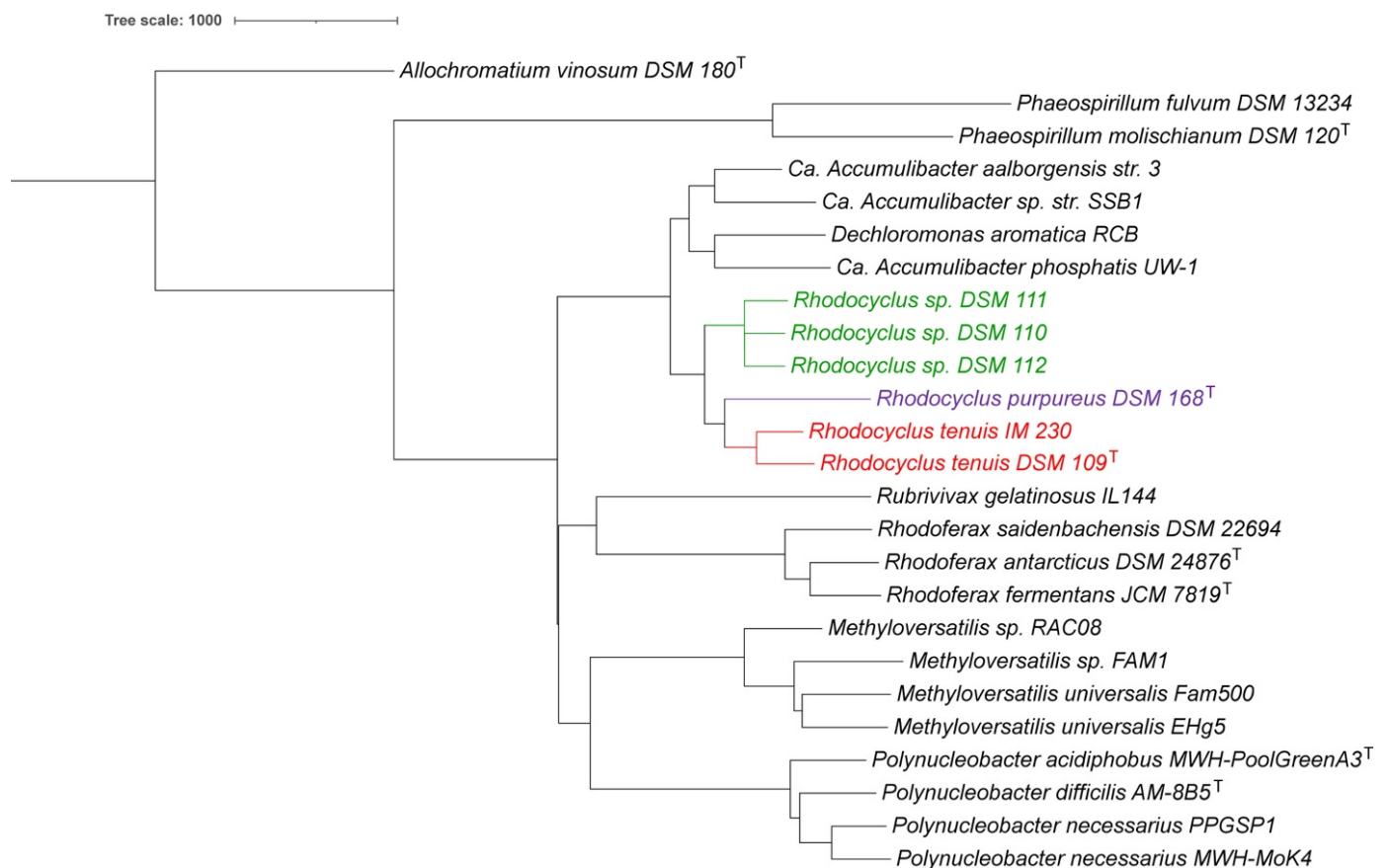
Figure 3. 16S rRNA phylogenetic tree for *Rhodocyclus* and related species. The phylogenetic tree was calculated by the neighbor-joining (NJ) method [36] within Jalview [37]. iTOL was used to draw the phylogenetic trees expressed in the Newick phylogenetic tree format [33]. *Allochromatium vinosum* DSM 180[T] was added as an outgroup. *Rhodocyclus* species were color-coded the same as in Figure 1.

3.2. Cytochrome and High Potential Iron Protein HiPIP Analysis

An interesting aspect of this study is the apparent use of different electron donors to the photosynthetic reaction center. While HiPIP is the normal electron donor to most photosynthetic reaction centers in the *Gammaproteobacteria*, cytochrome c_2 fills this role in the *Alphaproteobacteria*. In phototrophic *Betaproteobacteria*, the situation is different. While HiPIP is the usual electron donor in *Rvi. gelatinosus*, a high potential cytochrome c_8, which is induced under aerobic growth, may also participate under some conditions [40]. HiPIP is the electron donor to reaction centers in *Rfx. fermentans* as well [41]. *Rcy. tenuis* is known to utilize both HiPIP and cytochrome c_8 in cyclic electron transfer depending on the growth conditions [42].

Soluble electron transfer proteins were previously characterized from strains DSM 109[T] and DSM 111 and found to be similar to one another but distinct from those of *Rcy. purpureus* [43]. These were described as cytochrome c_4 (minor component), cytochrome c_8, cytochrome c-552 (NirB), cytochrome c', and HiPIP. The latter appears to be absent in *Rcy. purpureus*. A multiple sequence alignment and phylogenetic tree of the HiPIP protein from all *Rhodocyclus tenuis* species (Figure 4) resulted in a phylogenetic relationship that is consistent with the whole genome and ANI comparisons described above. The HiPIP protein sequences from DSM 110, DSM 111, and DSM 112 clearly form a clade on the tree separate from the two sequences of *Rcy. tenuis* DSM 109[T] and IM230. *Rcy. tenuis* DSM 109[T] and *Rcy. gracilis* apparently utilize HiPIP as electron donor to the photosynthetic reaction center, but we have now shown that the HiPIP gene, as well as the soluble protein, is lacking in *Rcy. purpureus*, confirming the previous analysis and also suggesting that a

cytochrome, presumably c_8, fills the role of mediator between the cytochrome bc_1 complex and the PufLMC reaction center.
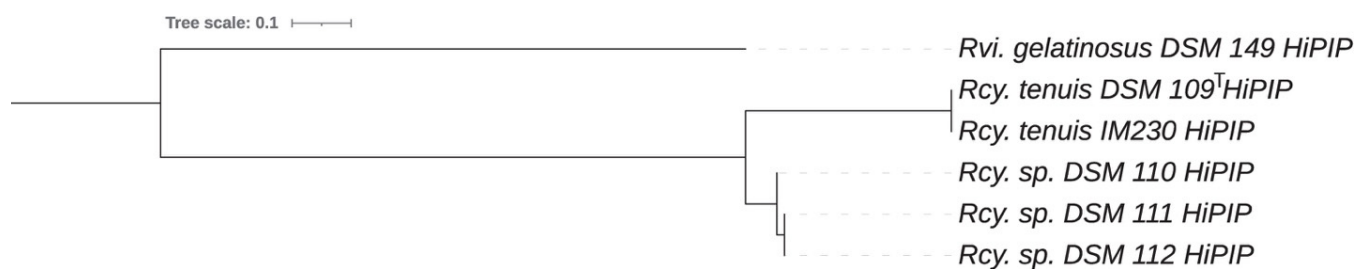
Figure 4. Phylogenetic tree of the HiPIP protein sequences obtained from the *Rhodocyclus* genomes. *Rubrivivax gelatinosus* DSM 149 HiPIP was used as an outgroup.

3.3. Nitrogen Metabolism

All species of *Rhodocyclus* apparently produce large amounts of the denitrifying diheme cytochrome NirB [43], which has been shown by the sequence of the protein from *Rcy. tenuis* DSM 109^T and DSM 111 [44] and by the fact that all of the *Rhodocyclus* genome sequences in the current study contain the *nir*B gene. However, none of these species appa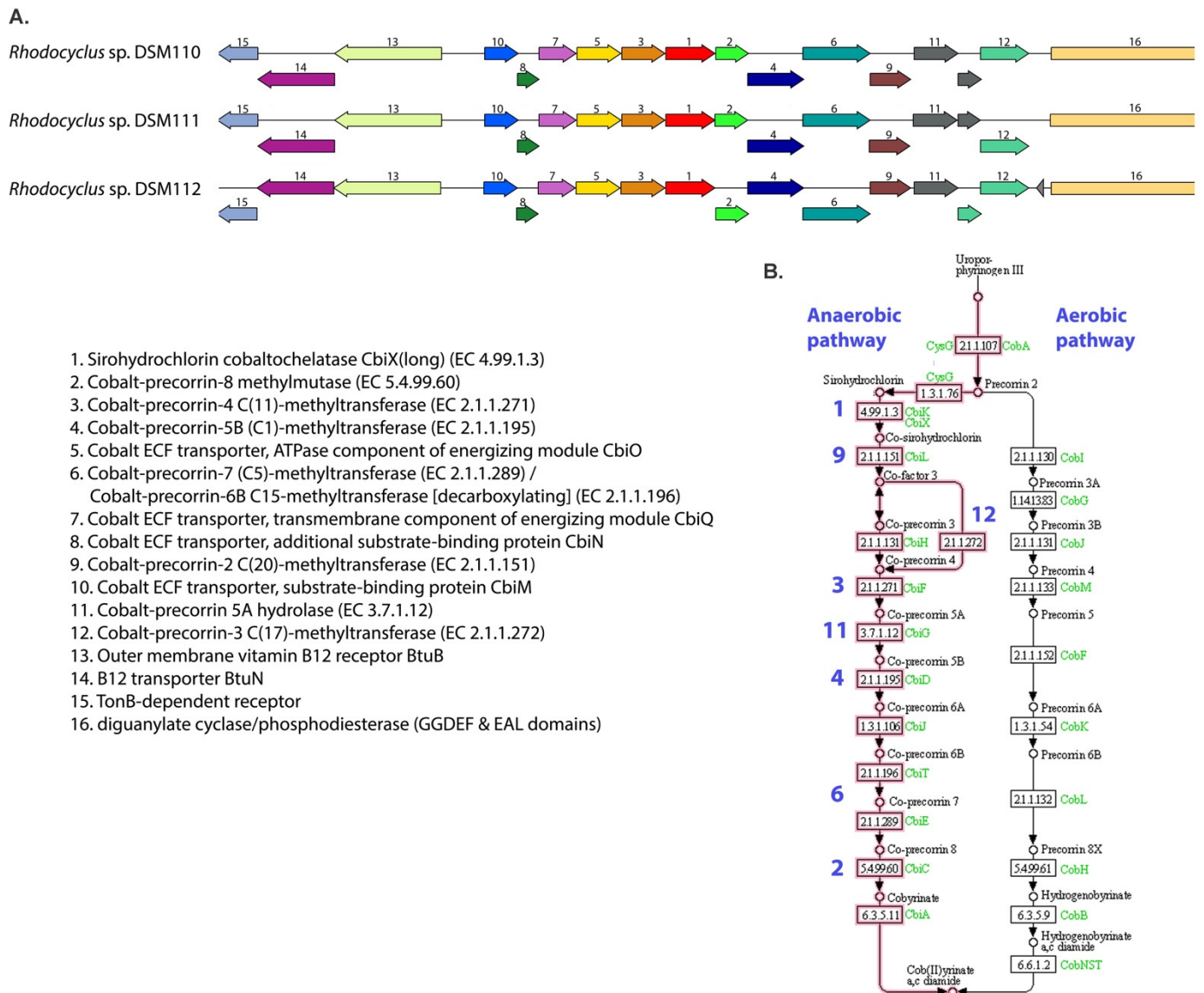rently have the corresponding denitrifying genes for nitrite, nitric, or nitrous oxide reductases, and NirB apparently assumes a different role. NirB was originally discovered in *Pseudomonas stutzeri* as part of the denitrification pathway, in which it forms a polycistronic mRNA along with the nitrite reductase, NirS (cytochrome cd$_1$), and the membrane-bound tetraheme cytochrome, NirT [45]. The implication is that NirT donates electrons to NirB, which in turn reacts with NirS to reduce nitrite to nitric oxide. The normal electron donor to NirS in other pseudomonads is NirM, or C8 as it is also called, and it may be involved in *Ps. stutzeri* as well with NirB enhancing the interaction. Perhaps the role of NirB in *Rhodocyclus* species is to facilitate the interaction of C8 with the cytochrome bc_1 complex and PufLMC. This deserves further study.

Consistent with the earlier observations that *Rcy. purpureus* is incapable of fixing molecular nitrogen, while *Rcy. tenuis* DSM 109^T showed nitrogenase activity [46], we identified a total of 16 nitrogenase-related PGFams that are absent in *Rcy. purpureus* but present in all of the other *Rhodocyclus* genomes. These include the (Fe-Fe) nitrogenase (alpha, beta, and delta chains), (Mo-Fe) nitrogenase (alpha and beta chains), nitrogen reductase and maturation proteins, two 4Fe-4S nitrogenase-associated ferredoxins, a nitrogenase transcriptional regulator, and several Fe-Mo cofactor assembly proteins.

3.4. Cobalamin Metabolism

When comparing the different genomes, it was found that the 3 *Rhodocylus* strains DSM 110, DSM 111, and DSM 112 all contain at least 12 unique genes related to anaerobic cobalamin (vitamin B$_{12}$) synthesis, which are all missing from the other 3 genomes. These genes code for cobalt–corrin metabolic enzymes and cobalt transporter subunits and are organized in a large gene cluster (Figure 5). The gene synteny of the cluster is conserved in all three strains. It has been known for decades that two pathways exist in nature for the de novo biosynthesis of vitamin B$_{12}$. The pathways differ in the first parts, which involves the corrin synthesis, in which one pathway is anaerobic (as found in *Salmonella typhimurium* and *Bacillus megaterium*) and the other is oxygen-dependent [47–50]. Figure 5 includes the KEGG pathways and shows that all the gene products necessary for anaerobic corrin synthesis (cobalt-containing modified tetrapyrrole component of vit. B$_{12}$) are present in the *Rhodocyclus* gene cluster. The ATP-dependent transport system encoded by the corrin biosynthetic operon in *S. typhimurium* (CbiMNQO), mediates transport of cobalt ions for the B$_{12}$ synthesis [47]. In addition, vitamin B$_{12}$ and other corrinoids are actively transported using the TonB-dependent outer membrane receptor BtuB in complex with the

ABC transport system BtuFCD [51]. The *Rhodocyclus* gene cluster also contains a unique outer membrane vitamin B$_{12}$ receptor BtuB and transporter BtuN (Figure 5).

1. Sirohydrochlorin cobaltochelatase CbiX(long) (EC 4.99.1.3)
2. Cobalt-precorrin-8 methylmutase (EC 5.4.99.60)
3. Cobalt-precorrin-4 C(11)-methyltransferase (EC 2.1.1.271)
4. Cobalt-precorrin-5B (C1)-methyltransferase (EC 2.1.1.195)
5. Cobalt ECF transporter, ATPase component of energizing module CbiO
6. Cobalt-precorrin-7 (C5)-methyltransferase (EC 2.1.1.289) /
   Cobalt-precorrin-6B C15-methyltransferase [decarboxylating] (EC 2.1.1.196)
7. Cobalt ECF transporter, transmembrane component of energizing module CbiQ
8. Cobalt ECF transporter, additional substrate-binding protein CbiN
9. Cobalt-precorrin-2 C(20)-methyltransferase (EC 2.1.1.151)
10. Cobalt ECF transporter, substrate-binding protein CbiM
11. Cobalt-precorrin 5A hydrolase (EC 3.7.1.12)
12. Cobalt-precorrin-3 C(17)-methyltransferase (EC 2.1.1.272)
13. Outer membrane vitamin B12 receptor BtuB
14. B12 transporter BtuN
15. TonB-dependent receptor
16. diguanylate cyclase/phosphodiesterase (GGDEF & EAL domains)

Figure 5. (**A**). Overview of the anaerobic cobalamin genetic pathway that was found in *Rhodocyclus* strains DSM 110, DSM 111, and DSM 112. Synteny plots were generated in PATRIC [22] and genes are colored based on enzymatic family. (**B**). Overview of the anaerobic and aerobic cobalamin metabolic pathways. Enzyme numbering is the same as corresponding gene numbers in (**A**).

In addition to the PATRIC PGFam comparison, we also checked for the presence of each of these cobalamin metabolic genes individually with BLAST and only found a truncated version of one of the enzymes (gene 6 in Figure 5) in the *Rcy. purpureus* and *Rcy. tenuis* DSM 109^T and IM230 genomes. This could be an indication that the anaerobic cobalamin biosynthetic pathway was lost during evolution, although further analysis that includes more evolutionary divergent species would be needed to confirm this.

The lack of this anaerobic cobalamin metabolism pathway explains why Pfennig originally described *Rcy. purpureus* as a vitamin B$_{12}$-requiring member of the *Rhodospirillaceae* [9]. According to this genome comparison, strains DSM 109^T and IM230 would also need cobalamin as a growth factor as they grow anaerobically; however, the need for this has not been described. Either way, the presence of this pathway exclusively in the three

Rcy. strains DSM 110, DSM 111, and DSM 112 further distinguishes them genetically from the other *Rhodocyclus* species.

3.5. Chemotaxis and Motility

Rcy. purpureus was found to be non-motile [9], while *Rcy. tenuis* was highly motile [10]. We found 31 flagella related PGFams that were absent from the *Rcy. purpureus* genome but present in all the other *Rhodocyclus* strains. These included all of the biosynthetic, structural and regulatory proteins for flagella assembly. In addition, there are at least 12 unique chemotaxis related PGFams present, including, 2 CheWs, CheA, CheB, CheR, CheZ, CheY, CheD, CheV and MCPs. This chemotaxis gene cluster is directly upstream of the flagella genes in all of these genomes. The presence of these genes confirms that all of the *Rhodocyclus* species, expect *Rcy. purpureus*, are motile, and are expected to perform chemotaxis.

3.6. RuBisCo

Genomes of *Rcy. tenuis* and the new species *Rhodocyclus gracilis* have two forms of RuBisCo. One gene encodes a single 463 aa. protein (PGFam_00048972) and is included in a gene cluster with *cbb*Y-*rbc*-*cbb*R-fructose-1,6-bisphosphatase- phosphoribulokinase. This cluster is also present in *Rcy. purpureus*. The second system of RuBisCo consists of a large (474 aa.; PGFam_00048973) and small subunit (118 aa.; PGFam_00048975), included in a gene cluster of *cbb*R-*rbc*L-*rbc*S-*cbb*Q-*cbb*O, which is absent from *Rcy. purpureus* DSM 168^T. The presence of multiple RuBisCo forms is not uncommon, and the sequences and structural gene organization indicate that the *Rhodocyclus* single rubisco gene belongs to Form II, while the rbcL and rbcS resemble Form I RuBisCo found in other proteobacteria [52].

The translated protein sequence of the single RuBisCo gene (PGFam_00048972) that is present in all *Rhodocyclus* genomes, was used for a phylogenetic comparison (Figure 6). The RuBisCo protein sequence of *Rcy. gracilis* strains is closer related to *Rcy. purpureus* DSM 168^T, with 94.1% identity (98.3% similarity) than to the *Rcy. tenuis* DSM 109^T with 91.6% identity (97.8% similarity). As expected, the *Rcy. tenuis* IM 230 RuBisCo shows a high protein identity to the one from *Rcy. tenuis* DSM 109^T (99.8% identity, 100% similarity), while the *Rcy. tenuis* DSM 109^T and *Rcy. purpureus* sequences were less similar (90.8% identity, 96.3% similarity). The phylogenetic tree based on the RuBisCo multiple protein sequence alignment (Figure 6) shows a similar topology as the whole genome, 16SrRNA and cytochrome-based comparisons, and further supports the proposed species differentiation.
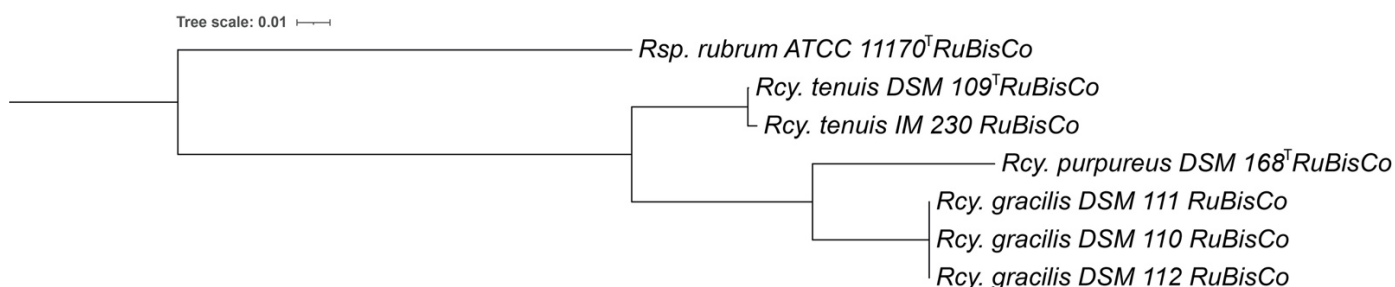
Figure 6. Phylogenetic tree based on multiple sequence alignment of the RuBisCo protein sequences, obtained from the *Rhodocyclus* genomes (PGFam_00048972). *Rhodospirillum rubrum* ATCC 11170^T RuBisCo was added as an outgroup.

3.7. Taxonomic Considerations

A limited number of strains of *Rhodocyclus* isolates have been studied and compared in several studies during the past decades. Strains were isolated from various freshwater lakes in Germany by Hanno Biebl and Norbert Pfennig and their properties compared [12]. Although two color variants were recognized, a brownish-red colored one, including the type strain of *Rcy. tenuis* DSM 109^T, and a red-colored or purple-violet one, including

strains now assigned to *Rcy. gracilis*, no clear differentiation of groups of strains was made [12].

The different colored cultures showed differences in absorption spectra in the carotenoid region (450–550 nm) [12] and also revealed a different carotenoid composition [13]. The type strain of *Rcy. tenuis* DSM 109[T] has carotenoids of the spirilloxanthin series with major portions of lycopene, rhodopin and anhydro-rhodovibrin and spirilloxanthin as the final product of this pathway. The purple-violet-colored DSM 110, DSM 111, and DSM 112 contain significant amounts of rhodopinal, rhodopinol, and lycopenal in addition to rhodopin and lycopene but lack spirilloxanthin, rhodovibrin, and anhydro-rhodovibrin [13], a property they share with *Rcy. purpureus* DSM 168[T] [13].

Studies on the lipopolysaccharides of eight isolates assigned to *Rhodocyclus tenuis* revealed a common pattern of sugars characterized by the presence of glycerol-mannoheptose, glucose, arabinose, 2-keto-3deoxyoctonoate and glucosamine in all studied strains [14]. While the presence of D-galactosamine was found only in those strains assigned now to *Rcy. gracilis* (DSM 110, DSM 111, DSM 112), the type strain of *Rcy. tenuis* DSM 109[T] (2761) lacked D-galactosamine and in turn had quinovosamine as a strain-specific sugar [14].

Based on the unique genomic and genetic features of the *Rhodocyclus* strains described above, it is clear that strains DSM 110, DSM 111, and DSM 112 belong into a single species which is separate from *Rcy. tenuis* and *Rcy. purpureus*. Of the strains compared in the present study, only strains DSM 109[T] and IM 203 remain as strains of *Rcy. tenuis*. For the strains recognized as a new species, the name *Rhodocyclus gracilis* sp. nov. is proposed.

The characteristic properties that distinguish strains of *Rcy. gracilis* from *Rcy. tenuis* are the utilization of ethanol [12] and the presence of carotenoids of the rhodopinal series [13], which coincides with the purple-violet color of *Rcy. gracilis*. The pH optimum of *Rcy. gracilis* is slightly lower, at pH 6.1–6.4, compared to the pH 6.7 of *Rcy. tenuis* [12]. The genomes are different in size, 2.93–2.98 Mb for *Rcy. gracilis* and 3.65–3.85 Mb for *Rcy. tenuis* (Table 1), and the G + C content is 64.5 mol% in *Rcy. gracilis* and 64.7 mol% in *Rcy. tenuis* (Table 1).

Due to the lack of a clear differentiation of groups of strains of *Rcy. tenuis* in previous studies, the properties of both *Rcy. gracilis* and *Rcy. tenuis* have been listed as the properties of *Rcy. tenuis* in the literature [12–14,53]. Consequently, the species description of *Rcy. tenuis* should be emended accordingly.

Characteristic for all *Rhodocyclus* species is the presence of phosphatidyl glycerol (PG), phosphatidyl ethanolamine (PE), diphosphatidyl glycerol (CL), and an ornithine lipid as major polar lipids; the dominance of C-16 fatty acids (33–36% C-16:0 and 43–50% C-16:1) and minor amounts of the C-18 fatty acids (<0.5% C-18:0 and 14–18% C-18:1); as well as ubiquinone Q-8 and menaquinone MK-8 as major quinone components [16].

3.7.1. Description of *Rhodocyclus gracilis* sp. nov

Rhodocyclus gracilis. gra'ci.lis M.L. neut. adj. *gracilis* Slender

Cells are weakly curved, 0.3–0.5 µm wide and 1.5–5 µm long, motile by polar flagella and divide by binary fission. Cultures grown anaerobically in the dark are red to purple-violet in color and have absorption maxima at 377–378, 469, 495–500, 529–533, 590–592, 798–801, and 856–858 nm. Photosynthetic pigments are bacteriochlorophyll-a and carotenoids of the rhodopinal series. Internal photosynthetic membranes are present as small finger-like intrusions of the cytoplasmic membrane.

Growth occurs preferably under phototrophic conditions anaerobically in the light. Under these conditions, organic carbon compounds are used as carbon and energy sources. The sources utilized are acetate, propionate, butyrate, valerate, caproate, lactate, pyruvate, fumarate, malate, succinate, and ethanol. Some strains may use pelargonate and yeast extract. Not utilized are tartrate, citrate, benzoate, methanol, glycerol, glucose, fructose, mannitol, alanine, glutamate, aspartate, arginine, thiosulfate, and sulfide. Sulfide is growth inhibitory at 2 mM. Chemotrophic growth under aerobic dark conditions is possible. Aerobically grown cells are colorless, and the aerobic Mg-protoporphyrin IX monomethyl

ester oxidative cyclase is absent. Photolithotrophic growth with hydrogen as an electron source may be possible. Ammonium chloride and dinitrogen are used as nitrogen sources. Growth factors may be required. Mesophilic freshwater bacterium with optimum growth at 30 °C and pH 6.1–6.4 (pH range 4.9–8.2). Habitats are freshwater lakes and peat bogs. The habitat of the type strain DSM 110 is a dystrophic pond in the Black Forest (Germany).

The type strain has a G + C content of the DNA of 64.5 mol% (genome analysis) and a genome size of 2.93 Mb. The type strain is deposited with the Deutsche Sammlung von Mikroorganismen und Zellkulturen as DSM 110^T (Pfennig 3760) and the Japan Collection of Microorganisms Riken BRC.

Gene bank accession number of the 16S rDNA sequence of the type strain OM179767 and of the genome WIXJ00000000.

3.7.2. Emended Description of *Rhodocyclus tenuis* Imhoff, Trüper and Pfennig 1984, 341.^{VP} (*Rhodospirillum tenue* Pfennig 1969, 619.^{AL})

te'nu.is. L. masc. adj. *tenuis* Slender, Thin

Cells are weakly curved spirals, highly motile by polar flagella. They are 0.3–0.5 μm wide and 1.5–6.0 μm long, sometimes even longer. One complete turn of a spiral is about 0.8–1.0 μm wide and 3 μm long. Photosynthetically grown cells are brownish-red and have absorption maxima at 378–380, 465, 492–495, 528, 592–594, 799–801, and 868–871 nm. Photosynthetic pigments are bacteriochlorophyll-a esterified with phytol and carotenoids of the spirilloxanthin series.

Growth occurs preferably under anoxic conditions in the light with organic carbon compounds as carbon and electron sources. Photolithotrophic growth with molecular hydrogen is possible. Chemotrophic growth is possible under microoxic to oxic conditions in the dark. Aerobically grown cells are colorless or pale red. Under phototrophic growth conditions organic carbon compounds are used as carbon and energy sources. The sources utilized are acetate, butyrate, valerate, caproate, lactate, pyruvate, fumarate, malate, and succinate. Pelargonate and propionate may be used by some strains. Not utilized are formate, ethanol, tartrate, citrate, benzoate, cyclohexane carboxylate, methanol, glycerol, glucose, fructose, mannitol, alanine, glutamate, aspartate, arginine, thiosulfate, and sulfide. Sulfide is growth-inhibitory at 2 mM. The nitrogen sources utilized are aspartate, glutamate, glutamine, ammonia, and dinitrogen and also casamino acids, peptone, yeast extract, alanine, arginine, lysine, methionine, serine, threonine, and urea. Sulfate, glutathione, cysteine, thiosulfate, and also sulfite and sulfide at low concentrations can serve as assimilatory sulfur sources. Growth factors are not required. Growth is stimulated, however, in the presence of complex organic nutrients or yeast extract and some strains may need vitamin B$_{12}$.

Mesophilic freshwater bacterium with optimum growth at 30 °C and pH 6.7–7.4.

Habitat: freshwater ponds, sewage ditches.

The type strain has a G + C of the DNA of 64.7 mol% (genome analysis) and a genome size of 3.85 Mb.

Type strain: ATCC 25093, DSM 109 (Pfennig: 2761, Grünenplan).

Gene bank accession number of the 16S rDNA sequence of the type strain: D16208.

Gene bank accession number of the genome sequence of the type strain: SSSP00000000.

Informed Consent Statement: Not applicable.

Data Availability Statement: This Whole Genome Shotgun project has been deposited at DDBJ/ENA/ GenBank under the accession numbers provided in Table 1.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hu, P.; Lang, J.; Wawrousek, K.; Yu, J.; Maness, P.-C.; Chen, J. Draft Genome Sequence of *Rubrivivax gelatinosus* CBS. *J. Bacteriol.* **2012**, *194*, 3262. [CrossRef] [PubMed]
2. Nagashima, S.; Kamimura, A.; Shimizu, T.; Nakamura-Isaki, S.; Aono, E.; Sakamoto, K.; Ichikawa, N.; Nakazawa, H.; Sekine, M.; Yamazaki, S.; et al. Complete genome sequence of phototrophic betaproteobacterium *Rubrivivax gelatinosus* IL144. *J. Bacteriol.* **2012**, *194*, 3541–3542. [CrossRef] [PubMed]
3. Mohammed, M.; Isukapatla, A.; Mekala, L.P.; Eedara Veera Venkata, R.P.; Chintalapati, S.; Chintalapati, V.R. Genome sequence of the phototrophic betaproteobacterium *Rubrivivax benzoatilyticus* strain JA2T. *J. Bacteriol.* **2011**, *193*, 2898–2899. [CrossRef] [PubMed]
4. Kim, Y.J.; Farh, M.E.A.; Yang, D. Genome sequencing of *Rhodoferax fermentans* JCM 7819. Genbank. Unpublished Work. 2017.
5. Baker, J.M.; Riester, C.J.; Skinner, B.M.; Newell, A.W.; Swingley, W.D.; Madigan, M.T.; Jung, D.; Asao, M.; Chen, M.; Loughlin, P.C.; et al. Genome Sequence of *Rhodoferax antarcticus* ANT.BRT; A Psychrophilic Purple Nonsulfur Bacterium from an Antarctic Microbial Mat. *Microorganisms* **2017**, *5*, 8. [CrossRef] [PubMed]
6. Imhoff, J.F.; Meyer, T.E.; Kyndt, J.A. The genome sequence of the giant phototrophic gammaproteobacterium *Thiospirillum jenense* gives insight into its physiological properties and phylogenetic relationships. *Arch. Microbiol.* **2021**, *203*, 97–105. [CrossRef] [PubMed]
7. Lee, S.H.; Choe, H.; Kim, S.G.; Park, D.S.; Nasir, A.; Kim, B.K.; Kim, K.M. Complete genome of biodegradable plastics-decomposing *Roseateles depolymerans* KCTC 42856(T) (=61A(T)). *J. Biotechnol.* **2016**, *220*, 47–48. [CrossRef] [PubMed]
8. Wang, Z.; Li, W.; Li, H.; Zheng, W.; Guo, F. Phylogenomics of Rhodocyclales and its distribution in wastewater treatment systems. *Sci. Rep.* **2020**, *10*, 3883. [CrossRef] [PubMed]
9. Pfennig, N. *Rhodocyclus purpureus* gen. nov. and sp. nov., a Ring-Shaped, Vitamin B$_{12}$-Requiring Member of the Family *Rhodospirillaceae*. *Int. J. Syst. Bacteriol.* **1978**, *28*, 283–288. [CrossRef]
10. Pfennig, N. *Rhodospirillum tenue* sp. n., a New Species of the Purple Nonsulfur Bacteria. *J. Bacteriol.* **1969**, *99*, 619–620. [CrossRef] [PubMed]
11. Imhoff, J.F.; Truper, H.G.; Pfennig, N. Rearrangement of the Species and Genera of the Phototrophic "Purple Nonsulfur Bacteria". *Int. J. Syst. Bacteriol.* **1984**, *34*, 340–343. [CrossRef]
12. Biebl, H. Ökologie einiger Seen in Norddeutschland: Phototrophe Nichtschwefel-Purpurbakterien. Ph.D. Thesis, University Freiburg, Freiburg im Breisgau, Germany, 1973.
13. Schmidt, K. Biosynthesis of Carotenoids. In *The Photosynthetic Bacteria*; Clayton, R.K., Sistrom, W.R., Eds.; Plenum Press: New York, NY, USA, 1978; Chapter 39; pp. 729–750.
14. Drews, G.; Weckesser, J.; Mayer, H. Cell Envelopes. In *The Photosynthetic Bacteria*; Clayton, R.K., Sistrom, W.R., Eds.; Plenum Press: New York, NY, USA, 1978; Chapter 4; pp. 61–77.
15. Imhoff, J.F. Occurrence and evolutionary significance of two sulfate assimilation pathways in the *Rhodospirillaceae*. *Arch. Microbiol.* **1982**, *132*, 197–203. [CrossRef]
16. Imhoff, J.F.; Bias-Imhoff, U. Lipids, Quinones and Fatty Acids of Anoxygenic Phototrophic Bacteria. In *Anoxygenic Photosyn-thetic Bacteria*; Blankenship, R.E., Madigan, M.T., Bauer, C.E., Eds.; Kluwer Academic Publ.: Alphen aan den Rijn, The Netherlands, 1995; pp. 179–205.
17. Imhoff, J.F. Anoxygenic Phototrophic Bacteria. In *Methods in Aquatic Bacteriology*; Austin, B., Ed.; John Wiley and Sons Ltd.: Hoboken, NJ, USA, 1988; Chapter 9; pp. 207–240.
18. Imhoff, J.F.; Rahn, T.; Künzel, S.; Neulinger, S.C. New insights into the metabolic potential of the phototrophic purple bacterium *Rhodopila globiformis* DSM 161T from its draft genome sequence and evidence for a vanadium-dependent nitrogenase. *Arch. Microbiol.* **2018**, *200*, 847–857. [CrossRef] [PubMed]
19. Nurk, S.; Bankevich, A.; Antipov, D.; Gurevich, A.A.; Korobeynikov, A.; Lapidus, A.; Prjibelski, A.D.; Pyshkin, A.; Sirotkin, A.; Sirotkin, Y.; et al. Assembling Single-Cell Genomes and Mini-Metagenomes from Chimeric MDA Products. *J. Comput. Biol.* **2013**, *20*, 714–737. [CrossRef] [PubMed]
20. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comp. Biol.* **2012**, *19*, 455–477. [CrossRef] [PubMed]
21. Wick, R.R.; Judd, L.M.; Gorrie, C.L.; Holt, K.E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* **2017**, *13*, e1005595. [CrossRef] [PubMed]
22. Wattam, A.R.; Davis, J.J.; Assaf, R.; Boisvert, S.; Brettin, T.; Bun, C.; Conrad, N.; Dietrich, E.M.; Disz, T.; Gabbard, J.L.; et al. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Res.* **2017**, *45*, D535–D542. [CrossRef] [PubMed]

23. Aziz, R.K.; Bartels, D.; Best, A.A.; DeJongh, M.; Disz, T.; Edwards, R.A.; Formsma, K.; Gerdes, S.; Glass, E.M.; Kubal, M.; et al. The RAST server: Rapid annotations using subsystems technology. *BMC Genom.* **2008**, *9*, 75. [CrossRef] [PubMed]

24. Parks, D.H.; Imelfort, M.; Skennerton, C.T.; Hugenholtz, P.; Tyson, G.W. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **2015**, *25*, 1043–1055. [CrossRef] [PubMed]

25. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]

26. Morgulis, A.; Gertz, E.M.; Schäffer, A.A.; Agarwala, R. A Fast and Symmetric DUST Implementation to Mask Low-Complexity DNA Sequences. *J. Comput. Biol.* **2006**, *13*, 1028–1040. [CrossRef] [PubMed]

27. Boetzer, M.; Henkel, C.V.; Jansen, H.J.; Butler, D.; Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **2011**, *27*, 578–579. [CrossRef] [PubMed]

28. Kang, D.D.; Froula, J.; Egan, R.; Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **2015**, *3*, e1165. [CrossRef] [PubMed]

29. Bushnell, B. BBMap: A Fast, Accurate, Splice-Aware Aligner. 2014. United States. Available online: https://www.osti.gov/servlets/purl/1241166 (accessed on 29 May 2017).

30. Richter, M.; Rosselló-Móra, R.; Oliver Glöckner, F.O.; Peplies, J. JSpeciesWS: A web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* **2016**, *32*, 929–931. [CrossRef] [PubMed]

31. Stamatakis, A.; Hoover, P.; Rougemont, J. A Rapid Bootstrap Algorithm for the RAxML Web Servers. *Syst. Biol.* **2008**, *57*, 758–771. [CrossRef] [PubMed]

32. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [CrossRef] [PubMed]

33. Letunic, I.; Bork, P. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **2019**, *47*, W256–W259. [CrossRef] [PubMed]

34. Meier-Kolthoff, J.P.; Göker, M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat. Commun.* **2019**, *10*, 2182. [CrossRef] [PubMed]

35. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539. [CrossRef] [PubMed]

36. Saitou, N.; Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **1987**, *4*, 406–425. [CrossRef] [PubMed]

37. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.A.; Clamp, M.; Barton, G.J. Jalview Version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009**, *25*, 1189–1191. [CrossRef]

38. Skennerton, C.T.; Barr, J.; Slater, F.R.; Bond, P.; Tyson, G.W. Expanding our view of genomic diversity in Candidatus *Accumulibacter clades*. *Environ. Microbiol.* **2015**, *17*, 1574–1585. [CrossRef] [PubMed]

39. Stackebrandt, E.; Ebers, J. Taxonomic parameters revisited: Tarnished gold standards. *Microbiol. Today* **2006**, *33*, 152–155.

40. Menin, L.; Yoshida, M.; Jaquinod, M.; Nagashima, K.V.P.; Matsuura, K.; Parot, P.; Verméglio, A. Dark Aerobic Growth Conditions Induce the Synthesis of a High Midpoint Potential Cytochrome c8 in the Photosynthetic Bacterium *Rubrivivax gelatinosus*. *Biochemistry* **1999**, *38*, 15238–15244. [CrossRef] [PubMed]

41. Hochkoeppler, A.; Zannoni, D.; Ciurli, S.; Meyer, T.E.; Cusanovich, M.A.; Tollin, G. Kinetics of photo-induced electron transfer from high-potential iron-sulfur protein to the photosynthetic reaction center of the purple phototroph *Rhodoferax fermentans*. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 6998–7002. [CrossRef] [PubMed]

42. Menin, L.; Schoepp, B.; Parot, P.; Verméglio, A. Photoinduced Cyclic Electron Transfer in *Rhodocyclus tenuis* Cells: Participation of HiPIP or Cyt c8 Depending on the Ambient Redox Potential. *Biochemistry* **1997**, *36*, 12183–12188. [CrossRef] [PubMed]

43. Meyer, T.E.; Kamen, M.D. Structural and Functional Diversity among Bacterial Electron Transport Protein. In *Electron Transport and Oxygen Utilization*; Ho, C., Ed.; Elsevier North Holland: New York, NY, USA, 1982; pp. 33–41.

44. Devreese, B.; Brigé, A.; Backers, K.; Van Driessche, G.; Meyer, T.E.; Cusanovich, M.A.; Van Beeumen, J.J. Primary Structure Characterization of a *Rhodocyclus tenuis* Diheme Cytochrome c Reveals the Existence of Two Different Classes of Low-Potential Diheme Cytochromes c in Purple Phototropic Bacteria. *Arch. Biochem. Biophys.* **2000**, *381*, 53–60. [CrossRef] [PubMed]

45. Hartig, E.; Zumft, W.G. Kinetics of nirS expression (cytochrome cd1 nitrite reductase) in *Pseudomonas stutzeri* during the tran-sition from aerobic respiration to denitrification: Evidence for a denitrification-specific nitrate- and nitrite-responsive regulatory system. *J. Bacteriol.* **1999**, *181*, 161–166. [CrossRef] [PubMed]

46. Masters, R.A.; Madigan, M. Nitrogen metabolism in the phototrophic bacteria *Rhodocyclus purpureus* and *Rhodospirillum tenue*. *J. Bacteriol.* **1983**, *155*, 222–227. [CrossRef] [PubMed]

47. Roth, J.R.; Lawrence, J.G.; Rubenfield, M.; Kieffer-Higgins, S.; Church, G. Characterization of the cobalamin (vitamin B$_{12}$) biosynthetic genes of *Salmonella typhimurium*. *J. Bacteriol.* **1993**, *175*, 3303–3316. [CrossRef] [PubMed]

48. Blanche, F.; Thibaut, D.; Debussche, L.; Hertle, R.; Zipfel, F.; Müller, G. Parallels and Decisive Differences in Vitamin B$_{12}$ Biosyntheses. *Angew. Chem. Int. Ed.* **1993**, *32*, 1651–1653. [CrossRef]

49. Raux, E.; Lanois, A.; Rambach, A.; Warren, M.; Thermes, C. Cobalamin (vitamin B$_{12}$) biosynthesis: Functional characterization of the *Bacillus megaterium cbi* genes required to convert uroporphyrinogen III into cobyrinic acid a,c-diamide. *Biochem. J.* **1998**, *335*, 167–173. [CrossRef] [PubMed]

50. Moore, S.J.; Lawrence, A.D.; Biedendieck, R.; Deery, E.; Frank, S.; Howard, M.J.; Rigby, S.E.J.; Warren, M.J. Elucidation of the anaerobic pathway for the corrin component of cobalamin (vitamin B$_{12}$). *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 14906–14911. [CrossRef] [PubMed]

51. Cadieux, N.; Bradbeer, C.; Reeger-Schneider, E.; Köster, W.; Mohanty, A.K.; Wiener, M.C.; Kadner, R.J. Identification of the Periplasmic Cobalamin-Binding Protein BtuF of *Escherichia coli*. *J. Bacteriol.* **2002**, *184*, 706–717. [CrossRef] [PubMed]

52. Badger, M.R.; Bek, E.J. Multiple Rubisco forms in proteobacteria: Their functional significance in relation to CO2 acquisition by the CBB cycle. *J. Exp. Bot.* **2008**, *59*, 1525–1541. [CrossRef] [PubMed]

53. Imhoff, J.F. Genus *Rhodocyclus*. In *Bergey's Manual of Systematic Bacteriology*, 2nd ed.; Brenner, D.J., Krieg, N.R., Staley, J.T., Eds.; Springer: New York, NY, USA, 2005; Volume 2, Part C; pp. 887–890.

Article

Phylogenomic Analyses and Molecular Signatures Elucidating the Evolutionary Relationships amongst the *Chlorobia* and *Ignavibacteria* Species: Robust Demarcation of Two Family-Level Clades within the Order *Chlorobiales* and Proposal for the Family *Chloroherpetonaceae* fam. nov

Sarah Bello [1], Mohammad Howard-Azzeh [2], Herb E. Schellhorn [3] and Radhey S. Gupta [1,*]

1 Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, ON L8N 3Z5, Canada; bellos1@mcmaster.ca
2 Department of Population Medicine, Ontario Veterinary College, University of Guelph, Guelph, ON N1G 2W1, Canada; mhowarda@uoguelph.ca
3 Department of Biology, McMaster University, Hamilton, ON L8S 4K1, Canada; schell@mcmaster.ca
* Correspondence: gupta@mcmaster.ca

Abstract: Evolutionary relationships amongst *Chlorobia* and *Ignavibacteria* species/strains were examined using phylogenomic and comparative analyses of genome sequences. In a phylogenomic tree based on 282 conserved proteins, the named *Chlorobia* species formed a monophyletic clade containing two distinct subclades. One clade, encompassing the genera *Chlorobaculum, Chlorobium, Pelodictyon,* and *Prosthecochloris,* corresponds to the family *Chlorobiaceae,* whereas another clade, harboring *Chloroherpeton thalassium,* Candidatus Thermochlorobacter aerophilum, Candidatus Thermochlorobacteriaceae bacterium GBChlB, and *Chlorobium* sp. 445, is now proposed as a new family (*Chloroherpetonaceae* fam. nov). In parallel, our comparative genomic analyses have identified 47 conserved signature indels (CSIs) in diverse proteins that are exclusively present in members of the class *Chlorobia* or its two families, providing reliable means for identification. Two known *Ignavibacteria* species in our phylogenomic tree are found to group within a larger clade containing several *Candidatus* species and uncultured *Chlorobi* strains. A CSI in the SecY protein is uniquely shared by the species/strains from this "larger *Ignavibacteria* clade". Two additional CSIs, which are commonly shared by *Chlorobia* species and the "larger *Ignavibacteria* clade", support a specific relationship between these two groups. The newly identified molecular markers provide novel tools for genetic and biochemical studies and identification of these organisms.

Keywords: phylogenomic and comparative genomic analyses; conserved signature indels (CSIs); molecular signatures; class *Chlorobia* and the families *Chlorobiaceae* and *Chloroherpetonaceae*; *Ignavibacteria*; uncultured species/strains related to *Chlorobia/Ignavibacteria*

1. Introduction

Members of the class *Chlorobia*, also known as green sulfur bacteria, were, until recently, comprised of a group of strictly anaerobic, photosynthetic bacteria [1] that occupy a very narrow environmental niche consisting of anoxic aquatic settings in stratified lakes, where sunlight can penetrate [1–4]. These bacteria use reduced sulfur compounds as reductants and electron source to fix CO_2 using the reverse TCA cycle [4–6]. Based on their branching in phylogenetic trees for the 16S rRNA gene and FMO protein sequences, these bacteria have been placed in a separate phylum *Chlorobiota*, consisting of a single class "*Chlorobia*" containing a single order *Chlorobiales* and a single family *Chlorobiaceae* [3–5,7–10]. According to the List of Prokaryotic Names with Standing in Nomenclature (LPSN) server [11], the

family *Chlorobiaceae* contains six genera (*Ancalochloris, Chlorobaculum, Chlorobium, Chloroherpeton, Pelodictyon,* and *Prosthecochloris*) [4,5,12,13]. However, there is no information available for the sole species *Ancalochloris perfilievii* from the genus *Ancalochloris* [14] and most of the species from the family *Pelodictyon* have been transferred into the genus *Chlorobium* [12,15]. According to the LPSN server, the family *Chlorobiaceae* now contains 15 species (excluding synonyms and non-validly published species) [11]. However, most of these species were described prior to 2008 [11] and since then, except for a few *Candidatus* species [16,17], no new *Chlorobiaceae* species have been identified. Until 2013, members of the phylum *Bacteroidetes* (recently renamed *Bacteroidota*) [18] were the closest relatives of the *Chlorobia* species [3,5,12,19,20] and together with the *Fibrobacteres* they formed the *Fibrobacteres–Bacteroidetes–Chlorobia* (FBC) superphylum [20,21]. However, subsequent studies have led to the identification of two new species, viz., *Melioribacter roseus* [22,23] and *Ignavibacterium album* [7], which, based on phylogenetic trees for several gene/protein sequences, are considered the closest relatives of *Chlorobia* species [7,23,24]. Unlike the *Chlorobia* species, all of which are strictly anaerobic and obligate photolithotrophs [3–6], *Ignavibacterium* as well as *Melioribacter* are moderately thermophilic, facultatively anaerobic, and obligate chemoorganotrophs [7,23–26]. In earlier phylogenetic trees based on the 16S rRNA gene and other genes/proteins, a clade comprising *I. album* and *M. roseus* was separated from *Chlorobia* species by a long branch [7,23]. This led to the placement of these two species into a new phylum, recently renamed *Ignavibacteriota* [18], containing a single class *Ignavibacteria* and a single order *Ignavibacteriales* [23,24], which is a sister taxon of the *Chlorobia* (*Chlorobiota*) [23,24]. In addition to these two *Ignavibacteria* species, analyses of metagenomic sequences from several thermophilic habitats have led to the reconstruction of genomes for several uncultivated organisms that are related to *Chlorobia–Ignavibacteria* species [25–27]. Based on their genome sequences, some of these organisms are hypothesized to have biochemical–physiological characteristics overlapping with the properties of either members of the class *Chlorobia* or members of the class *Ignavibacteria* (i.e., aerobic photoheterotrophs possessing genes for bacteriochlorophyll biosynthesis and related proteins) [25–27].

Due to rapid advances in genome sequencing technology, and because of several major genomic sequence projects [28–30], whole genome sequences are now available for most of the species from the *Chlorobia/Ignavibacteria* groups. In addition, sequence information is also available for several unnamed *Chlorobia* species in the NCBI database (https://www.ncbi.nlm.nih.gov/genome/ (accessed on 15 January 2022) [31]. The available genomes provide a comprehensive resource for more reliably understanding the evolutionary relationships amongst these organisms based on multiple independent genomic approaches. Based on genome sequences, robust phylogenetic trees can be constructed based on a large dataset of genes/proteins, exhibiting a high degree of statistical support at different taxonomic levels [32–35]. Indeed, phylogenetic analyses based on 120 proteins that are ubiquitously found in different micro-organisms have led to the creation of a Genome Taxonomy database (GTDB), which has become an important reference resource for the classification of prokaryotic organisms [32,36]. In the GTDB taxonomy (https://gtdb.ecogenomic.org/ (accessed on 30 May 2022)), members of the classes *Chlorobia* and *Ignavibacteria* exhibit considerable genetic diversity, indicating that more detailed studies are needed to clarify the evolutionary relationships and taxonomy of these organisms.

With the aim of clarifying the evolutionary relationships amongst *Chlorobia/Ignavibacteria* species, we have carried out phylogenomic and comparative genomic studies on the genome sequences of these species. To this end, we have constructed phylogenetic trees for these two groups of species based on concatenated sequences for several sets of conserved proteins. These trees, along with the GTDB taxonomy [32] and phylogenetic trees reported in earlier studies [25–27], provide a phylogenetic framework for understanding the evolutionary relationships among the *Chlorobia* species and their relationship to the class *Ignavibacteria*. In parallel, we have also performed extensive studies on protein sequences from the *Chlorobia/Ignavibacteria* genomes to identify molecular signatures, consisting of conserved signature indels (CSIs), that are uniquely shared by species from the major clades of *Chloro-*

bia/Ignavibacteria observed in our phylogenetic trees. Molecular synapomorphies, such as CSIs, that are specific for a given group of organisms, provide strong evidence, independently of phylogenetic trees, for the genetic cohesiveness and common ancestry of a given group of organisms [21,37–41]. Hence, these molecular markers provide reliable means for the delineation of specific clades in molecular terms, and they have proven useful for taxonomic purposes [35,39,42,43]. Results presented here show that the named *Chlorobia* species (order *Chlorobiales*) form a strongly supported clade in trees based on different genes/proteins. The distinctness of this clade is independently strongly supported by 33 CSIs that are present in diverse proteins that are uniquely shared by the members of this class/order. Furthermore, members of the class *Chlorobia* (order *Chlorobiales*) form two distinct clades, which can be reliably distinguished from each other based upon several identified CSIs that are exclusively shared by the members of these two clades. One of these clades, encompassing all *Chlorobiales* genera except *Chloroherpeton*, corresponds to the revised family *Chlorobiaceae*, whereas a second clade grouping together *Chloroherpeton thalassium* and some *Candidatus* Thermochlorobacter strains/isolates and an uncultured *Chlorobium* sp. 445 (we will be referring to this group as thermophilic photoheterotrophs), is now proposed as a new family (*Chloroherpetonaceae* fam. nov). Results presented here also show that the class *Ignavibacteria* is genetically highly diverse and that several uncultured species, currently referred to as "*Chlorobi* bacterium" as well as some *Candidatus* species, are also related to this class/phylum of bacteria. Our work has also identified two new CSIs, which are uniquely shared by most of the *Chlorobia* and *Ignavibacteria* species, providing further evidence that these two groups of organisms are closely related.

2. Materials and Methods

2.1. Construction of Phylogenetic Trees

Protein sequences were downloaded for different available genomes from named *Chlorobia* species as well as several unnamed/uncultured *Chlorobi* species that are denoted by specific numbers in the NCBI genome database [31]. In addition, the sequences for *I. album* and *M. roseus* and some *Candidatus* species (viz., Cand. *Thermochlorobacter aerophilum* [27], Cand. *Thermochlorobacteriaceae* bacterium GBChlB, Cand. *Kapabacteria thiocyanatum*, and Cand. *Kryptonium thompsoni*), which according to the GTDB taxonomy are related to the *Chlorobia/Ignavibacteria* taxa, were also downloaded. In addition, genome sequences for several *Bacteroidetes/Fibrobacteres* species (viz., *Rhodothermus* (R.) *marina*, *Salinibacter* (S.) *ruber*, *Cytophaga* (C.) *aurantiaca*, *Bacteroides* (B.) *fragilis,* and *Fibrobacter* (F.) *succinogenes*) were also downloaded to serve as outgroups in phylogenetic analysis. Using these genome sequences, a phylogenetic tree was constructed based on concatenated sequences of 282 conserved proteins that are a part of the phyloeco marker set for the FBC superphylum [20]. Two additional phylogenetic trees were constructed for these species based on the concatenated sequences for Gyrase A and Gyrase B proteins, as well as UvrD and PolA proteins, which are highly conserved proteins commonly employed for phylogenetic analysis [38].

The construction of the phylogenetic trees was done using an internally developed pipeline described by Adeolu et al. (2016) [40]. Briefly, the CD-HIT program and the profile Hidden Markov Models (HMMs) of the proteins that are part of the FBC–phyloeco set [34] were used to search for homologs of these proteins in the input genomes. The search parameters used required that the selected homologs of different proteins shared a minimum of 50% sequence identity and sequence length, and they were found in at least 80% of the input genomes. The Clustal Omega algorithm [44] was used to generate multiple sequence alignments (MSAs) of these protein families. The aligned protein families were trimmed with TrimAl [45] to remove poorly aligned regions before concatenation of the other core proteins. The final concatenated sequence alignments of these proteins used for tree construction consisted of 89,743 aligned amino acids. Based on this sequence alignment, an initial tree was constructed using FastTree 2 [46] based on the Whelan and Goldman model of protein sequence evolution [47] and optimized using RAxML 8 [48] based on the Le and Gascuel model of protein sequence evolution [49]. SH-like statistical support values

(which are similar to bootstrap scores) were calculated for each branch node using RAxML 8 [48]. The resultant phylogenetic tree was drawn using MEGA X [50]. Sequence alignment of the 282 core proteins was also used to calculate the pairwise average amino acid identity (AAI) [51] for the species belonging to the order *Chlorobiales*. In addition, based on genome sequences for the *Chlorobia* species, the pairwise percentage of conserved proteins (POCP) between different species was also determined [52,53].

We also constructed a 16S rRNA gene tree for the *Chlorobia* and *Ignavibacteriae* species based on sequences obtained from the SILVA ribosomal RNA [54] and the NCBI genome database (https://www.ncbi.nlm.nih.gov/ (accessed on 25 March 2022)). The sequences were aligned using the MUSCLE program in MEGA-X [50]. The non-conserved regions as well as regions with gaps were removed, leaving 1269 aligned positions in the final dataset. A maximum-likelihood phylogenetic tree based on this dataset was created using MEGA X [50], employing the Tamura–Nei model [55] based on 100 bootstrap replicates.

2.2. Identification of Conserved Signature Indels

The identification of conserved signature indels (CSIs) was carried out as described in earlier work [56,57]. Briefly, BLASTp searches using the NCBI non-redundant database were carried out on all proteins from the genomes of *Chlorobium limicola* and *Chloroherpeton thalassium*. Based on these BLASTp searches, protein sequences were obtained for 8–10 divergent *Chlorobia/Ignavibacteria* species (generally including representative species from all three main groups of interest, i.e., *Chlorobiaceae, Chloroherpeton,* and *Ignavibacteria*) as well as 8–10 species from other bacterial taxa (generally belonging to different families/orders of *Bacteroidetes*). The multiple sequence alignments (MSAs) of different proteins were created using ClustalX 2.1. However, other programs can also be used for the creation of multiple sequence alignments and will yield similar results [56,57] (unpublished results). Multiple sequence alignments were visually examined for insertions or deletions of fixed length that were present in conserved regions (i.e., flanked on both sides by minimally 4–5 conserved/identical amino acids (aa) in the neighbouring 40–50 aa) and shared by only the *Chlorobia/Ignavibacteria* species. The query sequences of interest containing the identified conserved indels and their flanking 30–50 aa (generally beginning and ending with a stretch of completely conserved amino acid residues) were reblasted against the NCBI non-redundant (nr) database and the top 250–500 hits were examined. Based on these BLASTp searches, conserved indels that were specifically shared by all or most of the species from the different main clades of *Chlorobia/Ignavibacteria* were identified and further formatted using SIG_CREATE and SIG_STYLE programs (available from Gleans.net (accessed on 25 March 2022) [56]. Due to space constraints, sequence information is presented in the main figures for only a limited number of representative species. However, unless otherwise stated, the CSIs described here are shared by and are exclusive to the indicated groups of *Chlorobia/Ignavibacteria* and absent in all other bacterial homologues in the top 250–500 BLASTp hits examined. More detailed information for different CSIs is provided in the Supplemental Figures.

3. Results

3.1. Phylogenetic Analysis of the Chlorobia/Ignavibacteria Species Based on Genome Sequences

The genome sequences for 36 available *Chlorobia/Ignavibacteria* species in the NCBI database were used to construct a rooted maximum-likelihood phylogenetic tree for these organisms based on concatenated sequences for 282 conserved proteins. The proteins used for tree construction are from the phyloeco set for the FBC group of bacteria and they are conserved and widely distributed in these bacteria [34]. The tree also includes sequences for some outgroup species (viz., *R. marina, S. ruber, C. aurantiaca,* and *B. fragilis*) and it was rooted using the sequence for *F. succinogenes.* The resulting maximum-likelihood distance tree based on 100 bootstrap replicates is shown in Figure 1. All major nodes in this phylogenomic tree are supported by 100% SH-support values (like bootstrap values), indicating that the evolutionary relationships observed here are robust.
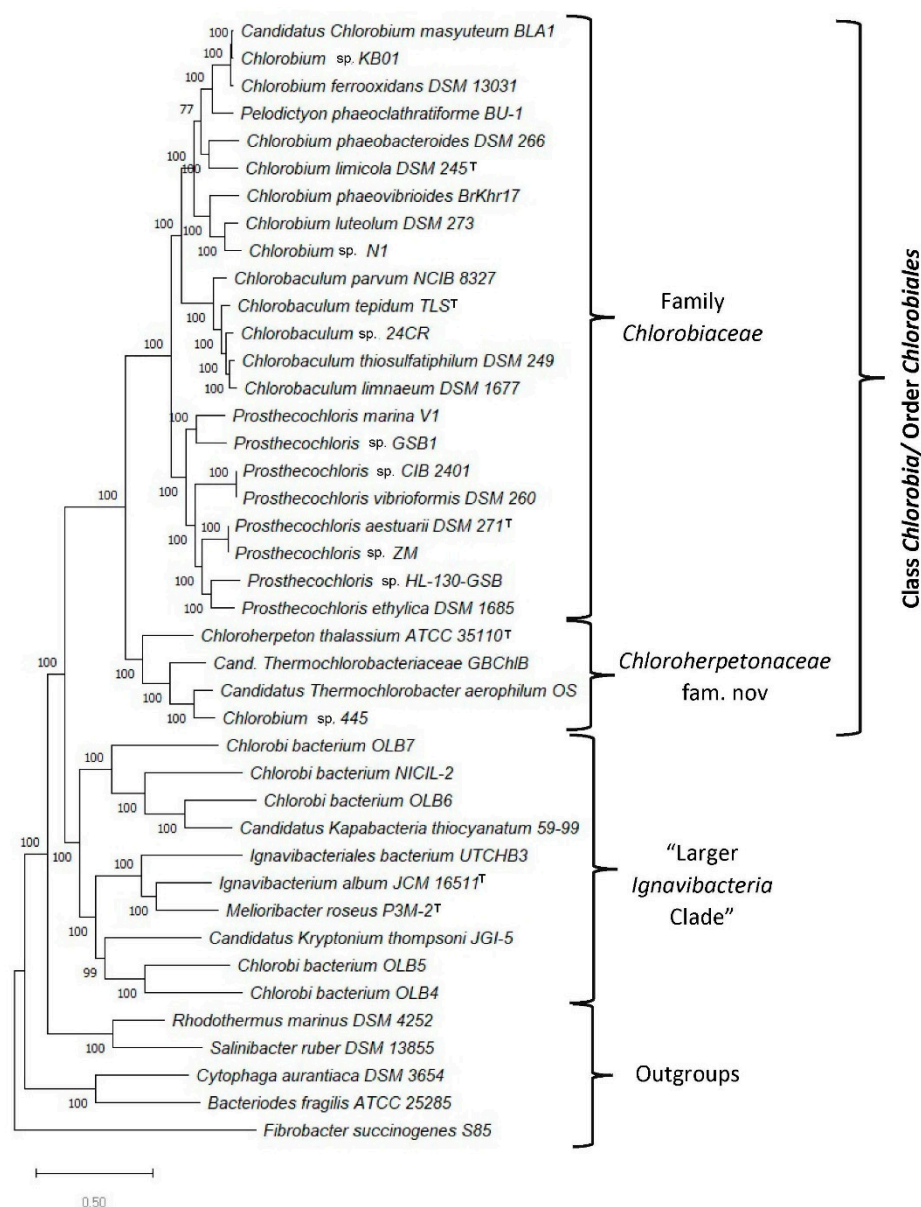
Figure 1. A bootstrapped maximum-likelihood tree for different genome-sequenced *Chlorobia/Ignavibacteria* and related species/strains based on concatenated sequences for 282 conserved proteins that are core proteins for these species. Statistical support values for different branches are indicated on the nodes. The tree was rooted using the sequences from the species *Fibrobacter succinogenes.* The type strains of different species are marked by the superscript[T]. Some species clades observed in this tree are marked.

In addition to the tree shown in Figure 1, we have also constructed phylogenetic trees for these species using concatenated sequences for GyrA and GyrB proteins as well as PolA and UvrD proteins. The results for these are shown in Supplementary Figures S1 and S2. The branching pattern as well as the grouping of species into different clades in these two trees are identical to that seen in Figure 1.

The branching pattern of species shown in Figure 1 (as well as Figures S1 and S2) is similar to that reported by Roy et al. [25]. Based on this tree, several inferences regarding the evolutionary relationships among the *Chlorobia/Ignavibacteria* species/strains can be drawn. First, all named *Chlorobia* species formed a strongly supported clade (labeled the class *Chlorobia* or the order *Chlorobiales* clade), which is separated from a clade harboring *I. album*, *M. roseus*, and several other uncultured species/strains by a long branch. Second,

the *Chlorobiales* clade comprises several distinct clades. One of these subclades groups together species from the genera *Chlorobaculum, Chlorobium, Pelodictyon,* and *Prosthecochloris*, whereas the second subclade consists of the species *Chloroherpeton thalassium* [58] and three uncultured thermophilic photoheterotrophic organisms related to *Candidatus* Thermochlorobacter aerophilum [25,27]. We have designated these two clades as the family *Chlorobiaceae* and the family *Chloroherpetonaceae* fam. nov., respectively. Third, within the *Chlorobiaceae* clade, species from the genera *Chlorobaculum* and *Prosthecochloris* group together, supporting the monophyly of these taxa. However, the genus *Chlorobium* is not monophyletic due to branching within *Pelodictyon phaeoclathratiforme,* indicating that this latter species is misclassified [12]. Fourth, the second major clade observed in Figure 1 consists of the species *I. album* and *M. roseus*, two *Candidatus* species (viz., Cand. Kapabacteria thiocyanatum [59] and Cand. Kryptonium thompsoni [60]), and several uncultured organisms annotated as *Chlorobi* bacterium (OLB4, OLB5, OLB6, OLB7, and NICIL-2). We have designated this clade as the "larger *Ignavibacteria* clade". Unlike the *Chlorobia* species clade, this larger *Ignavibacteria* clade shows greater genetic diversity, and it is made up of several subclades separated by short branches. We also constructed a phylogenetic tree for *Chlorobia/Ignavibacteria* species/strains based on 16S rRNA gene sequences (Figure 2).
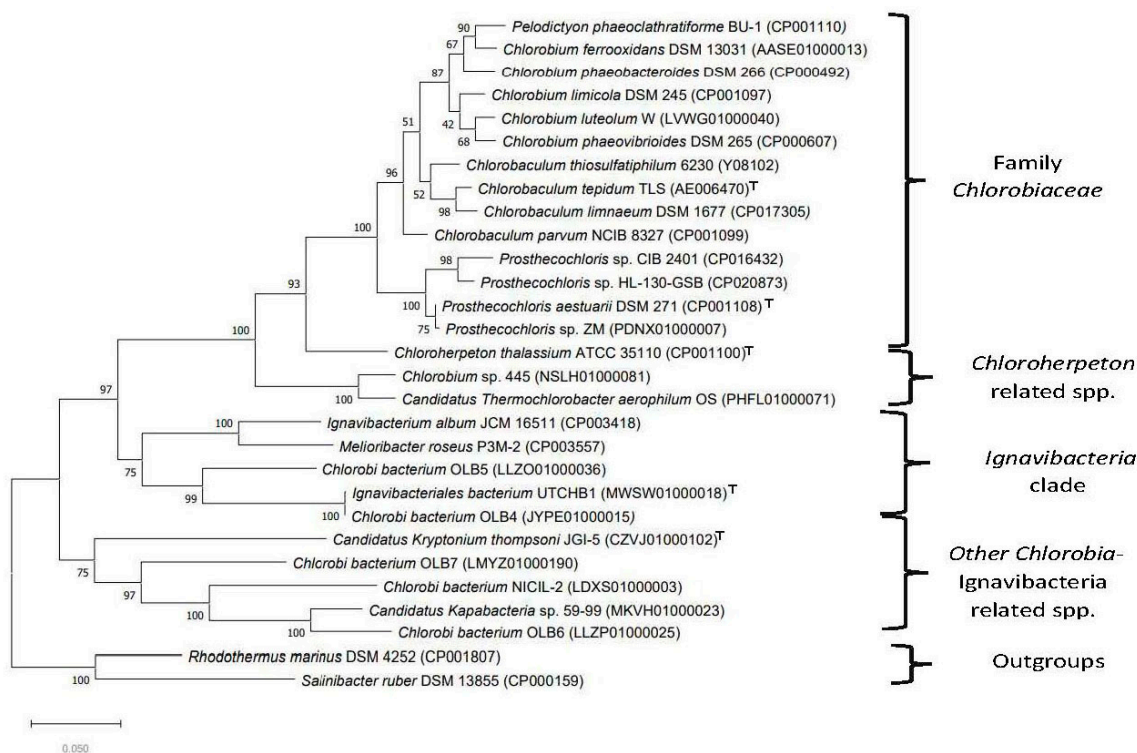
Figure 2. Maximum-likelihood phylogenetic tree based on 16S rRNA gene sequences for the type strains of all *Chlorobi* species. The tree was rooted using the sequences for *S. ruber* and *R. marinus*. Accession numbers of the 16S rRNA gene sequences are given within the bracket after species names in the tree. The main identified clades in this tree are marked.

Unlike the phylogenomic tree (Figure 1), the 16S rRNA gene tree is characterized by lower bootstrap scores for several branches. In spite of its somewhat poor resolution, all named *Chlorobia* species formed a distinct clade within this tree. Within this clade, the clade corresponding to the family *Chlorobiaceae* is strongly supported. However, unlike the core protein tree, this tree did not group together *C. thalassium* with *Candidatus* Thermochlorobacter aerophilum and *Chlorobium* sp. 445 and the latter two species branched more deeply than *C. thalassium.* Similar branching of *C. thalassium* and thermophilic phototrophic organisms has also been previously observed [25]. However, the 16S rRNA sequences from thermophilic organisms have higher G+C content, which may, artefactually, lead to more

deeper branching of thermophilic organisms in the phylogenetic tree [61]. Additionally, in the 16S rRNA tree a clade grouping the two *Ignavibacteria* species with several unnamed *Chlorobi* bacterium (OLB4, OLB5, OLB6, OLB7, and NICIL-2) and the two *Candidatus* species was not observed. These strains/species instead formed several clusters some with very long branches, branching at different positions in between the clades for *Chlorobia* and outgroup species.

Genome sequences for *Chlorobia* species were also used to calculate a pairwise matrix of percentage of conserved proteins (POCP) between different species/genomes. The POCP provides a whole-genome-based method for assessing the similarity and differences between species from related taxa [52,53]. It has been suggested that a POCP matrix is more useful than an average amino acid identity (AAI) matrix for discrimination of taxa at or above the genus rank [53]. In Figure 3, we show the POCP matrix for members of the class *Chlorobia*.
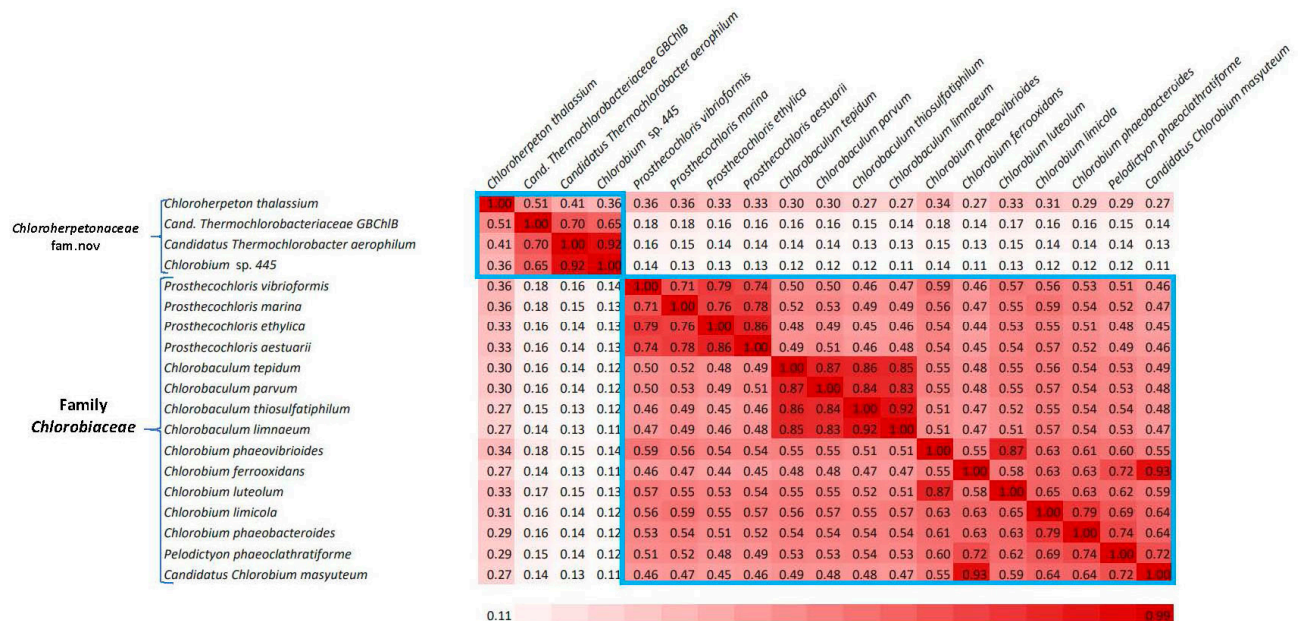
Figure 3. Pairwise comparison matrix showing the averages of the percentage of conserved proteins (POCP) between different genome-sequenced *Chlorobia* species. Genome pairs sharing higher POCP values are shaded more darkly (red). The regions of the matrix corresponding to the species from the two families are labeled.

As seen from this matrix, based on the POCP values, different species/strains from the order *Chlorobiales* are clearly separated into two clades corresponding to the families *Chlorobiaceae* and *Chloroherpetonaceae*. In terms of POCP values, the members of these two clades show no overlap and they are thus clearly distinct. In addition, a matrix based on pairwise AAI values between different *Chlorobia* species was also calculated based on core proteins using an internally developed pipeline [40]. This matrix is shown in Figure S3. While the AAI values for species within these two families were in the range of 0.72–0.87 (for *Chlorobiaceae*) and 0.65–0.80 (for *Chloroherpetonaceae*), the AAI values for interfamily comparisons ranged from 0.67 to 0.70. Although there is no established threshold value for the demarcation of family-level taxa based on AAI [51], the observed differences between the intrafamily and interfamily AAI values supported the overall distinction between the members of these two family-level clades.

3.2. Identification of Molecular Markers Specific for the Main Clades of Chlorobia Species

Results of our phylogenomic studies show that the named *Chlorobia* species form a strongly supported clade. Furthermore, within it, two family-level clades are observed. However, several uncultured species whose genome sequences are available have been

annotated as *Chlorobi* bacterium (OLB4, OLB5, OLB6, OLB7, and NICIL-2) branched outside of the main *Chlorobia* clade. Instead of grouping with the *Chlorobiales*, these sequences showed a closer relationship to the two *Ignavibacteria* species. Therefore, it is important to employ other means to reliably demarcate the family *Chlorobia* and its two family-level clades. With this objective, we have performed detailed comparative studies on protein sequences from *Chlorobia* genomes to identify molecular markers consisting of CSIs that are uniquely shared by members from the main clades of *Chlorobia* species. As noted in the introduction, CSIs in gene/protein sequences that are specifically shared by members from a given clade provide an important class of molecular markers for evolutionary and taxonomic studies [21,35,37,41,62,63]. Our analyses of protein sequences from *Chlorobia/Ignavibacteria* genomes have identified 50 new CSIs that are specific for the different main clades of these bacteria, providing important means for reliably demarcating these clades in molecular terms. The group-specificities and some characteristics of the identified CSIs are described below. Of the identified CSIs, 33 CSIs present in diverse proteins are commonly shared by all or most *Chlorobia* species for whom genome sequences are available. One example of a CSI specific for the class *Chlorobia*, as demarcated based on phylogenetic studies (Figures 1 and 2), is presented in Figure 4.

Sequence information for the 32 other CSIs that are also specific for the class *Chlorobi* (order *Chlorobiales*) is presented in Figures S4–S35 and some of their characteristics are summarized in Table 1.

Table 1. Conserved signature indels specific for members of the order *Chlorobiales* *,a.

Protein Name	Accession/GI Number	Figure No.	Indel Size	Indel Position
Phosphoribosylformylglycinamidine synthase II	WP_041465690	Figure 4	5 aa ins	270–320
Glycyl-tRNA synthetase	78185909	Figure S4	1 aa ins	248–314
RNA polymerase sigma-32 subunit	194337394	Figure S5	12 aa ins	24–79
Uroporphyrinogen decarboxylase	189347814	Figure S6	3 aa ins	284–333
Glutamyl-tRNA reductase	194336061	Figure S7	1 aa ins	211–250
Glutamate-1-semialdehyde aminotransferase	21674908	Figure S8	1 aa ins	60–109
fructose-1,6-bisphosphate aldolase, class II	189500266	Figure S9	11 aa del	233–282
A major facilitator superfamily protein	193212482	Figure S10	1 aa ins	325–365
DNA polymerase III subunit alpha	78187124	Figure S11	2–4 aa ins	294–340
Acetyl-CoA carboxylase carboxyltransferase subunit alpha	119356086	Figure S12	6 aa ins	103–137
N-acetyl-alpha-D-glucosaminyl L-malate synthase BshA	193214252	Figure S13	3 aa ins	88–132
Serine hydroxymethyltransferase	193215659	Figure S14	2 aa ins	228–280
Deoxyribonucleotide triphosphate pyrophosphatase	193215546	Figure S15	3 aa ins	94–144
Polynucleotide phosphorylase/polyadenylase	189346136	Figure S16	1 aa ins	37–69
Polynucleotide phosphorylase/polyadenylase	189346136	Figure S17	17 aa ins	264–335
7-carboxy-7-deazaguanine synthase QueE	500067752	Figure S18	3–7 aa ins	119–164
DNA polymerase III subunits gamma and tau	193213861	Figure S19	2 aa ins	47–93
DNA polymerase III subunits gamma and tau	193213861	Figure S20	4 aa ins	220–268
Cytidylate kinase	21673125	Figure S21	4 aa ins	146–199
4-hydroxy-3-methylbut-2-enyl diphosphate reductase	194333263	Figure S22	2–5 aa ins	264–311
molecular chaperone DnaK	193214979	Figure S23	1 aa del	70–119
Tryptophanyl-tRNA synthetase	78188056	Figure S24	8 aa ins	16–59
Tryptophanyl-tRNA synthase	493409794	Figure S25	5 aa ins	153–204
Succinyl-CoA synthetase subunit alpha	193216377	Figure S26	7 aa ins	31–87
A Gfo/Idh/MocA family oxidoreductase	21673819	Figure S27	4 aa ins	2–48

Table 1. *Cont.*

Protein Name	Accession/GI Number	Figure No.	Indel Size	Indel Position
A Gfo/Idh/MocA family oxidoreductase	21673819	Figure S28	2 aa ins	209–249
Ribonuclease R	193213033	Figure S29	7–8 aa ins	176–223
Ribonuclease R	193213033	Figure S30	4–5 aa ins	558–608
Ribonuclease R	193213033	Figure S31	4 aa ins	640–720
Bifunctional 5,10-methylene-THF dehydrogenase	78186789	Figure S32	2 aa ins	128–161
Lactoylglutathione lyase	78187149	Figure S33	2 aa ins	5–46
Thiazole synthase	78187419	Figure S34	1 aa ins	55–106
DNA mismatch repair protein *	193213813	Figure S35	1 aa del	599–644

* Except for an isolated exception all of these CSIs are specific for the *Chlorobia/Chlorobiales* species. [a] For some CSIs, homologs were not found in all *Chlorobia* species.

Our analyses have also identified multiple CSIs that are specific for the two family-level clades within the order *Chlorobiales*. Figure 5 presents a partial sequence alignment of the protein polyphosphate kinase-1 highlighting a one-amino-acid insert in a conserved region (boxed) that is exclusively shared by the members of the family *Chlorobiaceae*.

The polyphosphate kinase CSI in Figure 5 is commonly shared by all species/strains that are a part of the *Chlorobiaceae* family (see Figure 1), but it is absent in members of the family *Chloroherpetonaceae* as well as all other deeper branching species/strains including the *Ignavibacteria* species, unnamed *Chlorobi* bacteria, as well as different outgroup bacteria. In addition to this CSI, seven other CSIs identified in this study within other protein sequences are also specific for members of the family *Chlorobiaceae*. Sequence information for these other CSIs is presented in Figures S36–S42 and some of their characteristics are summarized in Table 2. Based on these CSIs, members of the family *Chlorobiaceae* can be clearly distinguished in molecular terms from all other bacteria.

Table 2. Conserved Signature Indels Specific for Members of the families *Chlorobiaceae* and *Chloroherpetonaceae* *,[a].

Protein Name	Accession/GI Number	Figure No.	Indel Size	Indel Position	Specificity
Polyphosphate kinase	WP_012466203	Figure 5	1 aa ins	25–65	
Peptide chain release factor 3	193212508	Figure S36	1 aa del	400–447	
Peptidase U32	493410285	Figure S37	6 aa ins	260–316	
MiaB-like tRNA modifying protein	145220314	Figure S38	5 aa ins	329–382	
Molecular chaperone HtpG	145219831	Figure S39	3 aa ins	259–314	Family *Chlorobiaceae*
DegT/DnrJ/EryC1/StrS aminotransferase	193212751	Figure S40	1 aa del	111–156	
biogenesis of lysosome-related organelles complex 1 subunit 2	194336576	Figure S41	1 aa del	238–279	
DNA gyrase subunit A	194335275	Figure S42	1 aa ins	234–268	
UDP-glucose/GDP-mannose dehydrogenase family protein	WP_012498759	Figure 6	2 aa ins	287–330	
hypoxanthine phosphoribosyltransferase	PIO48526	Figure S43	1 aa ins	57–94	
dihydrolipoyl dehydrogenase	PIO48610	Figure S44	1 aa ins	30–64	Family *Chloroherpetonaceae*
SDR family oxidoreductase	WP_012499003	Figure S45	4 aa ins	236–280	
RecQ family ATP-dependent DNA helicase	WP_012499407	Figure S46	2 aa ins	164- 203	
alkaline phosphatase family protein	RFM24133	Figure S47	5 aa ins	33–78	

* Except for an isolated exception, all of these CSIs are specific for the members of the indicated families. [a] For some CSIs, homologs were not identified in all members of a given family.
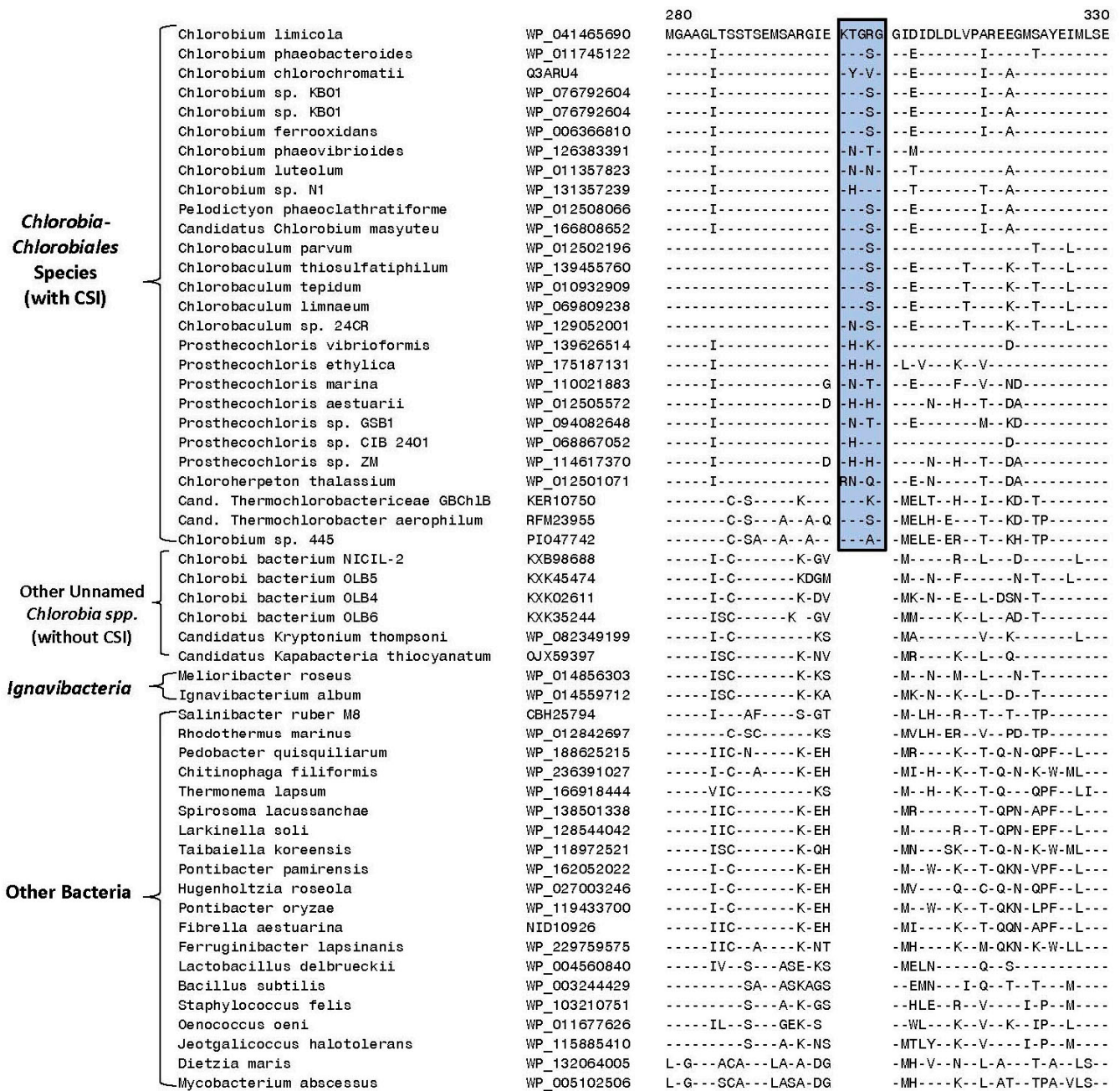
Figure 4. Partial sequence alignment of the protein phosphoribosylformylglycinamidine synthase II showing a five-amino-acid insertion (boxed) in a conserved region that is exclusively shared by all species/strains from the class *Chlorobia* (Figure 1). This insert is lacking in *Ignavibacteria* as well as other *Candidatus* and *Chlorobi* bacteria strains that do not group with the *Chlorobia* clade. The dashes (-) in this and all other sequence alignments indicate identity with the amino acids on the top line. Gaps in sequence alignment indicate that no amino acid is present in that position. Accession numbers for different sequences are indicated in the second column and the position of this sequence fragment within the protein is indicated above the sequences.
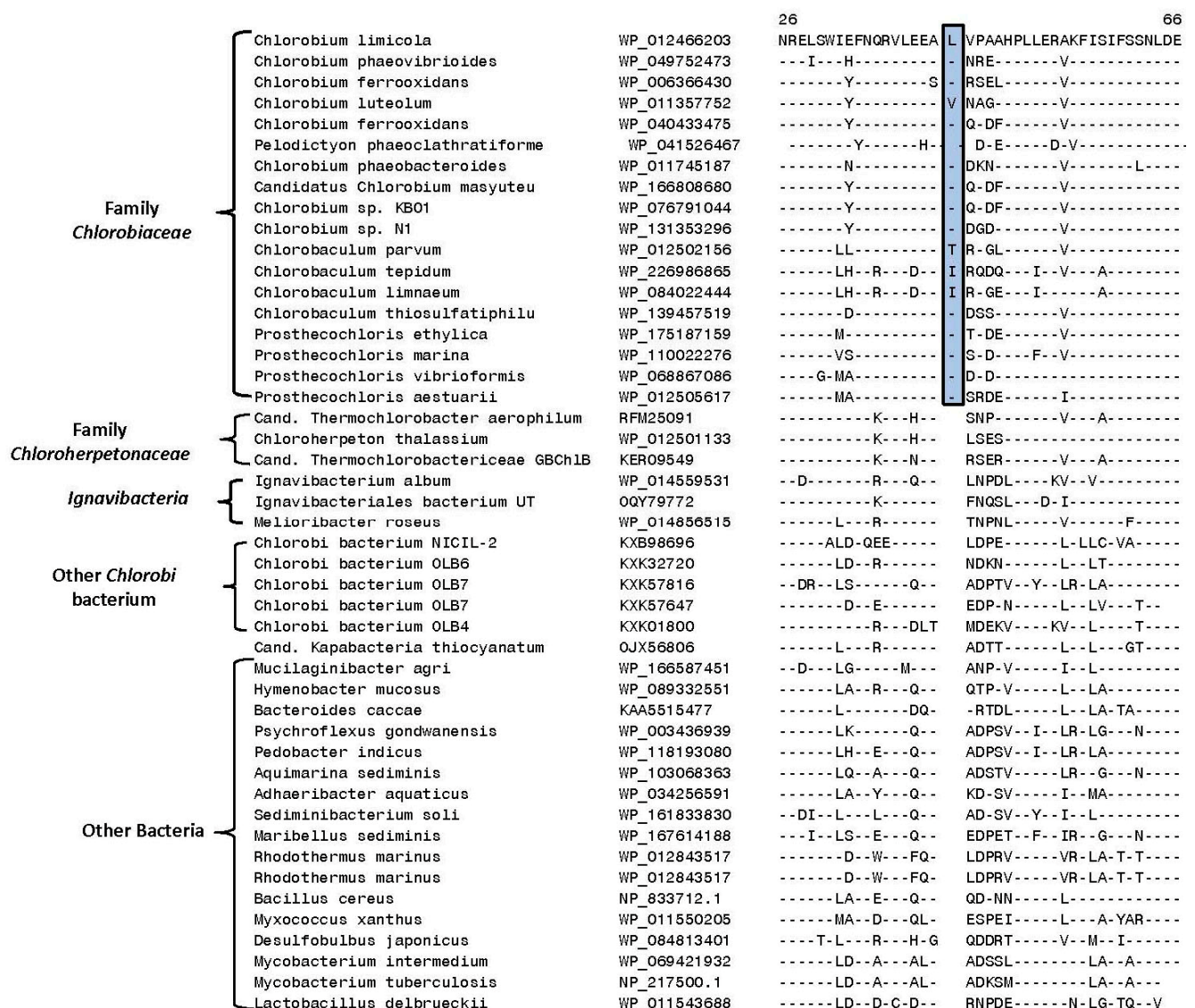
Figure 5. Excerpts from the sequence alignment of the protein polyphosphate kinase highlighting a one-amino-acid insertion (boxed) in a conserved region that is specifically shared by species/strains that are part of the family *Chlorobiaceae* (Figure 1). This insert is not found in members of the family *Chloroherpetonaceae*, class *Ignavibacteria*, other outgroup bacteria, as well as in other *Candidatus* and *Chlorobi* bacteria strains that do not group with the *Chlorobia* clade.

Six other CSIs identified in this work are specific for members of the family *Chloroherpetonaceae*. In Figure 6, we show a partial sequence alignment of the protein UDP-glucose GDP-mannose dehydrogenase, where a two-amino-acid insert is present in a conserved region (boxed) that is exclusively found in members of the family Chloroherpetonaceae but not found in any other *Chlorobia/Ignavibacteria* species or other bacteria. Sequence information for five other CSIs that are also specific for members of the family *Chloroherpetonaceae* is presented in Figures S43–S47 and some of their characteristics are summarized in Table 2. These CSIs provide strong and independent evidence for the distinctness of the family *Chloroherpetonaceae* from other *Chlorobiales* species and provide reliable means for the identification/demarcation of this clade.
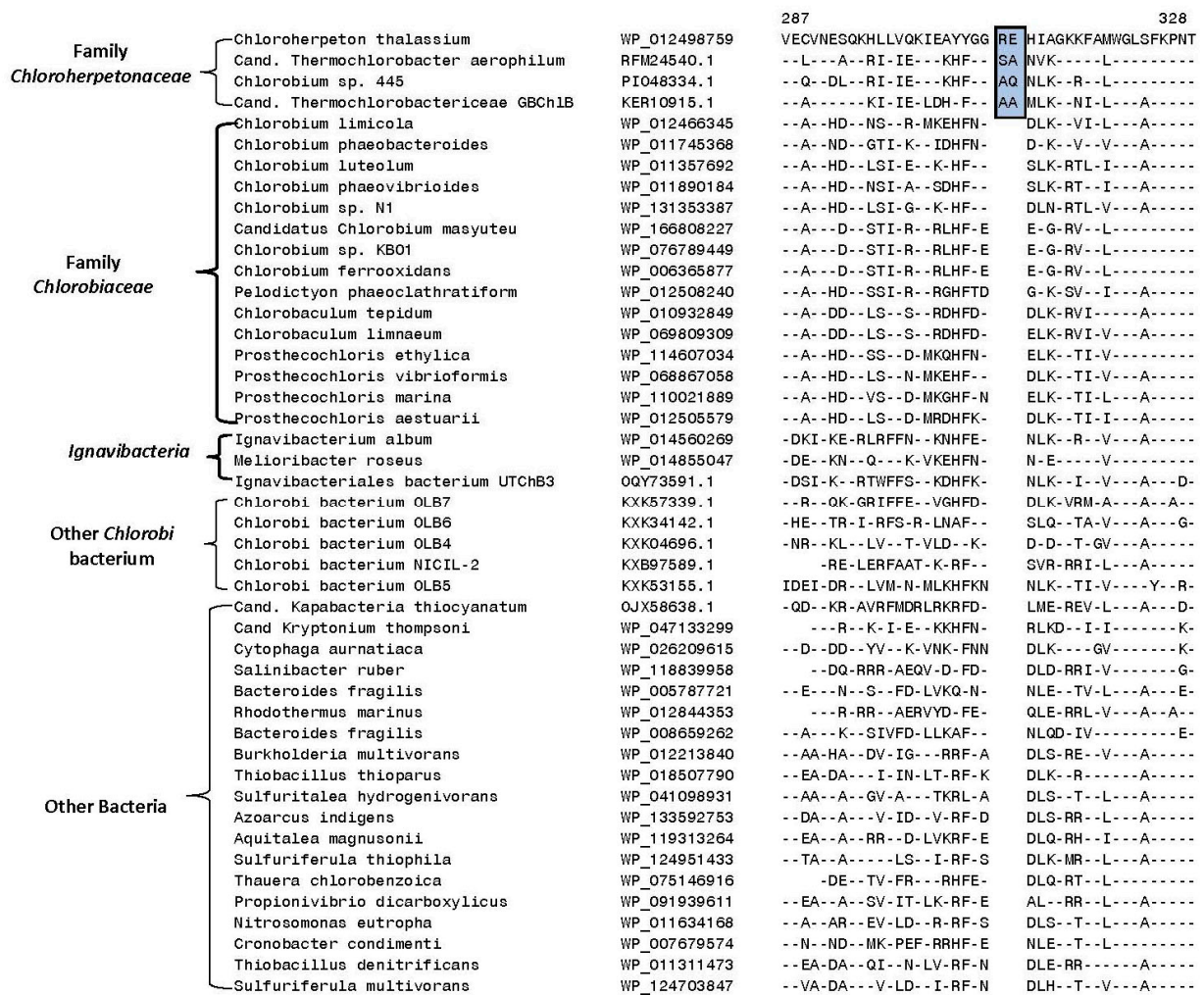
Figure 6. Partial sequence alignment of the protein UDP-glucose/GDP-mannose dehydrogenase showing a two-amino-acid insertion (boxed) in a conserved region that is exclusively shared by all species/strains that are part of the proposed family *Chloroherpetonaceae* (Figure 1). This insert is not found in members of the family *Chlorobiaceae*, class *Ignavibacteria*, as well as other *Candidatus* and *Chlorobi* bacteria strains that do not group with the *Chlorobia* clade. Sequence information for five other CSIs specific for the family *Chloroherpetonaceae* is presented in Figures S43–S47 and Table 2.

Our analysis has also identified one CSI that is commonly shared by most of the species/strains that are a part of the larger *Ignavibacteria* clade. Sequence information for this CSI consisting of a two-to-three amino acid insert in the protein preprotein translocase subunit SecY is shown in Figure 7. This CSI, in addition to the known *Ignavibacteria* species/strains, is also commonly shared by specific *Candidatus* species and unnamed *Chlorobi* bacterium strains that group with the *Ignavibacteria* clade in our core protein tree. However, this CSI is absent in all *Chlorobiales* species and the outgroup bacteria examined.
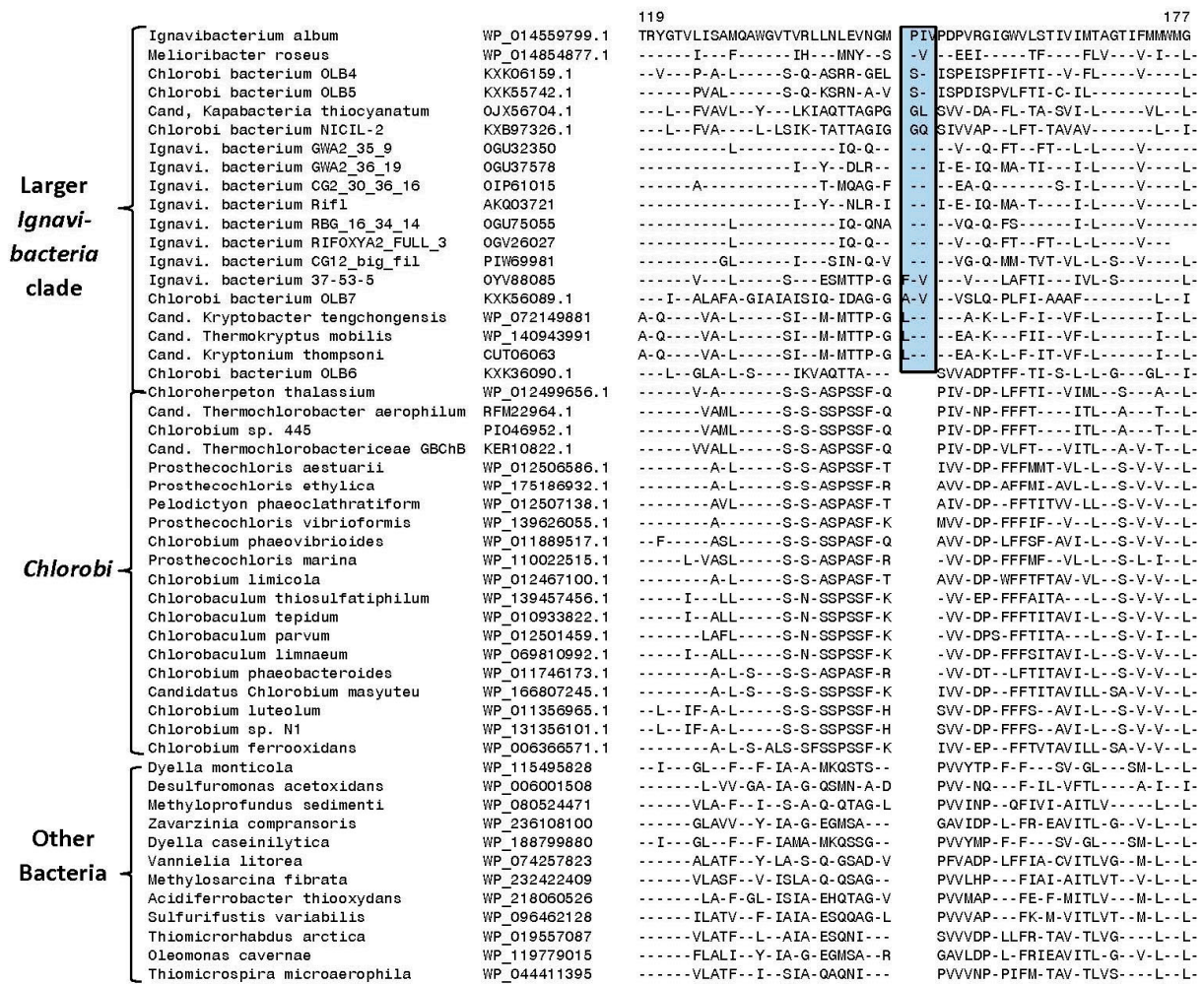
Figure 7. Partial sequence alignment of the protein preprotein translocase subunit SecY showing a two-to-three amino acid insertion (boxed) in a conserved region that is specifically shared by most of the species/strains that are a part of the larger *Ignavibacteria* clade (Figure 1). This insert is not found in members of the class *Chlorobia* or different outgroup bacteria examined, indicating that the genetic change giving rise to this CSI likely occurred in a common ancestor of the *Ignavibacteria* clade of species/strains.

Lastly, in phylogenetic trees, members of the class *Ignavibacteria* branch in the proximity of *Chlorobia* species (Figure 1) [25–27]. Our work has also identified two CSIs that are uniquely shared by all *Chlorobia* species and different *Ignavibacteria* species/strains. Figure 8 shows a partial sequence alignment of the protein methionine t-RNA ligase, where a four-amino-acid insert is present in a conserved region that is commonly shared by most species/strains from these two groups, but it is not present in different Bacteroidetes species examined or other closely related bacterial phyla in the top 500 BLASTp hits.
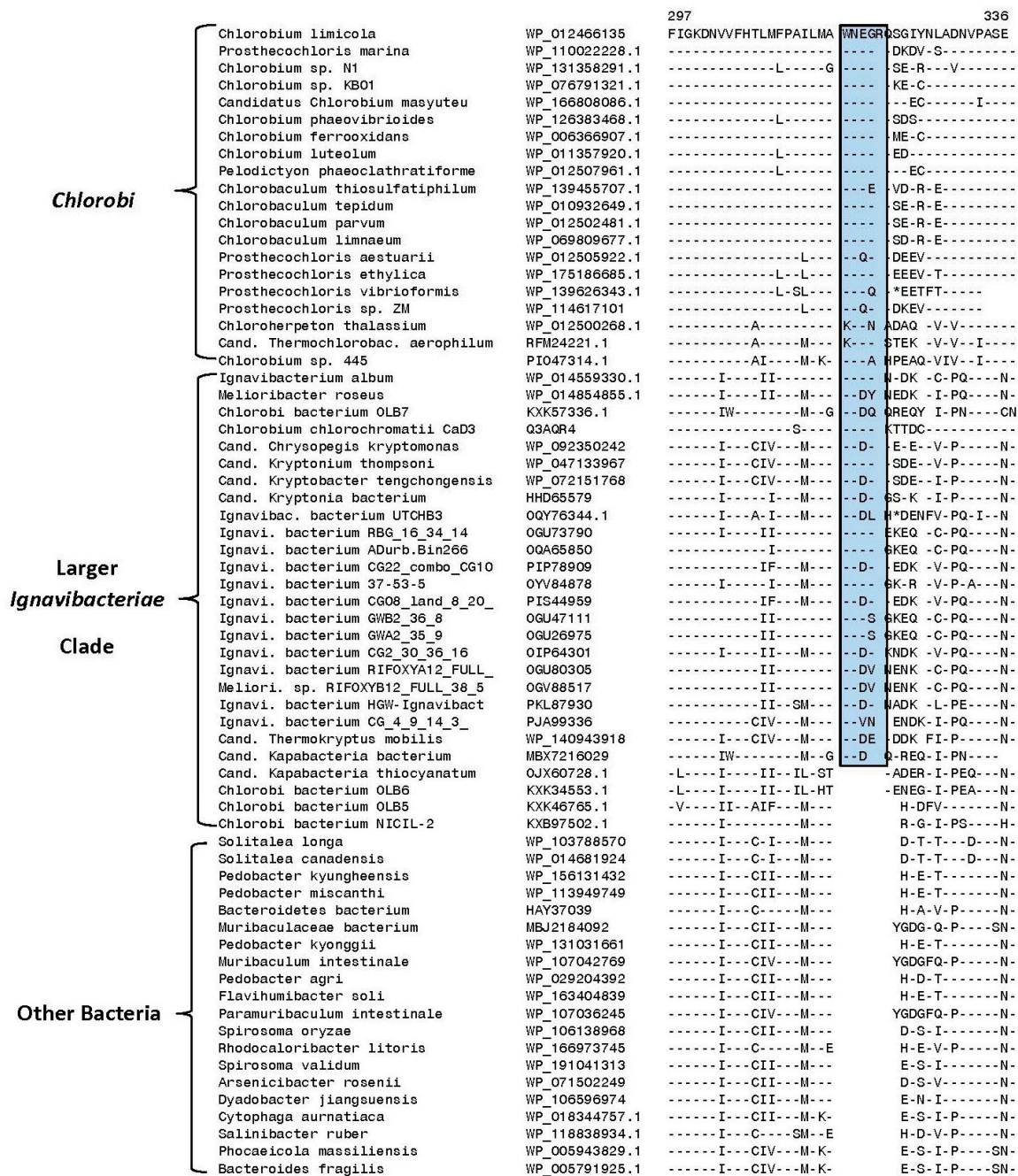
Figure 8. Partial sequence alignment of the protein methionine t-RNA ligase showing a four-amino-acid insertion (boxed) in a conserved region that is specifically shared by all Chlorobia species and most of the species/strains from the larger Ignavibacteria clade (Figure 1). However, this insert is not found in any other outgroup bacteria within the top 500 BLASTp hits. Sequence information for one additional CSI in the protein tRNA-dihydrouridine synthase that is also commonly shared by the species/strains from these two classes is presented in Figure S48.

The CSI shown in Figure 8 in addition to different Chlorobia and Ignavibacteria species is also shared by several Candidatus species, which group within the larger Ignavibacteria clade in our phylogenomic tree (Figure 1). However, several Chlorobi bacterium (viz., OLB5, OLB6, and NICIL-2), which also group within the larger Ignavibacteria clade (Figure 1), do not share this CSI, indicating that the grouping together of all these species/strains is not confirmed by the identified CSI. Sequence information for another CSI, consisting of a one-amino-acid deletion in the protein tRNA-dihydrouridine synthase that is commonly

shared by various Chlorobia and Ignavibacteria species/strains, is presented in Figure S48. This CSI is specific for the Chlorobia and Ignavibacteria species/strains, as homologs of this protein were not detected in other Candidatus species or Chlorobi bacterium strains.

4. Discussion

Members of the class *Chlorobia* constitute one of the seven discontinuous lineages of prokaryotic organisms that can carry out bacteriochlorophyll and chlorophyll-based photosynthesis [42,64,65]. As these bacteria are primarily found in a narrow environmental niche consisting of anoxic aquatic settings in stratified lakes, where sunlight can penetrate, identification and culturing of these bacteria are often not easy [3–5]. Thus, it is important to better understand the evolutionary relationships amongst these and related bacteria in addition to developing reliable means for their identification. Based on earlier work, members of the class *Ignavibacteria* are thought to be the closest relatives of *Chlorobia* [23–27,66]. Genome sequences are now available for most of the named *Chlorobia* and *Ignavibacteria* species as well as several other related uncultured species. In the present work, we have used these genome sequences to elucidate the evolutionary relationships among these organisms through several genome-sequence-based approaches. The approaches used to examine their evolutionary relationships include: (i) construction of a phylogenetic tree based on concatenated sequences of 282 core proteins from their genomes (Figure 1); (ii) construction of phylogenetic trees based on concatenated sequences of PolA–UvrD proteins (Figure S1), GyrA–GyrB proteins (Figure S2), and 16S rRNA gene sequences (Figure 2); (iii) determination of pairwise comparison matrices for *Chlorobia* species based on POCP (Figure 3) and AAI (Figure S3); and (iv) detailed analyses of protein sequences from *Chlorobia/Ignavibacteria* species, which have identified 50 novel conserved CSIs that are specific for different clades of these organisms. The CSIs in protein sequences result from rare genetic changes [21,41,56]. Hence, the shared presence of these molecular synapomorphies by a given group of species provides strong evidence that the species from that clade shared a common ancestor exclusive of other organisms and thus are specifically related to each other [21,41,56]. Additionally, earlier work on CSIs provides evidence that these molecular markers possess a high degree of predictive ability to be found in other unidentified or uncharacterized members of these clades [35,37,63].

Based on the results presented here, several consistent inferences can be drawn concerning the evolutionary relationships among *Chlorobia/Ignavibacteria* species. First, the results presented here confirm that the named *Chlorobia* species form a strongly supported clade, which can be reliably distinguished from *Ignavibacteria* and other bacteria by phylogenomic analysis and by 33 identified CSIs in diverse proteins that are uniquely shared (synapomorphies) by the members of this clade. The results from phylogenetic studies and clade specificities of the identified CSIs also reveal that several uncultured organisms that are referred to in the NCBI database as *Chlorobi* bacterium (OLB4, OLB5, OLB6, OLB7, and NICIL-2) are not related to the class *Chlorobia* and thus they are misclassified (or incorrectly annotated) as *Chlorobi* bacterium. Based on the grouping of these uncultured strains with the larger *Ignavibacteria* clade, rather than with the clade for *Chlorobia* species, it is suggested that in future these sequences should be referred to as the *Ignavibacteria* bacterium OLB4, OLB5, OLB6, OLB7, and NICIL-2. Second, the results presented here provide compelling evidence that the class *Chlorobia* or the order *Chlorobiales* comprises two distinct clades. The first of these clades grouping together different species from the genera *Chlorobaculum, Chlorobium, Pelodictyon,* and *Prosthecochloris* corresponds to the family *Chlorobiaceae*. The second clade harbors the species *Chloroherpeton thalassium, Candidatus* Thermochlorobacter aerophilum, *Candidatus* Thermochlorobacteriaceae bacterium GBChlB, and an unnamed *Chlorobium* sp. 445. Separation of the members of the order *Chlorobiales* into two family-level clades is also observed in other phylogenetic studies including the GTDB taxonomy [4,25,32]. Furthermore, the species from these two clades can also be distinguished from each other in pairwise matrices based on whole-genome comparison of POCP and AAI. However, the strongest evidence that the species from these two clades are

distinct from each other is provided by our identification of eight and six CSIs in diverse proteins, respectively, that are exclusively shared by the members of these two clades. These CSIs provide a novel and unambiguous means for the identification and demarcation of the members of these two clades in molecular terms. Based on the clear distinction seen between the members of these two clades, based on phylogenomic and molecular sequence-based characteristics, we propose that the species from the second clade consisting of *Chloroherpeton*-related organisms be referred to as a new family *Chloroherpetonaceae* fam. nov.

It should be noted that the species from the two main clades of *Chlorobiales* exhibit interesting clade-specific differences in growth, biochemical, and physiological characteristics [4,5,25–27,67,68]. Some of these differences are noted in Figure 9, which also summarizes the results from this study.
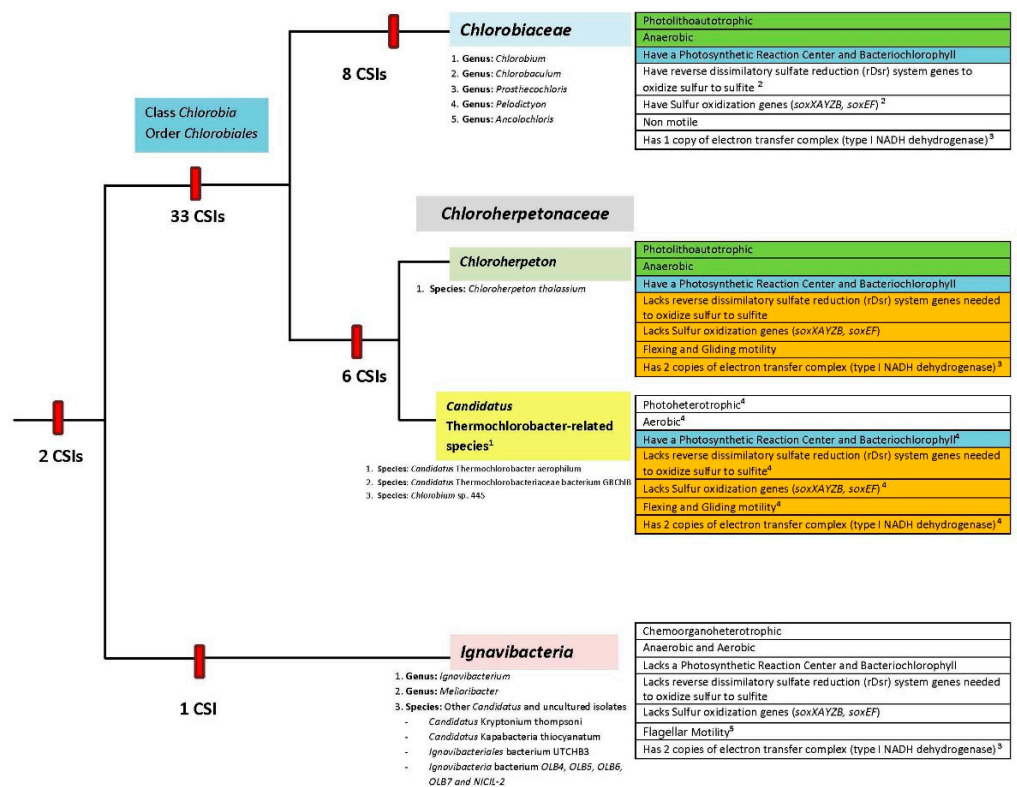
Figure 9. A conceptual diagram summarizing the evolutionary relationships among members of the phylum *Chlorobi* based on phylogenetic analysis and specific identified molecular signatures (synapomorphies). The numbers of CSIs that are specific for each clade or species-grouping are noted on the respective nodes. The species/genera present in each clade are listed underneath each label. The physiological and biochemical traits unique to each clade are listed on the right-hand side of the figure. Other notes: [1] These species have yet to be isolated and grown in pure culture; [2] *Chlorobium ferroxidans* is the sole exception; [3] Subunits NuoEFG missing from the sole copy in all *Chlorobiaceae* (11 subunits); *Chloroherpeton thalassium* NuoEFG missing from only one copy (14 and 11 subunits); *Candidatus* Thermochlorobacter aerophilum (12 and 11 subunits); *Ignavibacterium* album (14 and 11 subunits); [4] Inferred from genome sequences of *Chlorobium* sp. 445, *Candidatus* Thermochlorobacter aerophilum, and *Candidatus* Thermochlorobacteriaceae bacterium GBChlB; [5] Observations by light and electron microscopy suggested that *I. album* was not motile and lacked flagella (Iino et al., 2010 [7]). However, the *I. album* genome contains a nearly complete set of genes for flagella along with genes for chemotaxis and signal transduction. *Melioribacter roseus* is motile only during exponential growth, using flagella to move through water, but loses its flagella and mobility during stationary growth.

As noted in Figure 9, members of the family *Chlorobiaceae* are uniformly anaerobic, nonmotile, and photolithoautotrophic bacteria that use reduced sulfur compounds as reductants and electron source to fix CO_2. They contain reverse dissimilatory sulfate reduction (rDsr) system genes to oxidize sulfur to sulfite as well as sulfur oxidization genes (soxXAYZB, soxEF) [25,67,68]. In contrast, members of the proposed *Chloroherpetonaceae* family exhibit important differences in their biochemical and physiological characteristics. Among these, *Chloroherpeton thalassium* is an anaerobic photoautolithotroph like the members of the family *Chlorobiaceae*. However, this species differs from the other *Chlorobiaceae* species in that it lacks the genes for the reverse dissimilatory sulfate reduction (rDsr) system as well as the sulfur oxidization genes (soxXAYZB, soxEF) needed to oxidize sulfur to sulfite. It also exhibits flexing and gliding motility. All other members of this clade are uncultured species that, based on their genome sequences, have been inferred to be aerobic photoheterotrophs, containing genes for photosynthetic reaction centers and bacteriochlorophyll biosynthesis, that likely use photo-assimilated acetate and propionate as carbon sources [25–27]. Based on the biochemical and physiological characteristics of *Chloroherpeton*, which overlap with those of *Chlorobiaceae*, it has been suggested [25,27] that this species/genus should be placed into a separate family distinct from the other Cand. *Thermochlorobacter*-related photoheterotrophs. However, based on considerations discussed in this work, we propose to keep all members of this clade in one family rather than two separate families. The main reasons for doing this are as follows. In the different phylogenetic trees constructed in this work and in earlier studies [25,32], *Chloroherpeton* groups reliably with the Cand. *Thermochlorobacter*-related photoheterotrophs. Six CSIs identified in this work are also uniquely shared by these two groups, providing strong evidence that the species from these two groups are specifically related. On the other hand, no CSI was identified that was commonly shared by only the *Chloroherpeton* and *Chlorobiaceae*. The GTDB taxonomy, which now provides a widely used reference resource for prokaryotic taxonomy, also places these species into a single family. If we place Cand. *Thermochlorobacter*-related photoheterotrophs into a separate family, it will not be possible to publish it as a valid family in the List of Prokaryotic Names with Standing in Nomenclature [11], as it does not contain any cultured species. However, in future work, if additional information becomes available for this clade of species, including an isolated cultured species for the Cand. *Thermochlorobacter*-related photoheterotrophs, the possibility of dividing the proposed family *Chloroherpetonaceae* into two families could be re-evaluated.

In contrast to the monophyly and reliable demarcation of the clades corresponding to the order *Chlorobiales* (class *Chlorobia*) and the two families within this order, our results indicate that the class *Ignavibacteria* is genetically highly diverse, and it is difficult at present to unambiguously delineate this taxon either in phylogenetic terms or by means of identified molecular markers. The class *Ignavibacteria* presently contains only two named species (*M. roseus* and *I. album*), which have been placed into two separate families within the order *Ignavibacteriales* [7,23,24]. However, unlike the class *Chlorobia*, where very few new species have been identified in the past 10–15 years, in the same period several *Candidatus* and other uncultured species/strains have been described that branch in the proximity of two *Ignavibacteria* species (Figures 1 and 2) [17,27,59,60]. In the phylogenomic tree constructed in this work, the two *Ignavibacteria* species are part of a larger clade that includes several *Candidatus* species and uncultured *Chlorobi* species/strains. The GTDB taxonomy also indicates that these *Candidatus* and uncultured species/strains branch in the proximity of *Ignavibacteria* [32]. Although branching in a phylogenetic tree is affected by large numbers of variables, and can sometimes be misleading [69–71], a specific grouping or relationship of the species/strains that are a part of the larger *Ignavibacteria* clade is also independently supported by our identification of a CSI in the SecY protein that is uniquely shared by most of the species/strains from this clade (Figure 7). Two additional CSIs identified in this work, which are commonly shared by different *Chlorobia* species and several species/strains from this larger *Ignavibacteria* clade, provide further support that these specific *Candidatus* and uncultured species/strains are related to these bacteria.

These observations suggest that the circumscription of the class *Ignavibacteria* should be expanded to include these other *Candidatus* and uncultured *Chlorobi* species/strains (i.e., corresponding to the large *Ignavibacteria* clade (Figure 1)). However, we refrain from making a formal proposal, because the number of uncultured *Candidatus* species/strains related to this group is rapidly expanding in the public databases. Thus, more detailed future studies on *Ignavibacteria* and related species/strains need be conducted to develop a more precise and informative classification scheme for this group/class of bacteria, which will likely lead to its division into several different order/family-level taxa.

The present study has identified many molecular markers (CSIs) that are exclusively found in members of different clades within the *Chlorobia/Ignavibacteria* groups of bacteria. Extensive earlier work on CSIs specific for other prokaryotic taxa provides compelling evidence that these molecular markers possess a high degree of predictive ability to be found in other members from these clades whose genome sequences are currently unavailable, as well in novel and uncultured species that are related to these specific clades [39,57,62,63,72,73]. Due to the presence of these CSIs in conserved regions of different genes/proteins, they provide a novel means for the identification of other species from these taxa by either in silico analysis of genomic sequences (based on BLASTp searches checking for the presence or absence of these molecular sequences) or experimental means using PCR-based assays [62,74–76]. Most of the novel species/strains related to the *Chlorobia/Ignavibacteria* taxa reported in recent years consist of uncultured species that have been identified based on analyses of metagenomic sequences from diverse geological habitats [17,27,59,60]. In this context, the molecular markers described here, which are highly specific for these groups of bacteria, should prove helpful in the identification and characterization of other novel species/strains related to these taxa. Lastly, earlier work on CSIs provides evidence that these molecular characteristics are functionally important for the group of organisms for which they are specific, and several of these CSIs found in key chlorophyll–bacteriochlorophyll biosynthesis proteins and core centre proteins have provided important insights into the evolutionary relationships among different groups/phyla of photosynthetic bacteria [77–80]. Currently, very few reliable characteristics are known that are specific for the *Chlorobia/Chlorobiales* species [3,4]. Therefore, to incorporate the information for the CSIs that are specific for the order *Chlorobiales* and the family *Chlorobiaceae*, we also provide emended descriptions of these taxa. In addition, we provide a formal description of the class *Chlorobia* as this has not yet been validly published [11]. The descriptions of the emended and novel taxa are given below.

Description of the Class *Chlorobia* class nov. (Garrity and Holt 2001, 601[EP])

Chlorobia (Chlo.ro'bi.a. N.L. neut. n. *Chlorobium*, genus of the class, changing ending to denote a class; N.L. neut. pl. n. *Chlorobia*, class of the genus *Chlorobium*).

The class *Chlorobia* contains a single order (*Chlorobiales*) [8] and the description of this class is the same as that given below for the order *Chlorobiales.*

The type order is *Chlorobiales* Gibbons and Murray 1978 (Approved Lists 1980).

Emended Description of the order *Chlorobiales* Gibbons and Murray 1978 (Approved Lists 1980)

The order is composed of two families: *Chlorobiaceae* and *Chloroherpetonaceae.* Members of this order form a monophyletic clade in phylogenetic trees based on 16S rRNA gene sequences and trees based on several individual proteins (viz., Fmo, GyrA, GyrB, PolA, RecA, and UvrD) and large datasets of concatenated protein sequences. The order comprises species that are anaerobic and strict photolithoautotrophs, which use reduced sulfur compounds as an electron source to fix CO_2, as well as aerobic photoautotrophic organisms, which likely use photo-assimilated acetate and propionate as carbon sources. Species from this phylum generally contain genes encoding for photosynthetic reaction centers and bacteriochlorophyll biosynthesis. The members of this order can be reliably distinguished from all other bacteria based on the shared presence of 33 conserved signature indels (CSIs) in diverse proteins (Table 1) that are uniquely shared by the members

of this order. The CSIs specific for the order *Chlorobiales* are found in the following proteins: a major facilitator superfamily protein, a radical SAM-domain-containing protein, a Gfo/Idh/MocA family oxidoreductase, an acetyl-CoA carboxylase carboxyltransferase alpha subunit, a DNA mismatch repair protein, bifunctional 5,10-methylene-tetrahydrofolate dehydrogenase, cytidylate kinase, deoxyribonucleotide triphosphate pyrophosphatase, a DNA polymerase III alpha subunit, DNA polymerase III subunits gamma and tau, fructose-1,6-bisphosphate aldolase (class II), glutamate-1-semialdehyde aminotransferase, glutamyl-tRNA reductase, glutamate-1-semialdehyde aminotransferase, glycyl-tRNA synthetase, 4-hydroxy-3-methylbut-2-enyl diphosphate reductase, lactoylglutathione lyase, molecular chaperone DnaK, phosphoribosylformylglycinamidine synthase II, polynucleotide phosphorylase/polyadenylase, ribonuclease R, an RNA polymerase sigma-32 subunit, serine hydroxymethyltransferase, succinyl-CoA synthetase subunit alpha, thiazole synthase, tryptophanyl-tRNA synthetase, uroporphyrinogen decarboxylase, N-acetyl-alpha-D-glucosaminyl L-malate synthase BshA, and 7-carboxy-7-deazaguanine synthase QueE.

The type genus of this order is *Chlorobium* Nadson 1906 [81] (Approved Lists 1980).

Emended Description of the family *Chlorobiaceae* Copeland 1956 (Approved Lists 1980)

The family *Chlorobiaceae* contains the type genus *Chlorobium* [8,81] and the following validly published genera: *Ancalochloris* [8,14], *Chlorobaculum* [12], *Pelodictyon* Lauterborn [8], and *Prosthecochloris* [8,14]. The description of this family is partially based on that given by Imhoff [4]. Members of this family are non-motile photolithoautotrophs and they grow under anoxic conditions in the presence of limited light using reduced sulfur compounds as an electron source to fix CO_2. Thiosulfate, hydrogen, and ferrous iron can also be used as reductants. Most of the members of this family have been isolated from anoxic aquatic settings in stratified lakes, where sunlight can penetrate. All members of this family except *Chlorobium ferroxidans* possess a dissimilatory sulfite reductase (DSR) system. Members of this family also form a monophyletic clade in phylogenetic trees based on 16S rRNA gene sequences and trees constructed based on large datasets of concatenated protein sequences. The members of this order can be clearly distinguished from members of the family *Chloroherpetonaceae* and all other bacteria based on the shared presence of eight conserved signature indels (CSIs) found in diverse proteins (listed in Table 2) that are uniquely shared by the members of this family. The proteins containing the CSIs specific for this family are as follows: biogenesis of lysosome-related organelles complex 1 subunit 2, DegT/DnrJ/EryC1/StrS aminotransferase, DNA gyrase subunit A, molecular chaperone HtpG, MiaB-like tRNA modifying protein, peptidase U32, peptide chain release factor 3, and polyphosphate kinase.

The type genus of this family is *Chlorobium* Nadson 1906 [81] (Approved Lists 1980).

Description of the family *Chloroherpetonaceae* fam. nov.

Chloroherpetonaceae (Chlo.ro.her'pe.to.na.ce'ae. N.L. neut. n. *Chloroherpeton*, type genus of the family; -aceae ending to denote a family; N.L. fem. pl. n. *Chloroherpetonaceae* the *Chloroherpeton* family).

This family contains the type genus *Chloroherpeton* and another *Candidatus* genus Thermochlorobacter. In addition, two uncultured organisms with assembled genomes known by the names *Chlorobium* sp. 445 and *Candidatus* Thermochlorobacteriaceae GBChlB are also members of this family. Like the *Chlorobiaceae* family, members of this family contain type-I reaction centers and chlorosomes. The sole cultured species from this family is *Chloroherpeton thalassium*, which like members of the *Chlorobiaceae* family is a photolithoautotroph that grows under anoxic conditions in the presence of light using reduced sulfur compounds as an electron source to fix CO_2. However, unlike *Chlorobiaceae* species, which are nonmotile, cells of *Chloroherpeton thalassium* exhibit gliding mobility. In contrast to *Chloroherpeton*, the *Candidatus* species Thermochlorobacter aerophilum is an aerobic photoheterotroph that cannot oxidize sulfur compounds, cannot fix N_2, and does not fix CO_2 autotrophically [27]. It is considered to have gliding and flexing motility and two copies of the type-1 NADH dehydrogenase complex similarly to *Chloroherpeton*

thalassium. Members of this family are characterized by their lack of a dissimilatory sulfite reductase (DSR) system, which is a protein present in all members of *Chlorobiaceae* except for *Chlorobium ferroxidans*. Members of this family form a monophyletic clade in a phylogenetic tree based on concatenated sequences of core proteins from the genomes of *Chlorobiales* species. The members of this family are also clearly differentiated from the *Chlorobiaceae* species in a pairwise matrix of the percentage of conserved proteins in the genomes. Furthermore, members of this family can be reliably distinguished from members of the family *Chlorobiaceae* and all other bacteria based on the shared presence of six conserved signature indels (CSIs) found in diverse proteins (Table 2) that are uniquely shared by members of this family. The proteins containing the CSIs specific for this family are: alkaline phosphatase family protein, dihydrolipoyl dehydrogenase, hypoxanthine phosphoribosyltransferase, SDR family oxidoreductase, RecQ family ATP-dependent DNA helicase, and UDP-glucose/GDP-mannose dehydrogenase family protein.

The type genus for this family is *Chloroherpeton* Gibson et al. 1985 [13,82].

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/microorganisms10071312/s1, Figure S1. A bootstrapped maximum likelihood tree for different genome-sequenced *Chlorobia/Ignavibacteria* and related species/strains based on concatenated sequences of UvrD and PolA proteins. Figure S2. A bootstrapped maximum-likelihood tree for different genome-sequenced *Chlorobia/Ignavibacteria* and related species/strains based on concatenated sequences of GyrA and GyrB proteins. Figure S3. AAI matrix indicating the pairwise percentage average amino acid identities based on core proteins for the type species of different genera within the order *Chlorobiales*. Figure S4. Partial sequence alignment of the protein glycine-tRNA synthetase, showing a 1 aa insertion (boxed) that is specific for the order *Chlorobiales*. Figure S5. Partial sequence alignment of the protein RNA polymerase sigma factor RpoD/SigA, showing a 12 aa insertion (boxed) that is exclusive to all members belonging to the order *Chlorobiales*. Figure S6 Partial sequence alignment of the protein uroporphyrinogen decarboxylase, showing a 3 aa insertion (boxed) that is exclusive to all members belonging to the order *Chlorobiales*. Figure S7. Partial sequence alignment of the protein glutamyl-tRNA reductase, showing a 1 aa insertion (boxed) that is exclusive to all members belonging to the order *Chlorobiales*. Figure S8. Partial sequence alignment of the protein glutamate-1-semialdehyde 2,1-aminomutase, showing a 1 aa insertion (boxed) that is found in all members belonging to the order *Chlorobiales*. Figure S9. Partial sequence alignment of the protein class II fructose-1,6-bisphosphate aldolase, showing an 11 aa deletion (boxed) that is found in all members belonging to the order *Chlorobiales*. Figure S10. Partial sequence alignment of the protein major facilitator superfamily transporter, showing a 1 aa insertion (boxed) that is specific for members of the order *Chlorobiales*. Figure S11. Partial sequence alignment of the protein DNA polymerase III subunit alpha, showing a 2–4 aa insert (boxed) that is specific for all members belonging to the order *Chlorobiales*. Figure S12. Partial sequence alignment of the protein acetyl-CoA carboxylase carboxyltransferase subunit alpha, showing a 6 aa insertion (boxed) that is found in all members belonging to the order *Chlorobiales*. Figure S13. Partial sequence alignment of the protein N-acetyl-alpha-D-glucosaminyl L-malate synthase BshA, showing a 3 aa insertion (boxed) that is found in all members belonging to the order *Chlorobiales*. Figure S14. Partial sequence alignment of the protein serine hydroxymethyltransferase, showing a 2 aa insertion (boxed) that is found in all members belonging to the order *Chlorobiales*. Figure S15. Partial sequence alignment of the protein deoxyribonucleotide triphosphate pyrophosphatase, showing a 3 aa insertion (boxed) that is found in all members belonging to the order *Chlorobiales*. Figure S16. Partial sequence alignment of the protein polynucleotide phosphorylase/polyadenylase, showing a 1 aa insertion (boxed) that is found in all members belonging to the order *Chlorobiales*. Figure S17. Partial sequence alignment of the protein polynucleotide phosphorylase/polyadenylase, showing a 17 aa insertion (boxed) that is found in all members belonging to the order *Chlorobiales*. Figure S18. Partial sequence alignment of the protein 7-carboxy-7-deazaguanine synthase QueE, showing a 3–7 aa insertion (boxed) that is found in all members belonging to the order *Chlorobiales*. While other *Chlorobia* species contain a 7 aa insertion, *Chlorobaculum* species have a 3 aa insertion in this position. Figure S19. Partial sequence alignment of the protein DNA polymerase III subunits gamma and tau, showing a 2 aa insertion (boxed) that is found in all members belonging to the order *Chlorobiales*. Figure S20. Partial sequence alignment of the protein DNA polymerase III subunits gamma and tau, showing a 4 aa insertion (boxed) that is

found in all members belonging to the order *Chlorobiales.* Figure S21. Partial sequence alignment of the protein cytidylate kinase, showing a 4 aa insertion (boxed) that is found in all members belonging to the order *Chlorobiales*. Figure S22. Partial sequence alignment of the protein 4-hydroxy-3-methylbut-2-enyl diphosphate reductase, showing a 2–5 aa insertion (boxed) that is found in all members belonging to the order *Chlorobiales*. Figure S23. Partial sequence alignment of the protein molecular chaperone DnaK, showing a 1 aa deletion (boxed) that is found in all members belonging to the order *Chlorobiales*. Figure S24. Partial sequence alignment of the protein tryptophanyl-tRNA synthetase, showing an 8 aa insert (boxed) that is found in all members belonging to the order *Chlorobiales*. Figure S25. Partial sequence alignment of the protein tryptophanyl-tRNA synthetase, showing a 5 aa insert (boxed) that is found in all members belonging to the order *Chlorobiales*. Figure S26. Partial sequence alignment of the protein succinyl-CoA synthetase subunit alpha, showing a 7 aa insert (boxed) that is found in all members belonging to the order *Chlorobiales*. Figure S27. Partial sequence alignment of the protein Gfo/Idh/MocA family oxidoreductase, showing a 4 aa insert (boxed) that is found in all members belonging to the order *Chlorobiales*. Figure S28. Partial sequence alignment of the protein Gfo/Idh/MocA family oxidoreductase, showing a 2 aa insert (boxed) that is found in all members belonging to the order *Chlorobiales*. Figure S29. Partial sequence alignment of the protein ribonuclease R, showing a 7–8 aa insert (boxed) that is found in all members belonging to the order *Chlorobiales*. Figure S30. Partial sequence alignment of the protein ribonuclease R, showing a 4–5 aa insert (boxed) that is found in all members belonging to the order *Chlorobiales.* Figure S31. Partial sequence alignment of the protein ribonuclease R, showing a 4 aa insert (boxed) that is found in all members belonging to the order *Chlorobiales*. Figure S32. Partial sequence alignment of the protein bifunctional 5,10-methylene-tetrahydrofolate dehydrogenase, showing a 2 aa insert (boxed) that is found in all members belonging to the order *Chlorobiales*. Figure S33. Partial sequence alignment of the protein lactoylglutathione lyase, showing a 2 aa insert (boxed) that is found in all members belonging to the order *Chlorobiales*. Figure S34. Partial sequence alignment of the protein thiazole synthase, showing a 1 aa insert (boxed) that is found in all members belonging to the order *Chlorobiales*. Figure S35. Partial sequence alignment of the protein DNA mismatch repair protein (MutS), showing a 1 aa deletion (boxed) that is found in all members belonging to the order *Chlorobiales*. Figure S36. Partial sequence alignment of the protein peptide chain release factor 3, showing a 1 aa deletion (boxed) that is specific for members of the family *Chlorobiaceae*. Figure S37. Partial sequence alignment of the protein U32 family peptidase, showing a 6 aa insertion (boxed) that is specific for members of the family *Chlorobiaceae*. Figure S38. Partial sequence alignment of the MiaB-like tRNA modifying protein, showing a 5 aa insertion (boxed) that is specific for members belonging to the family *Chlorobiaceae*. Figure S39. Partial sequence alignment of the molecular chaperone HtpG protein, showing a 3 aa insertion (boxed) that is specific for members of the family *Chlorobiaceae*. Figure S40. Partial sequence alignment of the DegT/DnrJ/EryC1/StrS family aminotransferase protein, showing a 1 aa deletion (boxed) that is specific for the family *Chlorobiaceae*. Figure S41. Partial sequence alignment of the biogenesis of lysosome-related organelles complex 1 subunit 2 protein, showing a 1 aa deletion (boxed) that is specific for members of the family *Chlorobiaceae*. Figure S42. Partial sequence alignment of the protein DNA gyrase subunit A, showing a 1 aa insertion (boxed) that is specific for members belonging to the family *Chlorobiaceae*. Figure S43. Partial sequence alignment of the protein hypoxanthine phosphoribosyltransferase, showing a 1 aa insertion (boxed) that is exclusively shared by all members of the family *Chloroherpetonaceae*. Figure S44. Partial sequence alignment of the protein dihydrolipoyl dehydrogenase containing a 1 aa insertion (boxed) that is specific for members of the family *Chloroherpetonaceae*. Figure S45. Partial sequence alignment of the protein SDR family oxidoreductase, showing a 4 aa insertion (boxed) that is specific for members of the family *Chloroherpetonaceae*. Figure S46. Partial sequence alignment of the protein RecQ family ATP-dependent DNA helicase, showing a 2 aa insertion (boxed) that is specific for members of the family *Chloroherpetonaceae*. Figure S47. Partial sequence alignment of the alkaline phosphatase family protein, showing a 5 aa insertion (boxed) that is specific for members of the family *Chloroherpetonaceae*. Figure S48. Partial sequence alignment of the protein tRNA dihydrouridine synthase DusB, showing a 1 aa insertion (boxed) that is specific for members of the order *Chlorobiales* and the "larger *Ignavibacteriae* clade".

References

1. Pfennig, N. Chlorobium. In *Bergey's Manual of Systematics of Archaea and Bacteria*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2015. [CrossRef]
2. Pfennig, N. Chlorobium phaeobacteroides nov. spec. and C. phaeovibrioides nov. spec., two new species of green sulfur bacteria. *Arch. Mikrobiol.* **1968**, *63*, 224–226. [CrossRef]
3. Imhoff, J.F.; Thiel, V. Phylogeny and taxonomy of Chlorobiaceae. *Photosynth. Res.* **2010**, *104*, 123–136. [CrossRef] [PubMed]
4. Imhoff, J.F. The Family *Chlorobiaceae*. In *The Prokaryotes—Other Major Lineages of Bacteria and the Archaea*, 4th ed.; Rosenberg, E., De Long, E.F., Lory, S., Stackebrandt, E., Thompson, F., Eds.; Springer: Berlin/Heidelberg, Germany, 2014. [CrossRef]
5. Overmann, J. The Family Chlorobiaceae. In *The Prokaryotes*, 3rd ed.; Springer: New York, NY, USA, 2003.
6. Kondratieva, E.N.; Pfennig, N.; Truper, H.G. The Phototrophic Prokaryotes. In *The Prokaryotes*, 2nd ed.; Balows, A., Truper, H.G., Dworkin, M., Harder, W., Schleifer, K.H., Eds.; Springer-Verlag: New York, NY, USA, 1992; Volume 1, pp. 312–330.
7. Iino, T.; Mori, K.; Uchino, Y.; Nakagawa, T.; Harayama, S.; Suzuki, K. *Ignavibacterium album* gen. nov., sp. nov., a moderately thermophilic anaerobic bacterium isolated from microbial mats at a terrestrial hot spring and proposal of Ignavibacteria classis nov., for a novel lineage at the periphery of green sulfur bacteria. *Int. J. Syst. Evol. Microbiol.* **2010**, *60*, 1376–1382. [CrossRef]
8. Skerman, V.B.D.; McGowan, V.; Sneath, P.H.A. Approved lists of bacterial names. *Int. J. Syst. Bacteriol.* **1980**, *30*, 225–420. [CrossRef]
9. Truper, H.G.; Pfennig, N. The Family Chlorobiaceae. In *The Prokaryotes*, 2nd ed.; Balows, A., Truper, H.G., Dworkin, M., Harder, W., Schleifer, K.H., Eds.; Springer: New York, NY, USA, 1992; pp. 3583–3592.
10. Gibbons, N.E.; Murray, R.G.E. Proposals Concerning the Higher Taxa of Bacteri. *Int. J. Syst. Bacteriol.* **1978**, *28*, 1–6. [CrossRef]
11. Parte, A.C. LPSN—The List of Prokaryotic Names with Standing in Nomenclature. *Int. J. Syst. Evol. Microbiol.* **2018**, *68*, 1825–1829. [CrossRef] [PubMed]
12. Imhoff, J.F. Phylogenetic taxonomy of the family Chlorobiaceae on the basis of 16S rRNA and fmo (Fenna-Matthews-Olson protein) gene sequences. *Int. J. Syst. Evol. Microbiol.* **2003**, *53*, 941–951. [CrossRef]
13. Gibson, J.; Pfennig, N.; Waterbury, J.B. Chloroherpeton thalassium gen. nov. et spec. nov., a non-filamentous, flexing and gliding green sulfur bacterium. *Arch. Microbiol.* **1984**, *138*, 96–101. [CrossRef]
14. Gorlenko, V.M.; Lebedeva, E.V. New green sulfur bacteria with apophyses. *Mikrobiologiia* **1971**, *40*, 1035–1039.
15. Alexander, B.; Andersen, J.H.; Cox, R.P.; Imhoff, J.F. Phylogeny of green sulfur bacteria on the basis of gene sequences of 16S rRNA and of the Fenna-Matthews-Olson protein. *Arch. Microbiol.* **2002**, *178*, 131–140. [CrossRef]
16. Panwar, P.; Allen, M.A.; Williams, T.J.; Haque, S.; Brazendale, S.; Hancock, A.M.; Paez-Espino, D.; Cavicchioli, R. Remarkably coherent population structure for a dominant Antarctic Chlorobium species. *Microbiome* **2021**, *9*, 231. [CrossRef] [PubMed]
17. Tsuji, J.M.; Tran, N.; Schiff, S.L.; Venkiteswaran, J.J.; Molot, L.A.; Tank, M.; Hanada, S.; Neufeld, J.D. Anoxygenic photosynthesis and iron-sulfur metabolic potential of Chlorobia populations from seasonally anoxic Boreal Shield lakes. *ISME J.* **2020**, *14*, 2732–2747. [CrossRef] [PubMed]
18. Oren, A.; Garrity, G.M. Valid publication of the names of forty-two phyla of prokaryotes. *Int. J. Syst. Evol. Microbiol.* **2021**, *71*, 005056. [CrossRef] [PubMed]
19. Gupta, R.S.; Lorenzini, E. Phylogeny and molecular signatures (conserved proteins and indels) that are specific for the *Bacteroidetes* and *Chlorobi* species. *BMC Evol. Biol.* **2007**, *7*, 71. [CrossRef]
20. Gupta, R.S. The Phylogeny and Signature Sequences characteristics of *Fibrobacters, Chlorobi* and *Bacteroidetes*. *Crit. Rev. Microbiol.* **2004**, *30*, 123–143. [CrossRef]
21. Gupta, R.S. Impact of genomics on the understanding of microbial evolution and classification: The importance of Darwin's views on classification. *FEMS Microbiol. Rev.* **2016**, *40*, 520–553. [CrossRef]
22. Kadnikov, V.V.; Mardanov, A.V.; Podosokorskaya, O.A.; Gavrilov, S.N.; Kublanov, I.V.; Beletsky, A.V.; Bonch-Osmolovskaya, E.A.; Ravin, N.V. Genomic analysis of Melioribacter roseus, facultatively anaerobic organotrophic bacterium representing a novel deep lineage within Bacteriodetes/Chlorobi group. *PLoS ONE* **2013**, *8*, e53047. [CrossRef]

23. Podosokorskaya, O.A.; Kadnikov, V.V.; Gavrilov, S.N.; Mardanov, A.V.; Merkel, A.Y.; Karnachuk, O.V.; Ravin, N.V.; Bonch-Osmolovskaya, E.A.; Kublanov, I.V. Characterization of Melioribacter roseus gen. nov., sp. nov., a novel facultatively anaerobic thermophilic cellulolytic bacterium from the class Ignavibacteria, and a proposal of a novel bacterial phylum Ignavibacteriae. *Environ. Microbiol.* **2013**, *15*, 1759–1771. [CrossRef]

24. Iino, T. Ignavibacteriae. In *Bergey's Manual of Systematics of Archaea and Bacteria*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2018; p. 1. [CrossRef]

25. Roy, C.; Bakshi, U.; Rameez, M.J.; Mandal, S.; Haldar, P.K.; Pyne, P.; Ghosh, W. Phylogenomics of an uncultivated, aerobic and thermophilic, photoheterotrophic member of Chlorobia sheds light into the evolution of the phylum Chlorobi. *Comput. Biol. Chem.* **2019**, *80*, 206–216. [CrossRef]

26. Hiras, J.; Wu, Y.W.; Eichorst, S.A.; Simmons, B.A.; Singer, S.W. Refining the phylum Chlorobi by resolving the phylogeny and metabolic potential of the representative of a deeply branching, uncultivated lineage. *ISME J.* **2016**, *10*, 833–845. [CrossRef]

27. Liu, Z.; Klatt, C.G.; Ludwig, M.; Rusch, D.B.; Jensen, S.I.; Kuhl, M.; Ward, D.M.; Bryant, D.A. 'Candidatus Thermochlorobacter aerophilum:' an aerobic chlorophotoheterotrophic member of the phylum Chlorobi defined by metagenomics and metatranscriptomics. *ISME J.* **2012**, *6*, 1869–1882. [CrossRef] [PubMed]

28. Mukherjee, S.; Seshadri, R.; Varghese, N.J.; Eloe-Fadrosh, E.A.; Meier-Kolthoff, J.P.; Goker, M.; Coates, R.C.; Hadjithomas, M.; Pavlopoulos, G.A.; Paez-Espino, D.; et al. 1003 reference genomes of bacterial and archaeal isolates expand coverage of the tree of life. *Nat. Biotechnol.* **2017**, *35*, 676–683. [CrossRef] [PubMed]

29. Whitman, W.B. Genome sequences as the type material for taxonomic descriptions of prokaryotes 1. *Syst. Appl. Microbiol.* **2015**, *38*, 217–222. [CrossRef]

30. Wu, L.; McCluskey, K.; Desmeth, P.; Liu, S.; Hideaki, S.; Yin, Y.; Moriya, O.; Itoh, T.; Kim, C.Y.; Lee, J.S.; et al. The global catalogue of microorganisms 10K type strain sequencing project: Closing the genomic gaps for the validly published prokaryotic and fungi species. *Gigascience* **2018**, *7*, giy026. [CrossRef] [PubMed]

31. Sayers, E.W.; Agarwala, R.; Bolton, E.E.; Brister, J.R.; Canese, K.; Clark, K.; Connor, R.; Fiorini, N.; Funk, K.; Hefferon, T.; et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2019**, *47*, D23–D28. [CrossRef]

32. Parks, D.H.; Chuvochina, M.; Waite, D.W.; Rinke, C.; Skarshewski, A.; Chaumeil, P.A.; Hugenholtz, P. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **2018**, *36*, 996–1004. [CrossRef]

33. Rokas, A.; Williams, B.L.; King, N.; Carroll, S.B. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **2003**, *425*, 798–804. [CrossRef]

34. Wu, D.; Jospin, G.; Eisen, J.A. Systematic identification of gene families for use as "markers" for phylogenetic and phylogeny-driven ecological studies of bacteria and archaea and their major subgroups 1. *PLoS ONE* **2013**, *8*, e77033. [CrossRef]

35. Gupta, R.S.; Patel, S.; Saini, N.; Chen, S. Robust demarcation of 17 distinct *Bacillus* species clades, proposed as novel *Bacillaceae* genera, by phylogenomics and comparative genomic analyses: Description of *Robertmurraya kyonggiensis* sp. nov. and proposal for an emended genus *Bacillus* limiting it only to the members of the Subtilis and Cereus clades of species. *Int. J. Syst. Evol. Microbiol.* **2020**, *70*, 5753–5798.

36. Hugenholtz, P.; Chuvochina, M.; Oren, A.; Parks, D.H.; Soo, R.M. Prokaryotic taxonomy and nomenclature in the age of big sequence data. *ISME J.* **2021**, *15*, 1879–1892. [CrossRef]

37. Gupta, R.S. Microbial Taxonomy: How and Why Name Changes Occur and Their Significance for (Clinical) Microbiology. *Clin. Chem.* **2021**, *68*, 134–137. [CrossRef] [PubMed]

38. Patel, S.; Gupta, R.S. A Phylogenomics and Comparative Genomic Framework for Resolving the Polyphyly of the Genus *Bacillus*: Proposal for six new genera of *Bacillus* species, *Peribacillus* gen. nov., *Cytobacillus* gen. nov., *Mesobacillus* gen. nov., *Neobacillus* gen. nov., *Metabacillus* gen. nov. and *Alkalihalobacillus* gen. nov. *Int. J. Syst. Evol. Microbiol.* **2020**, *70*, 406–438.

39. Gupta, R.S.; Lo, B.; Son, J. Phylogenomics and Comparative Genomic Studies Robustly Support Division of the Genus *Mycobacterium* into an Emended Genus *Mycobacterium* and Four Novel Genera. *Front. Microbiol.* **2018**, *9*, 67. [CrossRef] [PubMed]

40. Adeolu, M.; Alnajar, S.; Naushad, S.; Gupta, R.S. Genome-based phylogeny and taxonomy of the 'Enterobacteriales': Proposal for *Enterobacterales* ord. nov. divided into the families *Enterobacteriaceae*, *Erwiniaceae* fam. nov., *Pectobacteriaceae* fam. nov., *Yersiniaceae* fam. nov., *Hafniaceae* fam. nov., Morganellaceae fam. nov., and *Budviciaceae* fam. nov. *Int. J. Syst. Evol. Microbiol.* **2016**, *66*, 5575–5599.

41. Rokas, A.; Holland, P.W. Rare genomic changes as a tool for phylogenetics. *Trends Ecol. Evol.* **2000**, *15*, 454–459. [CrossRef]

42. Gupta, R.S. Molecular signatures for the main phyla of photosynthetic bacteria and their subgroups. *Photosynth. Res.* **2010**, *104*, 357–372. [CrossRef]

43. Gupta, R.S.; Chander, P.; George, S. Phylogenetic framework and molecular signatures for the class Chloroflexi and its different clades; proposal for division of the class Chloroflexia class. nov. [corrected] into the suborder Chloroflexineae subord. nov., consisting of the emended family Oscillochloridaceae and the family Chloroflexaceae fam. nov., and the suborder Roseiflexineae subord. nov., containing the family Roseiflexaceae fam. nov. *Antonie Van Leeuwenhoek* **2013**, *103*, 99–119.

44. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Soding, J.; et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539. [CrossRef]

45. Capella-Gutierrez, S.; Silla-Martinez, J.M.; Gabaldon, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **2009**, *25*, 1972–1973. [CrossRef]

46. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS ONE* **2010**, *5*, e9490. [CrossRef]

47. Whelan, S.; Goldman, N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **2001**, *18*, 691–699. [CrossRef] [PubMed]

48. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [CrossRef] [PubMed]

49. Le, S.Q.; Gascuel, O. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **2008**, *25*, 1307–1320. [CrossRef] [PubMed]

50. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef]

51. Thompson, C.C.; Amaral, G.R.; Campeao, M.; Edwards, R.A.; Polz, M.F.; Dutilh, B.E.; Ussery, D.W.; Sawabe, T.; Swings, J.; Thompson, F.L. Microbial taxonomy in the post-genomic era: Rebuilding from scratch? 3. *Arch. Microbiol.* **2015**, *197*, 359–370. [CrossRef]

52. Gupta, R.S. Distinction between Borrelia and Borreliella is more robustly supported by molecular and phenotypic characteristics than all other neighbouring prokaryotic genera: Response to Margos' et al. "The genus Borrelia reloaded" (*PLoS ONE* 13(12): e0208432). *PLoS ONE* **2019**, *14*, e0221397. [CrossRef]

53. Qin, Q.L.; Xie, B.B.; Zhang, X.Y.; Chen, X.L.; Zhou, B.C.; Zhou, J.; Oren, A.; Zhang, Y.Z. A proposed genus boundary for the prokaryotes based on genomic insights. *J. Bacteriol.* **2014**, *196*, 2210–2215. [CrossRef]

54. Yilmaz, P.; Parfrey, L.W.; Yarza, P.; Gerken, J.; Pruesse, E.; Quast, C.; Schweer, T.; Peplies, J.; Ludwig, W.; Glockner, F.O. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res.* **2014**, *42*, D643–D648. [CrossRef]

55. Tamura, K.; Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **1993**, *10*, 512–526.

56. Gupta, R.S. Identification of Conserved Indels that are Useful for Classification and Evolutionary Studies. In *Bacterial Taxonomy, Methods in Microbiology Volume 41*; Goodfellow, M., Sutcliffe, I.C., Chun, J., Eds.; Elsevier: London, UK, 2014; pp. 153–182.

57. Naushad, H.S.; Lee, B.; Gupta, R.S. Conserved signature indels and signature proteins as novel tools for understanding microbial phylogeny and systematics: Identification of molecular signatures that are specific for the phytopathogenic genera Dickeya, Pectobacterium and Brenneria. *Int. J. Syst. Evol. Microbiol.* **2014**, *64*, 366–383. [CrossRef]

58. Gibson, J. Chloroherpeton. In *Bergey's Manual of Systematics of Archaea and Bacteria*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2015. [CrossRef]

59. Kantor, R.S.; van Zyl, A.W.; van Hille, R.P.; Thomas, B.C.; Harrison, S.T.; Banfield, J.F. Bioreactor microbial ecosystems for thiocyanate and cyanide degradation unravelled with genome-resolved metagenomics. *Environ. Microbiol.* **2015**, *17*, 4929–4941. [CrossRef] [PubMed]

60. Eloe-Fadrosh, E.A.; Paez-Espino, D.; Jarett, J.; Dunfield, P.F.; Hedlund, B.P.; Dekas, A.E.; Grasby, S.E.; Brady, A.L.; Dong, H.; Briggs, B.R.; et al. Global metagenomic survey reveals a new bacterial candidate phylum in geothermal springs. *Nat. Commun.* **2016**, *7*, 10476. [CrossRef] [PubMed]

61. Bhandari, V.; Gupta, R.S. Phylum Thermotogae. In *The Prokaryotes—Other Major Lineages of Bacteria and the Archaea*, 4th ed.; Rosenberg, E., DeLong, E., Lory, S., Stackebrandt, E., Thompson, F., Eds.; Springer: New York, NY, USA, 2014; Volume 11, pp. 989–1015.

62. Jiang, L.; Wang, D.; Kim, J.S.; Lee, J.H.; Kim, D.H.; Kim, S.W.; Lee, J. Reclassification of genus *Izhakiella* into the family *Erwiniaceae* based on phylogenetic and genomic analyses. *Int. J. Syst. Evol. Microbiol.* **2020**, *70*, 3541–3546. [CrossRef] [PubMed]

63. Dobritsa, A.P.; Samadpour, M. Reclassification of *Burkholderia insecticola* as *Caballeronia insecticola* comb. nov. and reliability of conserved signature indels as molecular synapomorphies. *Int. J. Syst. Evol. Microbiol.* **2019**, *69*, 2057–2063. [CrossRef]

64. Hohmann-Marriott, M.F.; Blankenship, R.E. Evolution of photosynthesis. *Annu. Rev. Plant. Biol.* **2011**, *62*, 515–548. [CrossRef]

65. Bryant, D.A.; Costas, A.M.; Maresca, J.A.; Chew, A.G.; Klatt, C.G.; Bateson, M.M.; Tallon, L.J.; Hostetler, J.; Nelson, W.C.; Heidelberg, J.F.; et al. Candidatus Chloracidobacterium thermophilum: An aerobic phototrophic Acidobacterium. *Science* **2007**, *317*, 523–526. [CrossRef]

66. Liu, Z.; Frigaard, N.-U.; Vogl, K.; Iino, T.; Ohkuma, M.; Overmann, J.; Bryant, D.A. Complete Genome of Ignavibacterium album, a Metabolically Versatile, Flagellated, Facultative Anaerobe from the Phylum Chlorobi. *Front. Microbiol.* **2012**, *3*, 185. [CrossRef]

67. Gregersen, L.H.; Bryant, D.A.; Frigaard, N.U. Mechanisms and evolution of oxidative sulfur metabolism in green sulfur bacteria. *Front. Microbiol.* **2011**, *2*, 116. [CrossRef]

68. Holkenbrink, C.; Barbas, S.O.; Mellerup, A.; Otaki, H.; Frigaard, N.U. Sulfur globule oxidation in green sulfur bacteria is dependent on the dissimilatory sulfite reductase system. *Microbiology* **2011**, *157*, 1229–1239. [CrossRef]

69. Felsenstein, J. *Inferring Phylogenies*; Sinauer Associates, Inc.: Sunderland, MA, USA, 2004.

70. Baldauf, S.L. Phylogeny for the faint of heart: A tutorial. *Trends Genet.* **2003**, *19*, 345–351. [CrossRef]

71. Gupta, R.S. Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. *Microbiol. Mol. Biol. Rev.* **1998**, *62*, 1435–1491. [CrossRef] [PubMed]

72. Barbour, A.G.; Adeolu, M.; Gupta, R.S. Division of the genus *Borrelia* into two genera (corresponding to Lyme disease and relapsing fever groups) reflects their genetic and phenotypic distinctiveness and will lead to a better understanding of these two groups of microbes (Margos et al. (2016) There is inadequate evidence to support the division of the genus *Borrelia*. Int. J. Syst. Evol. Microbiol. doi: 10.1099/ijsem.0.001717). *Int. J. Syst. Evol. Microbiol.* **2017**, *67*, 2058–2067. [CrossRef] [PubMed]

73. Gupta, R.S. Protein signatures (molecular synapomorphies) that are distinctive characteristics of the major cyanobacterial clades. *Int. J. Syst. Evol. Microbiol.* **2009**, *59*, 2510–2526. [CrossRef] [PubMed]

74. Gao, B.; Gupta, R.S. Conserved indels in protein sequences that are characteristic of the phylum Actinobacteria. *Int. J. Syst. Evol. Microbiol.* **2005**, *55*, 2401–2412. [CrossRef] [PubMed]

75. Ahmod, N.Z.; Gupta, R.S.; Shah, H.N. Identification of a *Bacillus anthracis* specific indel in the yeaC gene and development of a rapid pyrosequencing assay for distinguishing *B. anthracis* from the *B. cereus* group. *J. Microbiol. Methods* **2011**, *87*, 278–285. [CrossRef]

76. Wong, S.Y.; Paschos, A.; Gupta, R.S.; Schellhorn, H.E. Insertion/deletion-based approach for the detection of Escherichia coli O157:H7 in freshwater environments. *Environ. Sci. Technol.* **2014**, *48*, 11462–11470. [CrossRef]

77. Singh, B.; Gupta, R.S. Conserved inserts in the Hsp60 (GroEL) and Hsp70 (DnaK) proteins are essential for cellular growth. *Mol. Genet. Genom.* **2009**, *281*, 361–373. [CrossRef] [PubMed]

78. Khadka, B.; Gupta, R.S. Identification of a Conserved 8 aa Insert in the PIP5K Protein in the Saccharomycetaceae family of Fungi and the Molecular Dynamics Simulations and Structural Analysis to investigate its Potential Functional Role. *Proteins* **2017**, *85*, 1454–1467. [CrossRef]

79. Khadka, B.; Adeolu, M.; Blankenship, R.E.; Gupta, R.S. Novel insights into the origin and diversification of photosynthesis based on analyses of conserved indels in the core reaction center proteins. *Photosynth. Res.* **2016**, *131*, 159–171. [CrossRef]

80. Gupta, R.S. Moelcular Markers for Photosynthetic Bacteria and Insights into the Origin and Spread of Photosynthesis. *Adv. Bot. Res.* **2013**, *66*, 37–66.

81. Nadson, G.A. The morphology of inferior Algae. III. Chlorobium limicola Nads., the green chlorophyll bearing microbe. *Bull. Du Jard. Bot. St. Petersbourg* **1906**, *6*, 190.

82. Gibson, J.; Pfenning, N.; Waterbury, J.B. In Validation list no. 17. Validation of publication of new names and new combinations previously effectively published outside the IJSB. *Int. J. Syst. Bacteriol.* **1985**, *35*, 223–225.

Article

Structural Analyses of CrtJ and Its B$_{12}$-Binding Co-Regulators SAerR and LAerR from the Purple Photosynthetic Bacterium *Rhodobacter capsulatus*

Vladimira Dragnea, Giovanni Gonzalez-Gutierrez and Carl E. Bauer *

Molecular and Cellular Biochemistry Department, Indiana University, Bloomington, IN 47405, USA; vdragnea@indiana.edu (V.D.); giovgonz@indiana.edu (G.G.-G.)
* Correspondence: bauer@indiana.edu

Abstract: Among purple photosynthetic bacteria, the transcription factor CrtJ is a major regulator of photosystem gene expression. Depending on growing conditions, CrtJ can function as an aerobic repressor or an anaerobic activator of photosystem genes. Recently, CrtJ's activity was shown to be modulated by two size variants of a B$_{12}$ binding co-regulator called SAerR and LAerR in *Rhodobacter capsulatus*. The short form, SAerR, promotes CrtJ repression, while the longer variant, LAerR, converts CrtJ into an activator. In this study, we solved the crystal structure of *R. capsulatus* SAerR at a 2.25 Å resolution. Hydroxycobalamin bound to SAerR is sandwiched between a 4-helix bundle cap, and a Rossman fold. This structure is similar to a AerR-like domain present in CarH from *Thermus termophilus*, which is a combined photoreceptor/transcription regulator. We also utilized AlphaFold software to predict structures for the LAerR, CrtJ, SAerR-CrtJ and LAerR-CrtJ co-complexes. These structures provide insights into the role of B$_{12}$ and an LAerR N-terminal extension in regulating the activity of CrtJ.

Keywords: PpsR ortholog; AerR photoreceptor; light regulation; photosynthesis gene regulators

1. Introduction

The presence of dioxygen has long been known to repress the synthesis of the structural and pigment components of the *Rhodobacter* photosystem [1]. Several well-characterized transcription factors, such as the RegA-RegB signaling cascade, FnrL, and CrtJ, are known to regulate the expression of photosynthesis and other genes in response to changes in cellular redox [1–5].

Orthologs of the transcription regulator CrtJ are present in many purple bacterial species, where they are responsible for controlling the expression genes involved in the synthesis of heme, BChl, carotenoids, and photosystem apoproteins [5–9]. CrtJ's main action is to repress photosynthetic gene expression under aerobic conditions [5–9], although recent transcriptomic analysis showed that CrtJ also controls anaerobic gene expression [10,11]. CrtJ binds to promoters as a tetramer at the conserved palindrome sequence TGT-N12-ACA. Two copies of this palindrome are present as closely linked (typically separated by eight nucleotides) or as two distant sites that likely come together by DNA looping [12–15]. Until recently, it was presumed that the main control of CrtJ's activity involved the oxidation/reduction of two conserved Cys, one of which is present in the DNA binding domain [16,17]. However, in vivo transcriptome analysis shows that CrtJ binds to many promoters under both aerobic and anaerobic conditions. Furthermore, CrtJ binding to target promoters often coincides with its regulatory partner AerR [10]. While AerR itself does not exhibit DNA binding activity, it is cotranscribed with CrtJ [18] and appears to be the main controller of CrtJ ability to either repress or activate gene expression [10,11,17].

It has recently been shown that *Rba. capsulatus* synthesizes two isomeric forms of AerR, a short form called SAerR, and a longer variant called LAerR with a 40-amino

acid extension at its amino terminus [11]. SAerR predominates in cells grown under dark aerobic conditions, as well as in cells that are in the stationary phase of growth. In contrast, the longer variant LAerR predominates in cells actively growing under anaerobic photosynthetic conditions (Figure 1). Yamamoto et al. [11] demonstrated that SAerR is generated from a second internal translation initiation site present within in the LAerR coding sequence. SAerR has been shown to direct CrtJ to function as an aerobic repressor at the puc (LHII), *bchEJGP, CrtA-bchDIO* and *crtB* operons [11]. Interestingly, the LAerR variant that predominates under photosynthetic conditions converts CrtJ into an activator of photosystem gene expression [11]. In vitro (DNase I protection) and in vivo (ChIP-exo) footprint studies on the *bchC* promoter show that CrtJ complexed with LAerR forms a greatly extended interaction with the *bchC* promoter that extends ~120 bases beyond which is protected by CrtJ alone [10]. The current model suggests that SAerR likely enhances CrtJ repression by increasing the binding of CrtJ to its recognition palindrome, while LAerR functions to induce an extended wrapped DNA complex around a CrtJ+LAerR co-complex that converts CrtJ into an activator (Figure 1) [11]. Unraveling details of the mechanism by which LAerR and SAerR can alter CrtJ DNA binding activity and CrtJ repressor versus activator functions is the goal of this study.
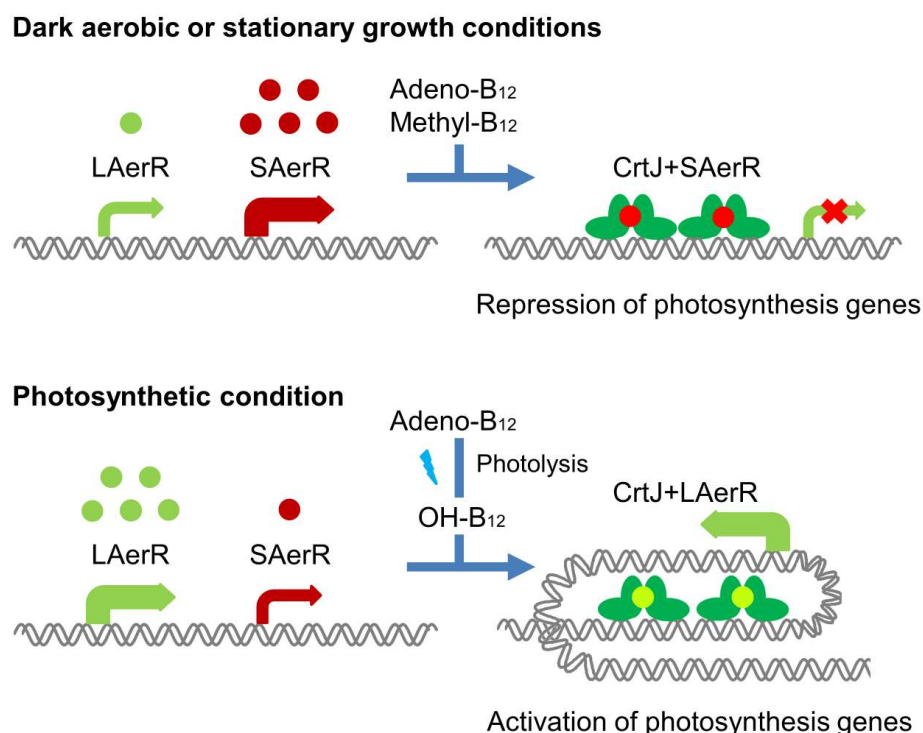
Figure 1. An updated model originally published in [11] that depicts the current understanding of the regulation of photosynthesis gene expression by CrtJ and its co-repressor SAerR (top) and its co-activator LAerR (bottom).

LAerR and SAerR also have very interesting novel functions as photoreceptors that uses B_{12} as a chromophore [11,17,19,20]. Studies have shown that LAerR tightly binds hydroxyl (OH-) cobalamin that is generated as a byproduct of light excitation of adeno or methyl-cobalamin. Cobalamin binding to LAerR involves the formation of a lower axial ligand to the corrin ring cobalt with His145 and the formation of an upper axial ligand to the same cobalt with His10 [17]. SAerR lacks the first 40 amino acids that are present in LAerR, so SAerR does not contain His10, which forms the upper axial ligand [11]. This alteration allows SAerR to exhibit less of a corrin ring selectivity, as this shorter variant binds adeno-, metyl, cyano- and hydroxyl-cobalamin irrespective of light exposure [11].

There are also interesting and informative variants in other species that exhibit similarities to the CrtJ-AerR regulatory mechanism. For example, *Thermus thermophilus* codes for a bacterial transcriptional regulator called CarH that has a B_{12}-binding AerR domain linked to a DNA-binding regulator of carotenoid biosynthesis genes [21]. In CarH, photoexcitation of adenosyl-B_{12} leads to the dissociation of the upper liganded adenosyl group from B_{12}, followed by the tight formation of a new upper ligand to a His in CarH [22,23]. CarH photochemistry is very similar to that observed with LAerR. Specifically, dark binding of adeno-B_{12} to CarH results in the formation of a CarH tetramer capable of binding and repressing carotenoid gene expression. Upon light excitation of the cobalamin ring, there is replacement of the upper adeno-Co ligand with a His that causes a conformational change in CarH which promotes dissociation into monomers incapable of binding DNA [22,23]. CarH's linked AerR-DNA binding motif thus provides *T. thermophilus* with a mechanism to induce carotenoid gene synthesis when cells are exposed to blue light. Furthermore, a crystal structure of CarH has been solved in various states, dark, free, or bound to DNA and also when light-excited [22,23]. A fast spectroscopy study [22] also provided insights into light signaling by B_{12}, which does not involve the formation of a radical pair, as is common in B_{12} photochemistry in solution or during enzymatic activity [24].

A second interesting variant occurs in *Rhodobacter sphaeroides*, which has a CrtJ ortholog called PpsR and an ortholog of AerR called PpaA [25]. The PpaA/AerR variant in this species has not been well studied beyond the observation that it also binds B_{12} in a light-dependent manner [19]. However, *Rba. sphaeroides* also codes for an AerR variant called AppA that has a flavin binding BLUF photoreceptor domain (blue light involving flavin) [26–28] linked to a AppA-like domain called the sensor containing heme instead of cobalamin (SCHIC) domain that binds heme instead of B_{12} [29,30]. Furthermore, AppA regulates PpsR activity in response to blue light absorption via FAD bound to the BLUF domain as well as heme availability sensed by the AppA-like domain [28,30,31]. As with CarH, there is a crystal structure of a PpsR deletion construct that lacks the DNA binding domain [32] as well as a crystal structure of the AppA FAD-binding BLUF domain [32,33], the SCHIC domain [30,32], and of a PpsR dimer bound to a monomer of AppA [32].

In this study, we provide a crystal structure of SAerR and models of LAerR and CrtJ that were constructed with Deep Mind AlphaFold software [34]. The LAerR and CrtJ *in-silco* models were validated by comparison with the known SAerR and PpsR structures. Docking of SAerR and LAerR to CrtJ as modeled on the solved PpsR2-AppA co-complex provides new structural insights as to the role of SAerR and LAerR in converting CrtJ from a repressor into an activator, respectively.

2. Materials and Methods

2.1. Protein Purification

We used previously described constructs pSUMO-SAerR and pSUMO-LAerR to express and purify both the short and long variants of AerR [11]. LAerR contains the full-length AerR sequence, while SAerR lacks forty N-terminal amino acids. For protein expression, the *E. coli* strain BL21(DE3) containing either pSUMO-SAerR or pSUMO-LAerR was grown on LB and kanamycin (25 μg/mL) at 37 °C to an optical density of $A_{600} = 0.6$. The flasks were then chilled to 16° C, and AerR expression was induced with the addition of IPTG to a final concentration of 50 μM with expression continued overnight. Cells were harvested by centrifugation at $10,000 \times g$ for 10 min with cell pellets stored at -80 °C until further use.

For AerR purification, cells were resuspended in buffer A: 20 mM Tris-HCl, pH 8.0, 0.3 M NaCl, 5 mM imidazole, and 10% glycerol to which 100 μM (final concentration) adenosyl cobalamin was added to the cell suspension. Cells were lysed with 3 passes through a French press with cell debris removed by centrifugation at $20,000 \times g$ for 10 min. The clarified supernatant was loaded into a superloop, briefly illuminated with strong white light (~1 min with Nikon High Intensity Illuminator NI-150 at maximum output). The sample was loaded onto the HisTrap FF column (ÄKTA FPLC) with unbound proteins

removed by washing with 20 column volumes of 45 mM imidazole in buffer A. Elution of AerR was performed with a step up of 100% buffer B (same as buffer A but containing 500 mM imidazole). In this process, 3 mL of the eluted B_{12} containing the AerR pink sample was cleaved from the SUMO tag by addition of 2 mM DTT and 0.2 μM SUMO protease, and then incubated overnight at 4 °C. The AerR sample was then desalted on Econo-Pack 10DG desalting column (Bio-Rad, Hercules, CA, USA) in buffer A. To remove the cleaved SUMO tag, the desalted sample was then passed through 1 mL of His resin (Bio-Rad) that was equilibrated with Buffer A. The sample was then concentrated to ~1 ml using 10 K Amicon concentrators and loaded onto the Superose 12 size-exclusion column pre-equilibrated with 20 mM Tris 8.0 and 200 mM NaCl. The eluted AerR protein was then concentrated to 6 mg/mL for crystallization trials.

For CrtJ purification, CrtJ was purified from a SUMO-CrtJ construct as described by Cheng et al. [35] using *E. coli* BL21DE3. After cleavage with Sumo protease, sample was applied directly on Superose 12 size-exclusion column as the SUMO tag (12 kDa) and CrtJ (>100 kDa) were easily size separated based on their different sizes.

2.2. Absorption Spectroscopy

UV-VIS spectra were recorded on Beckman DU-640 spectrophotometer.

2.3. DNase Footprint Analysis

bchC promoter (325 bp) was amplified with 6-carboxyfluorescein phosphoramidate (FAM)-labeled primers as described previously [36]. We used 10 nM of the PCR amplified *bchC* promoter DNA segment in buffer 20 mM Tris –HCl, pH 8.0, 2 mM $MgCl_2$, 0.5 mM $CaCl_2$, 0.1 mg/mL BSA, and 5 uM protein (either purified CrtJ or sAerR). After 15 min of incubation, DNase I was added to the reaction to a 4 U/μL final concentration. After 15 min of digestion, the reaction was stopped by the addition of 0.5 M EDTA. The reactions were analyzed using GENEWIZ service (Washington, DC, USA).

2.4. Crystallization

Crystallization trials of SAerR at a protein concentration of 6 mg/mL used a hanging drop vapor diffusion method 1:1 *v/v* of protein to Hampton Research Index screen trial buffer. Protein crystallized as small diamond-shaped pink crystals that appeared within a few days in Hampton Research Index Screen N. 11. (3 M NaCl and 0.1 M Hepes 7.5). A subsequent additive screen (Hampton Research) yielded larger crystals grown with 1,6-hexandiol. The largest crystals were obtained in 3.2 M NaCl, 0.1 M Hepes pH 7.8 with 8% 1,6-hexandiol. Crystals were harvested directly from the hanging drop and flash frozen in liquid nitrogen. Diffraction data were collected at 100 K at Beamline station 4.2.2 at the Advanced Light Source (Berkeley National Laboratory, Berkeley CA, USA) and indexed, integrated, and scaled using XDS [37].

2.5. Structure Calculations

Two complete datasets collected at wavelengths of 1.00003 Å and 1.5998 Å were in space group P6122 at a 2.25 Å resolution (Supplementary Table S1). The phase estimation was performed using single anomalous diffraction (SAD). Initially, weak phases were estimated using Autosol and one Co site, which led to a figure of merit (FOM) of ~0.3 and an incomplete model containing only 26 well-placed residues of an α-helix and Rwork/Rfree of 0.4291/0.4965. This incomplete model was used to improve phase estimation by MR-SAD with a final FOM of ~0.6. This model gave rise only to Rwork/Rfree of 0.3846/0.4643, which was sufficient to perform molecular replacement with the native dataset. Successive cycles of automatic building in Autobuild (PHENIX) and manual building in Coot, as well as refinement (PHENIX Refine), led to a model missing only 8 residues (PDB code 7TE2). The final model showed a Rwork/Rfree of 0.2308/0.2590, which is approximately the 70th percentile compared to structures with a similar 2.25 Å resolution with excellent

geometry and no Ramachandran outliers. All data collection statistics are shown in shown in Supplementary Table S1.

2.6. 3D Modeling Software

Alphafold CASP13 software [34] was used to model LAerR and CrtJ protein structures. Images were generated using PyMol.

3. Results

We initially set up hanging drop crystallization screens with both the short SAerR and long LAerR variants and a complex of these variants with CrtJ bound to DNA (26 or 27 bp oligomer of the *bchC* CrtJ binding sequence). LAerR and complexes with CrtJ/DNA/AerR did not crystallize in conditions that we tried. However, SAerR readily formed crystals under several conditions in Hampton Research Index Screen N.11. After optimization screening, we obtained well-diffracting pink diamond-shaped crystals that diffracted at a 2.25 Å resolution with Rwork/Rfree values of 0.2308/0.2590. Additional data collection and final refinement statistics are presented in Supplementary Table S1.

3.1. Structure of SAerR

The solved crystal structure of SAerR containing bound cobalamin has a four-helix bundle cap followed by an α/β Rossman fold. A flexible loop between the four-helix bundle and the α/β Rossman fold is not resolved in the structure (Figure 2A). In general, this SAerR structure closely resembles other cobalamin binding domains whose structures have been solved, such as the methyl cobalamin-binding domain of methionine synthase MetH [38] and the B_{12}-binding domain in the combined photoreceptor and transcriptional regulator CarH [23]. As shown in Figure 2A, B_{12} is nestled between the four-helix bundle cap on the left and the Rossman fold on the right. The corrin ring is anchored in place by His145 in the Rossman fold that functions as the lower ligand to the Co in cobalamin (Figure 2B).

Pseudo-covalent attachment of B_{12} in AerR, as well as in CarH, involves light excitation of the adeno- or methyl-B_{12} cobalamin ring, which catalyzes the replacement of the upper Co axial ligands adeno or methyl group with another residue [17,20,23,24]. In CarH, and in LAerR, the upper adeno or methyl ligand is replaced with a histidine (His10 in AerR) [17,20,23,24]. In SAerR, His10 is not present, as this variant is 40 AA shorter at the amino terminus relative to LAerR [11]. In this case, the best fit for an upper axial ligand in the SAerR electron density map is an atom of Cl (dappled ball in Figure 2B,C). SAerR crystals were grown in a high concentration of NaCl, so Cl appears to have replaced the upper ligand –OH that was originally present in hydroxycobalamin that was light-attached to SAerR. Two additional Cl atoms were found in the structure, as well as a molecule of the additive 1,6- hexanediol that we used for the optimization of crystal growth. Analysis of the region near the upper Cl axial ligand shows that it is protected by two aromatic residues, Trp94 and Phe101, in SAerR (Figure 2B,C). Trp94 is also conserved in CarH but is replaced by Phe in MetH where the upper axial ligand to B_{12} Co is a methyl leaving group.

Several hydrogen bonds stabilize cobalamin in place by interacting with its long tail (Figure 2D). Specifically, there are H-bonds between nitrogen in the 5,6-dimethylbenzimidazole ring of cobalamin and highly conserved Gly220, and with a hydroxyl group of Thr189. Thr189 is replaced in other B_{12}-binding domains such as MetH and CarH by Ser or Ala. Arg88 from the four-helix bundle and Glu155 from the Rossman fold also form a H-bond through two water molecules.
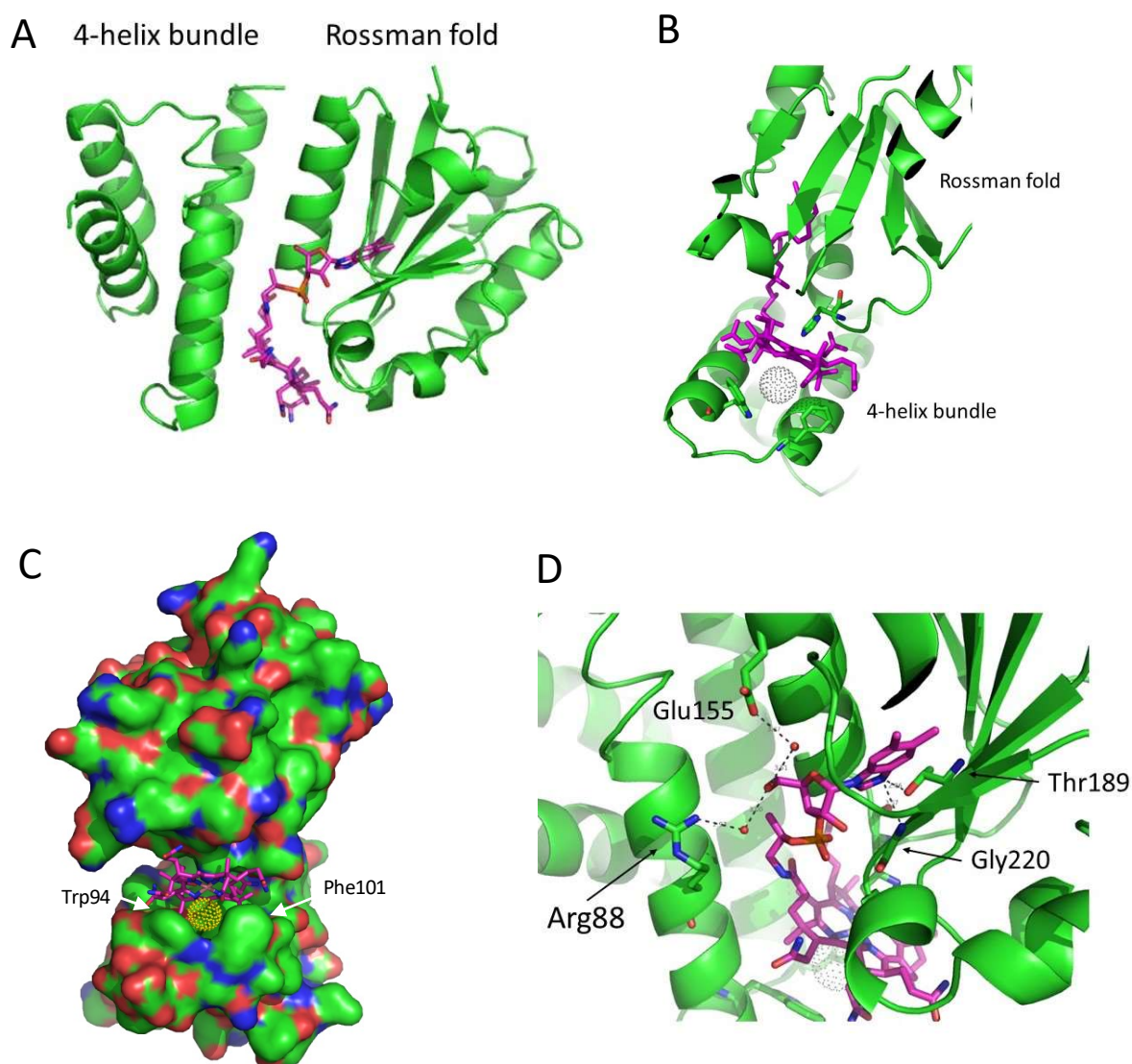
Figure 2. Ribbon representation of the SAerR crystal structure with bound cobalamin. (**A**) The flexible loop between the 4-helix bundle cap on the left and Rossman fold on the right. Missing in this structure is a flexible joining loop between these domains. (**B**) His145 forms the lower axial ligand while Cl (grey dotted circle) forms the upper axial ligand to cobalamin. Two aromatic residues Trp 94 and Phe101 in the 4-helix bundle also form a "cap" on the upper Cl ligand to cobalamin. (**C**) A space filling representation showing the cleft that holds cobalamin as well as the Cl atom (gold dotted circle) (**D**) H-bonding network with cobalamin. Red dots are water molecules. Note that the Cl atom has replaced the -OH upper ligand in the hydroxycobalamin attached to SAerR molecules with the 5-carbon sugar of cobalamin.

Interestingly, many residues forming an additional H-bond with B_{12} in the SAerR structure are not highly conserved. The few exceptions are the highly conserved residue Gly220 and the lower ligand of cobalamin His145 (marked by red boxes in Figure 2).

As discussed above, *Rba. sphaeroides* has a CrtJ orthlog called PpsR [8]. The activity of PpsR is regulated by PpaA, an AerR homolog [25], and by an additional photoreceptor/regulator called AppA [28,31]. AppA has a domain called the SCHIC domain that we previously crystallized [30]. This domain of AppA has structural features similar to that of B_{12}-binding proteins, including AerR with the caveat that the AppA SCHIC domain binds heme instead of cobalamin [29,30]. Overlaying the SAerR and AppA SCHIC domain structures shows that it aligns closely with a RMSD value of 6.1Å. Inspection of this overlay

shows several differences in the cobalamin binding region, the most obvious being an eight-amino acid loop in the SCHIC domain (residues Val329-Thr337) that likely hinders cobalamin binding (Figure 3A, blue). This loop connects the 3rd β-sheet and 3rd α-helix in the Rossman fold. In SAerR, this loop is much shorter (three residues: Val190-Met193), and in addition, the third β-sheet in SAerR is longer and tilted away from the B_{12} binding pocket providing more space for B_{12} in SAerR. His284 forms a Fe–heme ligand in SCHIC domain which corresponds to His145 in SAerR that forms the lower Co axial ligand (Figure 3B). However, the SCHIC His284 Fe ligand is shifted by two residues when aligned with other cobalamin binding domains. The "W94/F101 cap" protecting the B_{12} upper ligand in SAerR is also replaced in AppA SCHIC domain by L239 and R246, respectively. Collectively, these structural alterations between SAerR and the SCHIC domain of AppA appears to be responsible for the observed alteration of tetrapyrrole selectively between these two different photoreceptors.
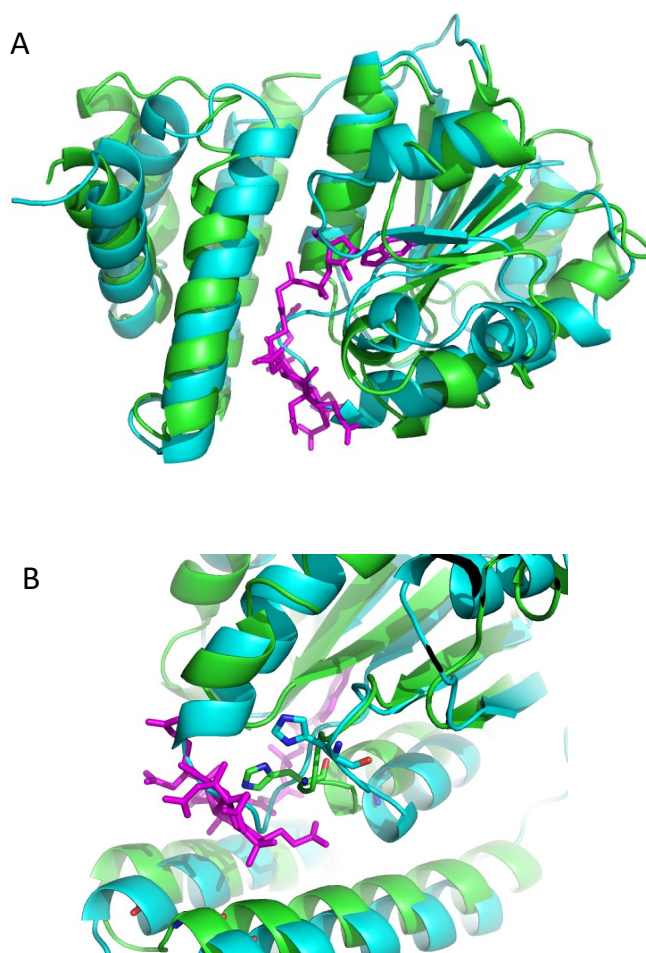
Figure 3. Structure of SAerR (green) containing cobalamin (purple) overlaying the SCHIC domain of AppA (blue). (**A**) A loop in the blue SCHIC domain which binds the heme occupies the place of cobalamin present in SAerR. (**B**) His284 in the AppA SCHIC domain is shifted by 2 positions (blue) from His145 in AerR that serves as lower axial ligand to heme and B_{12}, respectively.

3.2. Alphafold 3D Modeling of LAerR

We were not able to obtain crystals of LAerR, as this longer variant is much less soluble and more heterogeneous than SAerR, as indicated by dynamic light-scattering. However, the modeling software AlphaFold does generate predictive 3D structures of proteins from a primary sequence with high confidence [34]. Using AlphaFold, we generated several models of LAerR 3D structures with the three best superimposed over each other in Figure 4A.
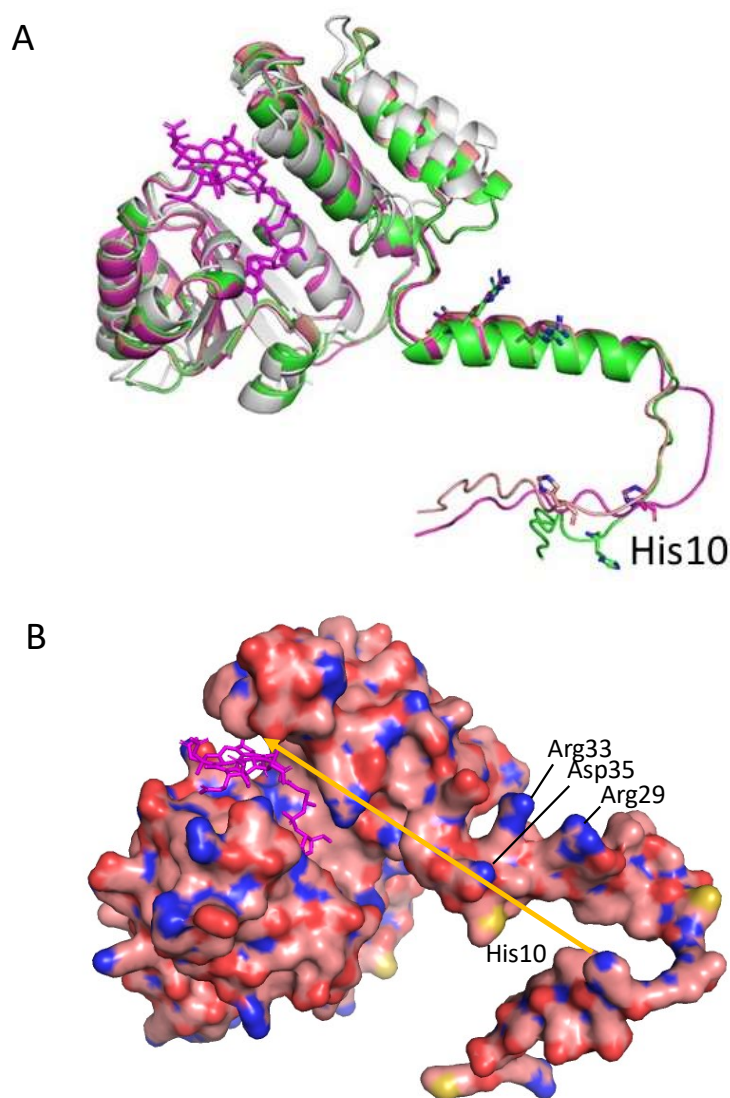
Figure 4. Crystal structure of SAerR (magenta) overlaid with the three best Alphafold predictions of LAerR. (**A**) The N-terminal 40 residues that are missing in SAerR are modeled in LAerR into an α1-helix preceded by a flexible tail that contains the upper Co ligand His10 shown as sticks. (**B**) Surface charges of LAerR with B_{12} (magenta; as positioned in SAerR). Oxygen atoms are red, nitrogen atoms are blue, and sulfur atoms are mustard yellow. The orange arrow shows the predicted movement of His10 that must occur to form the upper axial ligand to Co upon light excitation.

These predictive structures of LAerR contains a 40 amino terminal longer section as compared to our solved crystal structure of SAerR. As shown in Figure 4A, the LAerR model structures are highly superimposed over the solved SAerR structure as represented in magenta. The RMSD score between the same residues in the SAerR structure to the same residues in the LAerR structure is 2.8, which indicates an excellent model fit. The N-terminal's 40 amino acids in LAreR that are not present in SAerR are modelled into a long flexible tail (residues 1–18) followed by a long α-helix denoted as α1-helix (Phe19-Asn36). Mutational analysis of LAerR has shown that light excitation of adeno- or methyl-cobalamine allows His10 to form an upper ligand with the Co [20]. However, in this dark *in silico* model of LAerR (where no B_{12} would be attached), we see that His10 in the flexible amino-terminal tail segment is at a considerable distance from where an upper axial ligand could be formed with cobalamin (Figure 4A,B). One clue for how His10 forms the upper Co axial ligand can be derived from the crystal structure of CarH, which has a similar B_{12} binding domain. The crystal structure of dark CarH shows that the lower ligand is also

a His and that the upper ligand remains in the adeno group [22,23]. However, after light excitation of the corrin ring, the upper adeno axial ligand is released and replaced with a His, not unlike what occurs with LAerR. Furthermore, this light-mediated change in the upper axial ligand causes an ~8 Å movement of the 4-helix bundle cap, which then disrupts tetramerization and DNA binding of CarH [23]. We suspect that a similar movement of the same helix bundle cap occurs in LAerR that would allow His10 to form an axial ligand upon light excitation of B_{12}. Finally, there is a deep groove on one side of LAerR, which could potentially be involved in the docking of the α1 N-terminal helix once there is light-driven formation of a His10 Co axial ligand (Figure 4B).

3.3. Alphafold 3D Modeling of CrtJ

We also used AlphaFold to generate a predictive structure of CrtJ. CrtJ is a homolog of transcription regulator PpsR from *Rba. sphaeroides* that was previously crystallized without HTH DNA binding domain (PDB 4HH2.pdb) [32]. In Figure 5A, we show the five best in silico models of CrtJ aligned with the PpsR HTH crystal structure shown in green.
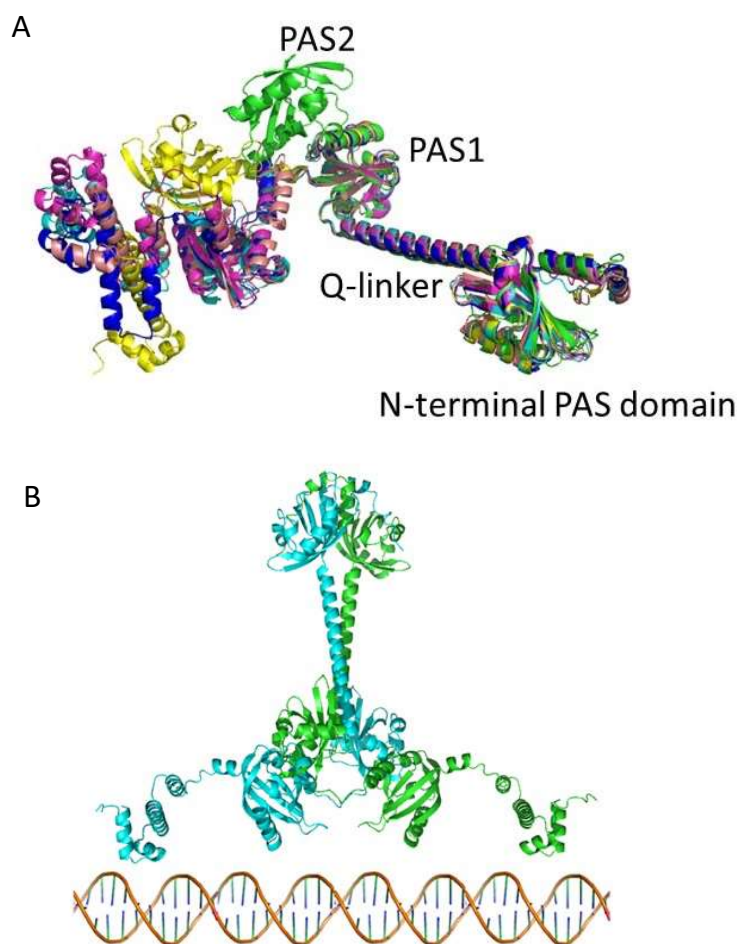
Figure 5. Crystal structure of PpsR from *Rba. sphaeroides* (in green) overlaid on the five best AlphaFold models of *Rba. capsulatus* CrtJ. (**A**) The crystallized portion of PpsR (right side of the figure) aligns very well with predicted CrtJ models, especially the N-terminal PAS domain, the long α-helix (Q linker) and the first PAS domain. The second PAS domain in PpsR is positioned differently from all of the CrtJ models. The models of the HTH DNA binding domain which is attached via a flexible region are also in variable positions. (**B**) Best CrtJ dimer model based on the *Rba. sphaeroides* PpsR dimer crystal structure (4HH2.pdb) CrtJ residues up to amino acid 261 are very well aligned with the PpsR crystal structure containing the N-terminal PAS domain, Q-linker a-helix and PAS1 with best RMSD = 5.5 Å at Cα atoms for CrtJ model in dark blue.

The last 200 residues of CrtJ containing PAS2 and HTH vary from PpsR and also among the five models. This portion is likely very flexible, preventing the crystallization of full-length CrtJ. In Figure 5B, a model is shown of a CrtJ dimer as based on the PpsR structure [32].

3.4. A Model of a 2CrtJ-AerR Regulatory Complex

Finally, Winkler et al. [32] also obtained a crystal structure of AppA bound to a PpsR dimer (PDB 4HH3.pdb). In that structure, the SCHIC domain, which is structurally similar to AerR (Figure 3), interacts with the long Q-linker helix that separates the N-terminal PAS domain from the PAS1 domain. Using this structure as a guide, we constructed models of SAerR and LAerR bound to a CrtJ dimer at a similar location as the SCHIC domain interaction with PpsR (Figure 6A,B). Regarding this potential site of interaction, there are six residues (Gln133, Leu136, Gln140, Glu144, Tyr147 and Arg151) in the CrtJ Q-linker region that are universally conserved in all known CrtJ orthologs that face and align with the fourth helix of the four-helix bundle of AerR/SCHIC. There are also three residues in this AerR/SCHIC interacting helix that could form potential bonds with conserved residues in the Q-linker helix in CrtJ. These potential hydrogen bond interactions are: AerR Thr105 potentially H-bonding with Gln133 of CrtJ, AerR Ser109 potentially H-binding with Gln140 of CrtJ and AerR Gln112 potentially H-binding with Glu144 of CrtJ.

As discussed in the Introduction, SAerR converts CrtJ into a repressor while LAerR converts CrtJ into an activator [11]. One clue for how this may occur is the results of extensive in vitro and in vivo footprint assays which show that CrtJ alone, or when complexed with SAerR, exhibit a small ~44 base footprint at the *bchC* promoter region that contains two CrtJ recognition palindromes (one spanning the −10 and the other the −35 promoter region) [10,11]. This is contrasted by in vitro and in vivo footprints of CrtJ complexed with LAerR, which show that LAerR-CrtJ binding to the *bchC* promoter extends an additional ~120 nucleotides beyond the CrtJ recognition palindrome [10]. This result has given rise to a model where DNA is wrapped around an LAerR-CrtJ complex that then promotes the activation of transcription [11].

The main structural difference between the SAerR and LAerR variants is a 40-amino acid extension at the amino terminus of LAerR that contains a flexible region with His10 that forms an upper axial ligand to light-excited B_{12} and the 17-residue α1-helix (Phe19-Asn36) [11]. These two structural features in LAerR must therefore be responsible for altering CrtJ activity into an activator. Note that the AlphaFold model of LAerR does not contain B_{12}, so the flexible amino-terminal region containing His10 is not positioned at the B_{12} binding pocket located on the narrow side of the Rossman fold. However, genetic evidence clearly demonstrates that His10 does form this upper axial ligand to B_{12} when the corrin ring is light-excited [17], so this region along with the α1-helix must move in light-excited LAerR to a position near the Rossman fold. There is also a flexible strand between the α1-helix and the four-helix cap that should allow α1-helix to rotate and move toward the Rossman fold upon the formation of a His10-Co axial ligand (Figure 4A). Furthermore, inspection of surface charges on the α1-helix shows that one side is composed of hydrophobic residues, while the other side has several positively charged Arg residues (Arg29, Arg33, Arg34, Asp36) (Figures 4B and 6C). We propose that when His10 forms an axial ligand to the cobalamin Co, then the hydrophobic side of the α1-helix is likely positioned along a groove between the Rossman fold and the alpha cap to position its positively charged Arg residues in a way that allows an extended interaction of a looped region of the DNA helix with LAerR.
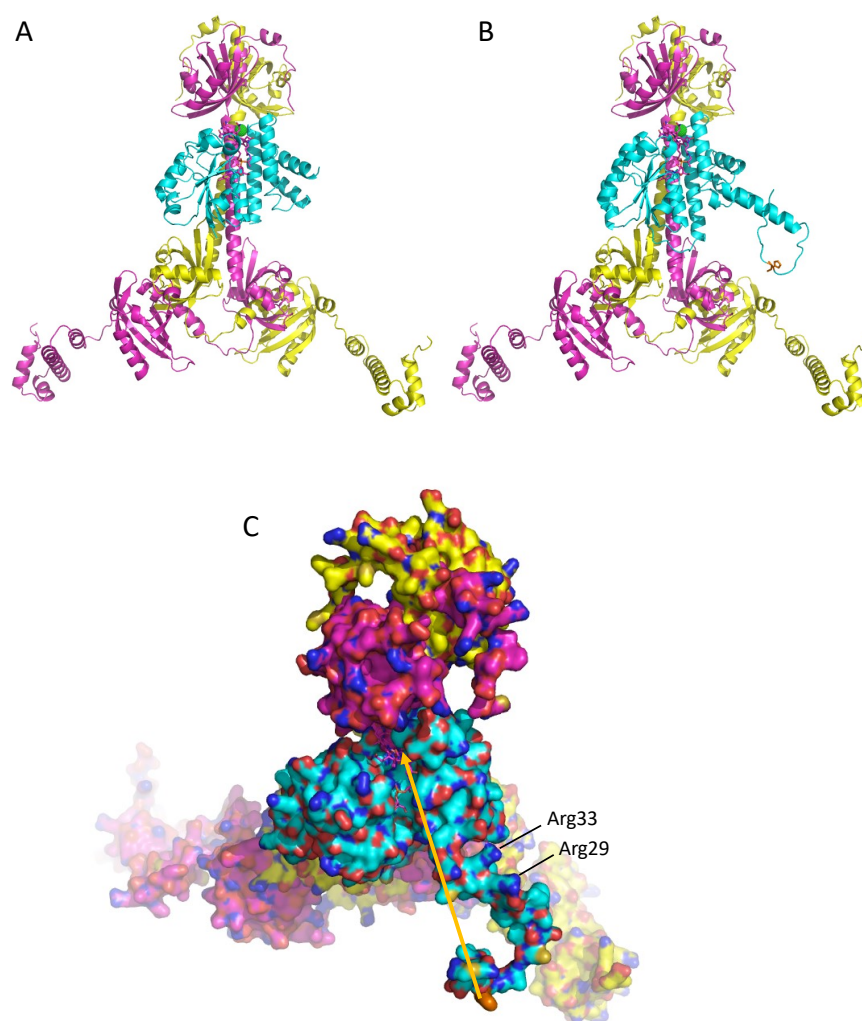
Figure 6. Model of SAerR and LAerR interactions with a CrtJ dimer. (**A**) Ribbon model of SAerR (blue) bound to the Q-linker helix region of a CrtJ dimer (magenta and yellow). (**B**) Ribbon model of LAerR (blue) bound to the Q-linker helix region of a CrtJ dimer (magenta and yellow. (**C**) Top orientation of a space filling model of LAerR (light blue) interacting with the Q-linker helix region of a CrtJ dimer (magenta and yellow). Oxygen atoms are red, nitrogen atoms are blue, and sulfur atoms are mustard yellow. The arrow shows where H10 must move upon light excitation. Positively charged Arg33 and Arg29 are also highlighted.

4. Discussion

In this work, we have solved a 2.25 Å resolution crystal structure of the short SAerR variant of AerR and used AlphaFold AI software [34] to obtain a predicted structure of the long variant LAerR. These two structures have a very high degree of similarity in regions of sequence conservation, specifically for the four-helix cap and Rossman fold that together contain a B_{12}-binding domain. Both AerR variant structures also show excellent similarities to the SCHIC domain crystal structure from AppA that is known to bind heme instead of B_{12} [29,30,32]. These structural similarities provide validity to the LAerR structural model as generated by AlphaFold. Likewise, AlphaFold also generated a structure of CrtJ that is strikingly similar to the known crystal structure of its *Rba. sphaeroides* ortholog PpsR [32], again showing a high degree of confidence on the CrtJ structure provided by AlphaFold.

With these structures, we obtain new insights into the mechanism by which SAerR functions as an anaerobic CrtJ co-repressor while LAerR functions as a photosynthetic (light anaerobic) CrtJ co-activator [11]. For example, the role of light excitation of B_{12} in LAerR appears to be as a catalyst to promote the formation of a B_{12} upper ligand with

His10. This event should pin the α1 helix to a specific position on the LAerR surface such that it promotes an additional positive charged region that can interact with DNA. Note that movement of the α1 helix would be a light-dependent event, which coincides with the role of LAerR to stimulate gene expression under photosynthetically logarithmic growth conditions [11]. The role of SAerR is to function as a dark/stationary phase co-repressor, so it likely that the binding of SAerR to CrtJ (and potentially also the binding of LAerR to CrtJ) likely helps to stabilize the dimerization of two CrtJ subunits, thereby increasing the binding affinity of this co-complex to target sequences. These are testable models that can be validated by future mutational analyses of critical protein–protein and protein–DNA interacting residues that are highlighted by this study.

Addressing the previously described regulatory role of the redox-active Cys in CrtJ was not a goal of this study. However, it is likely that these two AerR variants function as separate light-responding regulators that overlay their regulatory role over that of direct redox sensing by CrtJ. Specifically, prior studies have shown that a Cys in the HTH motif (Cys 420) can form a sulfenic acid modification in vivo when cells are exposed to oxygen [16,35]. Moreover, a Cys420 to Ala mutation leads to a ~60-fold reduction in DNA binding activity, while a Cys to Ser substitution at position 420 that mimics a cysteine sulfenic acid results in approximately a four-fold increase in DNA binding activity [35]. This Cys is located in the cluster of DNA recognition helices located at the carboxyl end of CrtJ. Additional structural analysis of CrtJ bound to DNA in oxidized and reduced forms both with and without SAerR and LAerR will be needed to obtain a unifying model of light and redox control of CrtJ repression and activation.

References

1. Cohen-Bazire, G.W.; Sistrom, W.R.; Stanier, R.Y. Kinetic studies of pigment synthesis by non-sulfur purple bacteria. *J. Cellular. Comp. Physiol.* **1957**, *49*, 25–68. [CrossRef] [PubMed]
2. Schindel, H.S.; Bauer, C.E. The RegA regulon exhibits variability in response to altered growth conditions and differs markedly between *Rhodobacter* species. *Microb. Genom.* **2016**, *2*, e000081. [CrossRef] [PubMed]
3. Else, S.; Swem, L.; Swem, D.; Bauer, C.E. RegB/RegA, a highly conserved redox-responding global regulatory system. *Microbio. Mol. Biol. Rev.* **2004**, *68*, 263–279. [CrossRef] [PubMed]
4. Kumka, J.E.; Bauer, C.E. Analysis of the FnrL regulon in *Rhodobacter capsulatus* reveals limited regulon overlap with orthologues from *Rhodobacter sphaeroides* and *Escherichia coli*. *BMC Genom.* **2015**, *16*, 895. [CrossRef]
5. Bauer, C.E.; Elsen, S.; Swem, L.R.; Swem, D.L.; Masuda, S. Redox and light regulation of gene expression in photosynthetic prokaryotes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2003**, *358*, 147–153. [CrossRef]
6. Ponnampalam, S.N.; Buggy, J.J.; Bauer, C.E. Characterization of an aerobic repressor that coordinately regulates bacteriochlorophyll, carotenoid, and light harvesting-II expression in *Rhodobacter capsulatus*. *J. Bacteriol.* **1995**, *177*, 2990–2997. [CrossRef]

7. Kovács, A.T.; Rákhely, G.; Kovács, K.L. Genes involved in the biosynthesis of photosynthetic pigments in the purple sulfur photosynthetic bacterium *Thiocapsa roseopersicina*. *Appl. Environ. Microbiol.* **2003**, *69*, 3093–3102. [CrossRef]

8. Penfold, R.J.; Pemberton, J.M. Sequencing, chromosomal inactivation, and functional expression in Escherichia coli of *ppsR*, a gene which represses carotenoid and bacteriochlorophyll synthesis in *Rhodobacter sphaeroides*. *J. Bacteriol.* **1994**, *176*, 2869–2876. [CrossRef]

9. Kovács, A.T.; Rákhely, G.; Kovács, K.L. The PpsR regulator family. *Res. Microbiol.* **2005**, *156*, 619–625. [CrossRef]

10. Fang, M.; Bauer, C.E. The vitamin B_{12}-dependent photoreceptor AerR relieves photosystem gene repression by extending the interaction of CrtJ with photosystem promoters. *mBio* **2017**, *8*, e00261-17. [CrossRef]

11. Yamamoto, H.; Fang, M.; Dragnea, V.; Bauer, C.E. Differing isoforms of the cobalamin binding photoreceptor AerR oppositely regulate photosystem expression. *eLife* **2018**, *7*, e39028. [CrossRef] [PubMed]

12. Ponnampalam, S.N.; Elsen, S.; Bauer, C.E. Aerobic repression of the *Rhodobacter capsulatus bchC* promoter involves cooperative interactions between CrtJ bound to neighboring palindromes. *J. Biol. Chem.* **1998**, *273*, 30757–30761. [CrossRef] [PubMed]

13. Elsen, S.; Ponnampalam, S.N.; Bauer, C.E. CrtJ bound to distant binding sites interacts cooperatively to aerobically repress photopigment biosynthesis and light harvesting II gene expression in *Rhodobacter capsulatus*. *J. Biol. Chem.* **1998**, *273*, 30762–30769. [CrossRef] [PubMed]

14. Nickens, D.G.; Bauer, C.E. Analysis of the *puc* operon promoter from *Rhodobacter capsulatus*. *J. Bacteriol.* **1998**, *180*, 4270–4277. [CrossRef] [PubMed]

15. Shimizu, T.; Cheng, Z.; Matsuura, K.; Masuda, S.; Bauer, C.E. Evidence that altered Cis element spacing affects PpsR mediated redox control of photosynthesis gene expression in *Rubrivivax gelatinosus*. *PLoS ONE* **2015**, *10*, e0128446. [CrossRef]

16. Masuda, S.; Dong, C.; Swem, D.; Setterdahl, A.T.; Knaff, D.B.; Bauer, C.E. Repression of photosynthesis gene expression by formation of a disulfide bond in CrtJ. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7078–7083. [CrossRef] [PubMed]

17. Cheng, Z.; Li, K.; Hammad, L.A.; Karty, J.A.; Bauer, C.E. Vitamin B_{12} regulates photosystem gene expression via the CrtJ antirepressor AerR in *Rhodobacter capsulatus*. *Mol. Microbiol.* **2014**, *91*, 649–664. [CrossRef]

18. Dong, C.; Elsen, S.; Swem, L.R.; Bauer, C.E. AerR, a second aerobic repressor of photosynthesis gene expression in *Rhodobacter capsulatus*. *J. Bacteriol.* **2002**, *184*, 2805–2814. [CrossRef]

19. Vermeulen, A.J.; Bauer, C.E. Members of the PpaA/AerR antirepressor family bind cobalamin. *J. Bacteriol.* **2015**, *197*, 2694–2703. [CrossRef]

20. Cheng, Z.; Yamomoto, H.; Bauer, C.E. Cobalamin's (Vitamin B_{12}) surprising function as a photoreceptor. *Trends Biochem. Sci.* **2016**, *41*, 647–650. [CrossRef]

21. Ortiz-Guerrero, J.M.; Polanco, M.C.; Murillo, F.J.; Padmanabhan, S.; Elías-Arnanz, M. Light-dependent gene regulation by a coenzyme B_{12}-based photoreceptor. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 7565–7570. [CrossRef] [PubMed]

22. Kutta, R.J.; Hardman, S.J.O.; Johannissen, L.O.; Bellina, B.; Messiha, H.L.; Ortiz-Guerrero, J.M.; Elias-Arnanz, M.; Padmanabhan, S.; Barran, P.; Scrutton, N.S.; et al. The photochemical mechanism of a B_{12}-dependent photoreceptor protein. *Nat. Commun.* **2015**, *6*, 7907. [CrossRef] [PubMed]

23. Jost, M.; Fernandez-Zapata, J.; Polanco, M.C.; Ortiz-Guerrero, J.M.; Chen, P.Y.; Kang, G.; Padmanabhan, S.; Elias-Arnanz, M.; Drennan, C.L. Structural basis for gene regulation by a B_{12}-dependent photoreceptor. *Nature* **2015**, *526*, 536–541. [CrossRef] [PubMed]

24. Jones, A.R.; Levy, C.; Hay, S.; Scrutton, N.S. Relating localized protein motions to the reaction coordinate in coenzyme B_{12}-dependent enzymes. *FEBS J.* **2013**, *280*, 2997–3008. [CrossRef]

25. Gomelsky, L.; Sram, J.; Moskvin, O.V.; Horne, I.M.; Dodd, H.N.; Pemberton, J.M.; McEwan, A.G.; Kaplan, S.; Gomelsky, M. Identification and in vivo characterization of PpaA, a regulator of photosystem formation in *Rhodobacter sphaeroides*. *Microbiology* **2003**, *149*, 377–388. [CrossRef] [PubMed]

26. Gomelsky, M.; Kaplan, S. AppA, a redox regulator of photosystem formation in *Rhodobacter sphaeroides* 2.4.1, is a flavoprotein. Identification of a novel FAD binding domain. *J. Biol. Chem.* **1998**, *273*, 35319–35325. [CrossRef] [PubMed]

27. Gomelsky, M.; Klug, G. BLUF: A novel FAD-binding domain involved in sensory transduction in microorganisms. *Trends Biochem. Sci.* **2002**, *27*, 497–500. [CrossRef]

28. Masuda, S.; Bauer, C.E. AppA is a blue light photoreceptor that antirepresses photosynthesis gene expression in *Rhodobacter sphaeroides*. *Cell* **2002**, *110*, 613–623. [CrossRef]

29. Moskvin, O.V.; Kaplan, S.; Gilles-Gonzalez, M.A.; Gomelsky, M. Novel heme-based oxygen sensor with a revealing evolutionary history. *J. Biol. Chem.* **2007**, *282*, 28740–28748. [CrossRef]

30. Yin, L.; Dragnea, V.; Feldman, G.; Hammad, L.A.; Karty, J.A.; Dann, C.E., 3rd; Bauer, C.E. Redox and light control the heme-sensing activity of AppA. *mBio* **2013**, *4*, e00563-13. [CrossRef]

31. Gomelsky, M.; Kaplan, S. Molecular genetic analysis suggesting interactions between AppA and PpsR in regulation of photosynthesis gene expression in *Rhodobacter sphaeroides* 2.4.1. *J. Bacteriol.* **1997**, *179*, 128–134. [CrossRef] [PubMed]

32. Winkler, A.; Heintz, U.; Lindner, R.; Reinstein, J.; Shoeman, R.L.; Schlichting, I. A ternary AppA-PpsR-DNA complex mediates light regulation of photosynthesis-related gene expression. *Nat. Struct. Mol. Biol.* **2013**, *20*, 859. [CrossRef] [PubMed]

33. Anderson, S.; Dragnea, V.; Masuda, S.; Ybe, J.; Moffat, K.; Bauer, C. Structure of a novel photoreceptor, the BLUF domain of AppA from *Rhodobacter sphaeroides*. *Biochemistry* **2005**, *44*, 7998–8005. [CrossRef] [PubMed]

34. Senior, A.W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Zidek, A.; Nelson, A.W.R.; Bridgland, A.; et al. Improved protein structure prediction using potentials from deep learning. *Nature* **2020**, *577*, 706–710. [CrossRef] [PubMed]

35. Cheng, Z.; Wu, J.; Setterdahl, A.; Reddie, K.; Carroll, K.; Hammad, L.A.; Karty, J.A.; Bauer, C.E. Activity of the tetrapyrrole regulator CrtJ is controlled by oxidation of a redox active cysteine located in the DNA binding domain. *Mol. Microbiol.* **2012**, *85*, 734–746. [CrossRef]

36. Willett, J.; Smart, J.L.; Bauer, C.E. RegA control of bacteriochlorophyll and carotenoid synthesis in *Rhodobacter capsulatus*. *J. Bacteriol.* **2007**, *189*, 7765–7773. [CrossRef]

37. Kabsch, W. XDS. *Acta Crystallographica. Sect. D Biol. Crystallogr.* **2010**, *66*, 125–132. [CrossRef]

38. Drennan, C.L.; Huang, S.; Drummond, J.T.; Matthews, R.G.; Lidwig, M.L. How a protein binds B_{12}: A 3.0 A X-ray structure of B_{12}-binding domains of methionine synthase. *Science* **1994**, *266*, 1669–1674. [CrossRef]

Article

The Response Regulator RegA Is a Copper Binding Protein That Covalently Dimerizes When Exposed to Oxygen

Nijia Ke [ID] **and Carl E. Bauer *** [ID]

Molecular and Cellular Biochemistry Department, Indiana University, Bloomington, IN 47405, USA; nijke@iu.edu
* Correspondence: bauer@iu.edu; Tel.: +1-812-855-6595

Abstract: In *Rhodobacter capsulatus*, the histidine kinase RegB is believed to phosphorylate its cognate transcriptional factor RegA only under anaerobic conditions. However, transcriptome evidence indicates that RegA regulates 47 genes involved in energy storage, energy production, signaling and transcription, under aerobic conditions. In this study, we provide evidence that RegA is a copper binding protein and that copper promotes the dimerization of RegA under aerobic conditions. Inductively coupled plasma mass spectrometry (ICP-MS) analysis indicates that RegA binds Cu^{1+} and Cu^{2+} in a 1:1 and 2:1 ratio, respectively. Through LC-MS/MS, ESI-MS and non-reducing SDS-PAGE gels, we show that Cu^{2+} stimulates disulfide bond formation in RegA at Cys156 in the presence of oxygen. Finally, we used DNase I footprint analysis to demonstrate that Cu^{2+}-mediated covalent dimerized RegA is capable of binding to the *ccoN* promoter, which drives the expression of cytochrome *cbb$_3$* oxidase subunits. This study provides a new model of aerobic regulation of gene expression by RegA involving the formation of an intermolecular disulfide bond.

Keywords: two-component system; RegA; aerobic; copper ion; disulfide bond; DNA binding; gene regulation

1. Introduction

Just as Ser/Thr/Tyr phosphorylation pathways are among the most common signaling pathways in eukaryotes, two-component systems (TCSs) constitute the most common signal transduction pathway in prokaryotes [1]. Receptor sensor kinases in TCSs sense various stimuli ranging from physical conditions such as light, temperature, redox state, osmolarity and concentration of chemicals such as nutrients and quorum signals [2]. TCSs control numerous cellular activities involved in metabolism, mobility and virulence [1].

In the purple non-sulfur bacterium *Rhodobacter capsulatus*, the RegB/RegA two-component system regulates a variety of physiological processes such as photosynthesis, hydrogen fixation, respiration, carbon fixation, nitrogen assimilation, denitrification, hydrogen uptake and aerotaxis [3]. Previous results indicate that RegB senses redox change through two sensing mechanisms. First, RegB has a ubiquinol/ubiquinone binding site located within a membrane-spanning region that monitors the redox state of the ubiquinol pool [4,5]. The binding of oxidized ubiquinone inhibits, while the binding of reduced ubiquinol stimulates, RegB kinase activity [4]. RegB has low affinity to both oxidized ubiquinone and reduced ubiquinol, allowing RegB to readily adjust its activity to the redox state of the ubiquinone pool, not unlike that of a rheostat [4,5]. The second level of RegB control involves direct inhibition of RegB kinase activity by dioxygen via oxidation of a conserved cytosolic cysteine residue (Cys265). The oxidation of Cys265 results in both the formation of a stable sulfenic acid derivative and disulfide bond formation that promotes stable tetramerization of RegB with both oxidized RedB derivatives exhibiting inactive kinase activity [6,7]. Collectively, these results support a long-standing model where RegB only undergoes autophosphorylation and subsequent transfer of the phosphoryl group to its cognate response regulator RegA under anaerobic conditions. Phosphorylated RegA

then binds to DNA as a transcriptional factor, modulating the expression of genes involved in numerous anaerobic energy-generating and energy-utilizing processes [8,9]. While the described model has broad acceptance, RNA-seq data from our laboratory indicate that this model may be too simplistic as RegA is still capable of aerobically regulating ~47 genes, which is a condition in which RegA would not be phosphorylated by RegB [10]. We have therefore investigated whether RegA is capable of regulating gene expression under aerobic unphosphorylated conditions.

One clue for a possible alternative mechanism for controlling the activity of RegA in the absence of phosphorylation by RegB may be the transcriptional co-localization of *regB* and *regA* with *senC*. In species that contain the group 1 RegB/RegA system, *regA* and/or *regB* is invariably cotranscribed with *senC* [11,12]. SenC is a known copper-binding protein with proposed copper chaperone activity [13–15]. In *R. capsulatus*, SenC has a role in the assembly of copper containing cbb_3 cytochrome oxidase [13–15]. However, SenC also has sequence similarity to a family of oxidoreductases that have thiol-disulfide bond forming activity [16]. Furthermore, group 1 RegA orthologs also contain an invariant conserved Cys (Cys156 in *R. capsulatus*) located immediately upstream of a DNA binding helix-turn-helix domain [12,17]. Thus, we explored whether the activity of RegA could be modulated by copper and/or the oxidation of Cys156. Our results indicate that RegA is capable of binding Cu^{2+} and that exposure of RegA-Cu^{2+} to O_2 results in disulfide bond formation at Cys156, which stimulates dimerization and DNA binding.

2. Materials and Methods

2.1. Overexpression of RegA

The construction of a RegA overexpression vector follows the same method of construction of a RegA* overexpression vector as described in Du et al. (1998) [18] with the exception that the 95th residue of the RegA* codon was mutated from Alanine to Serine, which is the wild-type RegA codon. The resulting plasmid pET29CBD::*regA*, which also contains a chitin-binding domain fused to the C-terminal of RegA, was then transformed into the overexpression strain BL21(DE3) and grown overnight by shaking at 250 rpm at 37 °C. The overnight culture was then subcultured (1:50 ratio) into 1 L of Terrific Broth (TB) containing 50 µg/mL kanamycin. When the OD_{600nm} reached 0.5 to 0.8, the expression of RegA was induced by the addition of 0.4 mL 1M IPTG per liter culture and then shaken at 250 rpm at 37 °C for 4 h. Cells were then harvested by centrifuging at 10,000× *g* for 15 min. Cell pellets were stored at −80 °C prior to RegA purification.

2.2. Purification of RegA from E. coli

Frozen cell pellets from 2 L of RegA culture were resuspended in 50 mL lysis/wash buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl) and digested with 400 µL 2 mg/mL DNaseI via homogenization on ice and rotation at 4 °C for an hour. The cell lysate was then disrupted by three passages through an emulsifier and centrifuged at 17,210× *g* for 30 min. RegA containing an intein chitin-binding domain was purified by passing through a 10 mL chitin affinity column after equilibrating the column with 10 column volumes of lysis buffer, and then washed with another 10 column volumes of lysis/wash buffer. The intein chitin-binding domain was cleaved off RegA by purging the chitin affinity column with 3 column volumes of cleavage buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 30 mM DTT) and incubated at room temperature overnight. Tagless RegA was then eluted with 3 column volumes of elution buffer (20 mM Tris-HCl (pH 8.0), 500 mM NaCl). Finally, reduced RegA monomer was treated with gel filtration chromatography using a Superose 12 column equilibrated with 20 mM Tris-HCl pH 8.0, 250 mM NaCl, 5 mM TCEP, 10 mM EDTA and 10% glycerol. Finally, RegA was concentrated to 2 to 3 mg/mL and stored at −80 °C.

2.3. Loading RegA with Metal through Buffer Exchange

An aliquot of purified RegA was concentrated to 100 μL in a 500 μL centrifugal filter (Amicon, Burlington, MA, USA) with the 10 kDa cutoff by centrifuging at 17,000× *g* for 8 min. RegA was then buffer exchanged into 20 mM Tris-HCl pH 8.0 250 mM NaCl by repeating the steps of adding 400 μL fresh buffer and centrifuging at 17,000× *g* five times with final centrifugation run until the sample volume reached 100 μL. The concentration of RegA was then measured using the Bradford method. The RegA sample was then anaerobically loaded with metals by incubating for 5 min with 400 μL 20 mM Tris-HCl pH 8.0 250 mM NaCl containing 3× metal to 1× RegA's concentration. CuCl and $CuSO_4$ solutions were prepared inside the anaerobic chamber using degassed ddH_2O for copper salt preparation. The protein buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl) used for buffer exchange was also degassed and purged with nitrogen before being brought into the anaerobic chamber. Excessive metal not bound to RegA was then removed from the buffer by repeating Amicon filter centrifugation eight times (maximum speed until ~100 μL concentrated sample was left) with each step involving the addition of 400 μL fresh buffer.

2.4. ICP-MS

The concentrations of protein were measured before analysis. At least 20 μL protein of 20 μM or 500 μL buffer was required to conduct inductive coupled plasma mass spectrometry (ICP-MS). Protein samples of RegA were diluted 100-fold with 2.5% HNO_3 (Sigma) into a final volume of 3 mL, while using Pure Plus Internal Standard Mix (100 ppb, PerkinElmer, Waltham, MA, USA) as an internal standard. Using hydrogen (^{55}Mn, ^{56}Fe, ^{59}Co, ^{60}Ni detection) or helium (^{24}Mg, ^{63}Cu, ^{66}Zn detection) as collision gases to remove possible interferences with the internal reference ^{45}Sc or ^{72}Ge, samples were analyzed with an Agilent 8800 QQQ ICP-MS device. Based on standard curves of all analyzed metals generated with Pure Plus Multi-Element Calibration Standard 3 (0.5–100 ppb, PerkinElmer), metal concentrations of protein or buffer samples were calculated.

2.5. ESI-MS

RegA were incubated with a gradient concentration of $CuSO_4$ in one-fold, three-fold, five-fold and ten-fold of RegA concentration, respectively. Measures of 30 μL of these Cu^{2+} treated RegA samples, apo-RegA and buffer were separately injected and sealed in glass vials and brought out of the anaerobic chamber. The electrospray ionization mass spectrometry (ESI-MS) spectra of the samples were generated using an LC (C_4 reverse phase)-MS (Synapt G2S HDMS) instrument and analyzed with Mass Lynx v4.1.

2.6. LC-MS/MS

Gel bands were diced into 1 mm cubes and incubated for 45 min at 57 °C. To maintain potential disulfide bonds, the samples were neither reduced nor alkylated. A solution containing 1 μg trypsin, in 25 mM ammonium bicarbonate was added and the samples were digested for 12 h at 37 °C. The resulting peptides were desalted using a ZipTip (Millipore, Billerica, MA, USA). The samples were dried down and injected into an Eksigent HPLC coupled to an LTQ Orbitrap XL (Thermo Fischer Scientific, Waltham, MA, USA). The peptides were separated using a 75 micron, 15 cm column packed in-house with C18 resin (Michrom Bioresources, Auburn, CA, USA) at a flow rate of 300 nl/min. A one-hour gradient was run from Buffer A (2% acetonitrile, 0.1% formic acid) to 60% Buffer B (100% acetonitrile, 0.1% formic acid). The Orbitrap was configured to acquire a survey scan over the mass range 300–2000 at a resolution of 30,000. This was followed by collision-induced dissociation mass spectrometry (CID MS/MS) on the top-five most-intense precursor ions above a threshold of 1000 counts.

The resulting MS/MS peaklists were searched against the Uniprot *Rhodobacter capsulatus* database using Protein Prospector (v5.10.14). Acetylation of the protein amino terminus, oxidation of methionine and pyroglutamine formation of peptide amino terminal glutamine and disulfide bond formation were set as variable modifications. A mass

tolerance of 20 ppm was used for precursor ions and 0.6 Da was used for fragment ions. Peptide expectation values were set to <0.05.

2.7. DNase I FootPrinting Assay

DNase I footprint analysis was performed as described by Willett et al. [19]. Briefly, a 299 bp DNA fragment containing the promoter region of *ccoN* was PCR amplified using a 6-FAM labeled forward primer and a HEX labeled reverse primer. A 50 nM DNA probe was incubated at room temperature for 15 min with a gradient concentration of purified proteins in 20 mM Tris-HCl pH 8.0, 2 mM $MgCl_2$, 0.5 mM $CaCl_2$, 0.1 mg/mL BSA of a total volume of 20 μL and then digested with 5 μL of DNaseI for another 15 min at room temperature. The reactions were quenched by the addition of 25 μL 0.5 M EDTA pH 8.0. The digested fragments were recovered using a Min Elute PCR purification kit (Qiagen, Hilden, Germany) and then eluted with 15 μL elution buffer. The samples were detected with a 3730 DNA Analyzer (Applied Biosystem, Waltham, MA, USA) using 500 LIZ^{TM} Size Standard and eventually analyzed with the Peak Scanner Software v2.0.

3. Results

3.1. RegA Binds Copper Anaerobically

We first addressed whether RegA can bind copper in the Cu^{1+} or Cu^{2+} states. Since Cu^{1+} is only stable in the absence of oxygen (oxygen oxidizes Cu^{1+} to Cu^{2+}), we assayed the ability of RegA to bind these two redox states of Cu under anaerobic conditions. For this analysis, we anaerobically exposed RegA to threefold molar excess of either Cu^{1+} or Cu^{2+} followed by removal of unbound Cu using centrifugal filtration mediated buffer exchange in an anaerobic chamber. After unbound Cu^{1+} and Cu^{2+} were removed, we measured the concentration of copper that remained bound to RegA using inductively coupled plasma mass spectrometry (ICP-MS). As shown in Figure 1, RegA anaerobically binds approximately stoichiometric amount of Cu^{1+}, which is indicated by a copper ion over RegA monomer ratio of 1.02 ± 0.21. In contrast, the ratio of Cu^{2+} over RegA is very close to 0.5 (0.45 ± 0.03), suggesting that two RegA's bind to one molecule of Cu^{2+} anaerobically. Taken together, these results indicate that RegA is indeed capable of binding copper.
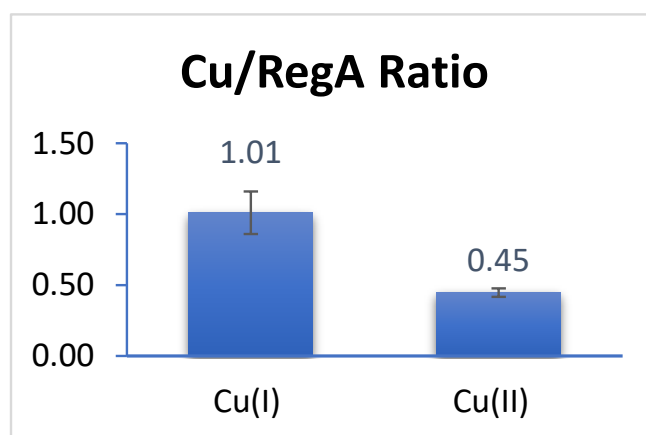
Figure 1. Ratio of concentration of copper ion over RegA anaerobically.

3.2. Cu^{2+} Promotes the Dimerization of RegA Aerobically

Since Cu^{2+} binds reduced RegA in a 0.5:1 ratio, we next investigated whether Cu^{2+} was capable of stimulating dimerization of RegA. For this analysis, monomeric RegA was loaded with Cu^{2+} via buffer exchange aerobically and immediately run through a non-reducing SDS-PAGE gel with no DTT added to the samples. As shown in Figure 2, the addition of Cu^{2+} indeed promotes the dimerization of RegA. As observed from the non-reducing SDS-PAGE gel in Figure 3A, Cu^{2+} mediated dimerization of RegA is also

quite rapid, with the majority of RegA dimerized after only one minute of Cu^{2+} treatment in the presence of air. In contrast, exposure of RegA to air in the absence of Cu^{2+} did not lead to any significant dimerization even after 4 h of exposure, confirming that Cu^{2+} in the presence of oxygen serves a critical role in mediating the dimerization of RegA (Figure 3B).
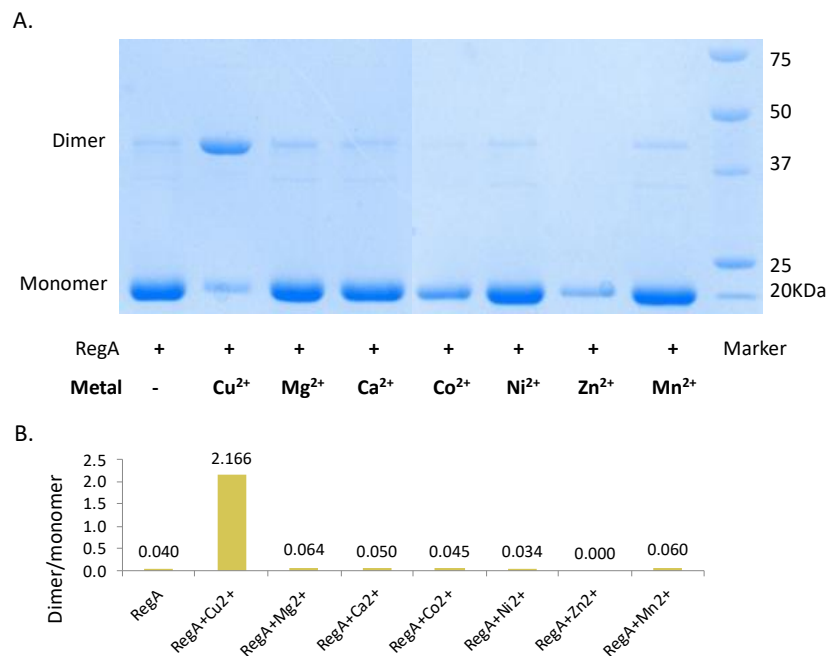
Figure 2. Cu^{2+} indispensably promotes the dimerization of RegA. (**A**) Non-reducing SDS-PAGE gel of apo-RegA and RegA treated with divalent metals. + denotes addition of RegA protein whereas –denotes where metal is not added. (**B**) Histogram of RegA dimer:monomer ratios determined by image J.
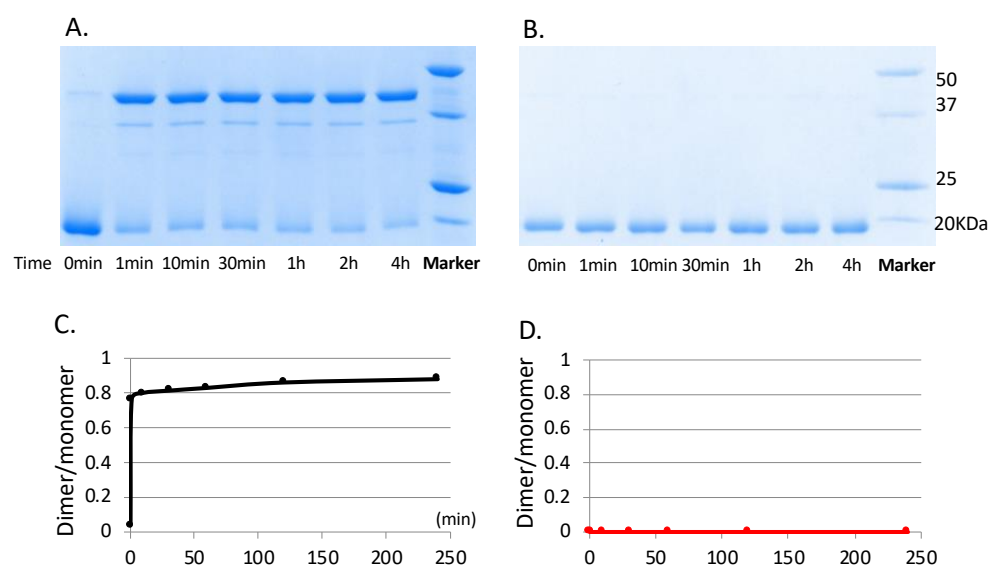
Figure 3. Copper catalyzes the dimerization of RegA likely with the help of oxygen. (**A**) Non-reducing SDS-PAGE gel of RegA treated with Cu^{2+} in the air after different time periods. (**B**) Non-reducing SDS-PAGE gel of RegA in the absence of Cu^{2+} in the air after different time periods. (**C**) Dimer:monomer ratio of RegA treated with Cu^{2+} in the air after different time periods. (**D**) Dimer:monomer ratio of RegA in the absence of Cu^{2+} in the air after different time periods. The dimer:monomer ratio was calculated based on the intensity of bands quantified through Image J v1.51.

To determine whether Cu^{2+} uniquely plays a role in promoting dimerization of RegA, we also treated RegA with six other divalent metals including Mg^{2+}, Ca^{2+}, Co^{2+}, Ni^{2+}, Zn^{2+} and Mn^{2+} following the same protocol. The non-reducing SDS-PAGE gel results in Figure 2 show that none of the other six divalent metals tested stimulated RegA dimerization, indicating that only Cu^{2+} is capable of promoting dimerization of RegA.

3.3. RegA Dimerization Requires Both Cu^{2+} and Molecular Oxygen

We next addressed whether molecular oxygen is a requirement for Cu^{2+} dimerization of RegA. For this analysis, we tested whether Cu^{2+} can promote the dimerization of RegA under anaerobic conditions by treating RegA monomers with Cu^{2+} in an anaerobic chamber with degassed buffers for 10 min. The reactions were quenched with EDTA (or passed through filters) to remove Cu^{2+} prior to analysis of the monomer/dimer state of RegA using electrospray ionization mass spectrometry (ESI-MS). As shown in Figure 4, the RegA anaerobically treated with Cu^{2+} and a monomer control without Cu^{2+} treatment both showed identical ESI-MS profiles exhibiting a monomer molecular weight of 20,207 daltons. This indicates that both Cu^{2+} and atmospheric levels of molecular oxygen are necessary to promote the dimerization of RegA.
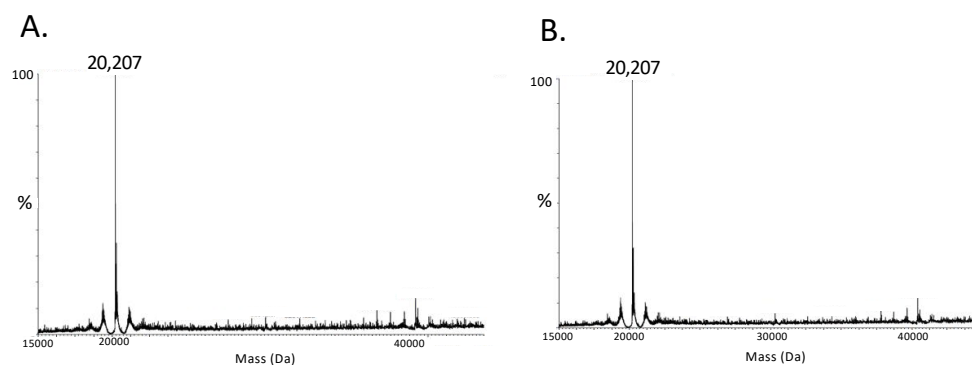
Figure 4. ESI-MS spectrum of RegA in the presence or absence of Cu^{2+} anaerobically. (**A**) RegA molecular weight determination by ESI-MS (20,207 daltons) in the presence of Cu^{2+} without oxygen. (**B**) RegA molecular weight determination by ESI-MS (20,207 daltons) in the absence of Cu^{2+} without oxygen.

We next addressed whether the dimer formed by Cu^{2+} treated RegA involves a covalent disulfide bond or instead is a non-covalent interaction. For this analysis, $Cu^{2+} + O_2$ mediated RegA dimer was isolated away from monomer RegA by gel filtration chromatography and then treated with either DTT or EDTA alone or both. As is indicated in Figure 5, the addition of excess EDTA, which would remove Cu^{2+}, did not affect dimerization. This is contrasted by the conversion of RegA dimers to its monomeric form by the addition DTT, which would indicate that Cu^{2+} likely stimulates disulfide bond formation. For confirmation, we also undertook liquid chromatography–tandem mass spectrometry (LC-MS/MS) of trypsin-digested Cu^{2+}-formed RegA dimer, which showed the presence of a peptide fragment VYELCDR containing an intermolecular disulfide bond at Cys156 (Figure 5B). Interestingly, RegA only has one cysteine residue at the 156th position, with this cysteine residue conserved among all the α-proteobacteria that contain RegA homologs. Finally, analysis of the presence of Cu in the isolated RegA dimers with ICP-MS shows no significant amounts of Cu present in covalently linked dimerized RegA.
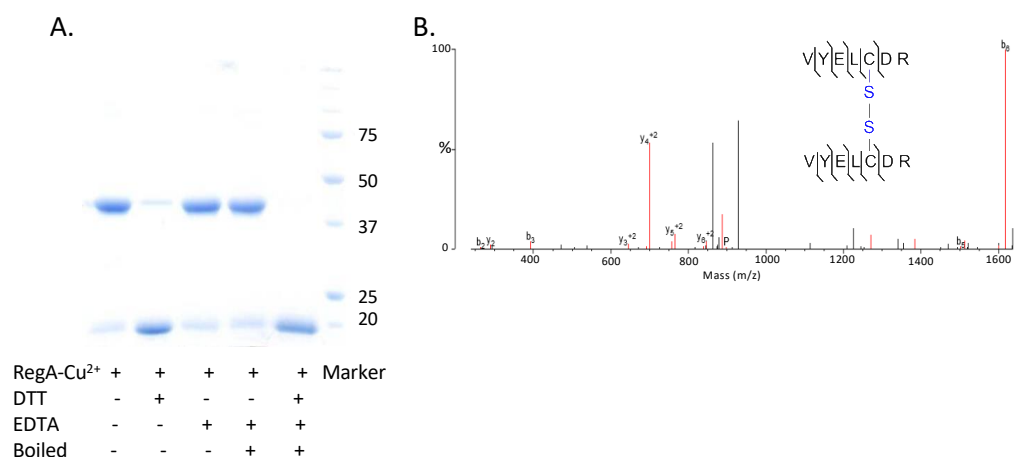
Figure 5. Cu^{2+} can promote RegA to dimerize with the formation of a disulfide bond in vitro. (**A**) Non-reducing SDS-PAGE gel of Cu^{2+} treated RegA dimer treated with EDTA or DTT. (**B**) Disulfide bond formation in Cu^{2+} treated RegA dimer observed in LC-MS/MS.

3.4. Covalently Dimerized RegA Is Capable of Binding DNA

Phosphorylation of response regulators by cognate sensor kinases is known to stimulate dimerization, which facilitates DNA binding. Given that unphosphorylated RegA monomers have low DNA binding activity [20], we next addressed whether RegA dimers formed by disulfide bond formation are also capable of binding to DNA. For this analysis, we performed DNase I footprint assays with Cu^{2+} stimulated RegA disulfide dimer to the *ccoN* promoter region that controls cytochrome cbb_3 oxidase expression. This promoter was chosen for analysis based on previous transcriptomic analyses, which showed that RegA regulates the expression of the *ccoN* promoter under aerobic conditions when RegB kinase activity is suppressed [3,10].

Footprint results shown in Figure 6 demonstrate that covalently dimerized unphosphorylated RegA can bind to the *ccoN* promoter region, as exhibited by excellent DNase I peak suppression with 2 µM of dimerized RegA. Two binding sites highlighted in red are readily observed on both the forward and the reverse strands (Figure 6B,D, respectively). The two forward strand protection sites observed with Cu^{2+} treated RegA dimers (−80 bp to −76 bp and −74 bp to −49 bp upstream of the transcription initiation site, respectively) are at the same location as previously reported with a constitutively dimerized and active RegA variant called RegA* [3]. On the reverse strand, RegA dimers also bind at two binding sites (−74 bp to −69 bp and −66 bp to −52 bp upstream of the transcription initiation site), which are just slightly narrower than the binding site observed with RegA* dimers (−74 bp to −49 bp) [3]. Footprint analysis with varying amounts of dimerized RegA shows that Cu^{2+} stimulated RegA dimer binds to the ccoN promoter with an apparent Kd of 1 µM ~ 4-fold lower than that observed with RegA* (0.25 µM) [3]. These results demonstrate that covalently dimerized unphosphorylated RegA indeed binds to the *ccoN* promoter at just a slightly lower affinity than phosphorylated noncovalent dimers do.
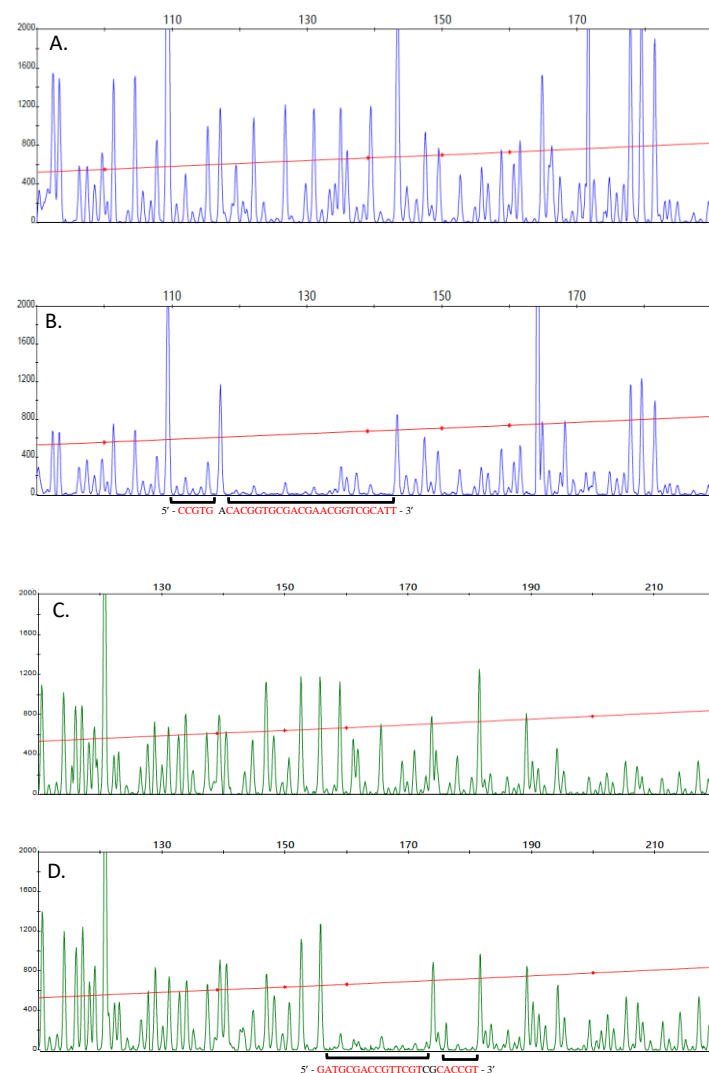
Figure 6. DNaseI footprinting of Cu^{2+}-treated RegA dimer to the *ccoN* promoter. (**A**) DNase I digestion of the *ccoN* promoter top strand without RegA. (**B**) DNase I digestion of the *ccoN* promoter top strand with 2 μM Cu^{2+} treated unphosphorylated RegA dimer. (**C**) DNase I digestion of the *ccoN* promoter bottom strand without RegA. (**D**) DNase I digestion of the *ccoN* promoter bottom strand with 2 μM Cu^{2+} treated unphosphorylated RegA dimer. Areas of RegA protection are highlighted with a black bar.

4. Discussion

The response regulator paradigm is that phosphorylation of a response regulator by its cognate histidine kinase promoted dimerization and subsequent DNA binding activity. This study shows that RegA from *R. capsulatus* binds copper and that bound Cu^{2+} in the presence of O_2 leads to intermolecular disulfide bond formation. Covalently dimerized RegA is capable of binding DNA at sub μM concentrations, which likely provides RegA the ability to aerobically regulate gene expression in the absence of phosphorylation.

The formation of disulfide bonds has been adopted by several transcription factors to control their DNA binding activities in response to redox changes. Some well-characterized examples are OxyR, which, in response to hydrogen peroxide or S-nitrosothiols, induces the expression of over ten genes encoding proteins [21] against oxidative and nitrosative stress; and *oxyS*, a gene that encodes a nontranslated RNA involved in DNA repair. In both cases, exposure to oxygen stimulates disulfide bond oxidation that affects DNA binding activity [22]. In *R. capsulatus*, CrtJ is an aerobic repressor of photosystem genes that has regulatory cysteines that are oxidized in the presence of molecular oxygen [23–25]. Besides

the use of disulfide bond formation, other transcription factors have utilized changes in the redox state of iron to sense the presence or absence of oxygen. For example, FNR, a global regulator of over a hundred genes in *E. coli*, contains a [4Fe-4S] iron–sulfur cluster that disassembles under aerobic conditions, which disrupts dimerization and subsequent binding to DNA [21,26,27]. In the case of RegA, not only does copper itself become oxidized by oxygen from Cu^{+1} to Cu^{+2}, but, in addition, the presence of both Cu^{2+} and oxygen are needed for disulfide bond formation. In some respects, this may be regarded as a hybrid of the above-described mechanisms of redox sensing.

4.1. A Model Describing How Copper and Oxygen Promote RegA Disulfide Bond Formation

Taking all of our results into consideration, we can propose a model of how Cu^{2+} and oxygen likely promote covalent dimerization of RegA as follows (Figure 7): (i) anaerobically, RegA monomer binds Cu^{1+}, possibly by accepting Cu^{1+} transfer from an intracellular copper chaperone such as from CopZ [28] or other copper chaperones. (ii) When the environment changes from anaerobic to aerobic conditions, oxygen oxidizes RegA bound Cu^{1+} into Cu^{2+}. Cu^{2+} with the help of O_2 then promotes subsequent disulfide bond formation. (iii) Disulfide bond dimerized RegA then binds to a subset of promoters.
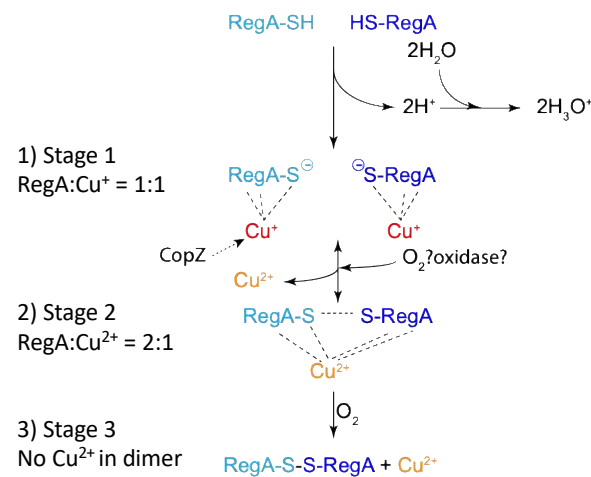
Figure 7. Proposed model of the mechanism of how Cu^{2+} promotes the dimerization of RegA in vivo.

Recently, Alphabet's DeepMind released the artificial intelligence program AlphaFold that predicts a protein's tertiary structures based on a protein's primary sequence [29]. Analysis of a putative RegA structure derived from AlphaFold shows two well-defined domains (Figure 8). The largest right-side domain in Figure 8 is very similar to response regulator receiver domains that have previously been solved by X-ray crystallography [1]. This phosphorylatable receiver domain is linked by a long, flexible linker region containing five prolines to a DNA binding domain shown on the left. The two smallest left-most helices exhibit a high degree of structural similarity to the RegA helix-turn-helix DNA binding helices as previously solved by NMR [17]. The larger helix after the flexible linker region just before the two small DNA binding helices contains the universally conserved Cys156 as shown in yellow. Given the flexibility of this domain from the receiver domain, this Cys should be available for oxidation into an intermolecular disulfide by molecular oxygen. We suspect that the role of Cu^{2+} may be a structural role where it binds and holds two RegA monomers in such a position that this helix from two RegA's are positioned in close proximity, thereby allowing oxygen to effectively promote disulfide bond formation.
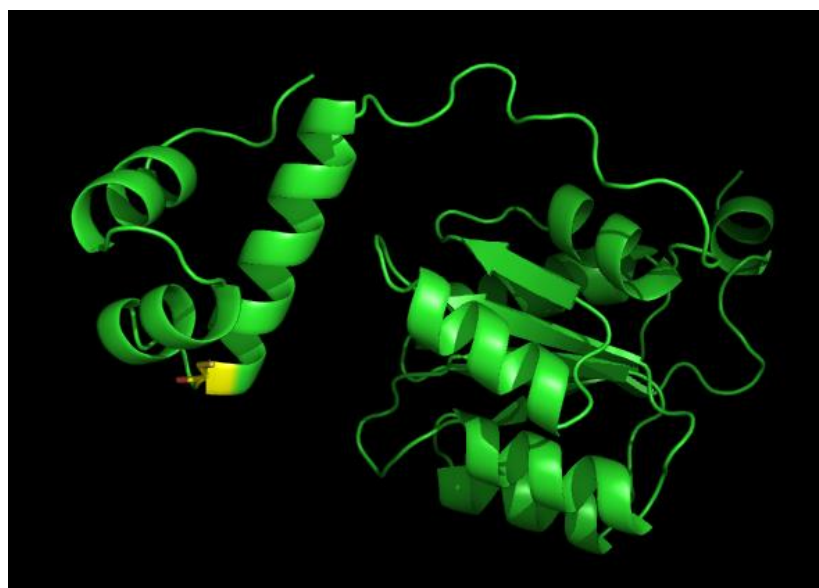
Figure 8. RegA tertiary structure as predicted from the DeepMind released artificial intelligence program AlphaFold. The DNA binding helix-turn-helix domain is on the left and the receiver domain is on the right. Cys156 that undergoes disulfide bond formation is highlighted in yellow.

Due to the capability of copper ions to undergo reversible oxidation of Cu^{1+} to Cu^{2+}, copper acts as an essential cofactor for several Cu metalloenzymes such as cytochrome cbb_3 oxidase encoded by the *ccoN* operon that RegA aerobically and anaerobically regulates and superoxide dismutase [30]. However, copper unbound to protein is also quite toxic. For example, Cu^{1+} tends to mediate the generation of radical oxygen species (ROS) through Fenton reaction [31,32], while Cu^{2+} can nucleate protein aggregation, leading to amyloid formation in mammals [30]. Hence, the transportation and distribution of copper is precisely regulated [33]. Apart from copper importers and exporters, copper ions are often delivered to copper-required proteins through copper chaperones [34] in the form of Cu^{1+}. It is therefore plausible that one or several copper chaperones may directly interact and transfer copper to RegA. SenC is unlikely to itself deliver copper to RegA as SenC's copper binding domain is located in the periplasm [14,15]. Besides SenC [13–15], the identified copper chaperones in *Rhodobacter* include a Cu^{1+} chaperone Cox11 [35] in *Rhodobacter sphaeroides*, a PCuAC-like periplasmic chaperone PccA [15], and a recently identified intracellular Cu^{1+} chaperone CopZ [28], in *Rhodobacter capsulatus*. Since RegA is an intracellular protein, and the copper-binding site of SenC is in the periplasm [14], we suspect that cytoplasmic CopZ most likely transfers Cu^{1+} to RegA in *R. capsulatus*.

4.2. Both Aerobic and Anaerobic Conditions Control RegA Activity

Most previous studies have focused on the role of RegA, and its cognate sensor kinase RegB, in controlling anaerobic gene expression [8,9]. These studies demonstrated that the phosphorylation activity of RegB is controlled by both redox changes in the ubiquinone pool as well as by oxidation of a conserved cystine (Figure 9) [4–7]. RegB has six membrane-spanning helices, three of which are known to be involved in binding both oxidized and reduced ubiquinone [4,5,36]. When bound to oxidized ubiquinone, the kinase activity is low, whereas when bound to reduced ubiquinol, the kinase activity is high [4] (Figure 9). Additionally, there is a redox-active Cys in the four-helix bundle that links the membrane-spanning domain to the kinase domain [6,7]. In the presence of oxygen, this Cys undergoes oxidation either to a sulfonic acid derivative or to a disulfide, either of which inhibits kinase activity (Figure 9) [6,7].
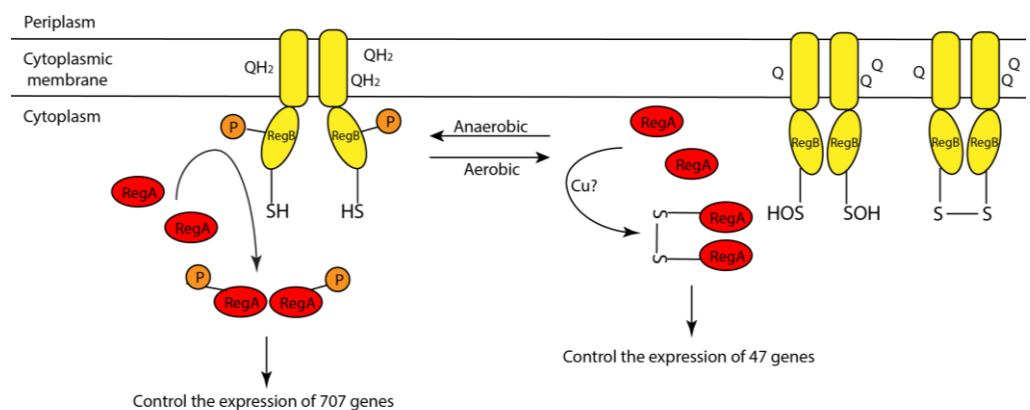
Figure 9. Model of RegB/RegA two-component system in sensing redox changes and regulating gene expressions anaerobically (**left**) and hypothetic model of how RegA regulates gene expressions aerobically (**right**).

Evidence that RegA itself may undergo redox control was recently revealed when it was observed that RegA controls the expression of ~47 genes under aerobic growth conditions when RegB kinase activity is inhibited [10]. Interestingly, all 47 genes that are aerobically regulated by unphosphorylated RegA also belong to the group of 707 genes that are anaerobically regulated by phosphorylated RegA [10]. We suspect that there must be a subtle difference between aerobic disulfide dimerized unphosphorylated RegA and anaerobic phosphorylated RegA that differentiates which promoters that these two variations of RegA function with. For example, the binding of Cu^{2+} and O_2 dimerized RegA to the *ccoN* promoter appears to be slightly lower than that observed with the constitutively active variant RegA*. Even such a subtle difference in binding affinity may have significant effects on the promoters that RegA controls. There could also be differences by which these two differently dimerized variants interact with RNA polymerase. Further additional comparative studies will have to be undertaken to reveal how promoter activity is differentially regulated by these variants.

References

1. Gao, R.; Stock, A.M. Biological insights from structures of two-component proteins. *Ann. Rev. Microbiol.* **2009**, *63*, 133–154. [CrossRef] [PubMed]
2. Stock, A.M.; Robinson, V.L.; Goudreau, P.N. Two-component signal transduction. *Ann. Rev. Biochem.* **2000**, *69*, 183–215. [CrossRef] [PubMed]
3. Swem, L.R.; Elsen, S.; Bird, T.H.; Swem, D.L.; Koch, H.G.; Myllykallio, H.; Daldal, F.; Bauer, C.E. The RegB/RegA two-component regulatory system controls synthesis of photosynthesis and respiratory electron transfer components in *Rhodobacter capsulatus*. *J. Mol. Biol.* **2001**, *309*, 121–138. [CrossRef] [PubMed]

4. Swem, L.R.; Gong, X.; Yu, C.-A.; Bauer, C.E. Identification of a ubiquinone-binding site that affects autophosphorylation of the sensor kinase RegB. *J. Biol. Chem.* **2006**, *281*, 6768–6775. [CrossRef]

5. Wu, J.; Bauer, C.E. RegB kinase activity is controlled in part by monitoring the ratio of oxidized to reduced ubiquinones in the ubiquinone pool. *mBio* **2010**, *1*, e00272-10. [CrossRef]

6. Wu, J.; Cheng, Z.; Reddie, K.; Carroll, K.; Hammad, L.A.; Karty, J.A.; Bauer, C.E. RegB kinase activity is repressed by oxidative formation of cysteine sulfenic acid. *J. Biol. Chem.* **2013**, *288*, 4755–4762. [CrossRef]

7. Swem, L.R.; Kraft, B.J.; Swem, D.L.; Setterdahl, A.T.; Masuda, S.; Knaff, D.B.; Zaleski, J.M.; Bauer, C.E. Signal transduction by the global regulator RegB is mediated by a redox-active cysteine. *EMBO J.* **2003**, *22*, 4699–4708. [CrossRef]

8. Wu, J.; Bauer, C.E. RegB/RegA, a global redox-responding two-component system. *Adv. Exp. Med. Biol.* **2008**, *631*, 131–148.

9. Wu, J.; Dragnea, V.; Bauer, C.E. Redox Responding Sensor Kinases. In *Two Component Systems in Bacteria*; Gross, R., Beier, D., Eds.; Caister Academic Press: Würzburg, Germany, 2012; pp. 41–56.

10. Schindel, H.; Bauer, C.E. The RegA regulon exhibits variability in response to altered growth conditions and differs markedly between *Rhodobacter* species. *Microb. Genom.* **2016**, *2*, e000081. [CrossRef]

11. Masuda, S.; Matsumoto, Y.; Nagashima, K.V.P.; Shimada, K.; Inoue, K.; Bauer, C.E.; Matsuura, K. Structural and functional analyses of photosynthetic regulatory genes *regA* and *regB* from *Rhodovulum sulfidophilum*, *Roseobacter denitrificans*, and *Rhodobacter capsulatus*. *J. Bacteriol.* **1999**, *181*, 4205–4215. [CrossRef]

12. Elsen, S.; Swem, L.R.; Swem, D.L.; Bauer, C.E. RegB/RegA, a highly conserved redox-responding global two-component regulatory system. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 263–279. [CrossRef] [PubMed]

13. Swem, D.L.; Swem, L.R.; Setterdahl, A.; Bauer, C.E. Involvement of SenC in assembly of cytochrome *c* oxidase in *Rhodobacter capsulatus*. *J. Bacteriol.* **2005**, *187*, 8081–8087. [CrossRef] [PubMed]

14. Lohmeyer, E.; Schröder, S.; Pawlik, G.; Trasnea, P.-I.; Peters, A.; Daldal, F.; Koch, H.-G. The ScoI homologue SenC is a copper binding protein that interacts directly with the *cbb₃*-type cytochrome oxidase in *Rhodobacter capsulatus*. *Biochim. Biophys. Acta* **2012**, *1817*, 2005–2015. [CrossRef] [PubMed]

15. Trasnea, P.I.; Utz, M.; Khalfaoui-Hassani, B.; Lagies, S.; Daldal, F.; Koch, H.G. Cooperation between two periplasmic copper chaperones is required for full activity of the *cbb₃* -type cytochrome *c* oxidase and copper homeostasis in *Rhodobacter capsulatus*. *Mol. Microbiol.* **2016**, *100*, 345–361. [CrossRef] [PubMed]

16. McEwan, A.G.; Lewin, A.; Davy, S.L.; Boetzel, R.; Leech, A.; Walker, D.; Wood, T.; Moore, G.R. PrrC from *Rhodobacter sphaeroides*, a homologue of eukaryotic Sco proteins, is a copper-binding protein and may have a thiol-disulfide oxidoreductase activity. *FEBS Lett.* **2002**, *518*, 10–16. [CrossRef]

17. Laguri, C.; Phillips-Jones, M.K.; Williamson, M.P. Solution structure and DNA binding of the effector domain from the global regulator PrrA (RegA) from *Rhodobacter sphaeroides*: Insights into DNA binding specificity. *Nucleic Acids Res.* **2003**, *31*, 6778–6787. [CrossRef] [PubMed]

18. Du, S.; Bird, T.H.; Bauer, C.E. DNA binding characteristics of RegA. A constitutively active anaerobic activator of photosynthesis gene expression in *Rhodobacter capsulatus*. *J. Biol. Chem.* **1998**, *273*, 18509–18513. [CrossRef] [PubMed]

19. Willett, J.; Smart, J.L.; Bauer, C.E. RegA Control of bacteriochlorophyll and carotenoid synthesis in *Rhodobacter capsulatus*. *J. Bacteriol.* **2007**, *189*, 7765–7773. [CrossRef]

20. Bird, T.H.; Du, S.; Bauer, C.E. Autophosphorylation, phosphotransfer, and DNA-binding properties of the RegB/RegA two-component regulatory system in *Rhodobacter capsulatus*. *J. Biol. Chem.* **1999**, *274*, 16343–16348. [CrossRef]

21. Bauer, C.E.; Elsen, S.; Bird, T.H. Mechanisms for redox control of gene expression. *Annu. Rev. Microbiol.* **1999**, *53*, 495–523. [CrossRef]

22. Kim, S.O.; Merchant, K.; Nudelman, R.; Beyer, W.F., Jr.; Keng, T.; DeAngelo, J.; Hausladen, A.; Stamler, J.S. OxyR: A molecular code for redox-related signaling. *Cell* **2002**, *109*, 383–396. [CrossRef]

23. Masuda, S.; Dong, C.; Swem, D.; Setterdahl, A.T.; Knaff, D.B.; Bauer, C.E. Repression of photosynthesis gene expression by formation of a disulfide bond in CrtJ. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7078–7083. [CrossRef] [PubMed]

24. Cheng, Z.; Wu, J.; Setterdahl, A.; Reddie, K.; Carroll, K.; Hammad, L.A.; Karty, J.A.; Bauer, C.E. Activity of the tetrapyrrole regulator CrtJ is controlled by oxidation of a redox active cysteine located in the DNA binding domain. *Mol. Microbiol.* **2012**, *85*, 734–746. [CrossRef] [PubMed]

25. Fang, M.; Bauer, C.E. The vitamin B₁₂ -dependent photoreceptor AerR relieves photosystem gene repression by extending the interaction of CrtJ with photosystem promoters. *mBio* **2017**, *8*, e00261-17. [CrossRef] [PubMed]

26. Kiley, P.J.; Beinert, H. Oxygen sensing by the global regulator, FNR: The role of the iron-sulfur cluster. *FEMS Microbiol. Rev.* **1998**, *22*, 341–352. [CrossRef] [PubMed]

27. Peuser, V.; Remes, B.; Klug, G. Role of the Irr protein in the regulation of iron metabolism in *Rhodobacter sphaeroides*. *PLoS ONE* **2012**, *7*, e42231. [CrossRef]

28. Utz, M.; Andrei, A.; Milanov, M.; Trasnea, P.; Marckmann, D.; Daldal, F.; Koch, H. The Cu chaperone CopZ is required for Cu homeostasis in *Rhodobacter capsulatus* and influences cytochrome *cbb₃* oxidase assembly. *Mol. Microbiol.* **2019**, *111*, 764–783. [CrossRef]

29. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [CrossRef]

30. Liu, T.; Ramesh, A.; Ma, Z.; Ward, S.K.; Zhang, L.; George, G.; Talaat, A.M.; Sacchettini, J.C.; Giedroc, D. CsoR is a novel Mycobacterium tuberculosis copper-sensing transcriptional regulator. *Nat. Chem. Biol.* **2007**, *3*, 60–68. [CrossRef]
31. Ekici, S.; Yang, H.; Koch, H.G.; Daldal, F. Novel transporter required for biogenesis of cbb3-type cytochrome c oxidase in Rhodobacter capsulatus. *mBio* **2012**, *3*, e00293-11. [CrossRef]
32. Ma, Z.; Cowart, D.M.; Scott, R.A.; Giedroc, D.P. Molecular Insights into the Metal Selectivity of the Copper(I)-Sensing Repressor CsoR from *Bacillus subtilis*. *Biochemistry* **2009**, *48*, 3325–3334. [CrossRef] [PubMed]
33. Ekici, S.; Turkarslan, S.; Pawlik, G.; Dancis, A.; Baliga, N.S.; Koch, H.-G.; Daldal, F. Intracytoplasmic copper homeostasis controls cytochrome *c* oxidase production. *mBio* **2014**, *5*, e01055-13. [CrossRef] [PubMed]
34. Palumaa, P. Copper chaperones. The concept of conformational control in the metabolism of copper. *FEBS Lett.* **2013**, *587*, 1902–1910. [CrossRef] [PubMed]
35. Thompson, A.K.; Smith, D.; Gray, J.; Carr, H.S.; Liu, A.; Winge, D.R.; Hosler, J.P. Mutagenic analysis of Cox11 of *Rhodobacter sphaeroides*: Insights into the assembly of Cu(B) of cytochrome *c* oxidase. *Biochemistry* **2010**, *49*, 5651–5661. [CrossRef]
36. Mosley, C.S.; Suzuki, J.Y.; Bauer, C.E. Identification and molecular genetic characterization of a sensor kinase responsible for coordinately regulating light harvesting and reaction center gene expression in response to anaerobiosis. *J. Bacteriol.* **1994**, *176*, 7566–7573. [CrossRef]

Article

Proteomic Time-Course Analysis of the Filamentous Anoxygenic Phototrophic Bacterium, *Chloroflexus aurantiacus*, during the Transition from Respiration to Phototrophy

Shigeru Kawai [1,2,*]**, Shigeru Shimamura** [1]**, Yasuhiro Shimane** [1,3] **and Yusuke Tsukatani** [1,*]

1 Institute for Extra-Cutting-Edge Science and Technology Avant-Garde Research (X-star), Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Yokosuka 237-0061, Japan; shimas@jamstec.go.jp (S.S.); yshimane@jamstec.go.jp (Y.S.)
2 Department of Biological Sciences, Tokyo Metropolitan University, Hachioji 192-0397, Japan
3 Research Institute of Industrial Technology, Toyo University, Kawagoe 350-8585, Japan
* Correspondence: kawais@jamstec.go.jp (S.K.); tsukatani@jamstec.go.jp (Y.T.)

Abstract: *Chloroflexus aurantiacus* is a filamentous anoxygenic phototrophic bacterium that grows chemotrophically under oxic conditions and phototrophically under anoxic conditions. Because photosynthesis-related genes are scattered without any gene clusters in the genome, it is still unclear how this bacterium regulates protein expression in response to environmental changes. In this study, we performed a proteomic time-course analysis of how *C. aurantiacus* expresses proteins to acclimate to environmental changes, namely the transition from chemoheterotrophic respiratory to photoheterotrophic growth mode. Proteomic analysis detected a total of 2520 proteins out of 3934 coding sequences in the *C. aurantiacus* genome from samples collected at 13 time points. Almost all proteins for reaction centers, light-harvesting chlorosomes, and carbon fixation pathways were successfully detected during the growing phases in which optical densities and relative bacteriochlorophyll *c* contents increased simultaneously. Combination of proteomics and pigment analysis suggests that the self-aggregation of bacteriochlorophyllide *c* could precede the esterification of the hydrophobic farnesyl tail in cells. Cytoplasmic subunits of alternative complex III were interchanged between oxic and anoxic conditions, although membrane-bound subunits were used for both conditions. These data highlight the protein expression dynamics of phototrophy-related genes during the transition from respiration to phototrophy.

Keywords: *Chloroflexus aurantiacus*; proteomic analysis; photosynthesis; respiration; bacteriochlorophyll; chlorosome; alternative complex III

1. Introduction

Chloroflexus aurantiacus is a thermophilic filamentous anoxygenic phototrophic bacterium in the phylum *Chloroflexota* [1]. This bacterium grows chemoheterotrophically under oxic dark conditions and photoheterotrophically under anoxic light conditions [2]. *Chloroflexus* species and the closely related *Roseiflexus* species are widely distributed in microbial mats in hot springs, where thermophilic cyanobacteria usually coexist [3,4]. Because cyanobacterial oxygenic photosynthesis causes drastic changes in O_2 concentration over a diel cycle, co-existing microbes need to switch their metabolisms for adaptation to surrounding environments [5–8]. Therefore, the regulatory systems of protein expression are essential for the growth of *Chloroflexus* species in natural environments.

The photosynthetic apparatuses of *C. aurantiacus* consist of type II photochemical reaction center (RC) complexes, peripheral membrane-bound light-harvesting complexes, and organelle-like light-harvesting vesicles called chlorosomes [9]. The type II RC of *C. aurantiacus* is composed of core PufLM subunits, which are bound to bacteriochlorophyll (BChl) *a*, and a cytochrome *c* subunit, PufC, containing four hemes. A cryo-EM study

using the RC from *Roseiflexus castenholzii* revealed that the RC core complex structure is similar to those of purple bacterial RCs [10]. Chlorosomes are composed of self-aggregated BChl *c* surrounded by a lipid monolayer envelope. A long axis of the chlorosome vesicle forms a crystalline-like structure, called a baseplate, consisting of a large number of CsmA proteins. The lipid monolayer envelope of chlorosomes also contains several other proteins, e.g., CsmMNPOY in the case of *C. aurantiacus*, CsmBCDEFHIJX in *Chlorobaculum tepidum* (*Chlorobi*), and CsmVRSTU in *Chloracidobacterium thermophilum* (*Acidobacteria*) [11–13]. It is noted that CsmA is the sole common chlorosome protein that shows sequence similarities among the three phototrophic phyla. It has been unclear how BChl *c* self-aggregates are packed inside chlorosome envelopes. Phototrophic members in *Chlorobi* and *Acidobacteria* possess the BChl *a*-containing antenna protein, called the FMO protein, which is located between chlorosomes and RC complexes, but *C. aurantiacus* does not have the FMO protein [14–17]. Instead, *C. aurantiacus* utilizes peripheral antenna proteins, called LH complexes, which attach to RC complexes in membranes. The LH complex is composed of two transmembrane helices, α- and β-peptides, which are encoded by the *pufA* and *pufB* genes, respectively.

BChl pigments are synthesized through multistep enzymatic reactions in cells of phototrophic micro-organisms [18,19]. However, the coexistence of chlorophyllous pigments and O_2 causes the generation of reactive oxygen species. Therefore, in the case of purple bacteria (*Proteobacteria*) genes involved in BChl biosynthesis (*bch*) as well as in carotenoid biosynthesis (*crt*) and RC complexes (*puf*) are clustered together in a single location in the genome for the control of gene transcription under certain environments. The location in the genome is called a photosynthetic gene cluster (PGC), and gene expression in a PGC is suppressed under oxic conditions in purple bacteria [20,21]. In contrast, genome analysis of *C. aurantiacus* revealed that photosynthesis-related genes demonstrate scattered distribution in the genome of the bacterium [22]. It is not yet known how *C. aurantiacus* regulates the expression of photosynthesis-related genes in response to environmental changes or even whether the expression of the scattered genes is controlled concertedly.

Alternative complex III (ACIII) was found as a substitute for cytochrome bc_1 complex in *C. aurantiacus* [23] and in the nonphotosynthetic bacterium *Rhodothermus marinus* [24,25]. Thus far, ACIII has been found in many other bacteria that do not possess cytochrome bc_1 complex [26]. While nonphotosynthetic bacteria such as *R. marinus* use ACIII in the respiratory electron transfer chain, *C. aurantiacus* utilizes ACIII in both the respiratory and photosynthetic electron transfer chains [27]. In *Chloroflexus* species, two paralog gene sets for ACIII have been found, and it has been proposed that *C. aurantiacus* uses a protein complex called ACIII-C_p for photosynthetic electron transport and another protein complex, called ACIII-C_r, for the respiratory chain [22,28,29].

In this study, we performed a proteomic time-course analysis of *C. aurantiacus* in the transition from respiration to phototrophy to investigate the expression dynamics of proteins involved in the biogenesis of the photosynthetic apparatus, electron transport chains, and central carbon metabolisms. Moreover, we analyzed the growth curve, absorption spectra, and pigment compositions of the bacterium during cultivation.

2. Materials and Methods

2.1. Bacterial Growth Conditions

C. aurantiacus strain J-10-fl[T] (=DSM635) was precultured heterotrophically under oxic conditions in the dark at 48 °C with vigorous shaking in AY medium [30]. Ten milliliters of preculture were transferred to 5 L of fresh AY medium equipped in a 10 L thermostat fermenter. The oxic dark cultivation in the fermenter was performed by stirring at 350 rpm with air bubbling at 48 °C for 2 days. The fermenter was then illuminated by incandescent light with irradiance of 20 μmol photons s^{-1} m^{-2} from the front. Sodium acetate (7.5 g per 5 L) was aseptically added as the nutrient for photoheterotrophic growth. In the preliminary experiment, oxygen in the headspace was replaced by N_2 to achieve strict anoxic conditions, but no growth was observed for at least 30 days. Therefore, 2% O_2 (*v/v*) was supplemented

after the N_2 flush, after which the photoheterotrophic cultivation was started with stirring at 50 rpm at 48 °C. Absorption spectra of phototrophic cultures were measured using a UV-2600 spectrophotometer (Shimadzu, Kyoto, Japan). Cells were harvested by centrifugation on 13 days during 39 days of light cultivation. The cells harvested for proteome analysis and pigment analysis were washed three times with 100 mM triethylammonium bicarbonate (TEAB) buffer (pH 8.6) and 5 mM phosphate buffer (pH 7.5), respectively. The cell pellets were stored at −30 °C until use.

2.2. Protein Sample Preparations for LC-MS Proteomic Analysis

Cell pellets were suspended in 300 μL of 100 mM TEAB buffer containing 2 mM PMSF. The suspended cells were disrupted by sonication for 6 min (30-s sonication periods with 30-s intervals) using a Qsonica 700 sonicator (20 kHz frequency) equipped with a Qsonica 431C2 Cup Horn and a Qsonica Compact Recirculating Chiller (Qsonica, Newtown, CT, USA). The protein concentrations of the solution were determined using a Qubit Protein assay kit (Thermo Fisher Scientific, Waltham, MA, USA). The cell extracts corresponding to 10 μg of proteins were evaporated to dryness, and the dried proteins were dissolved in the denaturing buffer consisting of MPEX PTS Reagent Solutions A and B (GL Science, Tokyo, Japan). The denatured proteins were reduced by the addition of dithiothreitol (MS grade, Thermo Fisher Scientific) at a final concentration of 25 mM. The solution was incubated at 95 °C for 5 min and then at room temperature for 25 min. Iodoacetamide (MS grade, Thermo Fisher Scientific) was then added at a final concentration of 25 mM, and the solutions were incubated at room temperature for 30 min for alkylation. Protein digestion was initiated by the addition of 100 ng trypsin protease (MS grade, Thermo Fisher Scientific), and the solution was incubated at 37 °C for 3 h. Subsequently, 100 ng of trypsin protease was added again, and the proteins were further digested at 30 °C overnight. Detergents in the samples were removed using MPEX PTS Reagent Solutions C and D (GL Science) according to the manufacturer's instructions. Residual surfactants were removed using Pierce Detergent Removal Spin Column (Thermo Fisher Scientific). The digested protein solutions were evaporated to dryness. The dried proteins were resuspended in 2% acetonitrile and 0.1% trifluoroacetic acid in water and then subjected to the LC-MS/MS analysis as described below.

Peptides were separated using an Ultimate 3000 RSLCnano system (Thermo Fisher Scientific) with a reverse-phase Zaplous alpha Pep-C18 column (3 μm, 120 Å, 0.1 × 150 mm; AMR, Tokyo, Japan). The gradient was composed of solvent A (0.1% formic acid in water) and solvent B (100% acetonitrile). At the time of injection, the mobile phase was 5% solvent B at a flow rate of 500 nL min^{-1}. Solvent B was linearly increased to 45% over 100 min. The column temperature was set at 35 °C via a nano-electrospray ion source (Dream Spray, AMR, Tokyo, Japan). The Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific) was operated in the positive ion mode. For peptide ionization, an electrospray voltage of 1.7 kV was applied, and the ion transfer tube temperature was set at 250 °C. All MS spectra were acquired in the Orbitrap mass analyzer (*m/z* range: 350–1800, resolution: 120,000 FWHM) with EASY-IC internal mass calibration. MS/MS spectra resulting from collision-induced dissociation (CID) fragmentation were acquired in an ion trap mass analyzer (*m/z* range: auto, scan rate: rapid).

2.3. Data Analysis

Protein identification, label-free quantification of detected protein derivatives, and principal component analysis (PCA) were performed with the Proteome Discoverer ver. 2.2 software package (Thermo Fisher Scientific). The search for acquired spectra was carried out using the SEQUEST HT search algorithm against a list of CDSs identified in the genome of *C. aurantiacus*. The following search parameters were used: trypsin was specified as protease allowing two missed cleavage sites at the maximum; cysteine carbamidomethylation was set as a fixed modification; methionine oxidation was set as a variable modification; and maximum error tolerances in MS and MS/MS were set at

10 ppm and 0.6 Da, respectively. Peptides corresponding to a <1% protein false discovery rate (FDR) were used in the calculations.

Relative values of peptide detection for each protein are calculated for the comparison among 13 sampling dates. The calculated relative abundance was normalized with a percentage in each protein. The Sum PEP score of each protein was calculated in Proteome Discoverer ver. 2.2.

2.4. Pigment Extraction and HPLC Analysis

Pigments were extracted from the harvested cells with acetone–methanol (7:2, v/v). The solvent was evaporated to dryness by a stream of N_2 gas, and the dried pigments were dissolved in the HPLC solvent described below. The pigment solution was filtered using a Cosmonice filter (0.45 μm; Nacalai Tesque, Kyoto, Japan); 0.1 mM of ammonium acetate was added to the filtrated solution before injection. Reverse-phase HPLC measurements were performed using a Cosmosil 5C18-AR-II (5 μm, 4.6 × 250 mm, Nacalai Tesque) column and the mobile phase of solvent A (methanol:acetonitrile:water, 42:33:25, $v:v:v$) and solvent B (methanol:acetonitrile:ethyl acetate, 50:20:30, $v:v:v$). The elution was conducted by the following method as previously described, with a modification [31,32]. The initial composition of solvents was 30% solvent B. The linear gradient was increased to 100% solvent B for 52 min followed by 6 min of constant flow. The solvent composition was then gradually returned to 30% solvent B over 2 min and held for 5 min.

3. Results

3.1. Growth Profiles

Growth curves of *C. aurantiacus* after the illumination started are shown in Figure 1. The growth profiles were evaluated using two sets of data: cell densities by measuring absorbance at 600 nm (A_{600}) (Figure 1A) and relative BChl *c* contents (A_{740}/A_{600}) (Figure 1B). The A_{740} value is reflected from an absorption band from BChl *c* aggregates in chlorosomes. During the first 6 days, the A_{600} value decreased from 0.36 to 0.10 and relative BChl *c* contents did not change. The period of days 0 to 5 was designated the first lag phase (Figure 1, orange). During the next period, from days 6 to 10, the relative BChl *c* contents increased slightly while cell density values did not change. This period was designated the second lag phase (Figure 1, yellow). Significant increases in both cell densities and relative BChl *c* contents were observed from days 12 to 17 (designated as the first growing phase; Figure 1, light green). The doubling time during this period was 3.2 ± 0.5 days with the use of A_{600} values measured at days 12 to 17. After day 17, no increase in relative BChl *c* contents was observed, but cell densities in the A_{600} values kept increasing, with a doubling time of 11.8 ± 1.6 days (calculated by A_{600} values measured at days 20 to 32). This period was designated the second growing phase (Figure 1, dark green).
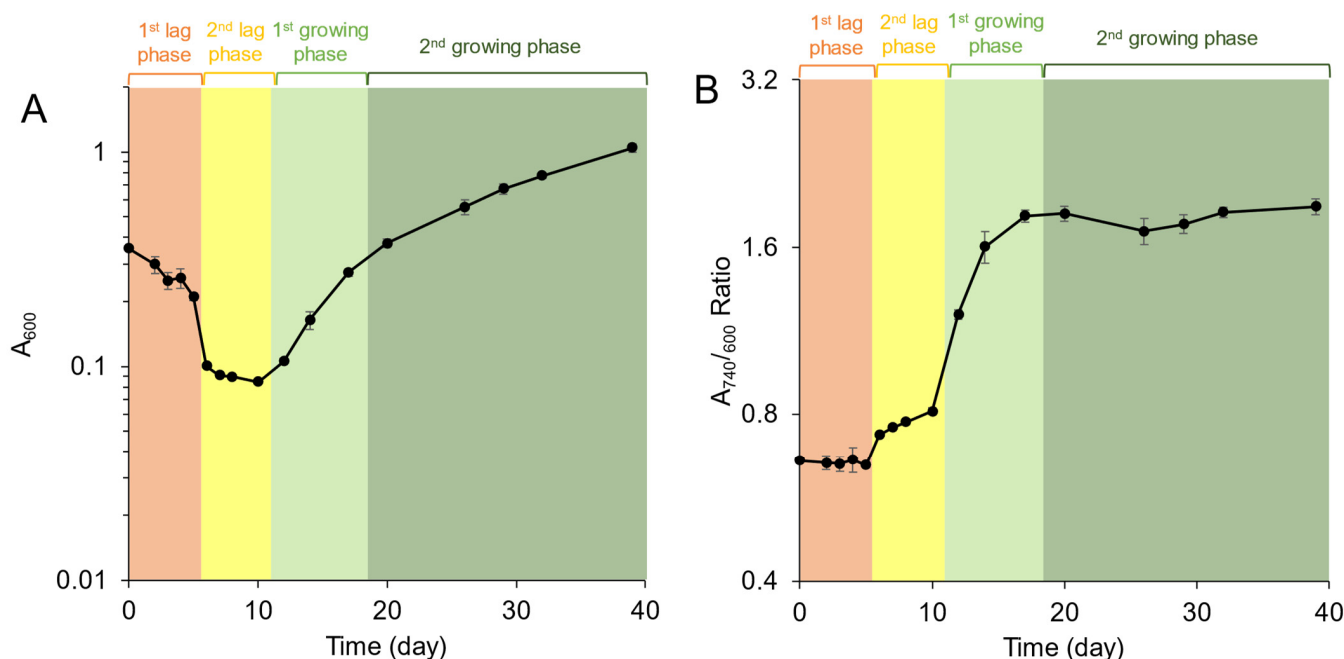
Figure 1. Photoheterotrophic growth profiles of *C. aurantiacus* represented by cell density values (**A**) and relative bacteriochlorophyll *c* 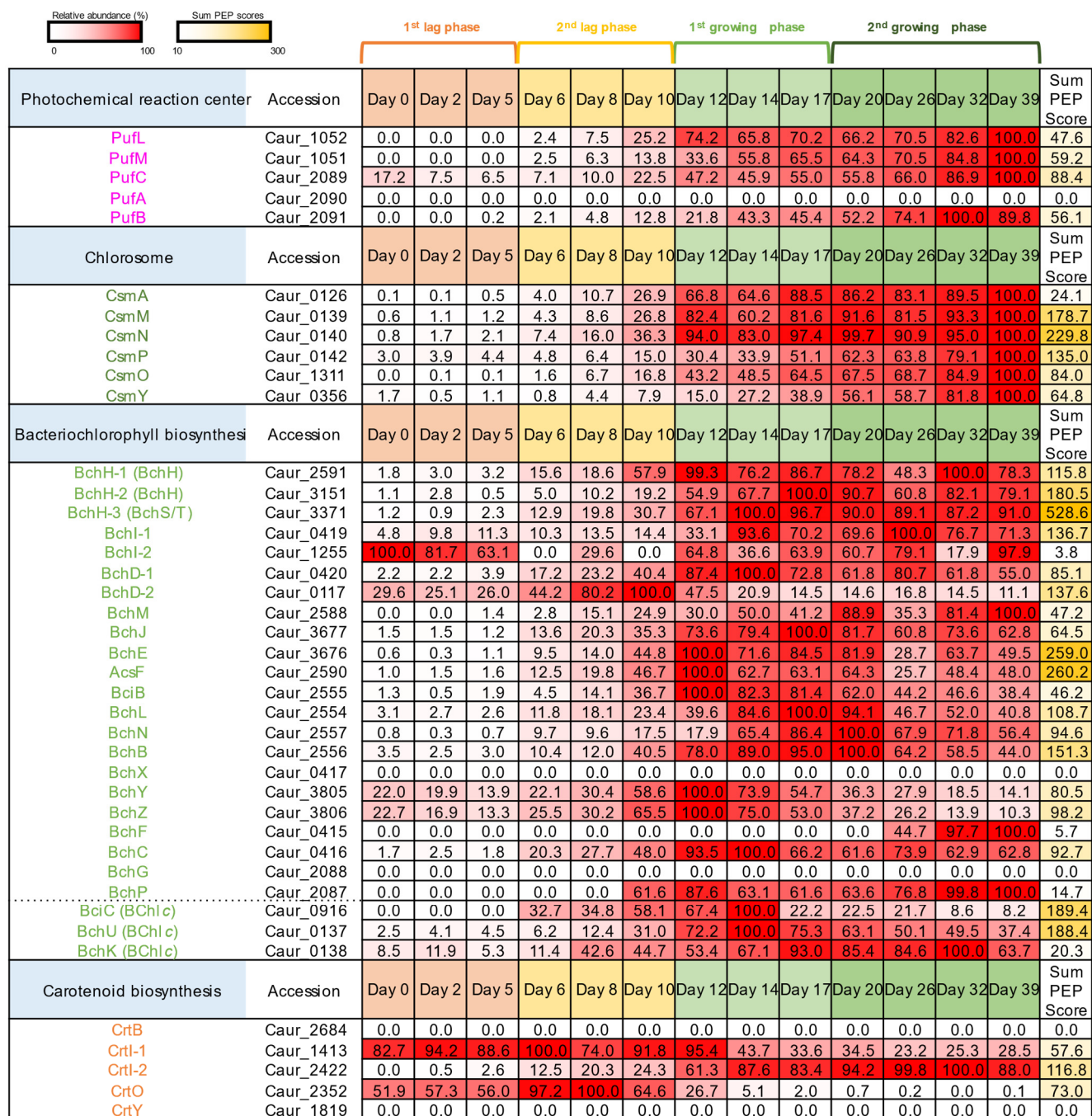contents (**B**). Measurements started after illumination. Orange, the first lag phase (days 0–5); yellow, the second lag phase (days 6–10); light green, the first growing phase (days 12–17); dark green, the second growing phase (days 20–39). The mean values of three measurements are shown with standard deviations.

3.2. Protein Expression of Photosynthetic Proteins in C. aurantiacus

Proteomic analysis of *C. aurantiacus* in the transition from respiration to phototrophy was performed using protein samples collected at 13 different time points after illumination (days 0–39). Peptides derived from a total of 2520 proteins out of 3934 CDSs in deduced genome sequences of *C. aurantiacus* were successfully detected from the LC-MS/MS proteomic analysis of 2 duplicates of the 13 sampling dates. All proteomic datasets were compared using PCA (Supplementary Figure S1). The PCA data showed that duplicates derived from the same sampling dates were clustered together, confirming the reproducibility of proteome data among duplicates.

Figure 2 shows the expression patterns of protein subunits and enzymes related to RC complexes, chlorosomes, and bacteriochlorophyll and carotenoid biosynthesis. Relative values of peptide detection for each protein among 13 sampling dates are shown in Figures 2 and 3 as percentages indicating abundance, where "100%" indicates the highest protein expression among all sampling dates. Sum PEP scores indicate the relative values of protein amounts for the comparison of protein expression levels between different CDSs. Almost all the photosynthetic proteins were successfully detected with high abundance in the growing phase (days 12–39) compared to the lag phase (days 0–10), except PufA, BchX, BchG, CrtB, and CrtY (Figure 2). The abundance of peptides for RC complexes and chlorosomes increased along with cellular growth, indicating that these proteins support phototrophic growth. Peptides derived from PufLMC subunits of the RC and all of the Csm proteins were detected during the first and second growing phases, with the highest abundance at the last date of sampling (Figure 1, day 39). On the other hand, peptides from Bch enzymes for BChls *a* and *c* biosynthesis were already detected in the second lag phase, and several Bch enzymes became the most abundant on day 12 or 14 of the first growing phase (BchHDEYZCU, BciBC, and AcsF). The expression patterns of Bch proteins were consistent with the observation of a slight increase in relative BChl *c* content in the second lag phase and rapid increase in the first growing phase (Figure 1B). These results suggest that BChl biosynthesis precedes the formation of the RC-LH complex and chlorosomes.

Photochemical reaction center	Accession	Day 0	Day 2	Day 5	Day 6	Day 8	Day 10	Day 12	Day 14	Day 17	Day 20	Day 26	Day 32	Day 39	Sum PEP Score
PufL	Caur_1052	0.0	0.0	0.0	2.4	7.5	25.2	74.2	65.8	70.2	66.2	70.5	82.6	100.0	47.6
PufM	Caur_1051	0.0	0.0	0.0	2.5	6.3	13.8	33.6	55.8	65.5	64.3	70.5	84.8	100.0	59.2
PufC	Caur_2089	17.2	7.5	6.5	7.1	10.0	22.5	47.2	45.9	55.0	55.8	66.0	86.9	100.0	88.4
PufA	Caur_2090	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
PufB	Caur_2091	0.0	0.0	0.2	2.1	4.8	12.8	21.8	43.3	45.4	52.2	74.1	100.0	89.8	56.1

Chlorosome	Accession	Day 0	Day 2	Day 5	Day 6	Day 8	Day 10	Day 12	Day 14	Day 17	Day 20	Day 26	Day 32	Day 39	Sum PEP Score
CsmA	Caur_0126	0.1	0.1	0.5	4.0	10.7	26.9	66.8	64.6	88.5	86.2	83.1	89.5	100.0	24.1
CsmM	Caur_0139	0.6	1.1	1.2	4.3	8.6	26.8	82.4	60.2	81.6	91.6	81.5	93.3	100.0	178.7
CsmN	Caur_0140	0.8	1.7	2.1	7.4	16.0	36.3	94.0	83.0	97.4	99.7	90.9	95.0	100.0	229.8
CsmP	Caur_0142	3.0	3.9	4.4	4.8	6.4	15.0	30.4	33.9	51.1	62.3	63.8	79.1	100.0	135.0
CsmO	Caur_1311	0.0	0.1	0.1	1.6	6.7	16.8	43.2	48.5	64.5	67.5	68.7	84.9	100.0	84.0
CsmY	Caur_0356	1.7	0.5	1.1	0.8	4.4	7.9	15.0	27.2	38.9	56.1	58.7	81.8	100.0	64.8

Bacteriochlorophyll biosynthesis	Accession	Day 0	Day 2	Day 5	Day 6	Day 8	Day 10	Day 12	Day 14	Day 17	Day 20	Day 26	Day 32	Day 39	Sum PEP Score
BchH-1 (BchH)	Caur_2591	1.8	3.0	3.2	15.6	18.6	57.9	99.3	76.2	86.7	78.2	48.3	100.0	78.3	115.8
BchH-2 (BchH)	Caur_3151	1.1	2.8	0.5	5.0	10.2	19.2	54.9	67.7	100.0	90.7	60.8	82.1	79.1	180.5
BchH-3 (BchS/T)	Caur_3371	1.2	0.9	2.3	12.9	19.8	30.7	67.1	100.0	96.7	90.0	89.1	87.2	91.0	528.6
BchI-1	Caur_0419	4.8	9.8	11.3	10.3	13.5	14.4	33.1	93.6	70.2	69.6	100.0	76.7	71.3	136.7
BchI-2	Caur_1255	100.0	81.7	63.1	0.0	29.6	0.0	64.8	36.6	63.9	60.7	79.1	17.9	97.9	3.8
BchD-1	Caur_0420	2.2	2.2	3.9	17.2	23.2	40.4	87.4	100.0	72.8	61.8	80.7	61.8	55.0	85.1
BchD-2	Caur_0117	29.6	25.1	26.0	44.2	80.2	100.0	47.5	20.9	14.5	14.6	16.8	14.5	11.1	137.6
BchM	Caur_2588	0.0	0.0	1.4	2.8	15.1	24.9	30.0	50.0	41.2	88.9	35.3	81.4	100.0	47.2
BchJ	Caur_3677	1.5	1.5	1.2	13.6	20.3	35.3	73.6	79.4	100.0	81.7	60.8	73.6	62.8	64.5
BchE	Caur_3676	0.6	0.3	1.1	9.5	14.0	44.8	100.0	71.6	84.5	81.9	28.7	63.7	49.5	259.0
AcsF	Caur_2590	1.0	1.5	1.6	12.5	19.8	46.7	100.0	62.7	63.1	64.3	25.7	48.4	48.0	260.2
BciB	Caur_2555	1.3	0.5	1.9	4.5	14.1	36.7	100.0	82.3	81.4	62.0	44.2	46.6	38.4	46.2
BchL	Caur_2554	3.1	2.7	2.6	11.8	18.1	23.4	39.6	84.6	100.0	94.1	46.7	52.0	40.8	108.7
BchN	Caur_2557	0.8	0.3	0.7	9.7	9.6	17.5	17.9	65.4	86.4	100.0	67.9	71.8	56.4	94.6
BchB	Caur_2556	3.5	2.5	3.0	10.4	12.0	40.5	78.0	89.0	95.0	100.0	64.2	58.5	44.0	151.3
BchX	Caur_0417	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
BchY	Caur_3805	22.0	19.9	13.9	22.1	30.4	58.6	100.0	73.9	54.7	36.3	27.9	18.5	14.1	80.5
BchZ	Caur_3806	22.7	16.9	13.3	25.5	30.2	65.5	100.0	75.0	53.0	37.2	26.2	13.9	10.3	98.2
BchF	Caur_0415	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	44.7	97.7	100.0	5.7
BchC	Caur_0416	1.7	2.5	1.8	20.3	27.7	48.0	93.5	100.0	66.2	61.6	73.9	62.9	62.8	92.7
BchG	Caur_2088	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
BchP	Caur_2087	0.0	0.0	0.0	0.0	0.0	61.6	87.6	63.1	61.6	63.6	76.8	99.8	100.0	14.7
BciC (BChl *c*)	Caur_0916	0.0	0.0	0.0	32.7	34.8	58.1	67.4	100.0	22.2	22.5	21.7	8.6	8.2	189.4
BchU (BChl *c*)	Caur_0137	2.5	4.1	4.5	6.2	12.4	31.0	72.2	100.0	75.3	63.1	50.1	49.5	37.4	188.4
BchK (BChl *c*)	Caur_0138	8.5	11.9	5.3	11.4	42.6	44.7	53.4	67.1	93.0	85.4	84.6	100.0	63.7	20.3

Carotenoid biosynthesis	Accession	Day 0	Day 2	Day 5	Day 6	Day 8	Day 10	Day 12	Day 14	Day 17	Day 20	Day 26	Day 32	Day 39	Sum PEP Score
CrtB	Caur_2684	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CrtI-1	Caur_1413	82.7	94.2	88.6	100.0	74.0	91.8	95.4	43.7	33.6	34.5	23.2	25.3	28.5	57.6
CrtI-2	Caur_2422	0.0	0.5	2.6	12.5	20.3	24.3	61.3	87.6	83.4	94.2	99.8	100.0	88.0	116.8
CrtO	Caur_2352	51.9	57.3	56.0	97.2	100.0	64.6	26.7	5.1	2.0	0.7	0.2	0.0	0.1	73.0
CrtY	Caur_1819	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Figure 2. The relative abundance of proteins involved in reaction centers, chlorosomes, and the biosynthesis of bacteriochlorophylls and carotenoids. Relative values of protein expression among 13 sampling dates are shown as percentages indicating abundance, where "100%" indicates the highest protein expression among all sampling dates. The red color becomes deeper as the relative abundance values increase. Sum PEP scores are shown for the comparison of protein expression levels between different CDSs. The yellow color becomes deeper as the Sum PEP scores increase.

In contrast, some enzymes for carotenoid biosynthesis (CrtO and CrtI-1) were expressed in the lag phase (Figure 2, days 0–10). CrtO is carotenoid ketolase, and ketolated carotenoids are known to scavenge reactive oxygen species [32,33]. The expression of CrtO ketolase was likely to be finished at around days 12–14, concomitant with the initiation of the expression of RCs and chlorosomes. This implies that oxygen added to the fermenter was consumed for up to 10–12 days, which may have initiated the phototrophic growth mode.

Alternative Complex III	Accession	Day 0	Day 2	Day 5	Day 6	Day 8	Day 10	Day 12	Day 14	Day 17	Day 20	Day 26	Day 32	Day 39	Sum PEP Score
actA (Cp)	Caur_0621	1.0	1.4	2.0	1.9	15.2	13.1	12.6	48.2	54.7	51.9	61.2	79.1	100.0	22.9
actB (Cp)	Caur_0622	2.7	4.0	3.8	12.2	20.4	29.3	46.0	46.2	56.0	63.2	63.7	82.7	100.0	596.0
actC (Cp)	Caur_0623	25.6	30.0	29.4	73.1	100.0	92.7	56.5	40.6	37.1	39.1	44.3	56.3	67.8	67.8
actD (Cp)	Caur_0624	21.8	24.3	18.6	54.3	100.0	98.3	29.1	45.4	44.9	58.9	64.4	81.7	80.4	83.3
actE (Cp)	Caur_0625	1.3	2.3	4.5	20.8	36.8	30.9	25.6	53.4	63.0	58.0	62.7	83.5	100.0	77.7
actF (Cp)	Caur_0626	5.7	26.2	38.3	100.0	87.5	19.7	11.4	41.2	34.9	31.8	30.3	55.6	70.2	53.2
actG (Cp)	Caur_0627	4.5	4.3	3.0	10.5	29.1	52.6	100.0	58.1	65.1	70.7	69.7	83.1	99.7	34.8
actB (Cr)	Caur_2136	32.1	32.0	31.9	80.2	100.0	76.3	30.8	9.3	4.9	4.5	3.5	3.1	4.3	465.4
actE (Cr)	Caur_2137	24.0	25.2	23.3	62.2	96.1	100.0	44.5	11.0	5.7	4.9	3.7	2.8	4.0	80.4
actA (Cr)	Caur_2138	39.6	100.0	29.3	0.0	16.2	24.4	22.5	0.0	0.0	0.0	0.0	0.0	0.0	26.3
actG (Cr)	Caur_2139	22.8	24.7	20.2	64.6	100.0	89.8	37.8	8.4	4.6	3.5	2.3	1.8	2.0	53.4
auracyanin A	Caur_3248	35.6	64.8	86.3	100.0	99.7	79.9	52.9	77.2	64.5	53.4	49.6	53.2	49.0	75.0
auracyanin B	Caur_1950	4.8	0.1	0.6	0.0	0.6	1.6	3.2	18.6	23.3	32.9	42.8	78.3	100.0	15.5

Cytochrome c oxidase	Accession	Day 0	Day 2	Day 5	Day 6	Day 8	Day 10	Day 12	Day 14	Day 17	Day 20	Day 26	Day 32	Day 39	Sum PEP Score
subunit II	Caur_2141	29.2	31.8	23.2	57.1	96.6	100.0	43.2	13.5	7.4	6.8	4.7	4.2	4.3	80.8
subunit I	Caur_2142	20.7	30.2	32.7	80.2	100.0	66.5	21.5	8.1	3.6	2.3	1.3	0.9	0.9	70.1
subunit III	Caur_2143	20.7	18.5	25.3	78.8	100.0	87.5	46.7	8.0	2.4	2.8	2.1	2.1	3.8	53.7
subunit IV	Caur_2144	6.3	6.9	8.6	33.5	80.1	100.0	67.9	6.4	2.3	0.1	0.5	1.1	0.0	9.7

Phases: 1st lag phase (Day 0, 2, 5); 2nd lag phase (Day 6, 8, 10); 1st growing phase (Day 12, 14, 17); 2nd growing phase (Day 20, 26, 32, 39). Relative abundance (%): 0–100. Sum PEP scores: 10–300.

Figure 3. The relative abundance of proteins involved in photosynthetic and respiratory electron transfer chains. Relative values of protein expression among 13 sampling dates and Sum PEP scores are shown as described in Figure 2.

3.3. Photosynthetic and Respiratory Electron Transfer Chain

ACIII of the photosynthesis version in *C. aurantiacus* comprises seven subunits: penta-heme subunit ActA, iron-sulfur binding subunit ActB, transmembrane subunits ActCDF, mono-heme subunit ActE, and membrane-bound subunit ActG [29]. *C. aurantiacus* has all of the seven genes for ACIII-C_p in a single operon (Caur_0621–0627). ActABEG (C_p) demonstrated the highest abundance in the second growing phase, whereas peptides derived from ActCDF (C_p) already reached the highest abundance in the second lag phase following relatively high abundance (30~80%) during the first and second growing phases (Figure 3). Peptides derived from auracyanin A and auracyanin B, which are copper-containing soluble electron carrier proteins assumed to be substitutes of soluble cytochrome *c*, were also detected. The expression of auracyanin A was relatively stable during the whole period, with the lowest abundance being 36%. In contrast, the peptide abundance of auracyanin B increased only at the growing phase (Figure 3), suggesting that auracyanin B functions in photosynthetic electron transport, probably between the RC and ACIII.

Genes encoding respiratory-type ACIII-C_r subunits are clustered together with genes for cytochrome *c* oxidase (called COX or Complex IV) in a single operon in the genome of *C. aurantiacus* [26,34]. The respiratory-type ActBEAG (C_r) peptides were detected during the lag phase and were less abundant during the growing phase (Figure 3). In a similar fashion, peptides derived from four subunits of cytochrome *c* oxidase were scarcely detected after 14 days (Figure 3). These results indicate the concerted expression of proteins for ACIII-C_r and terminal cytochrome *c* oxidase.

3.4. Central Carbon Metabolisms

Chloroflexus species use the oxidative tricarboxylic acid (TCA) cycle for NADH production and the 3-hydroxypropionate (3-OHP) bi-cycle for carbon fixation. The 3-OHP bi-cycle is found only in *Chloroflexota* bacteria [35,36]. Results of the proteomic analysis for proteins related to the central carbon metabolisms are shown as a pathway map in Figure 4. Most of the proteins related to the TCA cycle including glyoxylate bypass were abundant in the first lag phase (Figure 4, left-hand side). The result was consistent with the idea that the TCA cycle produces NADH for respiratory electron transfer under oxic conditions.
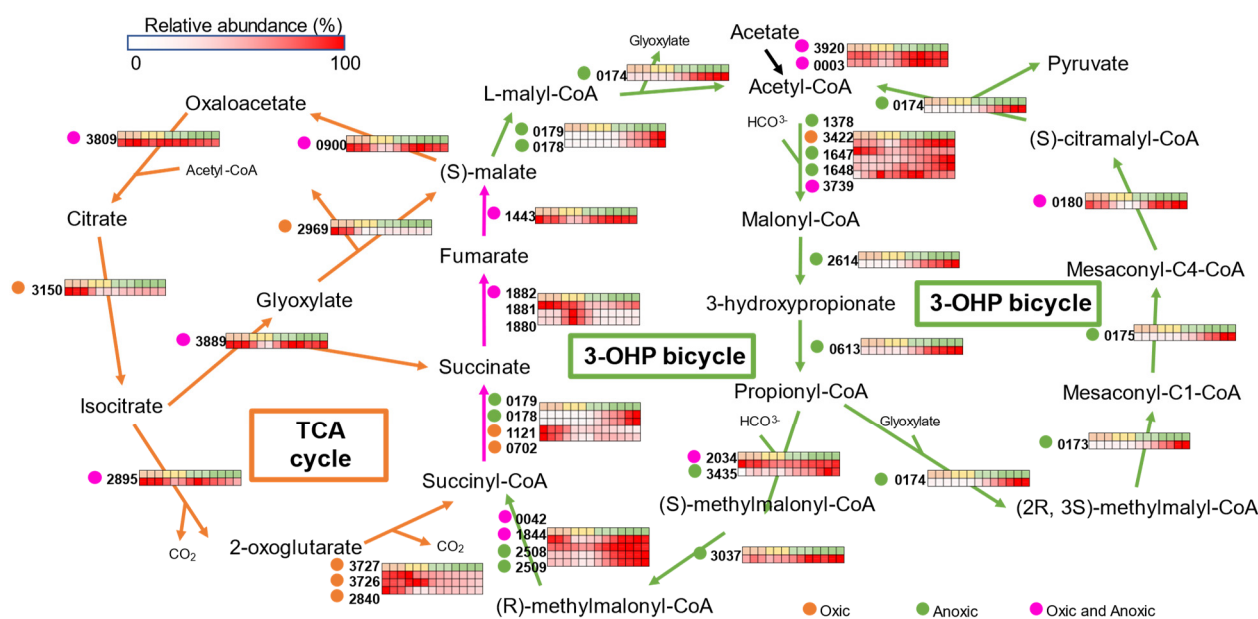
Figure 4. The heatmap showing the relative abundance of each protein involved in central carbon metabolisms in *C. aurantiacus*. The first row on the heat map corresponds to the number of cultivation days: orange, days 0–5; yellow, days 6–10; light green, days 12–17; dark green, days 20–39. Arrows indicate the enzymatic reactions for relevant metabolisms: orange, TCA cycle; green, 3-OHP bi-cycle; pink, both TCA and 3-OHP. Four-digit numbers on the left of heat maps represent locus tag numbers of *C. aurantiacus* genes.

Two key enzymes for the 3-OHP cycle, malonyl-CoA reductase (*mcr*, Caur_2614) and propionyl-CoA synthase (*psc*, Caur_0613), were abundant in the growing phase (Figure 4). This clearly indicates that the 3-OHP pathway is active during phototrophic growth, although no CO_2 was supplied to the culture medium.

The conversion of succinyl-CoA to malate is the common step in both the TCA cycle and the 3-OHP bi-cycle (Figure 4). Succinate is synthesized via two enzymatic reactions: CoA transfer from succinyl-CoA to malate by succinyl-CoA-L-malate CoA-transferase (Caur_0178–0179), and the reverse reaction by succinyl-CoA synthase (Caur_0702 and Caur_1121) with succinyl-CoA. Succinyl-CoA-L-malate CoA-transferase was abundant in the second growing phase, whereas succinyl-CoA synthase was abundant in the first lag phase. The synthesized succinate is oxidized by type B succinate dehydrogenase comprising an iron-sulfur subunit (*sdhB*, Caur_1880), a flavoprotein subunit (*sdhA*, Caur_1881), and a cytochrome *b* subunit (*sdhC*, Caur_1882) [37,38]. The expression patterns showed that SdhC was abundant in the first lag phase, whereas SdhAB was abundant in the second lag phase (Figure 4). Fumarate lyase (Caur_1443), which is involved in malate synthesis, was abundant in the first lag phase and in the phototrophic growing phase.

3.5. Pigment Compositions during the Cultivation

The pigment composition of *C. aurantiacus* cells was analyzed by reverse-phase HPLC. Two apparent peaks from carotenoids were detected at the retention times of 38.6 min and 42.8 min at days 0 to 10 (Figure 5A, peaks 1 and 2). These were temporarily assigned as 4-keto-γ-carotene and echinenone (ketolated β-carotene), based on their in-line absorption spectra (Supplementary Figure S2). This is consistent with the protein expression data showing that carotenoid ketolase CrtO was abundant until day 10 (Figure 2). After day 12, two other peaks appeared at retention times of 49.2 min and 52.1 min, and they were assigned as γ-carotene and β-carotene, based on the previous reports (Figure 5A, peaks 3 and 4) [1,39]. These results indicate that carotenoid compositions were different from cells grown under oxic and anoxic conditions.
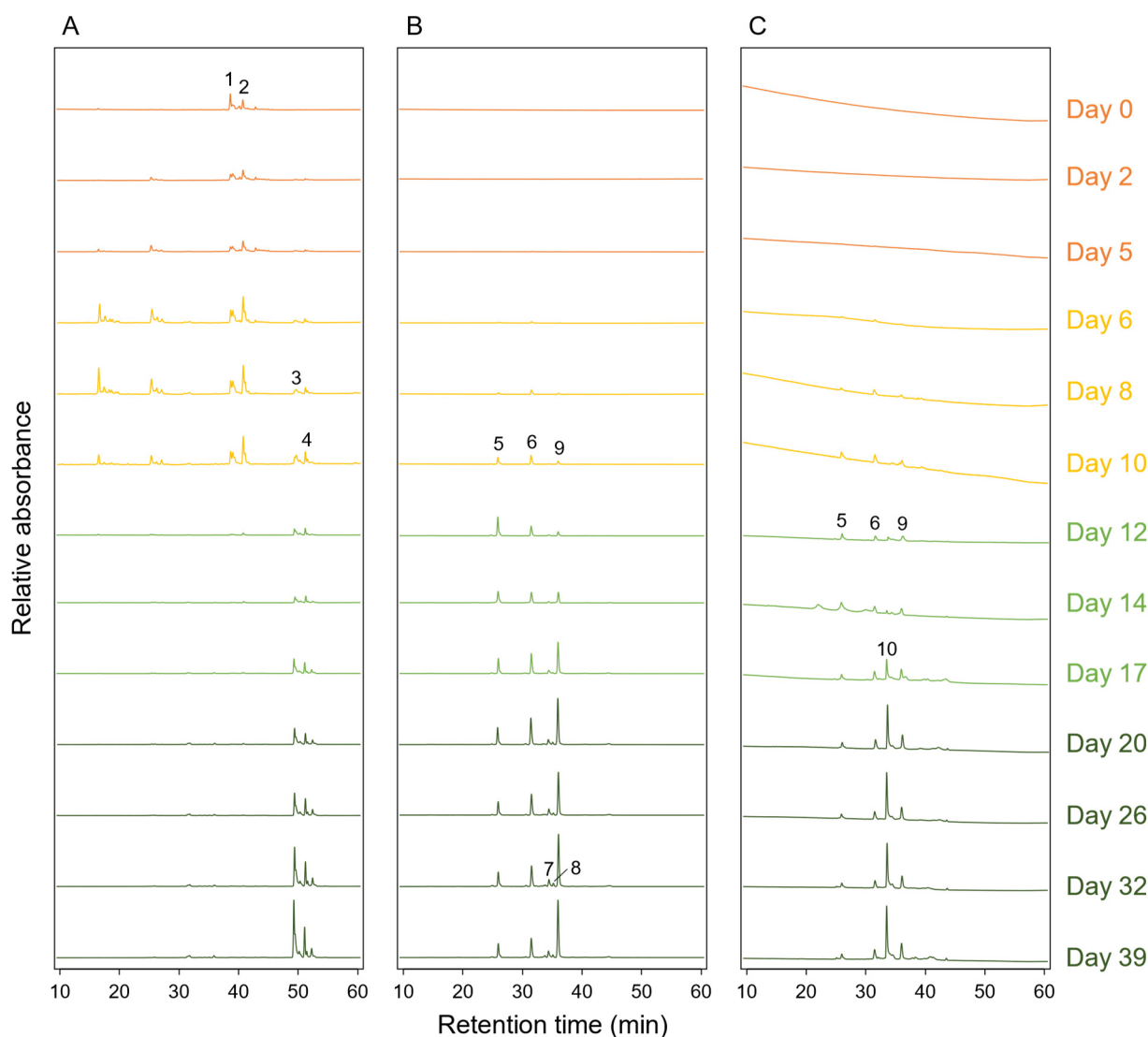
Figure 5. Reverse-phase HPLC elution profiles of pigments extracted from *C. aurantiacus* cells collected at the 13 sampling dates. Elution profiles when absorbance was measured at 490 nm, 667 nm, and 768 nm are shown in panels (**A–C**), respectively. The chromatograms in the first and second lag phases (days 0 to 10) were magnified five times for the clarity. The colors in elution profiles correspond to sampling dates: orange, days 0–5; yellow, days 6–10; light green, days 12–17; dark green, days 20–39. Putative assignments of peaks are as follows: peak 1, 4-keto-γ-carotene; peak 2, echinenone; peak 3, γ-carotene; peak 4, β-carotene; peaks 5, 6, 7, and 9, BChl *c*; peak 8, BChl *d*; peak 10, BChl *a*.

HPLC elution peaks from BChl *c* homologs were detected when measured at 667 nm (Figure 5B). Three peaks with retention times of 26.1, 31.6, and 36.1 were observed after day 6 (Figure 5B, peaks 5, 6, and 9). The in-line absorption spectra of the three elution peaks were identical to that of typical BChl *c* with Q_y/Soret absorbance maxima at 667/434 nm (Supplementary Figure S2). After day 20, two other small peaks could be seen at 34.7 and 35.2 min (Figure 5B, peaks 7 and 8). The in-line absorption spectrum of the small peak 8 showed Q_y/Soret absorbance maxima at 657/427 nm (Supplementary Figure S2), and therefore, the elution peak was ascribable to BChl *d*. Detection of the small amount of BChl *d* in *C. aurantiacus* was also reported previously [40].

Figure 5C shows the HPLC elution profile when absorbance was measured at 768 nm. Together with the elution peaks derived from BChl *c* (peaks 5, 6, and 9, as described above), a new substance was eluted at 33.7 min after day 12 (Figure 5C, peak 10). The in-line

absorption spectrum of the peak 10 was typical to that of BChl *a* (Supplementary Figure S2). These results indicate that BChl *a* started to be produced after day 12 in the growing phase, while BChl *c* already appeared after day 6 in the lag phase.

4. Discussion

Chloroflexus aurantiacus grows chemoheterotrophically under dark oxic conditions and photoheterotrophically under light anoxic conditions. Genes related to phototrophy are scattered in the genome of this organism [22]. This led to speculation that photosynthetic genes are not tightly regulated in *C. aurantiacus*. Proteomic analysis in this study revealed the expression dynamics of proteins involved in the photochemical RC complex, light-harvesting chlorosomes, pigment biosynthesis, electron transfer chains, and carbon fixation (Figures 2–4). The results suggest the photosynthesis-related genes are concertedly regulated in *C. aurantiacus* despite the scattered distribution.

CsmMN reached almost the highest expression level at day 12 of cultivation (82–94% compared to the highest abundance at day 39, Figure 2), whereas CsmPOY increased gradually during the growing phase and reached their highest abundance at day 39. CsmMN proteins of *C. aurantiacus* are distantly related to CsmC/CsmD family envelope proteins of *C. tepidum*, which are involved in the assembly and glycolipid composition of the chlorosome envelope [41]. This suggests that CsmMN proteins, and probably also CsmA baseplate protein, are the components in the formation of premature, developing chlorosomes. On the other hand, CsmPOY could be important for the maturation of chlorosomes rather than for the initial structural formation.

BciC is a recently characterized enzyme for the removal of C-13^2 methylcarboxyl groups which is the first committed step for BChl *c* biosynthesis [42]. Therefore, the enzymatic activity of BciC is essential for the formation of BChl *c* self-aggregates in chlorosomes. BciC reached its most abundant level at day 14 and then decreased to approx. one-fifth of its peak at day 17 (Figure 2). Simultaneously, the relative BChl *c* content in the cell did not increase after day 17 (Figure 1B). On the other hand, BchK was still abundant after day 17 of cultivation, and the highest abundance of BchK was at day 32. BchK is specific to BChl *c* biosynthesis and catalyzes esterification of the farnesyl tail at the final biosynthetic step of BChl *c*. The absorption peak of the monomer BChl *c* is at around 665 nm, while that of the aggregated BChl *c* is at around 740 nm. These results suggest that the self-aggregation of BChl *c* precedes the esterification of the hydrophobic farnesyl tail, and that BchK might work on the self-aggregated bacteriochlorophyllide *c*. *C. aurantiacus* is known to accumulate BChl *c* homologs with different alcohol chains at the carbon-17 position, usually geranylgeraniol, 9-octadecenol, and hexadecanol [43]. It was also reported that *C. aurantiacus* and *C. tepidum* can take up unusual alcohols for the long alkyl tail of BChl *c* when exogenous alcohols were supplied to the culture [40,44]. The low substrate specificity on the BChl tail esterification could be related to the speculated BchK enzymatic activity that the esterification occurred after BChl *c* self-aggregation.

There are two hypotheses for the process of the chlorosome biogenesis: "budding model" suggests the accumulation of BChl *c* and quinones in the cytoplasmic membrane triggers the chlorosome formation; "aggregation model" suggests CsmA-BChl *a* complexes are synthesized and assembled into a baseplate before the aggregation of other chlorosomal components and pigments (for the details of models, see references) [45–47]. The present study revealed that the BChl *c* production precedes the BChl *a* production in *C. aurantiacus* (Figures 2 and 5). The results in this study support the "budding model", at least in *C. aurantiacus*.

Mg-chelatase is composed of three subunits (BchH, BchI, and BchD). The *C. aurantiacus* genome contains three paralogs for *bchH* and two paralogs each for *bchI* and *bchD* genes. In *C. tepidum* in the phylum *Chlorobi*, three paralogs of *bchH* genes (CT1295, CT1955, and CT1957) were named *bchS*, *bchT*, and *bchH*, respectively. Two of the three *bchH* genes in *C. aurantiacus* (Caur_3151 and Caur_3371) are homologous to *bchH* of *C. tepidum*, and another (Caur_2591) is homologous to *bchS/T*, as determined by blastp analysis. Based

on the protein expression patterns, all three homolog proteins of BchH seem to be active during the phototrophic growing phase. Two paralogs for *bchI* and *bchD* were likely to be expressed in different ways. *bchI*-1 (Caur_0419) and *bchD*-1 (Caur_0420) are included in the same operon as *bchF*, *bchC*, and *bchX*, and proteins from these genes were expressed during the growing phase (Figure 2). In contrast, BchI-2 (Caur_1255) and BchD-2 (Caur_0117) were expressed mainly during the lag phase, suggesting that these enzymes had different roles than in the photosynthetic process.

HPLC analysis demonstrated that carotenoid compositions were different between oxic and anoxic conditions (Figure 5). Two paralogs of CrtI showed opposite protein expression patterns: CrtI-1 (Caur_1413) was expressed in the lag phase and CrtI-2 (Caur_2422) was expressed in the growing phase (Figure 2). Together, these results imply that CrtI-1 and CrtI-2 are involved in the synthesis of ketolated and nonketo carotenoids, respectively.

Both ACIII-C_p and ACIII-C_r in *C. aurantiacus* are presumed to consist of the ActABCDEFG subunits [27]. However, the gene operon for the C_r type includes *actABEG* but does not include *actCDF*. In this study, peptides from ActCDF, encoded by genes in the C_p operon, were stably expressed in both the lag and growing phases, suggesting that the ActCDF subunits are used for not only C_p type but also C_r type ACIII supercomplex formation (Figure 6). Cytoplasmic ActABE subunits of C_p and C_r types are likely to be interchanged between oxic and anoxic conditions (Figure 6). The putative membrane-bound ActG subunit is a novel, specific protein for the *C. aurantiacus* ACIII [27] and is not present in ACIII of *R. marinus*.
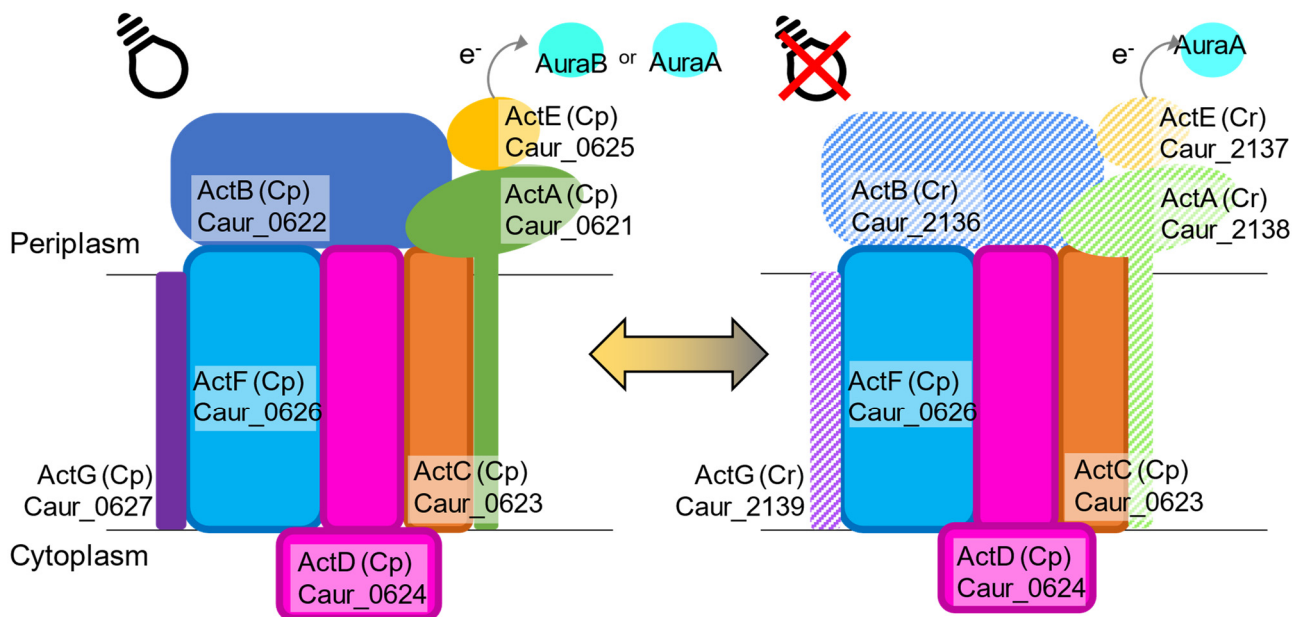
Figure 6. Schematic drawing of subunit compositions of alternative complexes III for photosynthesis (C_p, **left**) and respiration (C_r, **right**) in *C. aurantiacus*. The ActCDF subunits are commonly used in both ACIII-C_p and -C_r. Two sets of ActABEG subunits are used for oxic dark and anoxic light conditions, respectively. Auracyanin A functions as the electron carrier with both C_p- and C_r-type ACIII, but auracyanin B works with the C_p type only.

The expression patterns of two auracyanins indicate that auracyanin A works with both ACIII-C_p and -C_r, while auracyanin B works only with ACIII-C_p (Figure 3). During the phototrophic growing phase, the abundance of auracyanin B increased while that of auracyanin A decreased, suggesting that auracyanin B is more suitable than auracyanin A for photosynthetic electron transfer. *R. castenholzii* has a single auracyanin having auracyanin A-like N-terminal and auracyanin B-like C-terminal characteristics [48]. Proteome results showed that auracyanin B was specific to phototrophy in *C. aurantiacus*, implying that the

C-terminal structure of *R. castenholzii* auracyanin could work for photosynthetic electron transport given that this auracyanin is important for both phototrophy and respiration.

Most of the enzymes involved in the 3-OHP bi-cycle were abundant during the first and second growing phases, indicating the involvement of the carbon fixation pathway for the phototrophic growth of *C. aurantiacus* under the tested conditions. Calvin cycle mutants of *Rhodospirillum rubrum* and *Rhodopseudomonas palustris* are reported to lose their ability to grow under anoxic light conditions due to electron imbalance rather than to the accumulation of toxic ribulose-1,5-bisphosphate [49]. In aerobic anoxygenic phototrophic bacteria, it is speculated that the lack of phototrophic growing ability under anoxic conditions is caused by the absence of RubisCO genes in the genome [50]. The active expression of 3-OHP enzymes after day 12 suggests the necessity of carbon fixation activity for the phototrophic growth of *C. aurantiacus*. In other words, the 3-OHP bi-cycle also may contribute to relaxing the electron imbalance during phototrophic growth.

5. Closing Remarks

The proteomic analysis in this study revealed the expression dynamics of a total of 2520 proteins out of 3934 CDSs deduced from the *C. aurantiacus* genome. Photosynthesis-related proteins were concertedly expressed during phototrophic growth, although the corresponding genes were scattered in the genome. The present results lead to the following hypothesis of how *C. aurantiacus* expresses proteins to change metabolic modes from respiration to phototrophy. For the structural formation of light-harvesting chlorosomes, the expression of CsmAMN proteins is important at the initial stage. BChl *c* was produced before BChl *a*, although the expression patterns of enzymes specific for BChl *c* and *a* biosynthesis are not very different. Formation of the BChl *c* self-aggregates could precede the esterification of hydrophobic tail by BchK, BChl *c* synthase. CrtI paralogs expressed in the lag phase and growing phase are involved in the synthesis of carotenoids with different chemical structures. Transmembrane proteins for ACIII are common for both ACIII-C_p and -C_r, and relevant cytoplasmic subunits switch reversibly between respiratory and phototrophic conditions. Auracyanin A probably receives an electron from both ACIII-C_p and -C_r and transfers it to RC and cytochrome *c* oxidase, respectively. On the other hand, auracyanin B probably functions as the electron carrier between ACIII C_p and RC. The 3-OHP bi-cycle is important for not only photoautotrophic growth but also photoheterotrophic growth probably because it relaxes the electron imbalance. It is considered that functional genes for photosynthesis and carbon fixation in *Chloroflexota* members are obtained by horizontal gene transfers within the phylum and/or from other phyla [51–54]. Elucidation of the global regulatory system of genes involved in phototrophic growth will provide valuable insights into the identification of common features in nonphotosynthetic bacteria necessary for the acquisition of phototrophic ability through the horizontal gene transfer and will contribute to the understanding of the complex evolutionary process of phototrophic bacteria.

References

1. Pierson, B.K.; Castenholz, R.W. A phototrophic gliding filamentous bacterium of hot springs, *Chloroflexus aurantiacus*, gen. and sp. nov. *Arch. Microbiol.* **1974**, *100*, 5–24. [CrossRef] [PubMed]
2. Madigan, M.T.; Petersen, S.R.; Brock, T.D. Nutritional studies on *Chloroflexus*, a filamentous photosynthetic, gliding bacterium. *Arch. Microbiol.* **1974**, *100*, 97–103. [CrossRef]
3. Thiel, V.; Wood, J.M.; Olsen, M.T.; Tank, M.; Klatt, C.G.; Ward, D.M.; Bryant, D.A. The dark side of the Mushroom spring microbial mat: Life in the shadow of chlorophototrophs. I. Microbial diversity based on 16S rRNA gene amplicons and metagenomic sequencing. *Front. Microbiol.* **2016**, *7*, 919. [CrossRef] [PubMed]
4. Martinez, J.N.; Nishihara, A.; Lichtenberg, M.; Trampe, E.; Kawai, S.; Tank, M.; Kühl, M.; Hanada, S.; Thiel, V. Vertical distribution and diversity of phototrophic bacteria within a hot spring microbial mat (Nakabusa hot springs, Japan). *Microbes Environ.* **2019**, *34*, 374–387. [CrossRef]
5. Klatt, C.G.; Liu, Z.; Ludwig, M.; Kühl, M.; Jensen, S.I.; Bryant, D.A.; Ward, D.M. Temporal metatranscriptomic patterning in phototrophic *Chloroflexi* inhabiting a microbial mat in a geothermal spring. *ISME J.* **2013**, *7*, 1775–1789. [CrossRef]
6. Kawai, S.; Martinez, J.N.; Lichtenberg, M.; Trampe, E.; Kühl, M.; Tank, M.; Haruta, S.; Nishihara, A.; Hanada, S.; Thiel, V. In-situ metatranscriptomic analyses reveal the metabolic flexibility of the thermophilic anoxygenic photosynthetic bacterium *Chloroflexus aggregans* in a hot spring cyanobacteria-dominated microbial mat. *Microorganisms* **2021**, *9*, 652. [CrossRef]
7. Ruff-Roberts, A.L.; Kuenen, J.G.; Ward, D.M. Distribution of cultivated and uncultivated cyanobacteria and *Chloroflexus*-like bacteria in hot spring microbial mats. *Appl. Environ. Microbiol.* **1994**, *60*, 697–704. [CrossRef]
8. Ward, D.M.; Ferris, M.J.; Nold, S.C.; Bateson, M.M. A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. *Microbiol. Mol. Biol. Rev.* **1998**, *62*, 1353–1370. [CrossRef]
9. Blankenship, R.E.; Madigan, M.T.; Bauer, C.E. *Anoxygenic Photosynthetic Bacteria*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1995.
10. Xin, Y.; Shi, Y.; Niu, T.; Wang, Q.; Niu, W.; Huang, X.; Ding, W.; Yang, L.; Blankenship, R.E.; Xu, X.; et al. Cryo-EM structure of the RC-LH core complex from an early branching photosynthetic prokaryote. *Nat. Commun.* **2018**, *9*, 1568. [CrossRef]
11. Vassilieva, E.V.; Stirewalt, V.L.; Jakobs, C.U.; Frigaard, N.-U.; Inoue-Sakamoto, K.; Baker, M.A.; Sotak, A.; Bryant, D.A. Subcellular localization of chlorosome proteins in *Chlorobium tepidum* and characterization of three new chlorosome proteins: CsmF, CsmH, and CsmX. *Biochemistry* **2002**, *41*, 4358–4370. [CrossRef]
12. Frigaard, N.-U.; Bryant, D.A. Chlorosomes: Antenna organelles in photosynthetic green bacteria. In *Complex Intracellular Structures in Prokaryotes*; Springer: Berlin/Heidelberg, Germany, 2006; Volume 2, pp. 79–114.
13. Garcia Costas, A.M.; Tsukatani, Y.; Romberger, S.P.; Oostergetel, G.T.; Boekema, E.J.; Golbeck, J.H.; Bryant, D.A. Ultrastructural analysis and identification of envelope proteins of "*Candidatus* Chloracidobacterium thermophilum" chlorosomes. *J. Bacteriol.* **2011**, *193*, 6701–6711. [CrossRef] [PubMed]
14. Tsukatani, Y.; Wen, J.; Blankenship, R.E.; Bryant, D.A. Characterization of the FMO protein from the aerobic chlorophototroph, *Candidatus* Chloracidobacterium thermophilum. *Photosynth. Res.* **2010**, *104*, 201–209. [CrossRef] [PubMed]
15. Wen, J.; Tsukatani, Y.; Cui, W.; Zhang, H.; Gross, M.L.; Bryant, D.A.; Blankenship, R.E. Structural model and spectroscopic characteristics of the FMO antenna protein from the aerobic chlorophototroph, *Candidatus* Chloracidobacterium thermophilum. *Biochim. Biophys. Acta-Bioenerg.* **2011**, *1807*, 157–164. [CrossRef] [PubMed]
16. Hauska, G.; Schoedl, T.; Remigy, H.; Tsiotis, G. The reaction center of green sulfur bacteria. *Biochim. Biophys. Acta-Bioenerg.* **2001**, *1507*, 260–277. [CrossRef]
17. Olson, J.M.; Cole, P.; Cox, R.; Fenna, R.; Filmer, D.; Ke, B.; Melkozernov, A.; Miller, M.; Pearlstein, R.; Raymond, J.; et al. The FMO protein. *Photosynth. Res.* **2004**, *80*, 181–187. [CrossRef]
18. Chew, A.G.M.; Bryant, D.A. Chlorophyll biosynthesis in bacteria: The origins of structural and functional diversity. *Annu. Rev. Microbiol.* **2007**, *61*, 113–129. [CrossRef]
19. Tamiaki, H.; Teramura, M.; Tsukatani, Y. Reduction processes in biosynthesis of chlorophyll molecules: Chemical implication of enzymatically regio- and stereoselective hydrogenations in the late stages of their biosynthetic pathway. *Bull. Chem. Soc. Jpn.* **2016**, *89*, 161–173. [CrossRef]
20. Yen, H.C.; Marrs, B. Map of genes for carotenoid and bacteriochlorophyll biosynthesis in *Rhodopseudomonas capsulata*. *J. Bacteriol.* **1976**, *126*, 619–629. [CrossRef]
21. Pemberton, J.M.; Horne, I.M.; McEwan, A.G. Regulation of photosynthetic gene expression in purple bacteria. *Microbiology* **1998**, *144*, 267–278. [CrossRef]

22. Tang, K.H.; Barry, K.; Chertkov, O.; Dalin, E.; Han, C.S.; Hauser, L.J.; Honchak, B.M.; Karbach, L.E.; Land, M.L.; Lapidus, A.; et al. Complete genome sequence of the filamentous anoxygenic phototrophic bacterium *Chloroflexus aurantiacus*. *BMC Genom.* **2011**, *12*, 334. [CrossRef]

23. Yanyushin, M.F. Fractionation of cytochromes of phototrophically grown *Chloroflexus aurantiacus*. Is there a cytochrome *bc* complex among them? *FEBS Lett.* **2002**, *512*, 125–128. [CrossRef]

24. Pereira, M.M.; Refojo, P.N.; Hreggvidsson, G.O.; Hjorleifsdottir, S.; Teixeira, M. The alternative complex III from *Rhodothermus marinus*—A prototype of a new family of quinol:electron acceptor oxidoreductases. *FEBS Lett.* **2007**, *581*, 4831–4835. [CrossRef] [PubMed]

25. Pereira, M.M.; Carita, J.N.; Teixeira, M. Membrane-bound electron transfer chain of the thermohalophilic bacterium *Rhodothermus marinus*: A novel multihemic cytochrome *bc*, a new complex III. *Biochemistry* **1999**, *38*, 1268–1275. [CrossRef] [PubMed]

26. Refojo, P.N.; Sousa, F.L.; Teixeira, M.; Pereira, M.M. The alternative complex III: A different architecture using known building modules. *Biochim. Biophys. Acta-Bioenerg.* **2010**, *1797*, 1869–1876. [CrossRef]

27. Majumder, E.L.W.; King, J.D.; Blankenship, R.E. Alternative complex III from phototrophic bacteria and its electron acceptor auracyanin. *Biochim. Biophys. Acta-Bioenerg.* **2013**, *1827*, 1383–1391. [CrossRef]

28. Gao, X.; Xin, Y.; Blankenship, R.E. Enzymatic activity of the alternative complex III as a menaquinol:auracyanin oxidoreductase in the electron transfer chain of *Chloroflexus aurantiacus*. *FEBS Lett.* **2009**, *583*, 3275–3279. [CrossRef]

29. Gao, X.; Xin, Y.; Bell, P.D.; Wen, J.; Blankenship, R.E. Structural analysis of alternative complex III in the photosynthetic electron transfer chain of *Chloroflexus aurantiacus*. *Biochemistry* **2010**, *49*, 6670–6679. [CrossRef]

30. Izaki, K.; Haruta, S. Aerobic production of bacteriochlorophylls in the filamentous anoxygenic photosynthetic bacterium, *Chloroflexus aurantiacus* in the light. *Microbes Environ.* **2020**, *35*, ME20015. [CrossRef]

31. Frigaard, N.-U.; Takaichi, S.; Hirota, M.; Shimada, K.; Matsuura, K. Quinones in chlorosomes of green sulfur bacteria and their role in the redox-dependent fluorescence studied in chlorosome-like bacteriochlorophyll *c* aggregates. *Arch. Microbiol.* **1997**, *167*, 343–349. [CrossRef]

32. Maresca, J.A.; Graham, J.E.; Bryant, D.A. The biochemical basis for structural diversity in the carotenoids of chlorophototrophic bacteria. *Photosynth. Res.* **2008**, *97*, 121–140. [CrossRef]

33. Sandmann, G. Antioxidant protection from UV-and light-stress related to carotenoid structures. *Antioxidants* **2019**, *8*, 219. [CrossRef] [PubMed]

34. Refojo, P.N.; Teixeira, M.; Pereira, M.M. The alternative complex III: Properties and possible mechanisms for electron transfer and energy conservation. *Biochim. Biophys. Acta-Bioenerg.* **2012**, *1817*, 1852–1859. [CrossRef]

35. Klatt, C.G.; Bryant, D.A.; Ward, D.M. Comparative genomics provides evidence for the 3-hydroxypropionate autotrophic pathway in filamentous anoxygenic phototrophic bacteria and in hot spring microbial mats. *Environ. Microbiol.* **2007**, *9*, 2067–2078. [CrossRef] [PubMed]

36. Hügler, M.; Sievert, S.M. Beyond the calvin cycle: Autotrophic carbon fixation in the ocean. *Ann. Rev. Mar. Sci.* **2011**, *3*, 261–289. [CrossRef] [PubMed]

37. Hägerhäll, C. Succinate: Quinone oxidoreductases. *Biochim. Biophys. Acta-Bioenerg.* **1997**, *1320*, 107–141. [CrossRef]

38. Lemos, R.S.; Fernandes, A.S.; Pereira, M.M.; Gomes, C.M.; Teixeira, M. Quinol:fumarate oxidoreductases and succinate:quinone oxidoreductases: Phylogenetic relationships, metal centres and membrane attachment. *Biochim. Biophys. Acta-Bioenerg.* **2002**, *1553*, 158–170. [CrossRef]

39. Halfen, L.N.; Pierson, B.K.; Francis, G.W. Carotenoids of a gliding organism containing bacteriochlorophylls. *Arch. Mikrobiol.* **1972**, *82*, 240–246. [CrossRef]

40. Larsen, K.L.; Miller, M.; Cox, R.P. Incorporation of exogenous long-chain alcohols into bacteriochlorophyll *c* homologs by *Chloroflexus aurantiacus*. *Arch. Microbiol.* **1995**, *163*, 119–123. [CrossRef]

41. Tsukatani, Y.; Mizoguchi, T.; Thweatt, J.; Tank, M.; Bryant, D.A.; Tamiaki, H. Glycolipid analyses of light-harvesting chlorosomes from envelope protein mutants of *Chlorobaculum tepidum*. *Photosynth. Res.* **2016**, *128*, 235–241. [CrossRef]

42. Liu, Z.; Bryant, D.A. Identification of a gene essential for the first committed step in the biosynthesis of bacteriochlorophyll *c*. *J. Biol. Chem.* **2011**, *286*, 22393–22402. [CrossRef]

43. Larsen, K.L.; Cox, R.P.; Miller, M. Effects of illumination intensity on bacteriochlorophyll *c* homolog distribution in *Chloroflexus aurantiacus* grown under controlled conditions. *Photosynth. Res.* **1994**, *41*, 151–156. [CrossRef] [PubMed]

44. Nishimori, R.; Mizoguchi, T.; Tamiaki, H.; Kashimura, S.; Saga, Y. Biosynthesis of unnatural bacteriochlorophyll *c* derivatives esterified with α,ω-diols in the green sulfur photosynthetic bacterium *Chlorobaculum tepidum*. *Biochemistry* **2011**, *50*, 7756–7764. [CrossRef] [PubMed]

45. Orf, G.S.; Blankenship, R.E. Chlorosome antenna complexes from green photosynthetic bacteria. *Photosynth. Res.* **2013**, *116*, 315–331. [CrossRef] [PubMed]

46. Hohmann-Marriott, M.F.; Blankenship, R.E. Hypothesis on chlorosome biogenesis in green photosynthetic bacteria. *FEBS Lett.* **2007**, *581*, 800–803. [CrossRef] [PubMed]

47. Pedersen, M.O.; Linnanto, J.; Frigaard, N.-U.; Nielsen, N.C.; Miller, M. A model of the protein-pigment baseplate complex in chlorosomes of photosynthetic green bacteria. *Photosynth. Res.* **2010**, *104*, 233–243. [CrossRef]

48. Tsukatani, Y.; Nakayama, N.; Shimada, K.; Mino, H.; Itoh, S.; Matsuura, K.; Hanada, S.; Nagashima, K.V.P. Characterization of a blue-copper protein, auracyanin, of the filamentous anoxygenic phototrophic bacterium *Roseiflexus castenholzii*. *Arch. Biochem. Biophys.* **2009**, *490*, 57–62. [CrossRef]

49. Gordon, G.C.; McKinlay, J.B. Calvin cycle mutants of photoheterotrophic purple nonsulfur bacteria fail to grow due to an electron imbalance rather than toxic metabolite accumulation. *J. Bacteriol.* **2014**, *196*, 1231–1237. [CrossRef]

50. Swingley, W.D.; Blankenship, R.E.; Raymond, J. Evolutionary Relationships Among Purple Photosynthetic Bacteria and the Origin of Proteobacterial Photosynthetic Systems. In *The Purple Phototrophic Bacteria. Advances in Photosynthesis and Respiration*; Hunter, C.N., Daldal, F., Thurnauer, M.C., Beatty, J.T., Eds.; Springer: Dordrecht, The Netherlands, 2009; pp. 17–29.

51. Shih, P.M.; Ward, L.M.; Fischer, W.W. Evolution of the 3-hydroxypropionate bicycle and recent transfer of anoxygenic photosynthesis into the *Chloroflexi*. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 10749–10754. [CrossRef]

52. Ward, L.M.; Hemp, J.; Shih, P.M.; McGlynn, S.E.; Fischer, W.W. Evolution of phototrophy in the *Chloroflexi* phylum driven by horizontal gene transfer. *Front. Microbiol.* **2018**, *9*, 260. [CrossRef]

53. Bryant, D.A.; Liu, Z.; Li, T.; Zhao, F. Comparative and functional genomics of anoxygenic green bacteria from the taxa *Chlorobi*, *Chloroflexi*, and *Acidobacteria*. In *Functional Genomics and Evolution of Photosynthetic Systems*; Advances in Photosynthesis and Respiration; Springer: Dordrecht, The Netherlands, 2012; pp. 47–72. ISBN 978-94-007-1532-5.

54. Tsuji, J.; Shaw, N.; Nagashima, S.; Venkiteswaran, J.; Schiff, S.; Hanada, S.; Tank, M.; Neufeld, J. Anoxygenic phototrophic *Chloroflexota* member uses a type I reaction center. *bioRxiv* **2020**, 190934. [CrossRef]

Article

Persulfide-Responsive Transcription Factor SqrR Regulates Gene Transfer and Biofilm Formation via the Metabolic Modulation of Cyclic di-GMP in *Rhodobacter capsulatus*

Takayuki Shimizu [1],* , Toma Aritoshi [1], J. Thomas Beatty [2] and Tatsuru Masuda [1]

[1] Graduate School of Arts and Sciences, The University of Tokyo, Tokyo 153-8902, Japan; aritoshi-tohma403@g.ecc.u-tokyo.ac.jp (T.A.); cmasuda2@g.ecc.u-tokyo.ac.jp (T.M.)
[2] Department of Microbiology and Immunology, University of British Columbia, Vancouver, BC V6T 1Z3, Canada; j.beatty@ubc.ca
* Correspondence: ctshimizu@g.ecc.u-tokyo.ac.jp

Abstract: Bacterial phage-like particles (gene transfer agents—GTAs) are widely employed as a crucial genetic vector in horizontal gene transfer. GTA-mediated gene transfer is induced in response to various stresses; however, regulatory mechanisms are poorly understood. We found that the persulfide-responsive transcription factor SqrR may repress the expression of several GTA-related genes in the photosynthetic bacterium *Rhodobacter capsulatus*. Here, we show that the *sqrR* deletion mutant ($\Delta sqrR$) produces higher amounts of intra- and extracellular GTA and gene transfer activity than the wild type (WT). The transcript levels of GTA-related genes are also increased in $\Delta sqrR$. In spite of the presumption that GTA-related genes are regulated in response to sulfide by SqrR, treatment with sulfide did not alter the transcript levels of these genes in the WT strain. Surprisingly, hydrogen peroxide increased the transcript levels of GTA-related genes in the WT, and this alteration was abolished in the $\Delta sqrR$ strain. Moreover, the absence of SqrR changed the intracellular cyclic dimeric GMP (c-di-GMP) levels, and the amount of c-di-GMP was correlated with GTA activity and biofilm formation. These results suggest that SqrR is related to the repression of GTA production and the activation of biofilm formation via control of the intracellular c-di-GMP levels.

Keywords: transcriptional regulation; gene transfer; persulfide; redox signaling; cyclic GMP

1. Introduction

Bacteria do not perform sexual reproduction but are capable of acquiring exogenous DNA by horizontal gene transfer (HGT), which is important for genetic diversity and evolution [1]. Although HGT is classically mediated by transformation, conjugation and transduction, small bacteriophage-like particles called gene transfer agents (GTAs), which package random segments of cellular DNA, also mediate HGT [2,3]. GTA was originally discovered as a novel genetic vector in the alphaproteobacterium *R. capsulatus* [4], and subsequently in other diverse bacteria [5]. In particular, homologous GTAs are conserved in a number of families in the alphaproteobacterial order *Rhodobacterales*, which occupy over 25% the total prokaryotic community in some marine environments [6–8]. Moreover, gene transfer rates of antibiotic resistance markers via putative GTAs were, remarkably, one-million-fold higher than previous estimates of transformation and transduction rates in natural environments [9]. Therefore, it appears that GTAs are widely employed as an important genetic vector in bacterial HGT.

The GTAs of *R. capsulatus* (RcGTA) have been investigated in detail. It has been reported that gene transfer by RcGTA is induced at the stationary phase and by carbon starvation [10,11]. In these regulations, the GtaI-GtaR quorum-sensing system [12,13] and the CckA-ChpT-CtrA phosphorelay [14,15] are centrally important. Moreover, the transcription factor GafA functions as a direct activator of the RcGTA structural gene

cluster [16]. Quorum-sensing is a well-known cell-to-cell communication system in bacteria. Bacteria produce and release autoinducers such as acyl-homoserine lactones (AHL) during growth, sense the outside concentration of autoinducers to recognize high concentrations of related species, and consequently modulate the expression of certain genes [17]. *R. capsulatus* possesses an AHL synthase (GtaI), and its regulator (GtaR) indirectly regulates RcGTA production in response to AHL [13]. This regulation mechanism is thought to be involved in RcGTA production associated with the growth phase. The CckA-ChpT-CtrA phosphorelay is one widely studied phosphorelay that controls the cell cycle in many alphaproteobacteria. CtrA functions as a response regulator whose activity is controlled through the histidine kinase CckA [18] and the histidine phosphotransferase ChpT [19]. This phosphorelay system also regulates RcGTA production, switching from the production phase to the release phase dependent on the phosphorylation state of CtrA [14,15]. The PAS domain protein DivL promotes the phosphorylation of CtrA via the enhancement of CckA kinase activity, and thereby RcGTA production is controlled [20]. The RNA polymerase omega subunit is required for the RelA/SpoT-related stringent response induction of RcGTA production under carbon starvation, although its association with the CtrA pathway is unclear [11]. Another aspect of the CtrA regulation of RcGTA production is manifested by c-di-GMP, via the possible modulation of c-di-GMP metabolic enzymes [21]. Thus, the outline of a complex regulatory network has been revealed; however, the details of each regulatory pathway and their possible connections are unclear.

Recently, it was reported that persulfides, which are oxidized sulfur species generated from sulfide, are signaling molecules and modulate various physiological functions in both prokaryotes and eukaryotes [22–27]. Since hydrogen sulfide was abundant over oxygen in the prebiotic Earth period, and for much of the Archean Eon, it is considered that sulfide and persulfide contributed significantly to evolution in terms of energy metabolism and signal transduction [28]. We have identified the persulfide-responsive transcription factor SqrR as a regulator of sulfide-dependent photosynthesis in *R. capsulatus* [22]. SqrR binds to the promoter regions of target genes to repress their expression in the absence of persulfide. In the presence of persulfide, SqrR forms an intramolecular tetrasulfide crosslink between two cysteine residues and loses the ability to bind DNA, which results in the de-repression of target genes [22]. SqrR is the master regulator of the persulfide response: RNA-seq analysis has shown that contributions of SqrR regulate 45% of sulfide-responsive genes [22]. Interestingly, RNA-seq data have indicated that several RcGTA-related genes, such as the RcGTA capsid protein gene (rcc01687), *chpT* (rcc03000) and *divL* (rcc00042), were up-regulated between three and nine-fold in an *sqrR* deletion mutant (Δ*sqrR*) [22] (Table 1). Based on this observation, it appears that SqrR could contribute to the regulation of RcGTA production. Here, we provide evidence that SqrR regulates RcGTA production via H_2O_2 signaling. Because SqrR controls intracellular c-di-GMP levels, we suggest that SqrR modulates both RcGTA production and biofilm formation.

Table 1. Transcript levels of GTA-related genes affected by the loss of SqrR.

Gene	Annotation	Transcript Fold Change in Δ*sqrR*	*p*-Value
rcc00042	divL	9.276648	4.94×10^{-64}
rcc01687	capsid	3.318318	2.28×10^{-7}
rcc03000	chpT	2.984249	2.6×10^{-8}
rcc00645	c-di-GMP metabolic enzyme	118.6299	3.7×10^{-167}
rcc02857	c-di-GMP metabolic enzyme	0.497602	0.002762

2. Materials and Methods

2.1. Bacterial Strains, Media, and Growth Conditions

R. capsulatus strains B10 [29], DE442 [30] and mutants were grown under aerobic–dark (aerobically) or anaerobic–light (photosynthetically) conditions at 30 °C in PYS or YPS, a rich medium, or RCV, a minimum medium [29,31,32]. For photosynthetic growth, a light-emitting

diode (λ_{max} = 850 nm) (CCS) was provided. To establish anaerobic conditions, cultures in screw-capped test tubes were almost completely filled with the medium. Gentamycin and rifampicin were used at a concentration of 1.5 μg/mL and 75 μg/mL, respectively.

Escherichia coli strains were grown on Luria Bertani (LB) medium at 37 °C. Ampicillin and gentamycin were used at a concentration of 100 μg/mL and 10 μg/mL, respectively.

2.2. Cloning and Mutagenesis

sqrR disruption in *R. capsulatus* DE442 was performed using the plasmid pZJD29a::Δ*sqrR*, as previously described [22]. For the disruption of the *oxyR* gene, two ~500 bp DNA fragments consisting of the N-terminal and C-terminal regions of *oxyR* were amplified by polymerase chain reaction (PCR) with KOD one polymerase (TOYOBO). Two sets of primer pairs were used for the amplification; one was a forward primer, oxyR F1, and a reverse primer, oxyR R1, and the other was a forward primer, oxyR F2, and a reverse primer, oxyR R2 (Table 2). These two fragments were cloned into the *Bam*HI-cut pZJD29a [33] by an In-Fusion HD Cloning kit (Clontech, Mountain View, CA, USA). The obtained plasmids were introduced into *R. capsulatus* strains by conjugation with the *E. coli* strain S17-1/λ*pir*, and the subsequent homologous recombination events were induced as described [33]. The isolated mutants were analyzed by DNA sequencing to confirm a deletion.

Table 2. The list of primers used in this study.

Name	Sequence 5′–3′	Purpose
oxyR F1	CGACTCTAGAGGATCATTGCCGTATTTCTTCTTGATCGGC	
oxyR R1	CGAGAGGTTTATCATAATGAAAAACTATCGCAGGC	Cloning for gene disruption
oxyR F2	ATGATAAACCTCTCGGCGCGGGAGGCGTGAGGTCGGCGGGTTCGG	
oxyR R2	CGGTACCCGGGGATCGCGCATCCGCTGGCGCCCGAGACGC	
divL-F	CGGTACCCGGGGATCAGAATGCGCCGGTGCCGCGGCGCTC	
divL-R	CGACTCTAGAGGATCGACTGCAGCCCTCTCGCCTGTCCCG	
cckA-F	CGGTACCCGGGGATCCGCCGCAGCTATTCCCCGCGCGACG	
cckA-R	CGACTCTAGAGGATCGGGCTGATCGGGATGTACCACTGGC	
chpT-F	CGGTACCCGGGGATCAAGCTGCACCCGTCGCCCGTCGATC	
chpT-R	CGACTCTAGAGGATCGGGTCATGGTGGATCTCCCTTTCGG	
gafA-F	CGGTACCCGGGGATCGTAATCGCGCTGCCCGAAGCGTGCG	
gafA-R	CGACTCTAGAGGATCCTCCGGTCTCCCATCGACAGGCTGG	
capsid-F	CGGTACCCGGGGATCCACCGGCGGGCATGCTTTTGCCGAGA	Cloning for DNA probes of the gel shift assay
capsid-R	CGACTCTAGAGGATCGTCTTGCGTGACCCGCCTCTCATGC	
ctrA-F	CGGTACCCGGGGATCGCCGCCGAAAGAAACGCGTCGTTGG	
ctrA-R	CGACTCTAGAGGATCCCTGGGTTCTCCGCATTAATCCCTC	
02857-F	CGGTACCCGGGGATCTGGTGCCCCAGCCTAACCGCGGGAT	
02857-R	CGACTCTAGAGGATCCGGGAACGGACCCCTTCGAGTGGAT	
02630-F	CGGTACCCGGGGATCGTGCCCGGACCGGAGGCGGTTTTCC	
02630-R	CGACTCTAGAGGATCGGACCCTCCTCGCGCGGACCATAGC	
00620-F	CGGTACCCGGGGATCCTTGTCGGGGGGGATGACGCCGCTG	
uvrD qF	CAGAAGGAACACACGGTCAA	
uvrD qR	AAAGTGTCAGGCGGAATCTC	
divL qF	CCGACGCTTTATGCCTTTCT	
divL qR	CCTGTTCCAGTTCCGTCATCT	
cckA qF	GCGCATGATTTCAACAACTT	
cckA qR	TTCTGGCTGATCTGGTCAAG	For qRT-PCR
chpT qF	ACGGGGTGGAGTTGCTGAA	
chpT qR	AAAGGCGATGCGGAAGAA	
ctrA qF	TTTGGCGCCGATGATTAC	
ctrA qR	GGATGATCGACTGCGAATG	
gafA qF	GCTGAACGGCTGGATCTT	
gafA qR	TTCCAACAGCCGCTTCAA	
capsid qF	CGGTTGCCGAGGTGAAA	
capsid qR	CACACGCTCTTCCTGTTGTTG	

2.3. GTA Transduction Assay

GTA transduction assays were performed referring to previous study [34]. The rifampin-resistant donor strain was grown photosynthetically to the log phase (WT: OD_{660} = 1.2, DE*sqrR*: OD_{660} = 1.1) or stationary phase (WT: OD_{660} = 1.7, DE*sqrR*: OD_{660} = 1.4) in YPSm medium. Then, 1.5 mL culture normalized at OD_{660} = 1.0 with 20 mM Tris-HCl (pH 8.0) was centrifuged, and supernatant was filtered using a 0.45 μm membrane filter. Next, 100 μL of the obtained sample was mixed with 500 μL of rifampin-sensitive B10 recipient cells grown photosynthetically to mid log phase (OD_{660} = 0.5). The harvested cells were re-suspended with G buffer (10 mM Tris-HCl (pH 8.0), 1 mM $MgCl_2$, 1 mM $CaCl_2$, 1 mM NaCl, 250 μg/mL BSA). The mixture was incubated under aerobic shaking conditions at 30 °C for 1.5 h to undergo gene transfer to the recipient cells. The harvested cells were re-suspended with RCV medium and plated on rifampin-containing plates. The number of rifampin-resistant colony-forming units was determined.

2.4. Western Blotting of Capsid Protein

R. capsulatus was grown photosynthetically to the log phase or stationary phase in YPSm medium. Then, 1.5 mL culture normalized at OD_{660} = 1.0 with 20 mM Tris-HCl (pH 8.0) was separated by centrifugation, and the supernatant was filtered through a 0.45 μm pore-size membrane filter. Cell pellets were re-suspended in 500 μL of 20 mM Tris-HCl (pH 8.0) and disrupted by sonication. Proteins were then separated by 15% SDS-PAGE gels. After electrophoresis, proteins were blotted onto PVDF membrane and probed with commercially available RcGTA major capsid antiserum (AS08 365; Agrisera AB) according to the product's instructions. The secondary antibody was visualized by Clarity Western ECL substrate (Bio-Rad, Hercules, CA, USA).

2.5. RNA Isolation and Quantitative Real-Time PCR (qRT-PCR)

R. capsulatus was grown photosynthetically to the log phase or stationary phase in YPSm medium. For sulfide or H_2O_2 treatment, a final 0.2 mM of Na_2S or 1 mM H_2O_2 was added at the mid-log phase (OD_{660} = 0.7), and cells were further grown for 30 min. Then, 0.5 mL of cells were harvested, and the total RNA of each sample was extracted using NucleoSpin RNA (TaKaRa). The quality of the purified RNA was checked to confirm a typical OD_{260} to OD_{280} ratio of approximately 2.0. RNA samples were reverse transcribed using a PrimeScript RT Reagent kit (TaKaRa, Shiga, Japan). qRT-PCR reactions and detection were performed using the THUNDERBIRD Next SYBR qPCR Mix (TOYOBO, Osaka, Japan) and the CFX Connec Real-Time System (Bio-Rad). As an internal control, the house-keeping gene *uvrD* that encodes DNA helicase [16] was used with the gene-specific primers (Table 2).

2.6. Overexpression and Purification of SqrR

Recombinant SqrR was overexpressed in the *E. coli* strain BL21 (DE3) overexpression system utilizing pSUMO::SqrR plasmid, and was purified as previously described in [22].

2.7. Gel Mobility Shift Analysis

An FITC-labeled 200–300 bp DNA probe containing each promoter region was prepared by PCR amplification with the primer sets (Table 2). The amplified fragment was cloned into a pUC19-linearized vector (TaKaRa) using the In-Fusion HD Cloning kit (Clontech). The inserted DNA was amplified by PCR with a FITC-labeled primer, as described previously [35]. The amplified DNA was purified and utilized as a probe for the assays.

The gel shift assay was performed as described in [22]. The binding reaction buffer (25 mM Tris-HCl (pH 8.0), 100 mM NaCl, 2 mM $MgCl_2$, 6% glycerol, 0.5 mM DTT and 50 μg/mL heparin) was utilized for the reaction with the DNA probes and proteins. For the electrophoresis, 7% polyacrylamide gel was used in a buffer composed of 25 mM Tris-HCl (pH 8.0), 1 mM ethylenediaminetetraacetic acid (EDTA) and 144 mM glycine. After electrophoresis, the gel was analyzed using the Lumino Graph I (ATTO, Tokyo, Japan).

2.8. Quantification of c-di-GMP

R. capsulatus was grown photosynthetically to the log phase or stationary phase in YPSm medium. Cells from 1 mL culture were disrupted by sonication with 20 mM Tris-HCl (pH 8.0). After centrifugation, the supernatant was utilized for detecting c-di-GMP using the Cyclic di-GMP ELISA Kit (Cayman). The obtained values were normalized by the protein concentration of the supernatant.

2.9. Quantification of Biofilm Formation

The amount of biofilm formation was quantified based on crystal violet staining by the Biofilm Formation Assay Kit (DOJINDO, Kumamoto, Japan), according to the product's instructions. In short, *R. capsulatus* was grown photosynthetically to the stationary phase and dispensed onto a 96-well plate. The plate was covered by a 96-well peg lid containing protrusions that were immersed into the culture of the wells. Biofilm formed around this protrusion during overnight cultivation at 30 °C. The 96-well peg lid was washed with 20 mM Tris-HCl (pH 8.0) and the formed biofilm was stained by crystal violet solution. This staining was eluted by 100% ethanol after being washing twice with 20 mM Tris-HCl (pH 8.0). The absorbance at 600 nm of eluate was measured using the GloMax Multi Detection system (Promega, Madison, WI, USA). The obtained values were normalized by the OD_{660} of the culture.

3. Results

3.1. SqrR Contributes to GTA Production and Release

Because only ~3% of the cells in a population are responsible for the release of RcGTA in the WT *R. capsulatus* strain SB1003 [36,37], we constructed the Δ*sqrR* mutant using the *R. capsulatus* RcGTA overproducer strain DE442 as a background to verify whether SqrR contributes to RcGTA production and release. The DE442 strain had a mutation in the rcc00280 gene that increased the amount of RcGTA production [37], but was WT in terms of the *sqrR* gene. The growth rates of the parental wild-type strain (WT) and Δ*sqrR* were similar, although Δ*sqrR* showed a slightly longer log growth phase (Figure 1A). Both strains were cultivated using the YPSm-rich medium, and gene transduction was analyzed at the log and stationary phases to explore whether the deletion of *sqrR* affected the functional activity of RcGTA. In the WT strain, there was a low frequency of gene transduction in the log phase, with an induction in the stationary phase (Figure 1B). In contrast, the Δ*sqrR* mutant showed a significantly greater gene transduction frequency in both the log and stationary phases than the WT strain. We also measured the amount of RcGTA capsid protein by Western blotting using an anti-capsid antibody. A greater amount of RcGTA capsid protein was detected in Δ*sqrR* compared to the WT strains of both the pellet (cellular) and supernatant (extracellular) fractions at the stationary phase (Figure 1C). Although it was difficult to detect a difference at the log phase because of the weakness of the signal, the gene transduction frequency clearly showed a higher amount of RcGTA released from the Δ*sqrR* than the WT strain at the log phase (Figure 1B). These data indicate that SqrR contributes to RcGTA production and release.

We subsequently analyzed the transcript levels of RcGTA-related genes by real-time PCR (qRT-PCR). The transcript of the gene encoding the capsid protein was increased by *sqrR* deletion in correlation with the intracellular amount of GTA capsid protein at the log phase (Figure 2A). Moreover, the *gafA* gene, which encodes a direct activator of the RcGTA gene cluster, was also up-regulated in the Δ*sqrR* mutant. These data indicate that SqrR is needed to repress these genes at the log phase. In contrast, the transcripts of the phosphorelay-related regulators, *divL*, *cckA*, *chpT*, and *ctrA*, were unaffected (Figure 2A). We further tested the effect of sulfide on these transcriptional changes, because SqrR functions as a (per)sulfide-responsive transcriptional repressor. Unexpectedly, treatment with sulfide did not alter these transcript levels at the log phase of either the WT or Δ*sqrR* strains (Figure 2B). Since SqrR senses sulfide by persulfidation via the oxidation of cysteine residues, it was possible that another oxidant, such as hydrogen peroxide (H_2O_2), could

react with SqrR in vivo. Therefore, we subsequently tested the effect of H_2O_2 on the transcript levels of RcGTA-related genes. It was found that transcripts of the capsid and *gafA* genes were increased by the treatment of cells with H_2O_2, and this up-regulation did not occur in the Δ*sqrR* mutant (Figure 2C), indicating that SqrR contributes to the H_2O_2-induced transcription of RcGTA-related genes at the log phase. We also analyzed the transcript levels when cells were treated with mitomycin c, which induces the SOS response via DNA damage [38], in order to explore whether this effect of H_2O_2 on the transcript levels was due to a change in the intracellular redox state, or if it was an SOS response induced by H_2O_2 [39] that could be related to RcGTA regulation [40]. Although the *gafA* transcript was increased approximately two-fold by treatment with mitomycin c in the WT strain (Figure 2D), this change was a lot smaller than that resulting from the treatment with H_2O_2, indicating that the SqrR-related H_2O_2-induced transcriptional change in the RcGTA-related genes may have been due to a change in intracellular redox state.
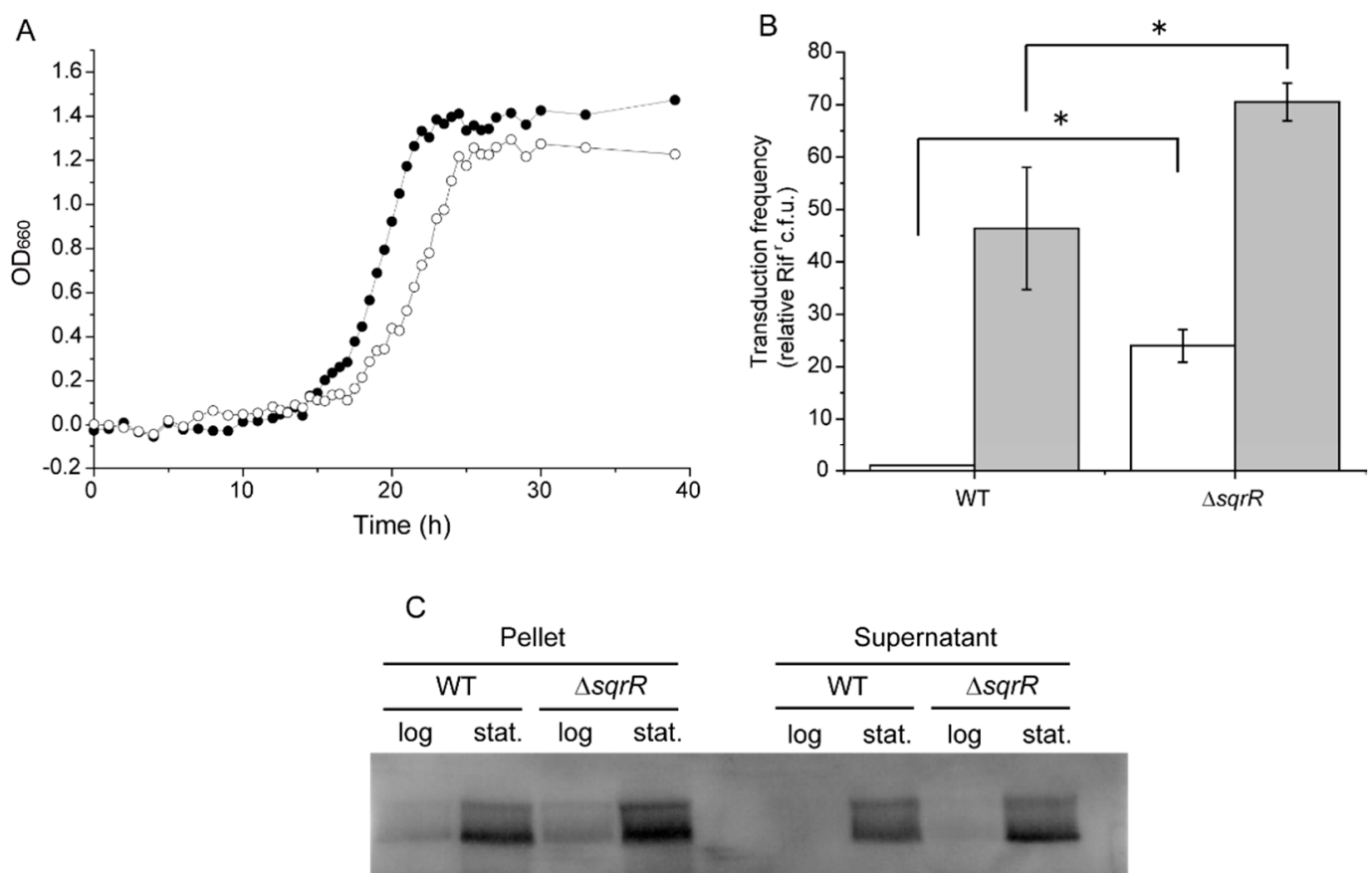
Figure 1. Effects of *sqrR* deletion on growth, gene transfer and GTA production and release. (**A**) Growth curves of the parental wild-type strain (filled circle) and the Δ*sqrR* strain (open circle) of *R. capsulatus* DE442 under anaerobic photosynthetic conditions. Data shown are the mean of triplicate culture tubes. (**B**) Transduction frequencies using GTA produced at the log phase (white bar) and stationary phase (gray bar) in the parental wild-type strain (WT) and the Δ*sqrR* strain. Bars show the mean, error bars show standard deviation of three biological replicates, and star (*) indicates a statistically significant difference (*t*-test, *p*-value < 0.05). (**C**) Western blots of the WT and Δ*sqrR* strain culture cell pellets and supernatant fractions at the log and stationary (stat.) phases, probed with *R. capsulatus* GTA capsid antiserum.
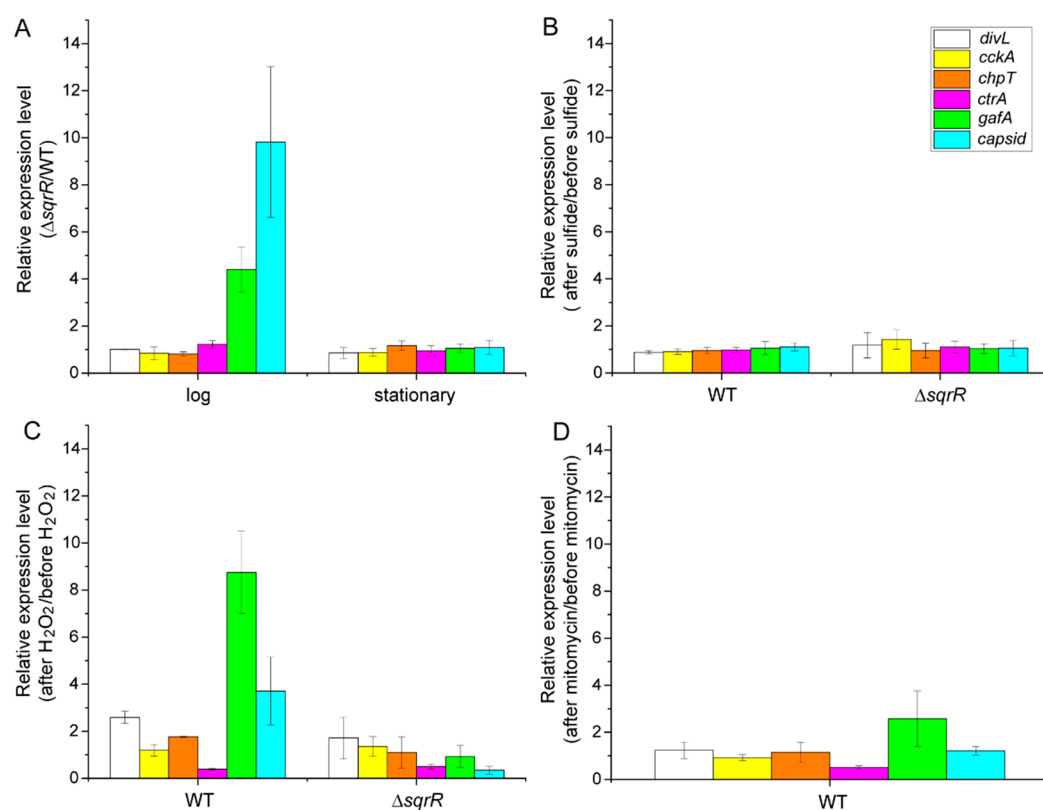
Figure 2. Relative level of transcripts of GTA-related genes. (**A**) Relative expression levels of Δ*sqrR* as compared with the parental wild-type strain (WT) at the log and stationary phases, respectively. Cells were grown under anaerobic photosynthetic conditions. (**B**–**D**) Changes in the relative transcript levels of the GTA-related genes after the addition of 0.2 mM sodium sulfide (**B**), 1 mM hydrogen peroxide (**C**), or 20 μg/mL mitomycin c (**D**) to WT and Δ*sqrR* strains. Cells were grown to mid-log phase under anaerobic photosynthetic conditions and each chemical was added at *t* = 0. Cells were harvested after 30 min and assayed for qRT-PCR. Bars show the mean, and error bars show the standard deviation of the three biological replicates.

3.2. Molecular Mechanism of the SqrR-Related H_2O_2-Induced Transcription of GTA-Related Genes

To verify whether SqrR directly regulates RcGTA-related genes, we performed gel mobility shift assays using SqrR recombinant protein and DNA probes containing the predicted promoter region of each gene. Unexpectedly, SqrR did not bind to the promoter regions of the *gafA* or capsid gene (upstream of the first gene of the structural gene cluster, g1). However, distinct binding to the *sqr* promoter region as a positive control was shown (Figure 3A). Therefore, the SqrR modulation of the RcGTA production appeared to require at least one additional factor. It is well known that the transcription factor OxyR functions as a master regulator of reactive oxygen species (ROS) signaling in bacteria [41,42]. As *R. capsulatus* also has OxyR, we explored the contribution of OxyR to the H_2O_2-induced transcription of RcGTA-related genes. To elucidate this, we constructed the *oxyR* single-deletion mutant (Δ*oxyR*), the *sqrR,* and the *oxyR* double-deletion mutant (Δ*sqrR*Δ*oxyR*), and measured the transcript levels of capsid and *gafA* after treatment with H_2O_2 (Figure 3B). The transcripts were slightly increased in the Δ*oxyR* strain compared to the Δ*sqrR* strain and, particularly, the transcript of the capsid protein significantly increased. Moreover, the Δ*sqrR*Δ*oxyR* double mutant showed significantly lower levels of these transcripts than every single mutant, except for the transcript of the capsid in Δ*sqrR*. These results indicate that SqrR and OxyR independently regulate the H_2O_2-induced transcription of RcGTA-related genes. Overall, SqrR appears to act downstream of the H_2O_2-induced regulation of RcGTA production, independently of OxyR, and indirectly mediates the transcriptional regulation of RcGTA-related genes to control RcGTA and gene transfer frequency.
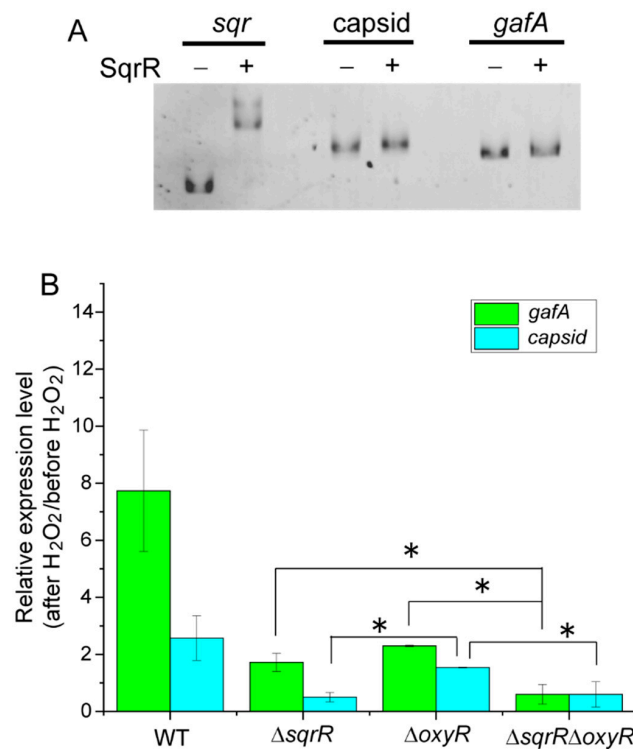
Figure 3. Molecular characterization of SqrR in the H$_2$O$_2$-induced transcriptional regulation of GTA-related genes. (**A**) Gel mobility shift assay using a DNA probe of the *sqr*, capsid gene and *gafA* promoter region under 5 mM DTT-reducing conditions without (−) or with (+) 0.5 mM DTT-reduced SqrR. (**B**) Changes in the relative transcript levels of the capsid gene (green) and *gafA* (blue) after the addition of 1 mM hydrogen peroxide to WT and each mutant culture. Cells were grown to mid-log phase under anaerobic photosynthetic conditions and each chemical was added at *t* = 0. Cells were harvested after 30 min and assayed for qRT-PCR. Bars show the mean, error bars show standard deviation of three biological replicates, and star (*) indicates a statistically significant difference (*t*-test, *p*-value < 0.05).

It has been reported that CtrA regulates the transcript levels of c-di-GMP metabolic enzymes (rcc00620, rcc00645, rcc02629 and rcc02857), and RcGTA production is negatively regulated by c-di-GMP [21]. Interestingly, our previous RNA-seq data indicated a contribution of SqrR to the expression of rcc00645 and rcc02857, which contain both GGDEF (diguanylate cyclase) and EAL (phosphodiesterase) domains [22] (Table 1). To explore the effect of SqrR on the level of c-di-GMP, we compared the intracellular amount of c-di-GMP between the WT and Δ*sqrR* strains. A significant decrease in the amount of c-di-GMP was detected in the Δ*sqrR* mutant compared with that in the WT strain at both the log and stationary phases (Figure 4A). We further investigated whether these changes in intracellular c-di-GMP levels were correlated to the transcript level changes in the c-di-GMP metabolic enzymes. Relative to the WT strain, the transcript of rcc00620, which encodes a c-di-GMP catabolic enzyme, was decreased at the log phase, and that of rcc02629, which encodes c-di-GMP synthase, was increased at the stationary phase in the Δ*sqrR* mutant (Figure 4B). These data do not correlate with a decrease in intracellular c-di-GMP levels. Moreover, SqrR did not bind to the promoter region of rcc00620 and rcc02629 (Figure 4C). These results indicate that SqrR does not modulate intracellular c-di-GMP levels via the transcriptional regulation of CtrA-regulated c-di-GMP metabolic enzymes, but by another of the ~14 possible genes encoding a diguanylate cyclase in the *R. capsulatus* genome.

It is well known that c-di-GMP positively controls biofilm formation [43]. In order to confirm the effect of the intracellular c-di-GMP levels, we analyzed biofilm formation in the WT and Δ*sqrR* strains. Biofilm formation was significantly lower in Δ*sqrR* than in WT (Figure 5). This observation was congruent with the lower intracellular c-di-GMP levels

in Δ*sqrR* compared to those in the WT strain (Figure 4A). Thus, the phenotype of biofilm formation activity also supports the regulation of c-di-GMP production by SqrR.
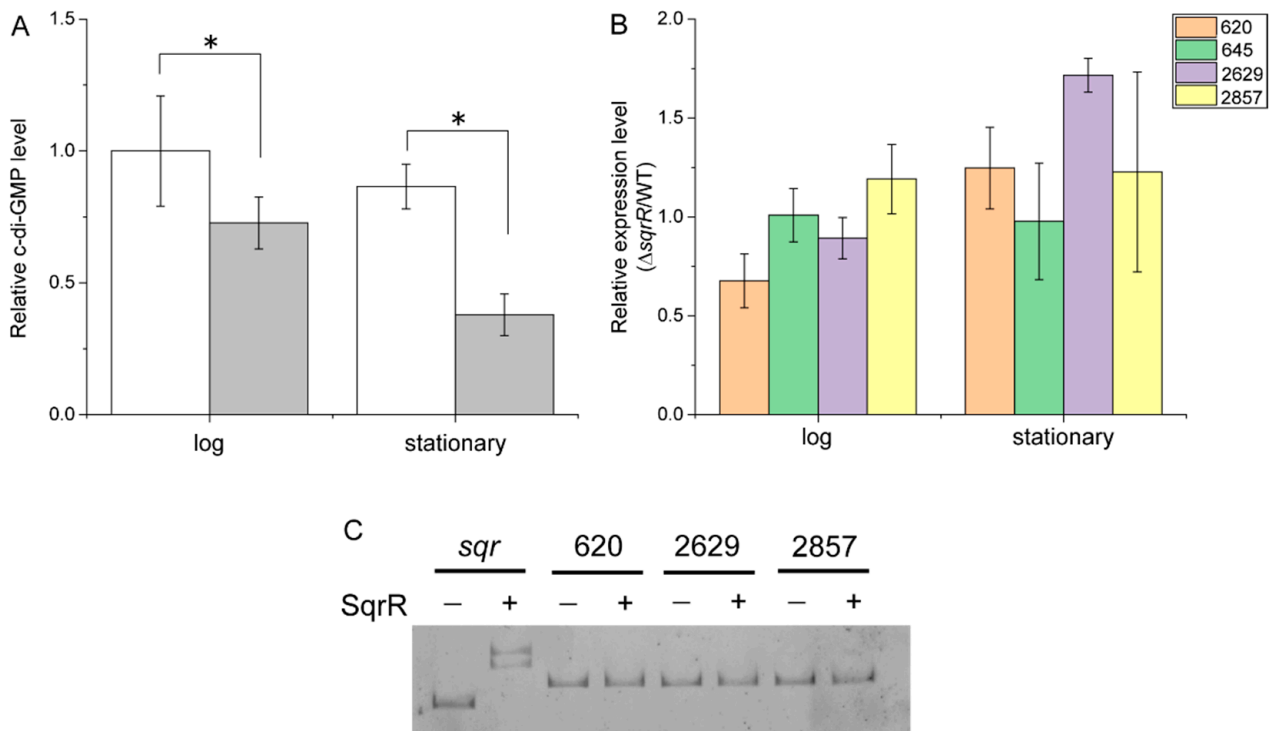
Figure 4. Effects of *sqrR* deletion on the intracellular c-di-GMP levels. (**A**) Relative c-di-GMP levels in the parental wild-type strain (WT) (white bar) and Δ*sqrR* (gray bar) at the log and stationary phases, compared to WT at the log phase. Error bars show the standard deviation of three biological replicates, and star (*) indicates a statistically significant difference (*t*-test, *p*-value < 0.05). (**B**) Relative expression levels in Δ*sqrR*, as compared with WT at the log and stationary phases. Cells were grown under anaerobic photosynthetic conditions. Error bars show standard deviation of three biological replicates. (**C**) Gel mobility shift assay using a DNA probe of the *sqr*, rcc00620, rcc02629 and rcc02857 promoter region under 5 mM DTT-reducing conditions without (−) or with (+) 5 mM DTT-reduced SqrR.
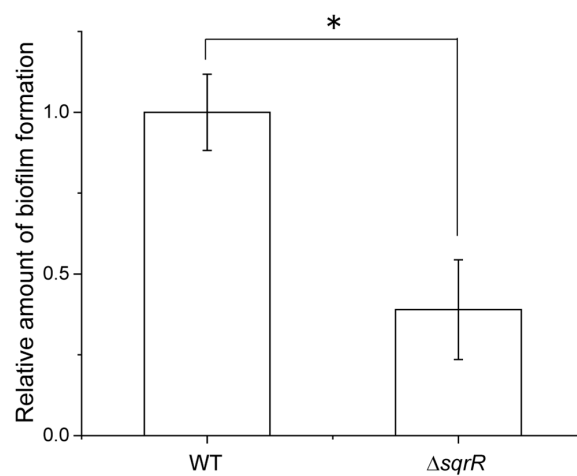
Figure 5. Effects of *sqrR* deletion on biofilm formation. Relative amount of biofilm formation in *sqrR* compared with the parental wild-type strain (WT). Bars show the mean, error bars show the standard deviation of three biological replicates, and star (*) indicates a statistically significant difference (*t*-test, *p*-value < 0.05).

4. Discussion

We studied the contribution of SqrR to gene transfer via RcGTA in order to explore the possibility of a novel regulatory process in RcGTA production. We have demonstrated that SqrR transduces the H_2O_2-mediated regulation of RcGTA production and biofilm formation by c-di-GMP as a novel regulatory pathway in *R. capsulatus*. This conclusion is based on the effect of *sqrR* deletion on RcGTA production and the transcript levels of RcGTA-related genes. The Δ*sqrR* mutant showed higher gene transfer frequency and greater amounts of intracellular and released RcGTA, as compared with the WT (Figure 1B,C). The transcript levels of the RcGTA capsid protein and the GafA direct activator of the RcGTA gene cluster transcription were up-regulated by *sqrR* deletion, and this regulation mediated H_2O_2 signaling (Figure 2). Although previous RNA-seq data have shown that *chpT* and *divL* are also regulated by SqrR (Table 1), these transcript levels were not changed by the deletion of *sqrR* (Figure 2A). This discordance could be due to different growth conditions, aerobic conditions in the previous study, or the anaerobic conditions in this study, because RcGTA production is affected by the oxygen tension of the culture. We note that RcGTA production was delayed from the induction of the transcripts at the log phase. It appears that transcription induction starts in the log phase, but high levels of RcGTA protein do not accumulate until the stationary phase. In addition, SqrR modulated the amount of intracellular c-di-GMP, which induced RcGTA production and inhibited biofilm formation (Figures 1, 4 and 5).

c-di-GMP is a ubiquitous bacterial second-messenger molecule that regulates many bacterial functions and behaviors [44]. It is thought that c-di-GMP affects the CckA-ChpT-CtrA phosphorelay by enhancing the phosphatase activity of CckA, and thereby gene transfer by RcGTA is negatively regulated by nonphosphorylated CtrA [21]. Indeed, gene transfer activity via RcGTA is affected by altering intracellular c-di-GMP levels by the overexpression of the exogenous c-di-GMP metabolic enzyme [21], and CckA phosphatase versus kinase activity is modulated by c-di-GMP [45]. Moreover, in positive regulatory systems of biofilm formation, c-di-GMP allosterically modulates regulators—such as effector kinases in two-component systems or transcription factors—to promote biofilm formation [46–48]. Our observations clearly show the alteration of RcGTA production and biofilm formation associated with intracellular c-di-GMP levels (Figures 1C, 4A and 5). Therefore, we suggest that SqrR regulates the levels of enzymes that metabolize c-di-GMP, resulting in changes in the levels of c-di-GMP which, in turn, regulate the levels of RcGTA [21]. It has been reported that c-di-GMP-mediated gene transfer is regulated by a two-component system composed of a sensor histidine kinase encoded by rcc00621 and a c-di-GMP catabolic enzyme encoded by rcc00620 as a response regulator [49]. Our findings indicate a new regulatory pathway of gene transfer via the control of c-di-GMP levels by SqrR. The regulatory network of RcGTA production comprises various response systems and the CtrA-dependent central regulation system to obtain a variety of abiotic stress responses [10–13,20,21]. Thus, SqrR functions as a redox stress-responsive c-d-GMP modulator, and may be valuable to appropriately regulate RcGTA production via the CtrA phosphorelay in response to one of several abiotic stresses.

SqrR should be basically employed as the persulfide-specific responsive regulator in *R. capsulatus* [22]. The promoter activity of the *sqr* gene, which encodes sulfide:quinone reductase, the transcription of which is repressed by SqrR, was accelerated not by H_2O_2, but by sulfide in vivo [22]. Moreover, previous MS-based reactivity profiling has revealed that the modification of cysteine residue results from a reaction with persulfide, but not other oxidants such as H_2O_2 [50]. Surprisingly, our data showed a distinct association between SqrR and the H_2O_2-induced regulation of RcGTA production and biofilm formation (Figures 2 and 5). It has been reported that the well-established ROS responsive regulator OxyR senses not only H_2O_2 via sulfene (–SOH) and/or the disulfide formation of cysteine residue(s), but also persulfide via the persulfidation (–SSH) of cysteine residues [51]. Given that persulfide might have been present earlier than ROS on the prebiotic Earth [28], and that persulfide plays an important role as a signaling molecule for organisms as well as ROS,

it is reasonable to suggest that the persulfide sensor protein detects both persulfide and ROS. Therefore, SqrR might have the ability to mediate ROS signal transduction in vivo, although the molecular mechanisms of how SqrR senses or mediates H_2O_2 signaling are unclear. One possible mechanism is that the heme-binding ability of SqrR is available for mediating H_2O_2 signaling. We have previously reported that SqrR can bind hemes, and the redox state of the heme iron affects the secondary structure of SqrR [52], suggesting that the SqrR repressor activity can be altered based on the redox state of the heme. However, we note that only 1.7% of SqrR can bind to a heme in vivo under normal growth conditions without abiotic stress. Oxidative stress induces free hemes by dissociation from hemoproteins [53,54], and heme-bound holo-SqrR showed lower DNA-binding affinity than apo-SqrR [52]. Therefore, another possibility is that SqrR senses increased free hemes under oxidative stress. Further analyses are needed to fully understand the mechanism of how SqrR mediates H_2O_2 signaling in RcGTA production.

5. Conclusions

We suggest that SqrR mediates the H_2O_2-induced regulation of RcGTA production by modulating c-di-GMP. Although SqrR appears to regulate RcGTA transcription, the effect is indirect, implicating another regulatory factor of RcGTA that remains unknown. However, our discovery of SqrR as a novel mediator of H_2O_2-induced RcGTA production allows for further elucidation of how the whole regulatory network functions in this model GTA-producing bacterium.

References

1. Boto, L. Horizontal gene transfer in evolution: Facts and challenges. *Proc. R. Soc. B Biol. Sci.* **2010**, *277*, 819–827. [CrossRef] [PubMed]
2. Stanton, T.B. Prophage-like gene transfer agents-Novel mechanisms of gene exchange for *Methanococcus*, *Desulfovibrio*, *Brachyspira*, and *Rhodobacter* species. *Anaerobe* **2007**, *13*, 43–49. [CrossRef] [PubMed]
3. Lang, A.S.; Beatty, J.T. Importance of widespread gene transfer agent genes in α-proteobacteria. *Trends Microbiol.* **2007**, *15*, 54–62. [CrossRef] [PubMed]
4. Marrs, B. Genetic recombination in *Rhodopseudomonas capsulata*. *Proc. Natl. Acad. Sci. USA* **1974**, *71*, 971–973. [CrossRef]
5. Lang, A.S.; Westbye, A.B.; Beatty, J.T. The Distribution, Evolution, and Roles of Gene Transfer Agents in Prokaryotic Genetic Exchange. *Annu. Rev. Virol.* **2017**, *4*, 87–104. [CrossRef]
6. Lang, A.S.; Zhaxybayeva, O.; Beatty, J.T. Gene transfer agents: Phage-like elements of genetic exchange. *Nat. Rev. Microbiol.* **2012**, *10*, 472–482. [CrossRef]
7. Buchan, A.; González, J.M.; Moran, M.A. Overview of the marine *Roseobacter* lineage. *Appl. Environ. Microbiol.* **2005**, *71*, 5665–5677. [CrossRef]
8. Fu, Y.; MacLeod, D.M.; Rivkin, R.B.; Chen, F.; Buchan, A.; Lang, A.S. High diversity of *Rhodobacterales* in the subarctic North Atlantic ocean and gene transfer agent protein expression in isolated strains. *Aquat. Microb. Ecol.* **2010**, *59*, 283–293. [CrossRef]
9. McDaniel, L.D.; Young, E.; Delaney, J.; Ruhnau, F.; Ritchie, K.B.; Paul, J.H. High frequency of horizontal gene transfer in the oceans. *Science* **2012**, *330*, 50. [CrossRef]
10. Fogg, P.C.M.; Westbye, A.B.; Beatty, J.T. One for all or all for one: Heterogeneous expression and host cell lysis are key to gene transfer agent activity in *Rhodobacter capsulatus*. *PLoS ONE* **2012**, *7*, e43772. [CrossRef]

11.   Westbye, A.B.; O'Neill, Z.; Schellenberg-Beaver, T.; Thomas Beatty, J. The *Rhodobacter capsulatus* gene transfer agent is induced by nutrient depletion and the RNAP omega subunit. *Microbiology* **2017**, *163*, 1355–1363. [CrossRef] [PubMed]

12.   Schaefer, A.L.; Taylor, T.A.; Beatty, J.T.; Greenberg, E.P. Long-chain acyl-homoserine lactone quorum-sensing regulation of *Rhodobacter capsulatus* gene transfer agent production. *J. Bacteriol.* **2002**, *184*, 6515–6521. [CrossRef] [PubMed]

13.   Leung, M.M.; Brimacombe, C.A.; Spiegelman, G.B.; Beatty, J.T. The GtaR protein negatively regulates transcription of the gtaRI operon and modulates gene transfer agent (RcGTA) expression in *Rhodobacter capsulatus*. *Mol. Microbiol.* **2012**, *83*, 759–774. [CrossRef] [PubMed]

14.   Lang, A.S.; Beatty, J.T. Genetic analysis of a bacterial genetic exchange element: The gene transfer agent of *Rhodobacter capsulatus*. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 859–864. [CrossRef]

15.   Mercer, R.G.; Quinlan, M.; Rose, A.R.; Noll, S.; Beatty, J.T.; Lang, A.S. Regulatory systems controlling motility and gene transfer agent production and release in *Rhodobacter capsulatus*. *FEMS Microbiol. Lett.* **2012**, *331*, 53–62. [CrossRef]

16.   Fogg, P.C.M. Identification and characterization of a direct activator of a gene transfer agent. *Nat. Commun.* **2019**, *10*, 595. [CrossRef]

17.   Fuqua, W.C.; Winans, S.C.; Greenberg, E.P. Quorum sensing in bacteria: The LuxR-LuxI family of cell density- responsive transcriptional regulators. *J. Bacteriol.* **1994**, *176*, 269–275. [CrossRef] [PubMed]

18.   Jacobs, C.; Domian, I.J.; Maddock, J.R.; Shapiro, L. Cell cycle-dependent polar localization of an essential bacterial histidine kinase that controls DNA replication and cell division. *Cell* **1999**, *97*, 111–120. [CrossRef]

19.   Biondi, E.G.; Reisinger, S.J.; Skerker, J.M.; Arif, M.; Perchuk, B.S.; Ryan, K.R.; Laub, M.T. Regulation of the bacterial cell cycle by an integrated genetic circuit. *Nature* **2006**, *444*, 899–904. [CrossRef]

20.   Westbye, A.B.; Kater, L.; Wiesmann, C.; Ding, H.; Yip, C.K.; Beatty, J.T. The protease ClpXP and the PAS domain protein DivL regulate CtrA and gene transfer agent production in *Rhodobacter capsulatus*. *Appl. Environ. Microbiol.* **2018**, *84*, e00275-18. [CrossRef]

21.   Pallegar, P.; Langille, E.; Gomelsky, M.; Lang, A.S. Cyclic di-GMP-Mediated Regulation of Gene Transfer and Motility in *Rhodobacter capsulatus*. *J. Bacteriol.* **2020**, *202*, e00554-19. [CrossRef]

22.   Shimizu, T.; Shen, J.; Fang, M.; Zhang, Y.; Hori, K.; Trinidad, J.C.; Bauer, C.E.; Giedroc, D.P.; Masuda, S. SqrR functions as a master regulator of sulfide-dependent photosynthesis. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 2355–2360. [CrossRef]

23.   Ida, T.; Sawa, T.; Ihara, H.; Tsuchiya, Y.; Watanabe, Y.; Kumagai, Y.; Suematsu, M.; Motohashi, H.; Fujii, S.; Matsunaga, T.; et al. Reactive cysteine persulfides and S-polythiolation regulate oxidative stress and redox signaling. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 7606–7611. [CrossRef]

24.   Nishida, M.; Sawa, T.; Kitajima, N.; Ono, K.; Inoue, H.; Ihara, H.; Motohashi, H.; Yamamoto, M.; Suematsu, M.; Kurose, H.; et al. Hydrogen sulfide anion regulates redox signaling via electrophile sulfhydration. *Nat. Chem. Biol.* **2012**, *8*, 714–724. [CrossRef]

25.   Cuevasanta, E.; Lange, M.; Bonanata, J.; Coitiño, E.L.; Ferrer-Sueta, G.; Filipovic, M.R.; Alvarez, B. Reaction of hydrogen sulfide with disulfide and sulfenic acid to form the strongly nucleophilic persulfide. *J. Biol. Chem.* **2015**, *290*, 26866–26880. [CrossRef]

26.   Yadav, P.K.; Martinov, M.; Vitvitsky, V.; Seravalli, J.; Wedmann, R.; Filipovic, M.R.; Banerjee, R. Biosynthesis and reactivity of cysteine persulfides in signaling. *J. Am. Chem. Soc.* **2016**, *138*, 289–299. [CrossRef]

27.   Rauch, B.J.; Klimek, J.; David, L.; Perona, J.J. Persulfide Formation Mediates Cysteine and Homocysteine Biosynthesis in *Methanosarcina acetivorans*. *Biochemistry* **2017**, *56*, 1051–1061. [CrossRef]

28.   Olson, K.R. Hydrogen sulfide, reactive sulfur species and coping with reactive oxygen species. *Free Radic. Biol. Med.* **2019**, *140*, 74–83. [CrossRef]

29.   Weaver, P.F.; Wall, J.D.; Gest, H. Characterization of *Rhodopseudomonas capsulata*. *Arch. Microbiol.* **1975**, *105*, 207–216. [CrossRef]

30.   Yen, H.C.; Hu, N.T.; Marrs, B.L. Characterization of the gene transfer agent made by an overproducer mutant of *Rhodopseudomonas capsulata*. *J. Mol. Biol.* **1979**, *131*, 157–168. [CrossRef]

31.   Sganga, M.W.; Bauer, C.E. Regulatory factors controlling photosynthetic reaction center and light-harvesting gene expression in *Rhodobacter capsulatus*. *Cell* **1992**, *68*, 945–954. [CrossRef]

32.   Wall, J.D.; Weaver, P.F.; Gest, H. Gene transfer agents, bacteriophages, and bacteriocins of *Rhodopseudomonas capsulata*. *Arch. Microbiol.* **1975**, *105*, 217–224. [CrossRef]

33.   Masuda, S.; Bauer, C.E. Null Mutation of HvrA Compensates for Loss of an Essential *relA/spoT*-Like Gene in *Rhodobacter capsulatus*. *J. Bacteriol.* **2004**, *186*, 235–239. [CrossRef]

34.   Solioz, M.; Yen, H.C.; Marrs, B. Release and uptake of gene transfer agent by *Rhodopseudomonas capsulata*. *J. Bacteriol.* **1975**, *123*, 651–657. [CrossRef]

35.   Shimizu, T.; Cheng, Z.; Matsuura, K.; Masuda, S.; Bauer, C.E. Evidence that altered *cis* element spacing affects PpsR mediated redox control of photosynthesis gene expression in Rubrivivax gelatinosus. *PLoS ONE* **2015**, *10*, e0128446. [CrossRef]

36.   Hynes, A.P.; Mercer, R.G.; Watton, D.E.; Buckley, C.B.; Lang, A.S. DNA packaging bias and differential expression of gene transfer agent genes within a population during production and release of the *Rhodobacter capsulatus* gene transfer agent, RcGTA. *Mol. Microbiol.* **2012**, *85*, 314–325. [CrossRef]

37.   Ding, H.; Grüll, M.P.; Mulligan, M.E.; Lang, A.S.; Thomas Beatty, J. Induction of *Rhodobacter capsulatus* gene transfer agent gene expression is a bistable stochastic process repressed by an extracellular calcium-binding rtx protein homologue. *J. Bacteriol.* **2019**, *201*, e00430-19. [CrossRef]

38. Barbé, J.; Vericat, J.A.; Cairó, J.; Guerrero, R. Further characterization of SOS system induction in *recBC* mutants of *Escherichia coli*. *Mutat. Res. DNA Repair Rep.* **1985**, *146*, 23–32. [CrossRef]

39. Goerlich, O.; Quillardet, P.; Hofnung, M. Induction of the SOS response by hydrogen peroxide in various *Escherichia coli* mutants with altered protection against oxidative DNA damage. *J. Bacteriol.* **1989**, *171*, 6141–6147. [CrossRef] [PubMed]

40. Kuchinski, K.S.; Brimacombe, C.A.; Westbye, A.B.; Ding, H.; Beatty, T.J. The SOS response master regulator LexA regulates the gene transfer agent of *Rhodobacter capsulatus* and represses transcription of the signal transduction protein CckA. *J. Bacteriol.* **2016**, *198*, 1137–1148. [CrossRef] [PubMed]

41. Christman, M.F.; Morgan, R.W.; Jacobson, F.S.; Ames, B.N. Oxidative Stress and Some Heat-Shock Proteins in Salmonella typhimurium. *Cell* **1985**, *41*, 753–762. [CrossRef]

42. Storz, G.; Tartaglia, L.A.; Ames, B.N. Transcriptional regulator of oxidative stress-inducible genes: Direct activation by oxidation. *Science* **1990**, *248*, 189–194. [CrossRef] [PubMed]

43. Pérez-Mendoza, D.; Sanjuán, J. Exploiting the commons: Cyclic diguanylate regulation of bacterial exopolysaccharide production. *Curr. Opin. Microbiol.* **2016**, *30*, 36–43. [CrossRef]

44. Hengge, R. Principles of c-di-GMP signalling in bacteria. *Nat. Rev. Microbiol.* **2009**, *7*, 263–273. [CrossRef]

45. Farrera-Calderon, R.G.; Pallegar, P.; Westbye, A.B.; Wiesmann, C.; Lang, A.S.; Beatty, J.T. The CckA-ChpT-CtrA phosphorelay controlling *Rhodobacter capsulatus* gene transfer agent production is bidirectional and regulated by cyclic di-GMP. *J. Bacteriol.* **2020**, *203*, e00525-20. [CrossRef] [PubMed]

46. Hsieh, M.L.; Hinton, D.M.; Waters, C.M. VpsR and cyclic di-GMP together drive transcription initiation to activate biofilm formation in *Vibrio cholerae*. *Nucleic Acids Res.* **2018**, *46*, 8876–8887. [CrossRef] [PubMed]

47. Pursley, B.R.; Maiden, M.M.; Hsieh, M.L.; Fernandez, N.L.; Severin, G.B.; Waters, C.M. Cyclic di-GMP regulates TfoY in *Vibrio cholerae* to control motility by both transcriptional and posttranscriptional mechanisms. *J. Bacteriol.* **2018**, *200*, e00578-17. [CrossRef] [PubMed]

48. Krasteva, P.V.; Jiunn, J.C.; Shikuma, N.J.; Beyhan, S.; Navarro, M.V.A.S.; Yildiz, F.H.; Sondermann, H. *Vibrio cholerae* VpsT regulates matrix production and motility by directly sensing cyclic di-GMP. *Science* **2010**, *327*, 866–868. [CrossRef]

49. Pallegar, P.; Canuti, M.; Langille, E.; Peña-Castillo, L.; Lang, A.S. A Two-Component System Acquired by Horizontal Gene Transfer Modulates Gene Transfer and Motility via Cyclic Dimeric GMP. *J. Mol. Biol.* **2020**, *432*, 4840–4855. [CrossRef]

50. Capdevila, D.A.; Walsh, B.J.C.; Zhang, Y.; Dietrich, C.; Gonzalez-gutierrez, G.; Giedroc, D.P. Structural basis for persulfide-sensing specificity in a transcriptional regulator. *Nat. Chem. Biol.* **2020**, *17*, 65–70. [CrossRef]

51. Hou, N.; Yan, Z.; Fan, K.; Li, H.; Zhao, R.; Xia, Y.; Xun, L.; Liu, H. OxyR senses sulfane sulfur and activates the genes for its removal in *Escherichia coli*. *Redox Biol.* **2019**, *26*, 101293. [CrossRef] [PubMed]

52. Shimizu, T.; Hayashi, Y.; Arai, M.; McGlynn, S.E.; Masuda, T.; Masuda, S. Repressor Activity of SqrR, a Master Regulator of Persulfide-Responsive Genes, Is Regulated by Heme Coordination. *Plant Cell Physiol.* **2021**, *62*, 100–110. [CrossRef] [PubMed]

53. Bunn, H.F.; Jandl, J.H. Exchange of heme among hemoglobins and between hemoglobin and albumin. *J. Biol. Chem.* **1968**, *243*, 465–475. [CrossRef]

54. Kvam, E.; Noel, A.; Basu-Modak, S.; Tyrrell, R.M. Cyclooxygenase dependent release of heme from microsomal hemeproteins correlates with induction of heme oxygenase 1 transcription in human fibroblasts. *Free Radic. Biol. Med.* **1999**, *26*, 511–517. [CrossRef]

Article

Molecular Evolution of Far-Red Light-Acclimated Photosystem II

Christopher J. Gisriel [1,†] ![ORCID]**, Tanai Cardona** [2,†] ![ORCID]**, Donald A. Bryant** [3] ![ORCID] **and Gary W. Brudvig** [1,4,*] ![ORCID]

1   Department of Chemistry, Yale University, New Haven, CT 06520, USA; christopher.gisriel@yale.edu
2   Department of Life Sciences, Imperial College London, London SW7 2AZ, UK; t.cardona@imperial.ac.uk
3   Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802, USA; dab14@psu.edu
4   Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA
*   Correspondence: gary.brudvig@yale.edu
†   These authors contributed equally to this work.

Abstract: Cyanobacteria are major contributors to global carbon fixation and primarily use visible light ($400-700$ nm) to drive oxygenic photosynthesis. When shifted into environments where visible light is attenuated, a small, but highly diverse and widespread number of cyanobacteria can express modified pigments and paralogous versions of photosystem subunits and phycobiliproteins that confer far-red light (FRL) absorbance ($700-800$ nm), a process termed far-red light photoacclimation, or FaRLiP. During FaRLiP, alternate photosystem II (PSII) subunits enable the complex to bind chlorophylls *d* and *f*, which absorb at lower energy than chlorophyll *a* but still support water oxidation. How the FaRLiP response arose remains poorly studied. Here, we report ancestral sequence reconstruction and structure-based molecular evolutionary studies of the FRL-specific subunits of FRL-PSII. We show that the duplications leading to the origin of two PsbA (D1) paralogs required to make chlorophyll *f* and to bind chlorophyll *d* in water-splitting FRL-PSII are likely the first to have occurred prior to the diversification of extant cyanobacteria. These duplications were followed by those leading to alternative PsbC (CP43) and PsbD (D2) subunits, occurring early during the diversification of cyanobacteria, and culminating with those leading to PsbB (CP47) and PsbH paralogs coincident with the radiation of the major groups. We show that the origin of FRL-PSII required the accumulation of a relatively small number of amino acid changes and that the ancestral FRL-PSII likely contained a chlorophyll *d* molecule in the electron transfer chain, two chlorophyll *f* molecules in the antenna subunits at equivalent positions, and three chlorophyll *a* molecules whose site energies were altered. The results suggest a minimal model for engineering far-red light absorbance into plant PSII for biotechnological applications.

Keywords: photosynthesis; cyanobacteria; ancestral sequence reconstruction; chlorophyll *f*; chlorophyll *d*; far-red light photoacclimation; *Synechococcus* sp. PCC 7335

1. Introduction

Photosystem II (PSII) is the water–plastoquinone photooxidoreductase that is the keystone enzyme of oxygenic photosynthesis [1]. By coupling solar energy to water oxidation, PSII generates the reducing equivalents that support photosynthetic carbon dioxide fixation by plants, algae, and cyanobacteria. These organisms use chlorophyll (Chl) *a*, which absorbs light in the visible region of the solar spectrum from 400 to 700 nm [2], as the primary light-harvesting and photochemically active pigment in PSII. However, by additionally synthesizing and inserting Chls *d* and *f* into far-red light-acclimated PSII (FRL-PSII), some terrestrial cyanobacteria can acclimate to thrive in shaded environments that are naturally enriched in lower energy, far-red light (FRL; λ = 700 to 800 nm) [3–6]. Because the absorbance spectra of Chls *d* and *f* are red-shifted relative to that of Chl *a* [7,8], these Chls allow lower energy FRL to support water oxidation by FRL-PSII [3,5,9,10]. Understanding the molecular basis of how and where Chls *d* and *f* are incorporated in PSII is of great

interest because it could provide design principles that might pave the way for engineering shade tolerance into crops [11,12].

The implementation of Chls *d* and *f* in FRL-PSII by some cyanobacteria is part of a wider acclimation process called far-red light photoacclimation, or FaRLiP [3,9]. FaRLiP-capable cyanobacteria are abundant and quite widespread [9,13–19], with members occurring in all five major taxonomic sections of the cyanobacteria [20]; thus, they substantively contribute to global carbon fixation. When grown under FRL, cyanobacteria that perform FaRLiP upregulate a FRL-specific cluster of 20 genes that encodes proteins involved in the biosynthesis of Chl *f*, as well as FRL-specific subunits found in peripheral phycobiliprotein antenna complexes and the two FRL-acclimated photosystems [3,13]. In PSII, FRL-specific isoforms of the four core subunits, PsbA (D1), PsbB (CP47), PsbC (CP43), and PsbD (D2), together with a peripheral subunit called PsbH, are incorporated into the FRL-PSII complex rather than isoforms found when the cells are grown in visible light (VL) [4]. For clarity, we hereafter refer to the PSII subunits by the name of their gene product without the designation or allele number. For example, the D1 subunit expressed during growth under VL will be referred to as VL-PsbA, and the corresponding isoform expressed during growth under FRL will be referred to as FRL-PsbA.

To uncover the architectural details of FRL-PSII, the structure of a monomeric FRL-PSII core complex from *Synechococcus* sp. PCC 7335 (hereafter, *Synechococcus* 7335) was recently obtained by cryo-electron microscopy (cryo-EM) [21]. Although the complex lacked the dimeric configuration typically observed in a PSII holocomplex and various peripheral subunits including FRL-PsbH, it retained nearly all Chl binding sites. This allowed for the assignment of four sites that bind Chl *f* molecules in the antenna subunits and one site, Chl_{D1}, that binds Chl *d* among the cofactors that comprise the electron transfer chain. It was found that the protein confers binding specificity due to FRL-specific residues that preferentially accommodate the structures of Chls *d* or *f*; these Chls exhibit polar formyl moieties at either the C2 position (Chl *f*) or the C3 position (Chl *d*) on the tetrapyrrole ring, whereas Chl *a* exhibits hydrophobic methyl or vinyl moieties at those locations, respectively. Additionally, the FRL-PsbD and FRL-PsbC subunits contained conserved FRL-specific residues on the stromal surface that are probably complementary to residues of FRL-specific allophycocyanin subunits of the peripheral phycobiliprotein complex [21,22].

Because FRL-PSII is such an important bioenergetic system in the contribution of cyanobacteria to global carbon fixation, understanding its evolution and diversity among extant organisms is also of great interest. Recent evolutionary studies have suggested that FaRLiP has been inherited in cyanobacteria mostly vertically from a common ancestor that likely existed during the early to mid-Proterozoic [14]. Building upon the basis of the structural data previously obtained [21], one can use sequence alignments and phylogenetics to gain detailed molecular insights into how FRL-PSII arose and subsequently evolved and diversified. Here, we have: (a) analyzed phylogenetic trees of PSII subunit sequences to derive the evolutionary events that led to modern FRL-PSII, (b) used all available protein sequences of FRL-PsbA, FRL-PsbB, FRL-PsbC, FRL-PsbD, and FRL-PsbH to reconstruct the ancestral sequence of each, and (c) performed a thorough structural analysis of the *Synechococcus* 7335 FRL-PSII structure with homology models of the ancestral sequences and PSII structures from cyanobacteria incapable of performing FaRLiP. These comparisons reveal insights into how the characteristics of modern FRL-PSII evolved over time and provide new insights into the diversity of FRL-PSII among modern cyanobacteria. The findings also provide a minimal model for FRL-specific characteristics that can be used as a starting point to engineer FRL absorbance into crops.

2. Materials and Methods

2.1. Phylogenetic Tree Construction

All non-redundant and complete amino acid sequences were downloaded from the National Center for Biotechnology Information (NCBI) RefSeq database using PSI-BLAST with default settings and restricted to the phylum Cyanobacteria. PsbA and PsbD sequences

were collected on the 17 August 2021, PsbB and PsbC sequences on the 6 January 2022, and PsbH sequences on the 14 February 2022. Redundancy was decreased to 98% sequence identity using the following tool https://web.expasy.org/decrease_redundancy/ (accessed on the same dates the sequence datasets were collected) from the Swiss Bioinformatics Resource Portal [23,24]. In total, 990 PsbA, 487 PsbB, 436 PsbC not including any "CP43-like" Chl-binding proteins, 274 PsbD, and 464 PsbH sequences were kept for further analysis. Sequence alignments were performed with Clustal Omega [25] using five combined guided trees and hidden Markov model iterations [26].

Maximum likelihood unrooted phylogenies were inferred using IQ-TREE multicore version 2.0.3 [27]. Best fit parameters were calculated using its in-built function ModelFinder [26]. Support values were estimated using ultrafastbootstrap with >1000 iterations until the correlation coefficient of split occurrence frequencies converged [28,29]. The average likelihood ratio test method to estimate branch support values was also used at the same time [30]. Ancestral sequences were calculated during tree inference by activating the function *-asr*. The ancestral sequence reconstruction function of IQ-TREE does not estimate ancestral sequence insertions or deletions; therefore, prior to homology modeling, all insertions were manually removed using the respective sequence from the *Synechococcus* 7335 FRL-PSII as the template. Trees were visualized using the software Dendroscope version 3.8.1 [31]. All sequence alignments, ancestral sequences and trees used in this study are provided in Supplementary Data S1.

2.2. Homology Modeling

To create homology models of ancestral FRL-PsbA, FRL-PsbB, FRL-PsbC, and FRL-PsbD, the subunit of interest was isolated from the cryo-EM structure of monomeric apo-FRL-PSII from *Synechococcus* 7335 deposited under PDB accession code 7SA3 using PyMOL [32]. For the above-mentioned ancestral sequences, each of these was used as the template to create homology models using the Swiss-Model server [33]. To create the homology model of ancestral FRL-PsbH, the PsbH subunit was isolated from the cryo-EM structure of the PSII holocomplex from *Synechocystis* sp. PCC 6803 (hereafter, *Synechocystis* 6803) [34], PDB accession code 7N8O.

3. Results

3.1. Phylogenetic Analysis

To provide insight into the evolution of the FRL-PSII subunits, we generated phylogenetic trees for each of the core subunits: PsbA, PsbD, PsbC, PsbB, and PsbH (Figures 1 and S1−S5). The evolution of PsbA has been studied in detail [35–38] and the tree shown in Figure 1 is generally consistent with those shown before. The tree includes two PsbA paralogs of importance for FaRLiP: (1) the Chl *f* synthase, also known as ChlF [39–41] or Group 1 (G1); and (2) FRL-PsbA, which enables water oxidation under FaRLiP. PsbA evolution is characterized and dominated by continuous gene duplications leading to the extensive diversity of paralogs in extant organisms. Some of these duplications are ancient and are suggested to have occurred before the most recent common ancestor (MRCA) of cyanobacteria [36,37], which is likely to have had multiple copies of PsbA, similar to extant cyanobacteria [42]. The earliest duplications are those that led to the emergence of the atypical PsbA variants (G0 to G2 sequences) and the so-called microoxic forms, Group 3. The PsbA tree calculated in this study, as well as that recently published by Sheridan et al. [35] shows that the duplication leading to FRL-PsbA is also likely to have occurred before the MRCA of cyanobacteria, prior to a duplication leading to another subgroup of sequences that was denoted in Sheridan et al. as D1^{INT} [35].

The other subunits of PSII have not left a record of gene duplications antedating the last common ancestor of cyanobacteria, with the potential exception of the duplication leading to the family of Chl-binding proteins derived from PsbC (e.g., IsiA), which are unable to support water oxidation or replace PsbC in active PSII complexes [38]. Therefore, the duplications leading to FRL-PsbB, FRL-PsbC, FRL-PsbD, and FRL-PsbH are likely to

have occurred after the MRCA of cyanobacteria, and consequently, after that leading to FRL-PsbA. FRL-PsbC and FRL-PsbD have a basal branching position. In contrast, the duplications leading to FRL-PsbB and PsbH appear to have occurred later, at a time mostly coincident with the radiation of cyanobacteria leading to most major groups [43].
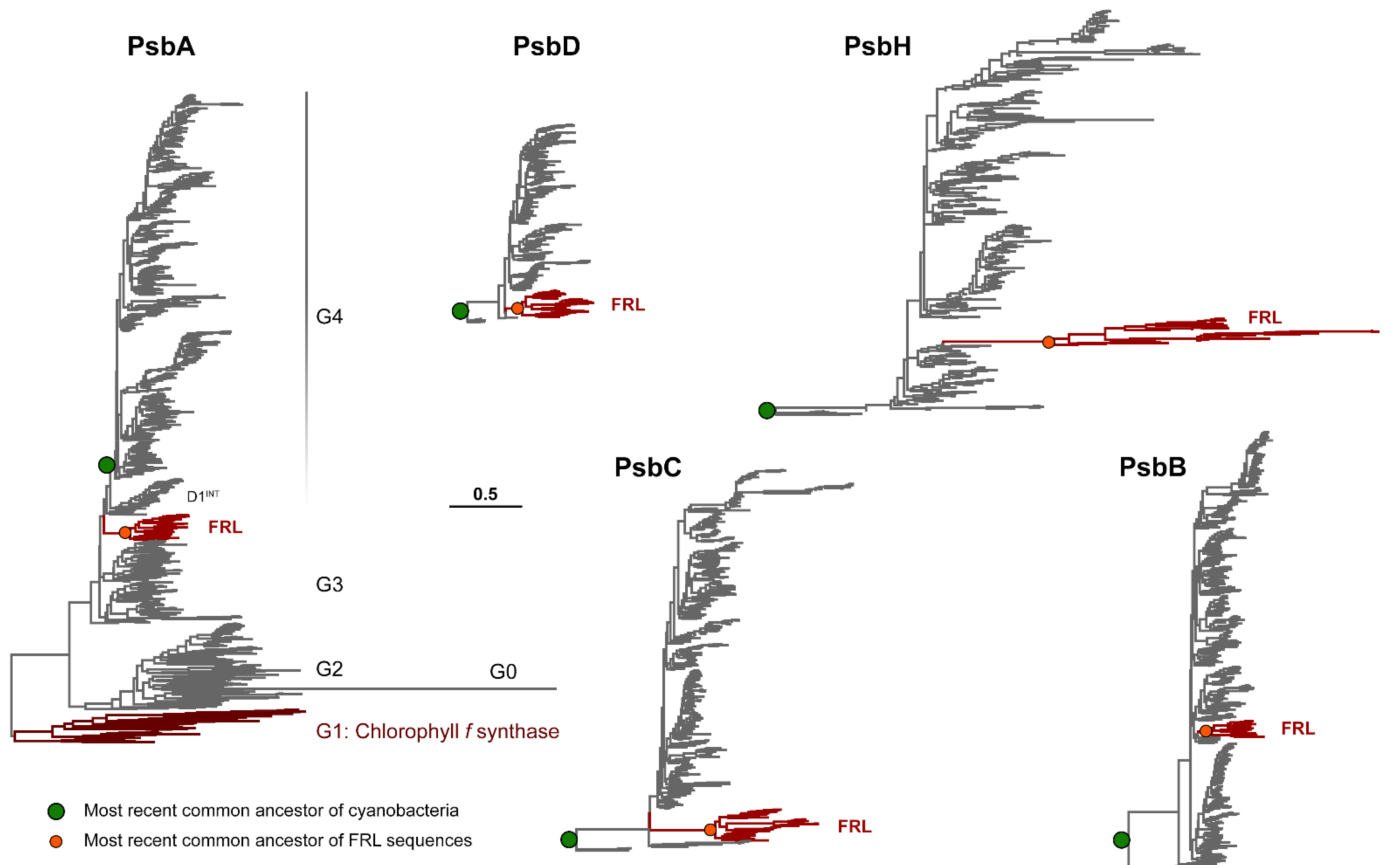
Figure 1. Maximum likelihood phylogenies of FRL-PSII subunits. The PsbA phylogeny has been rooted at the point of divergence of Group 1 sequences as categorized by Cardona et al. [37]. Together with Group 0 and Group 2, these make the "atypical" forms of PsbA characterized by the erosion of the ligand sphere of the water oxidizing cluster. Group 3 and Group 4 include all sequences with a conserved ligand sphere and that can support water oxidation function. In this phylogeny, Group 3 has been split into several subgroups, while the G0 sequences are artificially clustered within Group 2. The MRCA of cyanobacteria (green circle) is defined in this instance as the point in Group 4 PsbA, which contains all standard PsbAs, including those found in the basal genera *Gloeobacter* and close relatives (*Anthocerotibacter/Aurora*): order *Gloeobacterales.* For all other subunits, the trees have been rooted at the branching point of *Gloeobacterales*, and therefore the MRCA of cyanobacteria (green circle) is defined as the last shared ancestor between this and all other lineages. The MRCA of all FRL sequences is marked with an orange circle and it is defined as the last shared ancestor of all extant FRL sequences.

The trees in Figure 1 and Supplementary Figures S1−S5 also show that the different FRL subunits feature different rates of sequence change when compared with each other and within their respective groups, with FRL-PsbH and FRL-PsbC showing particularly longer branches. It is also worth noting that there is some variation in the length of the branch leading to the MRCA of each FRL subunit group, marked with an orange circle in Figure 1. This suggests that before the diversification of the FaRLiP gene cluster, each FRL subunit ancestor had accumulated a different number of FRL-specific substitutions since the point of duplication, with the FRL-PsbH ancestor having accumulated the most changes per amino acid position compared with the VL form, followed by the FRL-PsbC, FRL-PsbA,

FRL-PsbB, and FRL-PsbD ancestors. We note that none of the trees showed evidence that the FRL sequences had any close relationship to those in the Chl *d*-producing *Acaryochloris* sp. strains. This observation supports the hypothesis that the facultative FaRLiP response is unrelated to the constitutive expression of Chl *d* exhibited by *Acaryochloris* strains, and that the two strategies probably evolved convergently.

3.2. Sequence Annotations and Structural View

To explore the predicted sequence for each ancestor of the FRL-specific subunits, and to estimate the similarity of these subunits among species, we reconstructed the ancestral sequence for each (orange circle in Figure 1). We then compared these ancestral sequences with a subset of sequences from extant organisms to identify residues of interest more conveniently (Supplementary Figure S6). The subset of sequences comprises the FRL and VL sequences from representative cyanobacteria of different taxonomic sections including *Aphanocapsa* sp. GSE-SYN-MK-11-07L (*Aphanocapsa* GSE), *Pleurocapsa* sp. PCC 7327 (*Pleurocapsa* 7327), *Synechococcus* 7335, and *Fischerella thermalis* PCC 7521 (*Fischerella* 7521). These were realigned using Clustal Omega [25]. We annotated those residues that were conserved among the FRL sequences from the extant cyanobacteria but were also dissimilar from the corresponding VL sequences (green highlights in Supplementary Figure S6), and we annotated those residues from the ancestral sequence that were the same as the FRL-specific residues in extant cyanobacteria (yellow highlights in Supplementary Figure S6).

From these annotations (Supplementary Figure S6), we also calculated the percentage of residues in the FRL-specific subunits from *Synechococcus* 7335 that appear to be FRL-specific in the subset and are conserved with the corresponding ancestral sequence, and those that are not (Table 1). A relatively small fraction of the total residues in those subunits, only 74 of 1768 (4.2%), appear to be FRL specific. Most of these, 53 of the 74, are also conserved in the ancestral sequences. Of the four core subunits, three have FRL-specific residues that are all conserved in their corresponding ancestral sequences: FRL-PsbA, FRL-PsbC, and FRL-PsbD. This implies that these subunits have changed little in the history of FRL-PSII. The FRL-PsbD subunit has the smallest fraction of residues that appear FRL-specific, only 1.7%, which is consistent with it not binding Chls *d* or *f* [21]. It is also consistent with the short branch that separates the group in the phylogeny from all other PsbD sequences (Figure 1). FRL-PsbA and FRL-PsbC have 3.3% and 3.9% of their residues that appear FRL specific, respectively, and each binds a single FRL-absorbing Chl. FRL-PsbB has more FRL-specific residues than the other three core subunits, 6.3%, but only approximately one-third of those residues are conserved in the ancestral sequence. This indicates that FRL-PsbB presents greater lineage-specific differences than the other FRL-specific core subunits, despite this not being immediately obvious by inspecting the phylogenies alone. Finally, the FRL-PsbH is the only non-core subunit with a FRL isoform. It contains five FRL-specific residues that are all close to the N-terminus, and four of these are conserved in the ancestral sequence. This might indicate that FRL-PsbH associates with the FRL-allophycocyanin antenna. Generally, these observations suggest that the FRL-specific PSII subunits are under selective pressure to maintain a small number of specific sites that are important for FRL-based light absorption and photochemistry, but some extant FRL-PsbB sequences may contain some features that were not present in the ancestor of FRL-PSII.

To visualize the number and frequency of the FRL-specific residues from a structural perspective, we mapped those sites to the FRL-specific subunits from the cryo-EM structure of monomeric apo-FRL-PSII from *Synechococcus* 7335 [21], which contained a FRL-PsbH homology model (Figure 2) occupying the usual position based on the cryo-EM structure of the PSII holocomplex from *Synechocystis* 6803 [34] (Supplementary Figure S7). Consistent with the sequence alignment analysis (Supplementary Figure S6), residues are frequently observed that are FRL-specific and are conserved in the ancestral sequences (yellow spheres in Figure 2). Residues that are FRL-specific but are not conserved in the ancestral sequence are mostly found in FRL-PsbB and one in FRL-PsbH (green spheres in Figure 2). Notable

regions of FRL-specific residues that are maintained in the ancestral sequences are at the stromal surface of PsbC and PsbD, the cluster of conserved residues near the Chl *d* found in the electron transfer chain, and the N-terminal region of FRL-PsbH (Figure 2).

Table 1. Comparison of conserved FRL-specific residues in FRL-PSII of *Synechococcus* 7335 and the ancestral sequences.

FRL Sequence from *Synechococcus* 7335	FRL-Specific Positions Conserved in the Ancestral Sequence (%)	FRL-Specific Positions Conserved Only in the Extant Sequences (%)
FRL-PsbA	3.3	3.3
FRL-PsbB	2.4	6.3
FRL-PsbC	3.9	3.9
FRL-PsbD	1.7	1.7
FRL-PsbH	6.1	7.6
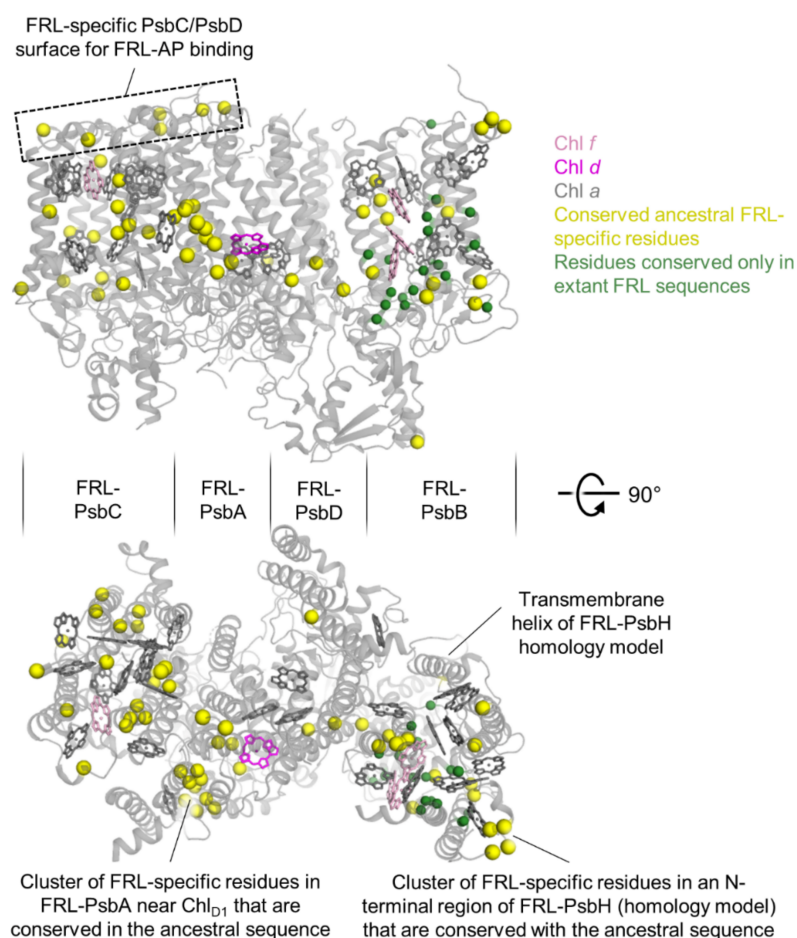
Figure 2. Structural view of conserved FRL-specific residues. The FRL-specific subunits of the apo-FRL-PSII cryo-EM structure from *Synechococcus* 7335 are shown (PDB 7SA3), together with a FRL-PsbH homology model, which was missing in that cryo-EM structure. The tetrapyrrole rings of each Chl molecule from the cryo-EM structure are shown, coloring the Chl *d* molecule in magenta, the Chl *f* molecules in pink, and the Chl *a* molecules in grey. Residues that appear FRL-specific between extant cyanobacteria and are also conserved in the FRL-ancestral sequences are shown as large yellow spheres. Residues that appear FRL-specific between extant cyanobacteria but are not conserved in the FRL-ancestral sequences are shown as smaller green spheres. The approximate regions of the four major core subunits of the two views are denoted by the middle sectioning boundaries. Note that the coloring of this figure corresponds to the highlights in Supplementary Figure S6.

3.3. Conserved Features of FRL-PsbA

All 12 of the conserved FRL-specific residues in the subset of FRL-PsbA sequences from extant cyanobacteria (Supplementary Figure S6) were estimated in the ancestor with posterior probabilities (PP) above 0.99. All form a cluster nearby the formyl moiety of the Chl *d* in the Chl$_{D1}$ site, which is consistent with previous observations [21] (Figures 2 and 3). This implies that the Chl *d* in the Chl$_{D1}$ site was present in the ancestor to extant FRL-PSII complexes, and that there is strong selective pressure to maintain those nearby FRL-specific residues. This cluster of conserved ancestral FRL-specific residues is found within an ~15 Å sphere between the formyl moiety of the Chl *d* and PsbI, adjacent to the PSII dimerization interface (Figure 3). Some notable residues are those that are found in an H-bonding network with the formyl moiety of Chl *d*, PsbA3-Thr155 and Tyr120, as reported previously [21]; however, other nearby residues are likely to participate in stabilizing the FRL-specific interactions or to facilitate Chl *d* insertion/binding during PSII assembly.
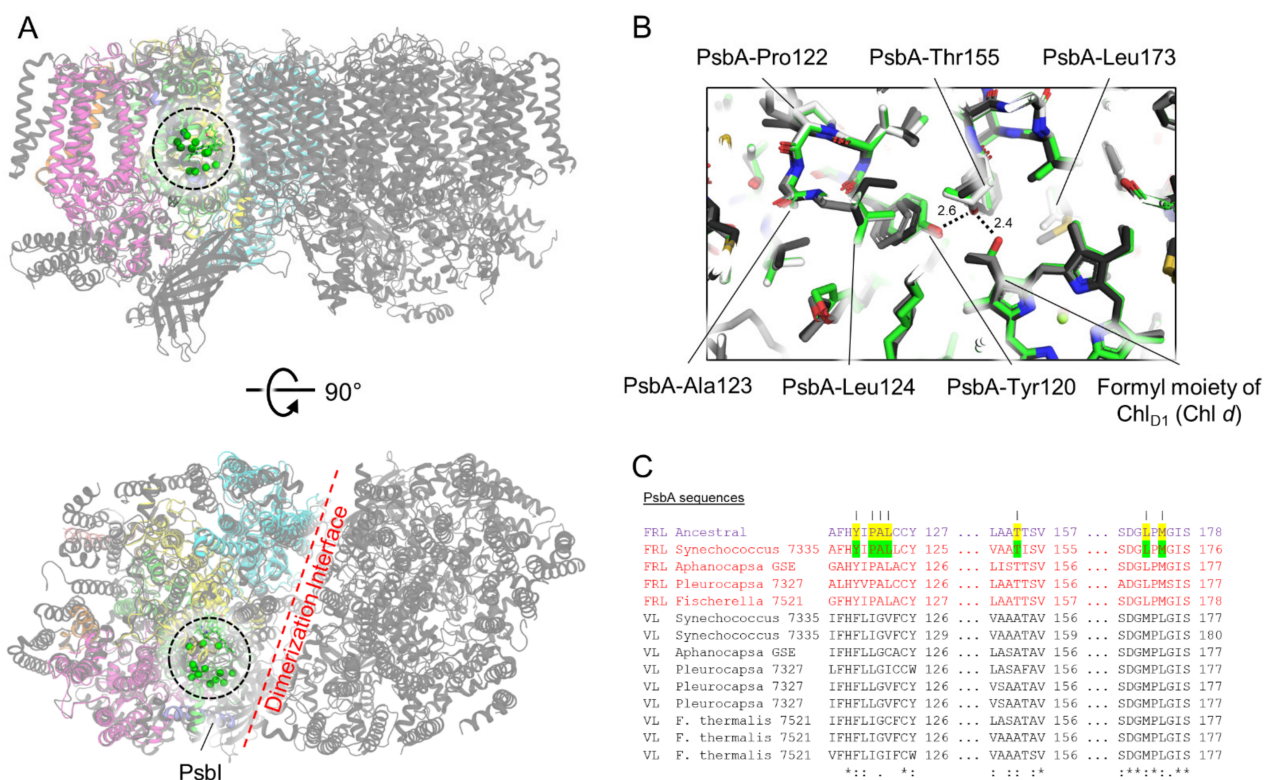
Figure 3. Conserved ancestral FRL-PsbA residues near the Chl$_{D1}$ site of the electron transfer chain. In (**A**), the structure of apo-FRL-PSII from *Synechococcus* 7335 (colors) is shown superimposed with the structure of the PSII holocomplex from *Synechocystis* 6803 (grey, PDB 7N8O). The C$_\alpha$ atom from each of the conserved residues in the cluster (black dashed line) near the Chl$_{D1}$ site is shown as a green sphere. (**B**) A magnified view of this region. The *Synechococcus* 7335 apo-FRL-PSII structure (green), the homology model of the FRL-ancestral sequence (white), and two non-FaRLiP holocomplex PSII structures (light and dark grey from *T. vulcanus* [PDB 3WU2] and *Synechocystis* 6803, respectively) are superimposed. H-bonding interactions involving the C3 formyl moiety of Chl *d* in dashed lines with distances in units of Å are also shown. In (**C**), partial sequence alignments are shown that include the FRL-PsbA ancestral sequence, and FRL- and VL-specific PsbA sequences from extant FaRLiP-capable cyanobacteria. FRL-specific residues conserved in extant cyanobacteria are highlighted green in the sequence from *Synechococcus* 7335. If the same position is conserved in the FRL ancestral sequence, it is highlighted yellow. Vertical lines above residue positions in (**C**) correspond to amino acids from the *Synechococcus* 7335 apo-FRL-PSII structure labeled in (**B**). The Clustal Omega sequence conservation identifiers are shown below the alignment.

3.4. Conserved Features of FRL-PsbD

All FRL-specific residues in the subset of FRL-PsbD sequences from extant organisms (Supplementary Figure S6) are conserved in the ancestral sequence with high confidence (PP $\geq$ 0.91). Two of the conserved residues are nearby the Chl *a* molecule in the P_{D2} site of the electron transfer chain (Supplementary Figure S8). As reported previously [21], the hydroxyl group of the PsbD3-Tyr191 sidechain donates an H-bond to the keto-oxygen atom of the 13^2 methoxycarbonyl moiety of P_{D2}. Interestingly, despite this being confidently predicted in the ancestor, out of the 37 FRL sequences in this dataset, there are six instances of independent reversal of the PsbD3-Tyr191 to Phe (Supplementary Figure S2), which could indicate that (a) the H-bond may not be strictly necessary for FRL function or (b) a different kind of H-bonding interaction is present in the organisms that have Phe at that position. Adjacent to PsbD3-Tyr191 is another conserved FRL-specific residue that is present in the FRL-ASR and strictly conserved in all FRL sequences, PsbD3-Met286. The position of the Met sidechain is shifted slightly away from position 191 (Supplementary Figure S8A), allowing enough space for the H-bonding interaction while maintaining a hydrophobic environment observed in the VL sequences (either Ile or Val, Supplementary Figure S8B). The strict conservation of this Met residue is consistent with the hypothesis that the six FRL-PsbD sequences that lack the H-bonding PsbD3-Tyr131 maintain some other kind of H-bonding interaction, perhaps with a water molecule.

Another region of FRL-PsbD with FRL-specific characteristics is the stromal surface loop between transmembrane helices 4 and 5. This loop is present in all cyanobacterial PSII structures and is unique in that it stretches over the D1 subunit and interacts with PsbC (Figure 4). In this loop, two residues are found in the subset of FRL-PsbD sequences and are also predicted in the FRL ancestor: Gly234 (PP = 0.99) and Ser236 (PP = 0.91) (Figure 4A,B). This is consistent with the suggestion that FRL-allophycocyanin binds to the stromal surface of FRL-PSII in extant FaRLiP-capable cyanobacteria, to the FRL-PsbC and FRL-PsbD subunits [4,21,22,44], possibly to stabilize the interaction and/or tune energy transfer to the core. It also suggests that PsbD-Ser228 and Gly234 were present and performing a similar function in the MRCA of FRL-PSII complexes well over a billion years ago. However, these two residues are not strictly conserved in all extant FRL sequences: Gly234 and Ser236 are found in 27 and 25 out of the 37 sequences, respectively, which suggests that the sites may be under weaker selective pressure compared to strictly conserved sites. Note that these FRL-specific surface residues are close to the Chl *f* in PsbC and a Chl *a* molecule in PsbC that exhibits a FRL-specific H-bond, as described below.

3.5. Conserved Features of FRL-PsbC

As was the case for FRL-PsbA and FRL-PsbD, all FRL-specific residues in the subset of FRL-PsbC sequences from extant cyanobacteria are conserved in the FRL-PsbC ancestral sequence (Supplementary Figure S6). FRL-PsbC contains one Chl *f* molecule that was suggested to serve as a bridging Chl for energy transfer from FRL-allophycocyanin to the electron transfer chain cofactors [21]. Similar to FRL-PsbD, there are FRL-specific residues on the stromal surface of FRL-PsbC that are found in the ancestral sequence (Figure 4). There is also a short deletion (dashed line in Figure 4) that corresponds to the FRL-PsbC sequence that is conserved with the FRL-PsbC ancestral sequence. Along with the conserved residues from the FRL-PsbD loop (Figure 4A), these observations are consistent with the idea that an ancestral FRL-allophycocyanin protein complex bound to the stromal surface of the FRL-PSII ancestor near the FRL-PsbC side of the complex similar to how this occurs in extant FRL-PSII complexes. Obtaining structural data for FRL-allophycocyanin in complex with PSII may aid in determining the roles of the individual conserved residues at their interface surfaces.
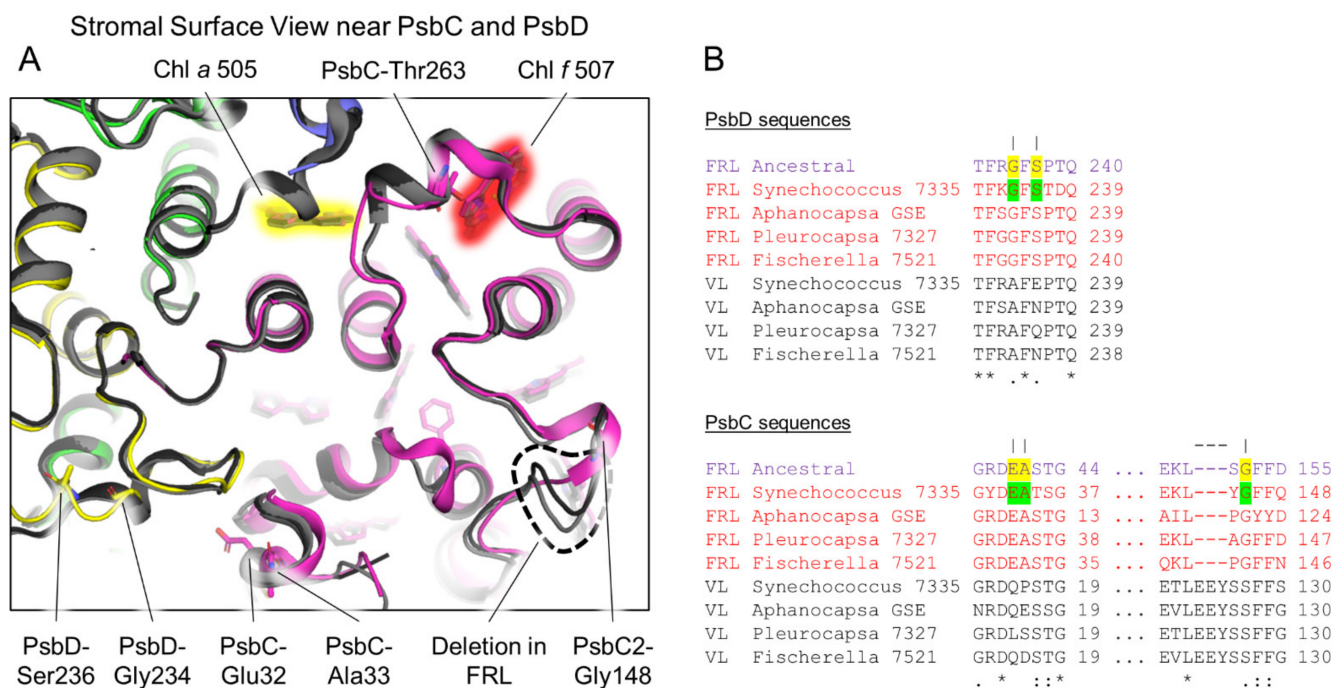
Figure 4. Conservation of stromal surface residues in FRL-PsbC and FRL-PsbD. In (**A**), the *Synechococcus* 7335 apo-FRL-PSII structure (colored) and two non-FaRLiP holocomplex PSII structures (light and dark grey from *T. vulcanus* and *Synechocystis* 6803, respectively) are superimposed. Conserved ancestral residues are shown in stick representation from the *Synechococcus* 7335 apo-FRL-PSII structure only. A PsbC loop absent in FRL-PSII is denoted with a black dashed line. Chl tetrapyrrole rings are shown highlighted in red or yellow. These correspond to the Chl *f* in PsbC and a Chl *a* that exhibits a FRL-specific H-bonding interaction in PsbC, respectively. In (**B**), partial sequence alignments are shown of PsbC and PsbD that include FRL ancestral sequences and FRL- and VL-specific sequences from extant FaRLiP-capable cyanobacteria. Conserved FRL-specific residues in extant cyanobacteria are highlighted green in the sequence from *Synechococcus* 7335. If the same position is conserved in the FRL ancestral sequence, it is highlighted yellow. Vertical lines above residue positions in (**B**) correspond to amino acids labeled in (**A**). The dashed line above the alignment corresponds to the deletion labeled in (**A**). The Clustal Omega sequence conservation identifiers are shown below the alignment.

An additional region with multiple FRL-specific residues in FRL-PsbC occurs in its fifth transmembrane helix. Here, four FRL-specific residues are strictly conserved and are predicted in the FRL-PsbC ancestral sequence (Figure 5) with high confidence (PP > 0.99). The backbone amide nitrogen atom of FRL-PsbC-Asn281 donates an H-bond to a water molecule that is, in turn, an H-bond donor to the formyl moiety of the single Chl *f* molecule in FRL-PsbC, Chl *f* 507 as reported previously [21]. Interestingly, there is also a FRL-specific interaction with an adjacent Chl *a* molecule, Chl *a* 505, which is ~5.1 Å from Chl *f* 507 (edge-to-edge). Specifically, the nitrogen atom of the sidechain for FRL-PsbC-Gln277 donates an H-bond to the 13^1-keto oxygen atom of Chl *a* 505 (Figure 5). H-bonding to this oxygen atom of Chl *a* is well-characterized to result in a bathochromic shift of the absorbance spectrum due to a decrease in site energy [45], which is likely also the case here, and is probably important for energy transfer to the core along with the presence of Chl *f* 507. In addition, FRL-PsbC-Gly284 is conserved in the extant and ancestral FRL sequences, but is instead a bulkier, hydrophobic Leu in the VL sequences. In *T. vulcanus* and *Synechocystis* 6803, this position is Met and Leu, respectively [34,46]. The extra volume available in the FRL-PSII structure allows Chl *a* 505 to be shifted slightly away from the Chl in position 507, allowing space for the water molecule that is important for the latter site to bind Chl *f*. This water is not found in the PSII structures from *T. vulcanus* or *Synechocystis* 6803. These observations

suggest that the Chl *f* in site 507, the H-bond to the 13^1-keto oxygen atom of Chl *a* 505, and the shift of Chl *a* 505 were all present in the FRL-PSII ancestor.
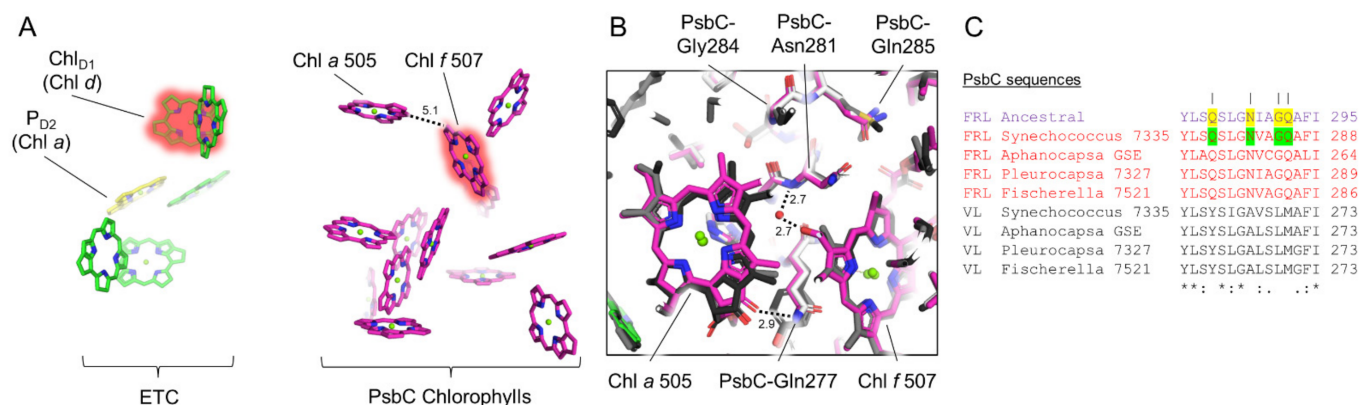
Figure 5. Conservation of FRL-PsbC interactions near Chl *f* 507 and Chl *a* 505. In (**A**), the Chls and pheophytins in the electron transfer chain and FRL-PsbC of the *Synechococcus* 7335 apo-FRL-PSII structure are shown as tetrapyrrole rings only from a stromal-side perspective. The Chls *d* and *f* are highlighted in red, and Chl *a* molecules with FRL-specific H-bonds are additionally labeled. In (**B**), the *Synechococcus* 7335 apo-FRL-PSII structure (magenta), the homology model of the FRL-PsbC ancestral sequence (white), and two non-FaRLiP holocomplex PSII structures (light and dark grey from *T. vulcanus* and *Synechocystis* 6803, respectively) are superimposed. Important H-bonding interactions with distances in units of Å are also shown. In (**C**), a partial sequence alignment is shown that includes the FRL-PsbC ancestral sequence, and FRL- and VL-specific sequences from extant FaRLiP-capable cyanobacteria. Conserved FRL-specific residues in extant cyanobacteria are highlighted in green in the sequence from *Synechococcus* 7335. If the same position is conserved in the FRL-PsbC ancestral sequence, it is highlighted in yellow. Vertical lines above residue positions in (**C**) correspond to amino acids labeled in (**B**). The Clustal Omega sequence conservation identifiers are shown below the alignment.

Another interesting observation regarding FRL-PsbC is the absence of three residues found in all VL-PsbC, with the exception of the basal *Gloeobacterales* clade. These are PsbC-Glu142, Tyr143, and Ser144 in the structure of *T. vulcanus*, located in the stromal surface between the second and third transmembrane helices. The presence of this unique gap does not appear to have any relationship to FaRLiP, but it is consistent with the phylogenetic placement of the FRL group indicating an origin from an early duplication event prior to the main cyanobacterial radiation.

3.6. Conserved Features of FRL-PsbB

Of the core subunits, FRL-PsbB has the largest percentage of FRL-specific residues conserved among the extant cyanobacteria in the subset, but not necessarily in the ancestral sequence (Supplementary Figure S6). This suggests that unlike the other core subunits, FRL-PsbB proteins have acquired more lineage-specific changes since their most recent common ancestor. Interestingly, only one of the Chl *f* sites in FRL-PsbB, Chl *f* 614, appears to be conserved based on the sequence alignments (Figure 6). Whereas all the FRL-specific residues near Chl *f* 614 are conserved in the FRL-PsbB ancestral sequence (PP > 0.99), the FRL-specific residues near Chl *f* 605 and Chl *f* 608 are not, including those near their C2 formyl moieties that confer binding specificity for Chl *f* over Chl *a*. This suggests that the FRL-PSII ancestor did not bind Chl *f* molecules at sites 605 and 608, and some extant representatives also might not. For example, the 18 FRL sequences in our dataset cluster in three distinct clades in the maximum likelihood phylogeny (Supplementary Figure S4). Out of these sequences, only six have PsbB-Trp268 needed in the H-bonding network to Chl *f* 608 (Figure 6C), grouping together as a monophyletic clade. This subgroup coincidentally includes, besides *Synechococcus* 7335, *Aphanocapsa* GSE, and *Pleurocapsa* 7327, the sequences

from the heterocystous *Fischerella* spp. and those from other very closely related strains with over 98% sequence identity (e.g., *Chlorogloeopsis*). The *Fischerella* spp. group is separate from other heterocyst-forming cyanobacteria (e.g., *Calothrix, Rivularia*, etc.), which instead maintain Phe in this position as found in the ancestral FRL-PsbB sequence as well as VL-PsbB sequences and the PsbB sequences from the FaRLiP-incapable *T. vulcanus* and *Synechocystis* 6803. The same is true for positions Phe33 and Leu62 near Chl *f* 605. It is plausible that this may represent a case of gene replacement via horizontal gene transfer, although convergent evolution cannot be entirely ruled out. It suggests that Chls *f* 605 and 608 may be unique to the subgroup containing *Synechococcus* 7335, which highlights the need for more structural data on FRL-PSII from other cyanobacterial species.
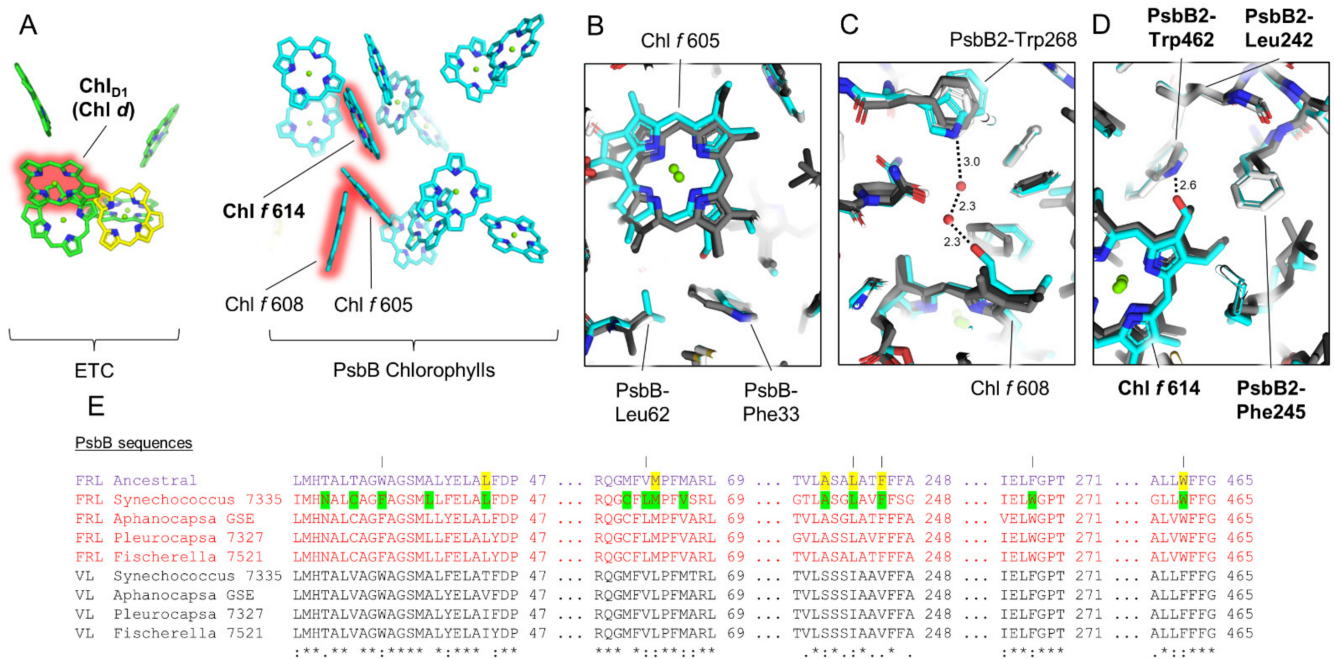
Figure 6. Conservation of FRL-PsbB interactions near the Chl *f* molecules bound to it. In (**A**), the Chls and pheophytins in the electron transfer chain and FRL-PsbB of the *Synechococcus* 7335 apo-FRL-PSII structure are shown as tetrapyrrole rings only from a membrane plane view. The Chls *d* and *f* are highlighted in red. In (**B**–**D**), the *Synechococcus* 7335 apo-FRL-PSII structure (cyan), the homology model of the FRL-PsbB ancestral sequence (white), and two non-FaRLiP holocomplex PSII structures (light and dark grey from *T. vulcanus* and *Synechocystis* 6803, respectively) are superimposed. (**B**–**D**) show the regions of the structures near the C2 formyl moiety of the three Chl *f* molecules bound to FRL-PsbB (sites 605, 608, and 614 based on the original assignment of Chl *a* sites from *T. vulcanus* PSII [46]), respectively. They also show H-bonding interactions involving the C2 formyl moieties of Chl *f* molecules in dashed lines with distances in units of Å where applicable. In (**A**–**D**), features that appear to be conserved in the ancestral FRL-PSII are labeled in bold font. In (**E**), a series of partial sequence alignments is shown that includes the FRL ancestral sequence and FRL- and VL-specific sequences from extant FaRLiP-capable cyanobacteria. Conserved FRL-specific residues in extant cyanobacteria are highlighted in green in the sequence from *Synechococcus* 7335. If the same position is conserved in the FRL ancestral sequence, it is highlighted in yellow. Vertical lines above residue positions in (**E**) correspond to amino acids labeled in (**B**–**D**). The Clustal Omega sequence conservation identifiers are shown below the alignment.

Another important observation occurs at position PsbB-Val244 in the sequence of *Synechococcus* 7335, which is found as Thr244 in 16 out of the 18 FRL sequences, and as Ala in VL-PsbB sequences (and also the PsbB sequences from the FaRLiP-incapable *Synechocystis* 6803 and *T. vulcanus*). The two sequences in the FRL dataset without Thr244 are that of *Synechococcus* 7335, from which the apo-FRL-PSII cryo-EM structure was solved, and one

from a strain of *Nodosilinia* also with Val at the same position. Based on homology modeling, the sidechain of PsbB-Thr244 was previously suggested to confer Chl *f*-binding in site 611 of FRL-PSII from *Fischerella* 7521 by donating an H-bond to a C2 formyl moiety [47]. This assignment was also supported by the lack of an axial-coordinating His sidechain for this site, which is more frequently observed for Chl *f*-containing sites than it is for Chl *a*-containing sites [48,49]. The ancestral sequence predicts a Thr in position 244, so if Chl *f* occupies this site in many extant FRL-PSII complexes, it is also likely to have been found in the FRL-PSII ancestor. To investigate this further, we modeled a Chl *f* into the Chl 611 site from the *Synechococcus* 7335 apo-FRL-PSII structure and modeled a Thr in position 244 instead of Val, occupying the identical rotamer position. Indeed, this approximation places the hydroxyl moiety of Thr244 within H-bonding distance of the C2 formyl moiety of Chl *f* 611 (Supplementary Figure S9). Additionally, the Chl in site 611 is related by pseudo-C_2 symmetry to the Chl *f* found in FRL-PsbC. Such symmetry in antenna Chl *f* sites was also observed in FRL-acclimated photosystem I structures [50], further supporting Chl site 611 binding Chl *f* in organisms where FRL-PsbB contains Thr at position 244. Thus, we propose that the Chl 611 site in most FRL-PSII complexes binds Chl *f*, and that this was a characteristic of the ancestral FRL-PSII complex.

3.7. Conserved Features of FRL-PsbH

PsbH is known to be involved in PSII assembly by binding to PsbB prior to insertion into the PsbA/PsbD/cytochrome b_{559} complex [51]. It is also involved in PSII repair, assisting the incorporation of newly synthesized PsbA into the complex and the binding of bicarbonate [51,52]. Unfortunately, FRL-PsbH is the only FRL-specific subunit in the FRL-acclimated photosystem complexes not yet resolved structurally, as it was not present in the apo-FRL-PSII structure from *Synechococcus* 7335 [21]. Despite its different sequence relative to VL-PsbH isoforms, FRL-PsbH probably binds to the same site relative to the core as is observed in PSII structures from non-FaRLiP strains because only a small sequence region at the N-terminus appears to be FRL specific (Figure 7).

In all PSII structures containing PsbH, the N-terminal region of PsbH extends across the stromal surface near the edge of PsbB where the dimerization interface is found. In the FRL subunit isoforms, including the ancestral sequence, the N-terminus is extended even further, by six or seven residues, and four FRL-specific residues are found that are conserved between FRL-PsbH of extant cyanobacteria and the FRL-PsbH ancestral sequence. This N-terminal region of PsbH surrounds a stromal side Chl *a* molecule in VL-PSII structures, Chl *a* 617, which suggests that FRL-PsbH may alter the site energy of this pigment relative to VL-PsbH. In PSII structures from cyanobacteria incapable of FaRLiP, an H-bond is donated from PsbH-Thr5, which results in red-shifted fluorescence emission from Chl *a* 617 [53], and this residue appears to be conserved in the VL-PsbH sequences from our sequence alignments (Figure 7). In the FRL-PsbH sequences from the subset of extant cyanobacteria and the FRL-PsbH ancestral sequence, a well conserved Ala is found at this position (Supplementary Figure S6). This suggests that the site energy of Chl *a* 617 is altered in FRL-PSII. In addition to this Ala residue, the three other FRL-specific residues are also close to this site, and all are also present in the FRL-PsbH ancestral sequence with high confidence. Again, their presence suggests that the site energy of Chl *a* 617 is altered in FRL-PSII compared to VL-PSII. It seems unlikely, however, that FRL-PsbH could confer Chl *f*-binding at site 617 because the C2 moiety is positioned away from the stromal surface (i.e., toward the hydrophobic interior), distant from the conserved FRL-specific FRL-PsbH residues. We note that the single Chl *f* site found in PsbB of the *Synechococcus* 7335 apo-FRL-PSII structure, which is conserved in the FRL-PSII ancestor (site 614), and the site that probably contains Chl *f* in many extant species and may have been present in the FRL-PSII ancestor (site 611) are also located close to the stromal side of the complex (Figure 7). These sites are ~20–30 Å from the N-terminal extension of FRL-PsbH, whereas the other two FRL-PsbB Chl *f* sites that were probably not present in the FRL-PSII ancestor are on the lumenal side of the complex (Figure 6). Thus, if the site energy of Chl *a* 617 is

altered in FRL, whether in extant or ancestral FRL-PSII complexes, it could be associated with energy transfer with the stromal Chl *f* molecules.
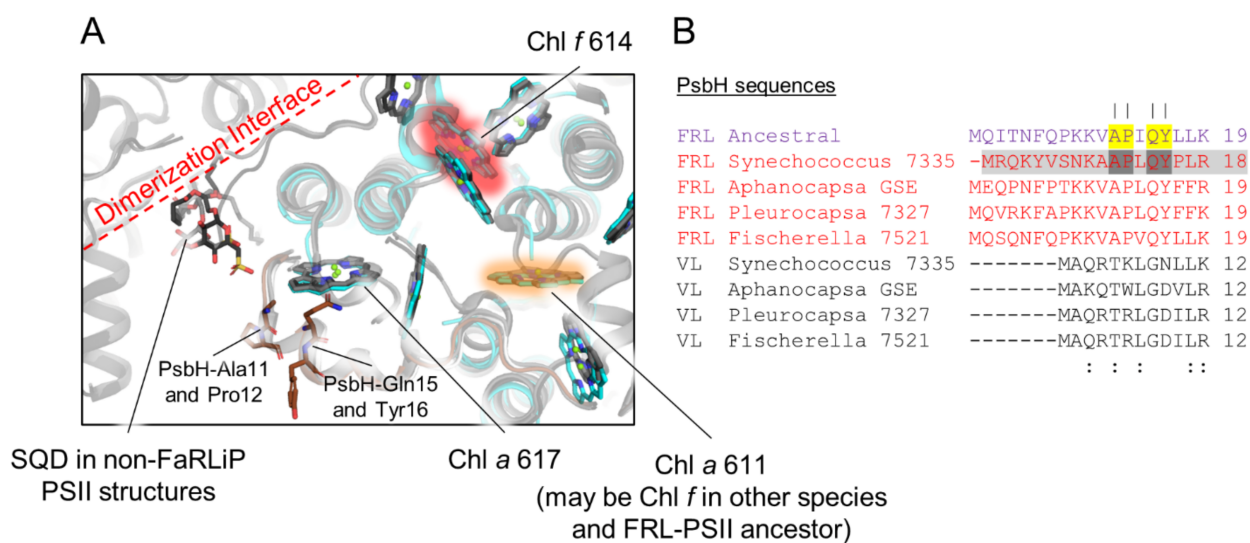
Figure 7. Conservation of FRL-PsbH residues. In (**A**), the *Synechococcus* 7335 apo-FRL-PSII structure (colored), the homology model of the FRL-PsbH ancestral sequence (white), and two non-FaRLiP holocomplex PSII structures (light and dark grey from *T. vulcanus* and *Synechocystis* 6803, respectively) are superimposed. Cartoons are shown with partial transparency and stick representations of applicable residues and cofactors are shown. For Chl molecules, only tetrapyrrole rings are shown. The Chl *f* assigned in PsbB of the *Synechococcus* 7335 apo-FRL-PSII structure is highlighted in red. The Chl *a* site that is suggested to bind Chl *f* in other species and the FRL-PSII ancestor is highlighted in orange. Additional views of the PsbH2 homology model can be found in Supplementary Figure S7. In (**B**), a partial sequence alignment is shown that includes the FRL-ancestral sequence and FRL- and VL-specific sequences from extant FaRLiP-capable cyanobacteria. Note that the sequence from *Synechococcus* 7335 is highlighted in grey to signify that there are presently no corresponding structural data on this subunit. Conserved FRL-specific residues in extant cyanobacteria are highlighted in dark grey in the sequence from *Synechococcus* 7335. If the same position is conserved in the FRL ancestral sequence, it is highlighted in yellow. Vertical lines above residue positions in (**B**) correspond to amino acids labeled in (**A**). The Clustal Omega sequence conservation identifiers are shown below the alignment.

Based on the known function of PsbH in non-FaRLiP cyanobacterial strains [51,52] and its N-terminal alterations in FRL isoforms (Figure 7), FRL-PsbH also seems likely to be involved in PSII assembly and repair and/or dimerization. In non-FaRLiP PSII holocomplex structures, a sulfoquinovosyl-diacylglycerol (SQD) is found in the dimerization interface, very near the N-terminal region of PsbH (Figure 7). The homology models cannot capture the positions of most of the N-terminal residues that comprise the FRL-specific extension of FRL-PsbH or that of its ancestral sequence, and even the residues that can be modeled in the N-terminal regions are likely positioned with low confidence, but the FRL-specific N-terminal extensions could alter the region of the SQD and/or the other interactions at the dimerization interface. Furthermore, it was surprising that the apo-FRL-PSII structure lacked FRL-PsbH, as PsbH is found in all other cryo-EM structures of PSII, even assembly intermediate-like-, apo-, and monomeric-PSII complexes [54–57]. These observations suggest that FRL-PsbH binds less tightly to the FRL-PSII complex relative to VL-PsbH, which could have implications for details of assembly and repair.

Another unique observation of FRL-PsbH is that out of the 31 sequences in our dataset, 6 contain a second transmembrane helix at the C-terminus not found in the other sequences (Supplementary Figure S5). These include the sequences from *Aphanocapsa* GSE, strains designated as *Leptolyngbya* and *Elainella*, and some unnamed strains. These sequences

make up the basal clades of the FRL-PsbH sequences and therefore indicate that this second helix of FRL-PsbH was likely present in the FRL-PSII ancestor. Further inspection of the sequence revealed that the additional helix has homology to another small subunit found in some, but not all, FaRLiP-capable cyanobacteria, including *Halomicronema hongdecloris* (XM38_010780) and some strains of *Nodosilinea* spp (e.g. WP_194026004.1), *Chroococcidiopsis* spp. (WP_199197231.1) and a few more. This is a 48-residue small subunit annotated as hypothetical, but it is also found within the FaRLiP gene cluster downstream of the *psbH2* gene, suggesting that this could be a novel small subunit of FRL-PSII originating from fission of the ancestral gene for the double helical form of FRL-PsbH.

4. Discussion

The phylogenetic analysis presented here suggests several possible transitions prior to the emergence of the MRCA of FaRLiP-capable cyanobacteria and the FaRLiP gene cluster (Figure 8). Taken at face value, the oldest duplication recorded within these PSII subunits is the one leading to ChlF, enabling the synthesis of Chl *f*. Previous work had suggested that ChlF evolved due to a *psbA* gene duplication that occurred before extant cyanobacteria began to diversify [37,38]. However, it is unclear at what point during the evolution of cyanobacteria this paralog evolved to allow Chl *f* synthesis before the MRCA of the FaRLiP-capable cyanobacteria. It is reasonable to assume that the ability to produce and bind, even with low specificity, a few red-shifted pigments alone may have been sufficient to allow survival in FRL-enriched environments prior to the duplications leading to the FRL-specific subunits that could use these pigments more optimally. The second oldest FRL-PSII subunit, as seen in the phylogeny, is FRL-PsbA, which, like ChlF, may have originated from a duplication occurring before the radiation of cyanobacteria. Based on these observations it could be hypothesized that the availability of Chl *f* pigments and the capacity to bind Chl *d* at position Chl_{D1} were some of the earliest innovations towards adapting oxygenic photosynthesis to lower-energy photons, beyond the visible range. It can also be hypothesized that some of the earliest cyanobacteria could have had some ability to use FRL prior to the emergence of the gene cluster as found today. Therefore, it could be predicted that within the more basal clades, which remain poorly sampled, novel FRL use strategies may still be found. It is worth noting that the Chl *d* synthase has not yet been identified [6,58]. While no evidence was found that the FRL sequences had any relationship to those used by *Acaryochloris* spp., there is still a possibility that the adaptations could be still related through the synthesis of Chl *d*. Once the protein(s) responsible for Chl *d* synthesis is(are) found in FaRLiP strains and *Acaryochloris* spp. it may be possible to resolve whether the two adaptations represent separate but convergent evolutionary processes for the use of FRL, or whether the *Acaryochloris* strategy and FaRLiP have a common root that followed diverging paths towards a similar adaptation.

Our analysis suggests that only after the early PsbA duplications, PsbC and PsbD paralogs were added to the emerging FRL-PSII. It is plausible that *psbC* and *psbD* duplicated at the same time since these two genes overlap in most cyanobacteria and in all plastids, including the basal branches, although the overlap is not conserved in the paralogous copies found in the FaRLiP cluster and there is also a stand-alone version of VL-*psbD* outside the FaRLiP cluster [38,59]. The ancestral FRL-PsbD contained a Tyr sidechain that provided a red-shifting H-bond to the 13^2 methoxycarbonyl moiety of P_{D2} analogous to PsbD-Tyr191 observed in the *Synechococcus* 7335 apo-FRL-PSII cryo-EM structure [21]. The ancestral FRL-PsbD also exhibited some residues on the stromal surface that were important for the binding of or energy transfer from FRL-allophycocyanin. Stromal residues in the ancestral FRL-PsbC served similar roles. The conservation of the stromal residues of these two subunits in the ancestral subunits implies that FRL-allophycocyanin emerged with or before the emergence of FRL-PsbC and FRL-PsbD. If FRL-allophycocyanin evolved first, FRL-PSII may have evolved to optimize energy transfer to it. In any case, the FRL-PsbC ancestor bound a single Chl *f* molecule in site 507 and exhibited a red-shifted site energy for the Chl *a* in site 505. The last stage in the evolution of FRL-PSII was the addition of

FRL-PsbB and FRL-PsbH, which may have been added at a later stage to extend FRL-PSII absorption further into the red and to enhance assembly. The FRL-PsbB ancestor contained two Chl *f*-binding sites, 614 and 611, the latter of which is related to the Chl *f* molecule in FRL-PsbC by symmetry with the core. The ancestor of FRL-PsbH appears to have played a role in altering the site energy of a Chl *a* molecule in PsbB, that in site 617, and may have influenced assembly dynamics of the ancestral FRL-PSII complex. We note that every antenna-related characteristic of the FRL-PSII ancestor we have described is present on the stromal side of the FRL-PSII complex (Figure 8). This may support the hypothesis that FRL-allophycocyanin evolved prior to FRL-PSII.
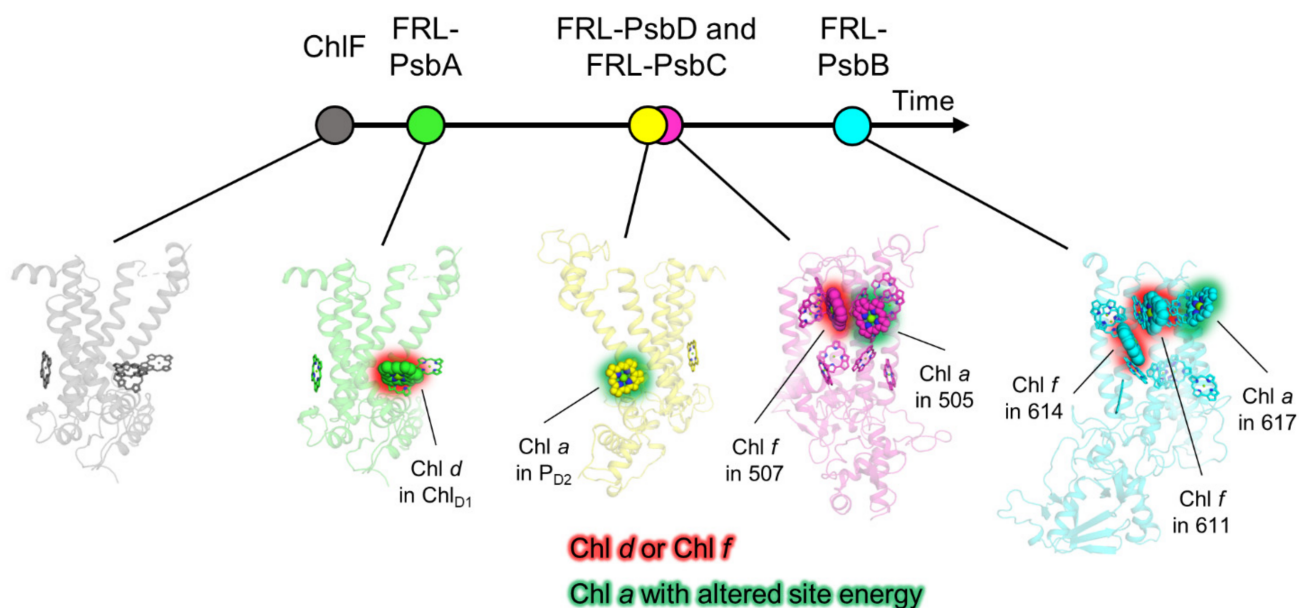
Figure 8. Proposed sequence of events during the emergence of FRL-PSII as described in the text.

Most of this description of ancestral FRL-PSII characteristics is maintained in FRL-PSII complexes from extant cyanobacteria, with only the exception of site 611 probably binding Chl *a* and Chl *f* variably among different extant cyanobacteria. This suggests that a minimal model for FRL-PSII may include one Chl *d* (in the Chl_{D1} site of the electron transfer chain), one Chl *f* in each antenna subunit, and possibly three Chl *a* sites (one in each antenna subunit and one at the P_{D2} site in the electron transfer chain) that are altered in their site energy. These modifications are achieved by altering a relatively small percentage of the total residues in the entire complex. These observations will be useful for engineering FRL absorption into crops with the aim of enhancing biomass production in the future.

References

1. Blankenship, R.E. *Molecular Mechanisms of Photosynthesis*, 3rd ed.; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2021; ISBN 978-1-119-80001-9.

2. Björn, L.O.; Papageorgiou, G.C.; Blankenship, R.E. Govindjee A Viewpoint: Why Chlorophyll *a*? *Photosynth. Res.* **2009**, *99*, 85–98. [CrossRef] [PubMed]

3. Gan, F.; Zhang, S.; Rockwell, N.C.; Martin, S.S.; Lagarias, J.C.; Bryant, D.A. Extensive Remodeling of a Cyanobacterial Photosynthetic Apparatus in Far-Red Light. *Science* **2014**, *345*, 1312–1317. [CrossRef] [PubMed]

4. Ho, M.-Y.; Niedzwiedzki, D.M.; MacGregor-Chatwin, C.; Gerstenecker, G.; Hunter, C.N.; Blankenship, R.E.; Bryant, D.A. Extensive Remodeling of the Photosynthetic Apparatus Alters Energy Transfer among Photosynthetic Complexes When Cyanobacteria Acclimate to Far-Red Light. *Biochim. Biophys. Acta Bioenerg.* **2020**, *1861*, 148064. [CrossRef] [PubMed]

5. Nürnberg, D.J.; Morton, J.; Santabarbara, S.; Telfer, A.; Joliot, P.; Antonaru, L.A.; Ruban, A.V.; Cardona, T.; Krausz, E.; Boussac, A.; et al. Photochemistry beyond the Red Limit in Chlorophyll *f*–Containing Photosystems. *Science* **2018**, *360*, 1210–1213. [CrossRef]

6. Bryant, D.A.; Hunter, C.N.; Warren, M.J. Biosynthesis of the Modified Tetrapyrroles-the Pigments of Life. *J. Biol. Chem.* **2020**, *295*, 6888–6925. [CrossRef]

7. Chen, M.; Blankenship, R.E. Expanding the Solar Spectrum Used by Photosynthesis. *Trends Plant Sci.* **2011**, *16*, 427–431. [CrossRef]

8. Chen, M. Chlorophyll Modifications and Their Spectral Extension in Oxygenic Photosynthesis. *Annu. Rev. Biochem.* **2014**, *83*, 317–340. [CrossRef]

9. Gan, F.; Bryant, D.A. Adaptive and Acclimative Responses of Cyanobacteria to Far-Red Light. *Environ. Microbiol.* **2015**, *17*, 3450–3465. [CrossRef]

10. Viola, S.; Roseby, W.; Santabarabara, S.; Nürnberg, D.; Assunção, R.; Dau, H.; Sellés, J.; Boussac, A.; Fantuzzi, A.; Rutherford, A.W. Impact of Energy Limitations on Function and Resilience in Long-Wavelength Photosystem II. *bioRxiv* **2022**. [CrossRef]

11. Wolf, B.M.; Blankenship, R.E. Far-Red Light Acclimation in Diverse Oxygenic Photosynthetic Organisms. *Photosynth. Res.* **2019**, *142*, 349–359. [CrossRef]

12. Hu, C.; Nawrocki, W.J.; Croce, R. Long-Term Adaptation of *Arabidopsis thaliana* to Far-Red Light. *Plant. Cell Environ.* **2021**. [CrossRef] [PubMed]

13. Gan, F.; Shen, G.; Bryant, D.A. Occurrence of Far-Red Light Photoacclimation (FaRLiP) in Diverse Cyanobacteria. *Life* **2014**, *5*, 4–24. [CrossRef] [PubMed]

14. Antonaru, L.A.; Cardona, T.; Larkum, A.W.D.; Nürnberg, D.J. Global Distribution of a Chlorophyll *f* Cyanobacterial Marker. *ISME J.* **2020**, *14*, 2275–2287. [CrossRef] [PubMed]

15. Averina, S.; Velichko, N.; Senatskaya, E.; Pinevich, A. Far-Red Light Photoadaptations in Aquatic Cyanobacteria. *Hydrobiologia* **2018**, *813*, 1–17. [CrossRef]

16. Ohkubo, S.; Miyashita, H. A Niche for Cyanobacteria Producing Chlorophyll *f* within a Microbial Mat. *ISME J.* **2017**, *11*, 2368–2378. [CrossRef]

17. Behrendt, L.; Brejnrod, A.; Schliep, M.; Sørensen, S.J.; Larkum, A.W.D.; Kühl, M. Chlorophyll *f*-Driven Photosynthesis in a Cavernous Cyanobacterium. *ISME J.* **2015**, *9*, 2108–2111. [CrossRef]

18. Behrendt, L.; Trampe, E.L.; Nord, N.B.; Nguyen, J.; Kühl, M.; Lonco, D.; Nyarko, A.; Dhinojwala, A.; Hershey, O.S.; Barton, H. Life in the Dark: Far-Red Absorbing Cyanobacteria Extend Photic Zones Deep into Terrestrial Caves. *Environ. Microbiol.* **2019**. [CrossRef]

19. Kühl, M.; Trampe, E.; Mosshammer, M.; Johnson, M.; Larkum, A.W.D.; Frigaard, N.-U.; Koren, K. Substantial Near-Infrared Radiation-Driven Photosynthesis of Chlorophyll *f*-Containing Cyanobacteria in a Natural Habitat. *Elife* **2020**, *9*, e50871. [CrossRef]

20. Rippka, R.; Deruelles, J.; Waterbury, J.B.; Herdman, M.; Stanier, R.Y. Generic Assignments, Strain Histories and Properties of Pure Cultures of Cyanobacteria. *Microbiology* **1979**, *111*, 1–61. [CrossRef]

21. Gisriel, C.J.; Shen, G.; Ho, M.-Y.; Kurashov, V.; Flesher, D.A.; Wang, J.; Armstrong, W.H.; Golbeck, J.H.; Gunner, M.R.; Vinyard, D.J.; et al. Structure of a Monomeric Photosystem II Core Complex from a Cyanobacterium Acclimated to Far-Red Light Reveals the Functions of Chlorophylls *d* and *f*. *J. Biol. Chem.* **2022**, *298*, 101424. [CrossRef]

22. Soulier, N.; Laremore, T.N.; Bryant, D.A. Characterization of Cyanobacterial Allophycocyanins Absorbing Far-Red Light. *Photosynth. Res.* **2020**, *145*, 189–207. [CrossRef] [PubMed]

23. Nguyen, L.-T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [CrossRef] [PubMed]

24. Duvaud, S.; Gabella, C.; Lisacek, F.; Stockinger, H.; Ioannidis, V.; Durinx, C. Expasy, the Swiss Bioinformatics Resource Portal, as Designed by Its Users. *Nucleic Acids Res.* **2021**, *49*, W216–W227. [CrossRef]

25. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; et al. Fast, Scalable Generation of High Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 1–6. [CrossRef]

26. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; von Haeseler, A.; Jermiin, L.S. ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates. *Nat. Methods* **2017**, *14*, 587–589. [CrossRef] [PubMed]

27. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; von Haeseler, A.; Lanfear, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534. [CrossRef]

28. Minh, B.Q.; Nguyen, M.A.T.; von Haeseler, A. Ultrafast Approximation for Phylogenetic Bootstrap. *Mol. Biol. Evol.* **2013**, *30*, 1188–1195. [CrossRef]

29. Hoang, D.T.; Chernomor, O.; von Haeseler, A.; Minh, B.Q.; Vinh, L.S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **2018**, *35*, 518–522. [CrossRef]

30. Anisimova, M.; Gascuel, O. Approximate Likelihood-Ratio Test for Branches: A Fast, Accurate, and Powerful Alternative. *Syst. Biol.* **2006**, *55*, 539–552. [CrossRef]

31. Huson, D.H.; Scornavacca, C. Dendroscope 3: An Interactive Tool for Rooted Phylogenetic Trees and Networks. *Syst. Biol.* **2012**, *61*, 1061–1067. [CrossRef]

32. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. Available online: https://pymol.org/2/ (accessed on 26 May 2022).

33. Guex, N.; Peitsch, M.C.; Schwede, T. Automated Comparative Protein Structure Modeling with SWISS-MODEL and Swiss-PdbViewer: A Historical Perspective. *Electrophoresis* **2009**, *30*, S162–S173. [CrossRef]

34. Gisriel, C.J.; Wang, J.; Liu, J.; Flesher, D.A.; Reiss, K.M.; Huang, H.-L.; Yang, K.R.; Armstrong, W.H.; Gunner, M.R.; Batista, V.S.; et al. High-Resolution Cryo-EM Structure of Photosystem II from the Mesophilic Cyanobacterium, *Synechocystis* sp. PCC 6803. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2116765118. [CrossRef] [PubMed]

35. Sheridan, K.J.; Duncan, E.J.; Eaton-Rye, J.J.; Summerfield, T.C. The Diversity and Distribution of D1 Proteins in Cyanobacteria. *Photosynth. Res.* **2020**, *145*, 111–128. [CrossRef] [PubMed]

36. Cardona, T.; Murray, J.W.; Rutherford, A.W. Origin and Evolution of Water Oxidation before the Last Common Ancestor of the Cyanobacteria. *Mol. Biol. Evol.* **2015**, *32*, 1310–1328. [CrossRef] [PubMed]

37. Cardona, T.; Sánchez-Baracaldo, P.; Rutherford, A.W.; Larkum, A.W. Early Archean Origin of Photosystem II. *Geobiology* **2019**, *17*, 127–150. [CrossRef] [PubMed]

38. Oliver, T.; Sánchez-Baracaldo, P.; Larkum, A.W.; Rutherford, A.W.; Cardona, T. Time-Resolved Comparative Molecular Evolution of Oxygenic Photosynthesis. *Biochim. Biophys. Acta Bioenerg.* **2021**, *1862*, 148400. [CrossRef]

39. Ho, M.Y.; Shen, G.; Canniffe, D.P.; Zhao, C.; Bryant, D.A. Light-Dependent Chlorophyll *f* Synthase Is a Highly Divergent Paralog of PsbA of Photosystem II. *Science* **2016**, *353*, aaf9178. [CrossRef]

40. Trinugroho, J.P.; Bečková, M.; Shao, S.; Yu, J.; Zhao, Z.; Murray, J.W.; Sobotka, R.; Komenda, J.; Nixon, P.J. Chlorophyll *f* Synthesis by a Super-Rogue Photosystem II Complex. *Nat. Plants* **2020**, *6*, 238–244. [CrossRef]

41. Shen, G.; Canniffe, D.P.; Ho, M.-Y.; Kurashov, V.; van der Est, A.; Golbeck, J.H.; Bryant, D.A. Characterization of Chlorophyll *f* Synthase Heterologously Produced in *Synechococcus* Sp. PCC 7002. *Photosynth. Res.* **2019**, *140*, 77–92. [CrossRef]

42. Saw, J.H.; Cardona, T.; Montejano, G. Complete Genome Sequencing of a Novel *Gloeobacter* Species from a Waterfall Cave in Mexico. *Genome Biol. Evol.* **2021**, *13*, evab264. [CrossRef]

43. Shih, P.M.; Wu, D.; Latifi, A.; Axen, S.D.; Fewer, D.P.; Talla ECalteau, A.; Cai, F.; de Marsac, N.T.; Rippka, R. Improving the Coverage of the Cyanobacterial Phylum Using Diversity-Driven Genome Sequencing. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 1053–1058. [CrossRef] [PubMed]

44. Ho, M.Y.; Gan, F.; Shen, G.; Bryant, D.A. Far-Red Light Photoacclimation (FaRLiP) in *Synechococcus* sp. PCC 7335. II. Characterization of Phycobiliproteins Produced during Acclimation to Far-Red Light. *Photosynth. Res.* **2017**, *131*, 187–202. [CrossRef] [PubMed]

45. Llansola-Portoles, M.J.; Li, F.; Xu, P.; Streckaite, S.; Ilioaia, C.; Yang, C.; Gall, A.; Pascal, A.A.; Croce, R.; Robert, B. Tuning Antenna Function through Hydrogen Bonds to Chlorophyll *a*. *Biochim. Biophys. Acta Bioenerg.* **2020**, *1861*, 148078. [CrossRef]

46. Umena, Y.; Kawakami, K.; Shen, J.-R.; Kamiya, N. Crystal Structure of Oxygen-Evolving Photosystem II at a Resolution of 1.9 Å. *Nature* **2011**, *473*, 55–60. [CrossRef] [PubMed]

47. Gisriel, C.J.; Huang, H.-L.; Reiss, K.M.; Flesher, D.A.; Batista, V.S.; Bryant, D.A.; Brudvig, G.W.; Wang, J. Quantitative Assessment of Chlorophyll Types in Cryo-EM Maps of Photosystem I Acclimated to Far-Red Light. *BBA Adv.* **2021**, *1*, 100019. [CrossRef]

48. Gisriel, C.J.; Wang, J.; Brudvig, G.W.; Bryant, D.A. Opportunities and Challenges for Assigning Cofactors in Cryo-EM Density Maps of Chlorophyll-Containing Proteins. *Commun. Biol.* **2020**, *3*, 408. [CrossRef] [PubMed]

49. Gisriel, C.J.; Shen, G.; Kurashov, V.; Ho, M.-Y.; Zhang, S.; Williams, D.; Golbeck, J.H.; Fromme, P.; Bryant, D.A. The Structure of Photosystem I Acclimated to Far-Red Light Illuminates an Ecologically Important Acclimation Process in Photosynthesis. *Sci. Adv.* **2020**, *6*, eaay6415. [CrossRef]

50. Gisriel, C.J.; Flesher, D.A.; Shen, G.; Wang, J.; Ho, M.-Y.; Brudvig, G.W.; Bryant, D.A. Structure of a Photosystem I-Ferredoxin Complex from a Marine Cyanobacterium Provides Insights into Far-Red Light Photoacclimation. *J. Biol. Chem.* **2022**, *298*, 101408. [CrossRef]

51. Komenda, J.; Tichý, M.; Eichacker, L.A. The PsbH Protein Is Associated with the Inner Antenna CP47 and Facilitates D1 Processing and Incorporation into PSII in the Cyanobacterium *Synechocystis* PCC 6803. *Plant Cell Physiol.* **2005**, *46*, 1477–1483. [CrossRef]

52. Komenda, J.; Lupínková, L.; Kopecký, J. Absence of the *PsbH* Gene Product Destabilizes Photosystem II Complex and Bicarbonate Binding on Its Acceptor Side in *Synechocystis* PCC 6803. *Eur. J. Biochem.* **2002**, *269*, 610–619. [CrossRef]

53. D'Haene, S.E.; Sobotka, R.; Bučinská, L.; Dekker, J.P.; Komenda, J. Interaction of the PsbH Subunit with a Chlorophyll Bound to Histidine 114 of CP47 Is Responsible for the Red 77K Fluorescence of Photosystem II. *Biochim. Biophys. Acta Bioenerg.* **2015**, *1847*, 1327–1334. [CrossRef] [PubMed]

54. Zabret, J.; Bohn, S.; Schuller, S.K.; Arnolds, O.; Möller, M.; Meier-Credo, J.; Liauw, P.; Chan, A.; Tajkhorshid, E.; Langer, J.D.; et al. Structural Insights into Photosystem II Assembly. *Nat. Plants* **2021**, *7*, 524–538. [CrossRef] [PubMed]

55. Graça, A.T.; Hall, M.; Persson, K.; Schröder, W.P. High-Resolution Model of *Arabidopsis* Photosystem II Reveals the Structural Consequences of Digitonin-Extraction. *Sci. Rep.* **2021**, *11*, 15534. [CrossRef] [PubMed]

56. Yu, H.; Hamaguchi, T.; Nakajima, Y.; Kato, K.; Kawakami, K.; Akita, F.; Yonekura, K.; Shen, J.-R. Cryo-EM Structure of Monomeric Photosystem II at 2.78 Å Resolution Reveals Factors Important for the Formation of Dimer. *Biochim. Biophys. Acta Bioenerg.* **2021**, *1862*, 148471. [CrossRef] [PubMed]

57. Huang, G.; Xiao, Y.; Pi, X.; Zhao, L.; Zhu, Q.; Wang, W.; Kuang, T.; Han, G.; Sui, S.-F.; Shen, J.-R. Structural Insights into a Dimeric Psb27-Photosystem II Complex from a Cyanobacterium *Thermosynechococcus vulcanus*. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2018053118. [CrossRef]

58. Bryant, D.A.; Shen, G.; Turner, G.M.; Soulier, N.; Laremore, T.N.; Ho, M.-Y. Far-Red Light Allophycocyanin Subunits Play a Role in Chlorophyll *d* Accumulation in Far-Red Light. *Photosynth. Res.* **2020**, *143*, 81–95. [CrossRef]

59. Carpenter, S.D.; Charite, J.; Eggers, B.; Vermaas, W.F.J. The PsbC Start Codon in *Synechocystis* sp. PCC 6803. *FEBS Lett.* **1990**, *260*, 135–137. [CrossRef]

60. Hessa, T.; Meindl-Beinker, N.M.; Bernsel, A.; Kim, H.; Sato, Y.; Lerch-Bader, M.; Nilsson, I.; White, S.H.; von Heijne, G. Molecular Code for Transmembrane-Helix Recognition by the Sec61 Translocon. *Nature* **2007**, *450*, 1026–1030. [CrossRef]

61. Emsley, P.; Lohkamp, B.; Scott, W.G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 486–501. [CrossRef]

Article

Comparative Genomic Analysis of the Marine Cyanobacterium *Acaryochloris marina* MBIC10699 Reveals the Impact of Phycobiliprotein Reacquisition and the Diversity of *Acaryochloris* Plasmids

Haruki Yamamoto [1,*]**, Kazuma Uesaka** [2]**, Yuki Tsuzuki** [1]**, Hisanori Yamakawa** [1]**, Shigeru Itoh** [3] **and Yuichi Fujita** [1]

[1] Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya 464-8601, Japan; duffko0614@gmail.com (Y.T.); yamakawa008@gmail.com (H.Y.); fujita@agr.nagoya-u.ac.jp (Y.F.)
[2] Center for Gene Research, Nagoya University, Nagoya 464-8601, Japan; uesaka.kazuma.p1@f.mail.nagoya-u.ac.jp
[3] Graduate School of Sciences, Nagoya University, Nagoya 464-8601, Japan; itoh@bio.phys.nagoya-u.ac.jp
* Correspondence: haruki@agr.nagoya-u.ac.jp; Tel.: +81-52-789-4090

Abstract: *Acaryochloris* is a marine cyanobacterium that synthesizes chlorophyll *d*, a unique chlorophyll that absorbs far-red lights. *Acaryochloris* is also characterized by the loss of phycobiliprotein (PBP), a photosynthetic antenna specific to cyanobacteria; however, only the type-strain *A. marina* MBIC11017 retains PBP, suggesting that PBP-related genes were reacquired through horizontal gene transfer (HGT). *Acaryochloris* is thought to have adapted to various environments through its huge genome size and the genes acquired through HGT; however, genomic information on *Acaryochloris* is limited. In this study, we report the complete genome sequence of *A. marina* MBIC10699, which was isolated from the same area of ocean as *A. marina* MBIC11017 as a PBP-less strain. The genome of *A. marina* MBIC10699 consists of a 6.4 Mb chromosome and four large plasmids totaling about 7.6 Mb, and the phylogenic analysis shows that *A. marina* MBIC10699 is the most closely related to *A. marina* MBIC11017 among the *Acaryochloris* species reported so far. Compared with *A. marina* MBIC11017, the chromosomal genes are highly conserved between them, while the genes encoded in the plasmids are significantly diverse. Comparing these genomes provides clues as to how the genes for PBPs were reacquired and what changes occurred in the genes for photosystems during evolution.

Keywords: cyanobacteria; *Acaryochloris*; comparative genome analysis

1. Introduction

Photosynthetic organisms on earth convert solar energy into chemical energy that can be used by living organisms. Many photosynthetic organisms, including land plants, use chlorophyll *a* (Chl *a*) as the major photosynthetic pigment. Additionally, there are several derivatives with different substituents, such as Chl *b*, *c*, *d*, and *f*, and each has an absorption spectrum different from that of Chl *a* [1]. Those Chl derivatives are auxiliary pigments to Chl *a*, and in higher plants with Chl *a* and Chl *b*, only Chl *a* functions as the reaction center Chl to ignite the photosynthetic electron transfer, while Chl *b* does as antenna Chl in the LHC and other complexes [2]. The marine cyanobacterium *Acaryochloris marina* is a unique cyanobacterium that synthesizes Chl *d* as its primary photosynthetic pigment. *Acaryochloris* retains both Chl *d* and Chl *a*, but more than 90% of its total Chl is composed of Chl *d*, and the Chl special pair in the photosynthetic reaction centers is also occupied by Chl *d* [3,4]. *Acaryochloris* is the only example of a Chl-*d*-centered photosystem that could be recognized as a specialization in utilizing far-red light for oxygenic photosynthesis [4,5].

Acaryochloris marina MBIC11017, which has become the most analyzed type-strain today, was first discovered in 1996 as a symbiont with ascidians in Palau Sea waters as the first cyanobacterium to biosynthesize Chl *d* [3]. Since then, *Acaryochloris* species have

been reported to be distributed in various environments, including marine stromatolites, a saline lake epilithic biofilm, and the aquatic plant rhizosphere of freshwater streams [6]. The complete genome of *A. marina* MBIC11017, reported in 2008, consists of a chromosome (6.5 Mb) and nine plasmids, giving it a total genome size of 8.3 Mb, still one of the largest genomes for unicellular cyanobacteria to date [7]. This huge genome and its extensibility could allow *Acaryochloris* to adaptively change into various niches. The fact that many *Acaryochloris* strains have been found in various environments confirms this hypothesis. Gene acquisition through horizontal gene transfer (HGT) acts as the primary driving force for adaptive evolution to a new environmental niche. A set of phycobiliprotein (PBP)-related genes is known as one typical example of HGT in *Acaryochloris* [8]. Among numerous *Acaryochloris* species, only *A. marina* MBIC11017 retains PBP as a photosynthetic antenna complex. PBP-binding bilin pigments as a photosynthetic antenna are distributed among cyanobacteria and red algae. Genes for PBPs appear to have been lost in the common ancestor of *Acaryochloris* and some marine cyanobacteria. However, *A. marina* MBIC11017 is the only *Acaryochloris* strain that retains PBP, which is thought to have reacquired the PBP genes through HGT [8]. Additionally, some *Acaryochloris* strains have been reported to acquire genes related to the nitrogen-fixing enzyme nitrogenase by HGT and are actually capable of nitrogen-fixing growth [6,9]. In both cases, the gene transfer event appears to have occurred in units of gene clusters containing many genes, suggesting that *Acaryochloris* may have a mechanism which allows it to acquire such a large gene cluster and provides an excellent model for investigating the mechanism of HGT of such large gene clusters. However, most *Acaryochloris* genomes reported so far have been sequenced with short-read types of next-generation sequencers, and registered as many contigs, with complete genome sequences limited to only three species.

In this paper, we report the complete genome of *Acaryochloris marina* MBIC10699 (previously registered as *Acaryochloris* sp. MBIC10699), a strain of *Acaryochloris* without PBP isolated from the same ocean as *A. marina* MBIC11017. Since *A. marina* MBIC10699 was found to be phylogenetically very closely related to *A. marina* MBIC11017, a genomic comparison of these two strains allows us to examine how the *Acaryochloris* lineage leading to *A. marina* MBIC11017 acquired the PBP genes through HGT and what changes in the photosynthetic reaction centers and antenna complexes occurred as a result of the PBP reacquisition. Comparative analysis showed that there is diversity in plasmid composition and plasmid-coding genes in these two *Acaryochloris* strains, and that phenotypic differences, including the presence or absence of PBP, can be explained by differences in plasmid gene composition. It suggests that *Acaryochloris* has acquired various phenotypes through HGT by using multiple giant plasmids as a gene pool.

2. Materials and Methods

2.1. Isolation and Cultivation of Acaryochloris

Acaryochloris marina MBIC10699 was purchased from Biological Resource Center, NITE (NBRC; Kisarazu, Japan), as *Acaryochloris* sp. NBRC102871. Contaminant bacteria, which showed small white spherical colonies on an agar plate in the original cultures were eliminated by repeated single-colony isolation. The isolated pure culture was maintained on agar plates in the laboratory. *Acaryochloris* strains were grown in a 0.5% (*v/v*) Daigo IMK medium with 3.6% (*w/v*) artificial seawater (Marine Art SF-1) with white fluorescent light exposure at 10 μE/m^2 s (FRL40SW; Hitachi, Tokyo, Japan) at 26 °C.

2.2. Pigment Extraction

Total pigments from *Acaryochloris* cells were extracted according to the previous report [10]. After adding methanol to the collected cells (final 90% (*v/v*)), the cells were disrupted by sonication (TOMY UD-201; output: 3, duty: 30, 10 s; TOMY, Tokyo, Japan) and then left on ice for 30 min for pigment extraction. The supernatant collected by centrifugation at 15,000 rpm for 20 min (MX-300, AR015-24; TOMY) was used for HPLC analysis. HPLC analysis (Shimadzu LC20-AD, Shimadzu, Kyoto, Japan) was performed

according to Zapata's method [11] and Chl *a* and *d* were detected by absorption at 440 nm and 690 nm, respectively. Quantification of Chl *a* and Chl *d* was performed by preparing standard curves of peak areas on HPLC chromatograms with standard pigments.

2.3. Extraction of Genomic DNA

Cells grown in 45 mL IMK medium were corrected and suspended in 500 µL TE buffer. After three cycles of freezing and thawing, 100 µL of 50 mg/mL lysozyme solution was added to the cell suspension, followed by 1 h incubation at 37 °C. After washing of the cells with TE, cells were suspended in 600 µL of DNA extraction buffer (DNA suisui; Rizo Inc., Tsukuba, Japan) with 3 µL of RNase solution and incubated at 70 °C for 10 min. After cooling it down to room temperature (RT), 600 µL of TE-saturated phenol was added and mixed by inverting the tube up and down, then slowly stirred at RT for 15 min. The upper phase was recovered after centrifugation (15,000 rpm, 5 min, RT) and then an equal volume of PCI solution (phenol/chloroform/isoamyl alcohol, 25:24:1) was added and stirred slowly for 15 min at RT. This wash step with PCI was repeated once, and the upper phase was mixed with an equal volume of CIA (chloroform/isoamyl alcohol, 24:1) and stirred slowly for 5 min at RT. The upper phase was recovered after centrifugation (15,000 rpm, 5 min, RT) and 1/10 volume of 3 M sodium acetate solution and an equal volume of 2-propanol were added, mixed gently, and incubated on ice for 10 min. After centrifugation (15,000 rpm, 10 min, 4 °C), the pellet was washed with 70% ethanol and dried to remove ethanol completely. Genomic DNA was dissolved in 30 µL of TE buffer.

2.4. Whole Genome Sequencing of A. marina MBIC10699

The obtained high molecular weight DNA was shipped to the Oral Microbiome Center in Taniguchi Dental Clinic in Japan for long-read and short-read sequencing. Briefly explained, the paired-end (2 × 150-bp) DNA library was prepared using the MGIEasy FS PCR-free DNA library prep set (MGI Tech., Shenzhen, China), according to the manufacturer's instructions. DNBSEQ-G400 sequencing yielded 1,860,686 paired-end reads. For long-read sequencing, DNA library was prepared using a ligation sequencing kit (SQK-LSK-109; Oxford Nanopore Technologies, Ltd. (ONT), Oxford, UK). An R9.4.1 flow cell (FLO-MIN106) was used to sequence a DNA library using a GridION X5 system (ONT).1-06). After a 24 h run, the FAST5 format file was base called using Guppy v.3.6.0 (ONT), which generated 56,313 reads.

2.5. Quality Control of Sequencing Reads

All raw sequencing data were pre-processed to reduce low-quality bases. For the BGI-seq reads, fastp 0.23.0 [12] pre-processor was used with the "-q 20 -t 1 -T 1 -l 20" option. For the Oxford Nanopore reads, nanofilt 2.7.0 [13] was used with the "-q 10 -l 500 –headcrop 75" option.

2.6. Genome Assembly and Gene Annotation

For the genome assembly, pre-processed short reads and long reads were hybrid assembled using Flye 0.2.8 [14]. The output of Flye was polished using Pilon [15] three times, generating a single circular sequence for the chromosome and another four circular plasmid sequences. The chromosome sequence was rotated to the first nucleotide of the 100 bp upstream of the *dnaA* gene. The genome sequence was then annotated using the annotation pipeline DFAST 1.5.0 [16] provided by DDBJ. DFAST automatic annotation predicted 6800 coding sequences and 6 rRNA genes and 72 tRNA genes. The sequencing depth of the chromosome and four plasmid sequences were calculated using BBmap (BBMap-BushnellB.-sourceforge.net/projects/bbmap/, accessed on 15 March 2021). The complete genome sequence and annotation of *A. marina* MBIC10699 was deposited at DDBJ under accession numbers AP026075 (Chr), AP026076 (pREC1), AP026077 (pREC2), AP026078 (pREC3), and AP026079 (pREC4). Raw sequencing data were deposited in the DDBJ SRA database under BioProject number PRJDB13468 and BioSample number SAMD00467986.

2.7. Comparative Genomics

A. marina MBIC10699 and *A. marina* MBIC11017 genome sequences were compared using pairwise average nucleotide identity (ANI) and dot plot analysis. For ANI calculation, orthoANI 0.5.0 [17] was used. For dot plot comparison, D-GENIES web server (http://dgenies.toulouse.inra.fr (accessed on 23 February 2022) was used [18]. Annotation of genes by KEGG was performed by GhostKOALA. Homologous genes were searched by blastp (e-value cut off: 1×10^{-9}), and the sequence alignment was visually confirmed. When a global alignment could not be confirmed, it was eliminated.

2.8. Core genome Phylogenic Analysis

To determine the phylogenic relationship between 26 *Acaryochloris* strains with out-group strains, core protein alignments were generated using roary 3.13.0 [19] with the following options: -e –mafft -r -qc -cd 90 -i 90 –group limit 70000. The maximum likelihood tree was constructed using IQ-TREE version 2.0.3 with 100 bootstraps using the best-fit model (LG + F + R10) determined in ModelFinder.

3. Results and Discussion

3.1. Comparative Analysis of Pigment Contents of Two Acaryochloris Strains

Acaryochloris sp. MBIC10699 (hereafter MB10699) differs significantly in cell coloration compared with the type-strain *A. marina* MBIC11017 (MB11017). Comparison of their cellular spectra showed a clear difference in the 550–650 nm absorption peak derived from phycobiliprotein (PBP), confirming the presence and absence of PBP in MB11017 and MB10699, respectively (Figure 1). HPLC analysis of methanol cell extracts showed that the two *Acaryochloris* strains had the same chlorophyll (Chl) and carotenoid compositions. The amount of cellular Chl *d* in MB10699 was about 1.5 times higher than that in MB11017, as reflected in the Chl *a/d* ratios: 0.044 in MB10699 and 0.052 in MB11017 (Figure 1). The major carotenoids common to MB10699 and MB11017 are thought to be zeaxanthin and *α*-carotene [20]. These results indicate that the two strains are nearly identical in photosynthetic pigment composition except for the presence or absence of PBP, even though there is a slight difference in Chl *d* content. MB10699, which does not have PBP as an antenna protein, is thought to employ another antenna complex, Chl-bound Pcb (prochlorophyte chlorophyll-binding protein), as the primary antenna [21,22]. Therefore, it is expected that MB10699 contains a larger amount of Pcb antennae than MB11017 to accommodate higher amounts of Chl *d*.

3.2. Overall View of the Complete Genome of A. marina MBIC10699

The complete genomic structure of MB10699 was determined by combining Oxford Nanopore long-read and MGI-seq short-read sequencing. The genome of MB10699 consists of one circular chromosome (Chr; 6.4 Mb) and four giant plasmids (pREC1, 393 kb; pREC2, 329 kb; pREC3, 303 kb; and pREC4, 205 kb) with a total genome size of 7.6 Mb (Figure 2), which is significantly less than that of MB11017 (8.3 Mb). The short-read coverage depths were almost the same for Chr and the four plasmids, suggesting that these plasmids exist in the same number of copies as Chr (data not shown). The GC content and coding capacity were 47.0% and 82.3%, respectively, which were almost identical to those of MB11017, with 47.0% and 84.1%, respectively. All 6 rRNAs and 72 tRNAs were encoded by Chr, and 3 CRISPRs were encoded by Chr (1) and pREC1 (2). The number of CDSs in the MB10699 genome was 6813, about 80% of the 8528 CDSs found in MB11017 (Table 1). Pairwise average nucleotide identity (ANI) calculations showed that MB10699 had a value of 98.0 relative to MB11017, indicating that these two *Acaryochloris* strains are closely related enough to be categorized as the same species [23,24]. A phylogenetic tree based on the 897 number of concatenated protein sequences among 26 *Acaryochloris* species with *Cyanothece* sp. PCC7425 and *A. thomasi* RCC1774 showed that *Acaryochloris* is divided into three major clades, and MB10699 and MB11017 are the most closely related (Figure 3, Table S1). Based on the results of ANI and this phylogenetic analysis, MB10699 was renamed *A. marina*

MBIC10699 as the same species as *A. marina* MBIC11017 in the taxonomy, and the genome information was registered as *A. marina* MBIC10699. Since MB11017 is the sole strain that retains PBP among the 26 *Acaryochloris* strains, including MB10699, genome comparison between MB10699 and MB11017 provides an excellent example to trace the evolution of the reacquisition of PBP genes through HGT.
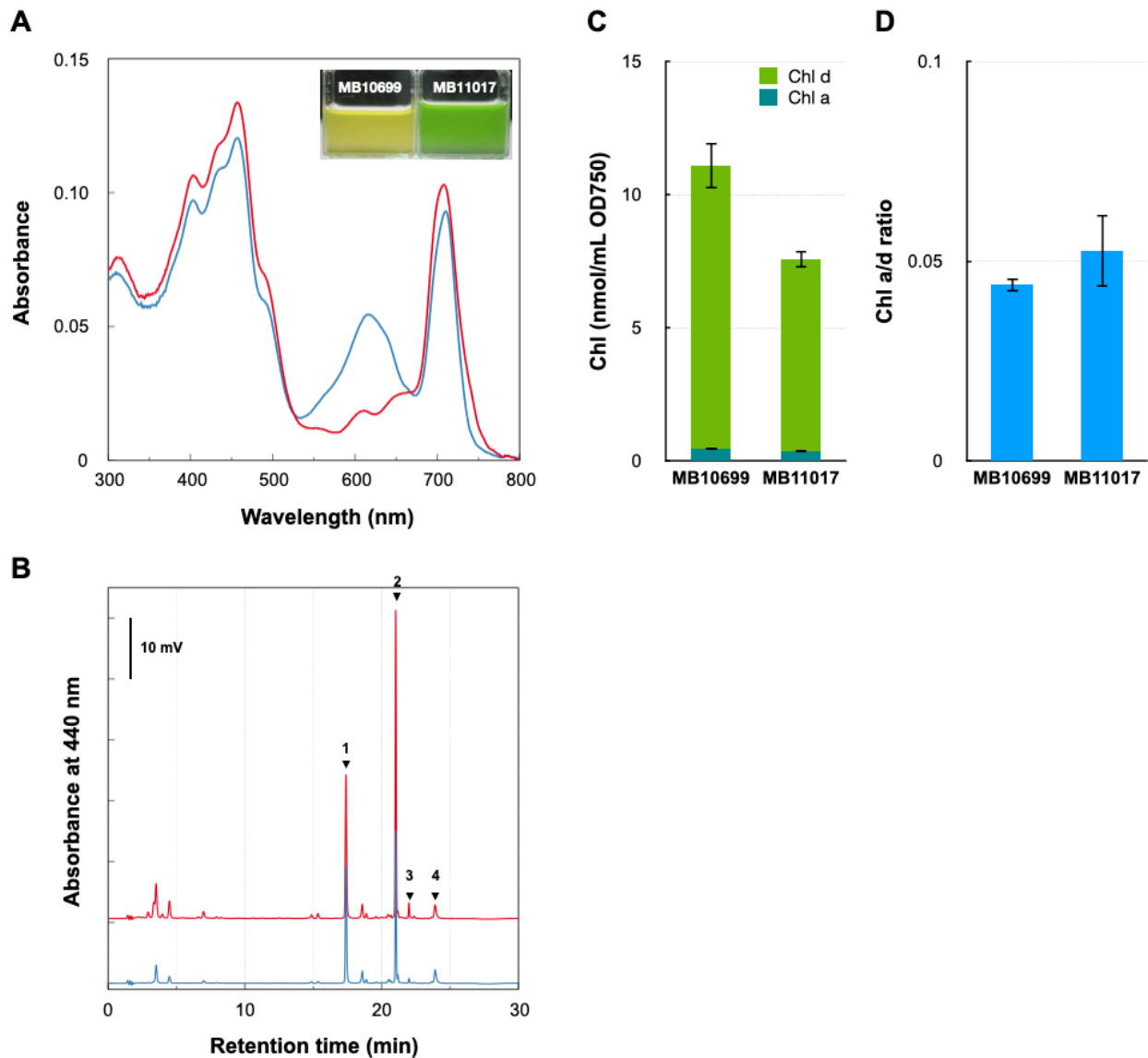
Figure 1. Comparison of cellular pigments between two *Acaryochloris* strains. (**A**) In vivo cellular spectra of *A. marina* MBIC10699 (red) and *A. marina* MBIC11017 (blue). Inset: picture of each culture. (**B**) HPLC profile of total pigments extracted with methanol from cells of *A. marina* MBIC10699 (red) and *A. marina* MBIC11017 (blue) monitored by absorption at 440 nm. Each arrow indicates zeaxanthin (1), Chl *d* (2), Chl *a* (3), and α-carotene (4), respectively. (**C**) The amounts of Chl *a* (dark green) and Chl *d* (green) were measured in two *Acaryochloris* strains. (**D**) The ratio of cellular Chl *a/d* amounts in the two *Acaryochloris* strains.
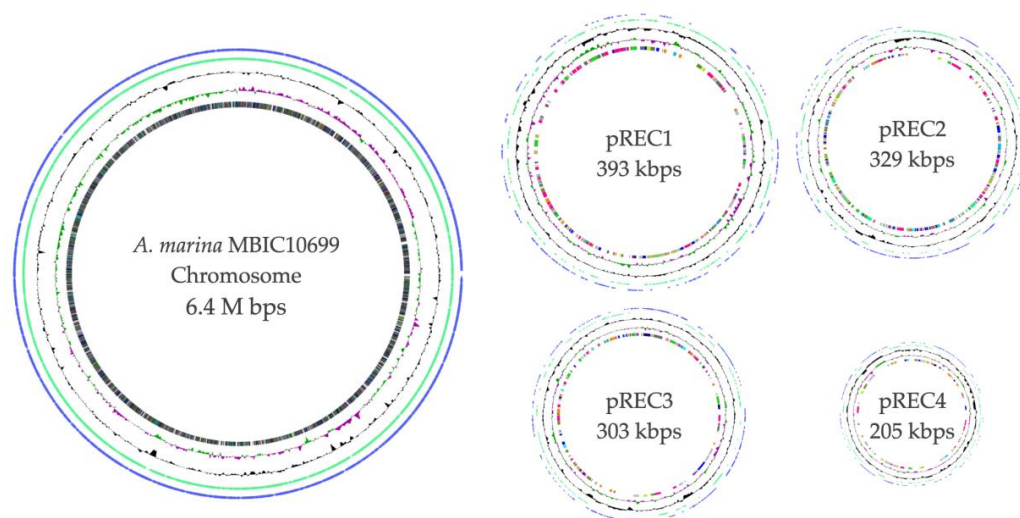
Figure 2. Genome structure of *A. marina* MBIC10699. The different rings represent, from outside to inside, all CDS (counterclockwise: blue, clockwise: green), deviation from the average GC content, GC skew, and CDS with colors corresponding to COG categories. For visualization, all plasmids are represented on a 10-fold scale relative to the chromosome.

Table 1. General features of the genome of *A. marina* MBIC10699.

	Chr	pREC1	pREC2	pREC3	pREC4	Total
Total Sequence Length (bp):	6,415,507	393,608	329,949	303,490	205,174	7,647,728
Number of Sequences:	1	1	1	1	1	5
Gap Ratio (%):	0.0	0.0	0.0	0.0	0.0	0.0
GC content (%):	47.3	45.1	46.7	46.1	43.2	47.0
Number of CDSs:	5674	366	283	280	193	6813
Coding Ratio (%):	83.4	76.0	75.3	78.1	68.7	82.3
Number of rRNAs:	6	0	0	0	0	6
Number of tRNAs:	72	0	0	0	0	72
Number of CRISPRs:	1	2	0	0	0	3

Dot plot analysis of MB10699 and MB11017 genomes revealed global sequence identity between their Chrs (Figure 4). In contrast, there was low homology and large divergence among plasmid sequences. In support of this, an examination of the genes annotated by KEGG in these two *Acaryochloris* strains showed that the composition of the genes encoded by Chr was almost identical, with MB10699 and MB11017 specific to each with only 0.6% (15) and 0.75% (12), respectively (Figure 5, Tables S2 and S3). In contrast, for plasmid-encoded genes, 25% (63) of the genes in MB10699 and 43% (146) of the genes in MB11017 were unique, with no other homologous genes in the other (Tables S4 and S5). There are four plasmids in MB10699 (pREC1-4) and nine in MB11017 (pREB1-9). Focusing on the homology of plasmids between the two strains, significant homology was observed only between pREC1 (MB10699) and pREB2 (MB11017) (Figure 4). In fact, characteristic metabolic genes such as genes for uridine kinase and aliphatic amidase are conserved in these plasmids, suggesting that these two plasmids have derived from the same origin. Additionally, a region of pREC2 showed significant homology to the sequence within pREB6. Although these plasmids may share the same origin, their size was quite different, 329 kb (pREC2) and 172 kb (pREB6), suggesting that many more genes have been acquired in pREC2 than in pREB6. No other combinations showed significant homology among the other plasmids.
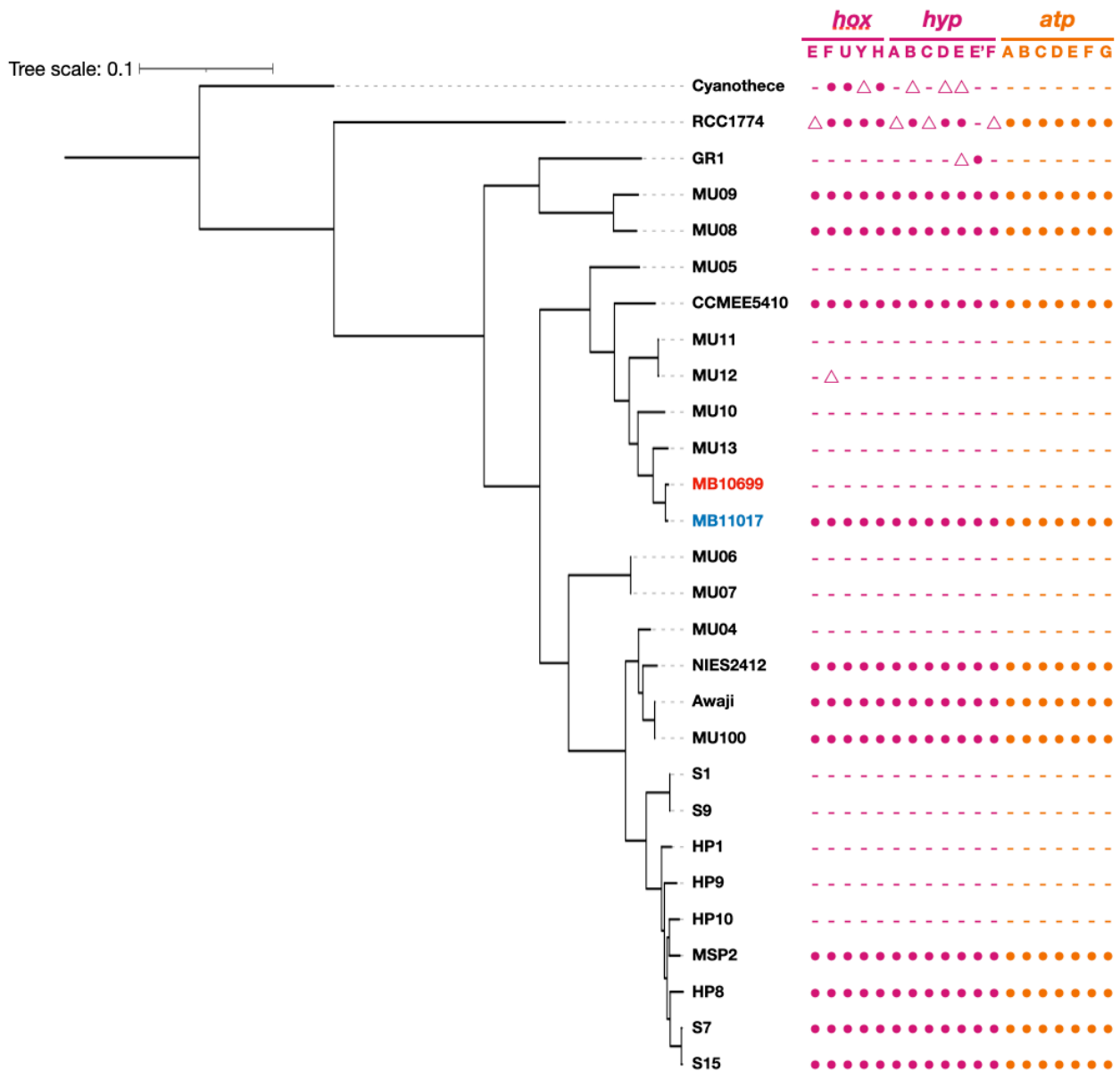
Figure 3. Phylogenic tree of *Acaryochloris* species. A maximum likelihood phylogenic tree of *Acaryochloris* strains was constructed from genome-wide concatenated protein sequences from single copy orthologous genes (*n* = 897). The tree was outgroup-rooted with *Cyanothece* sp. PCC7425 (Cyanothece). Another sister group of *Acaryochloris* strains, *Acaryochloris thomasi* RCC1774 (RCC1774), which biosynthesizes Chl *b* instead of Chl *d*, was used along with previously reported *Acaryochloris* strains. *A. marina* MBIC10699 and MBIC11017 are highlighted in red and blue, respectively. The symbols indicate the presence or absence of genes with significant homology to *hoxEFUYH/hypABCDEE'F* (purple) and *atpABCDEFG* (orange) genes encoded in pREB4 of MB11017 in each strain. Circles indicate the conservation of genes with significant homology (e-value < 1×10^{-100}), and triangles indicate genes that show some degree of homology with e-values between 1×10^{-100} and 1×10^{-50}.

It has been reported that plasmids contain mobilization factors such as *oriT* and relaxases as factors involved in their mobility from cell to other cells via conjugation [25,26]. Among these factors, it has been proposed that plasmids can be classified based on the sequences of relaxases, which are essential for mobility. The relaxase gene was found one by one on all four pREC plasmids in MB10699, and on all pREB plasmids except pREB9 in MB11017. Since the relaxase gene is a unique factor for each plasmid, the

correspondence of the plasmids between the two *Acaryochloris* strains was analyzed based on the homology of this relaxase gene, *mobF* (Figure 6). Phylogenetic analysis using the amino acid sequences of MobF revealed that they were classified into two major clades, each of which further branched into two subgroups. The phylogenetic relationship of relaxases suggested that pREC1 is of the same origin as pREB2, pREC2 as pREB6, pREC3 as pREB4,5, and pREC4 as pREB7,8, respectively. In fact, dot plot analysis and ANI comparisons indicated high similarities in the pairs of pREC1/pREB2 and pREC2/pREB6; thus, this plasmid classification based on phylogenetic analysis of relaxases is considered reliable. From these results, the pREC1/pREB2 pair and the pREC2/pREB6 pair are likely derived from common ancestral plasmids, respectively, and similarly the pair of pREC3/pREB4(5) and pREC4/pREB7(8) are also assumed to be derived from the same origins. There is no MB10699 plasmid in the clade corresponding to pREB1 and pREB3, suggesting that these plasmids were either acquired after the divergence of the two species or lost at MB10699 after the divergence. This suggests that HGT of PBP-related genes in pREB3 possibly occurred in plasmid units.
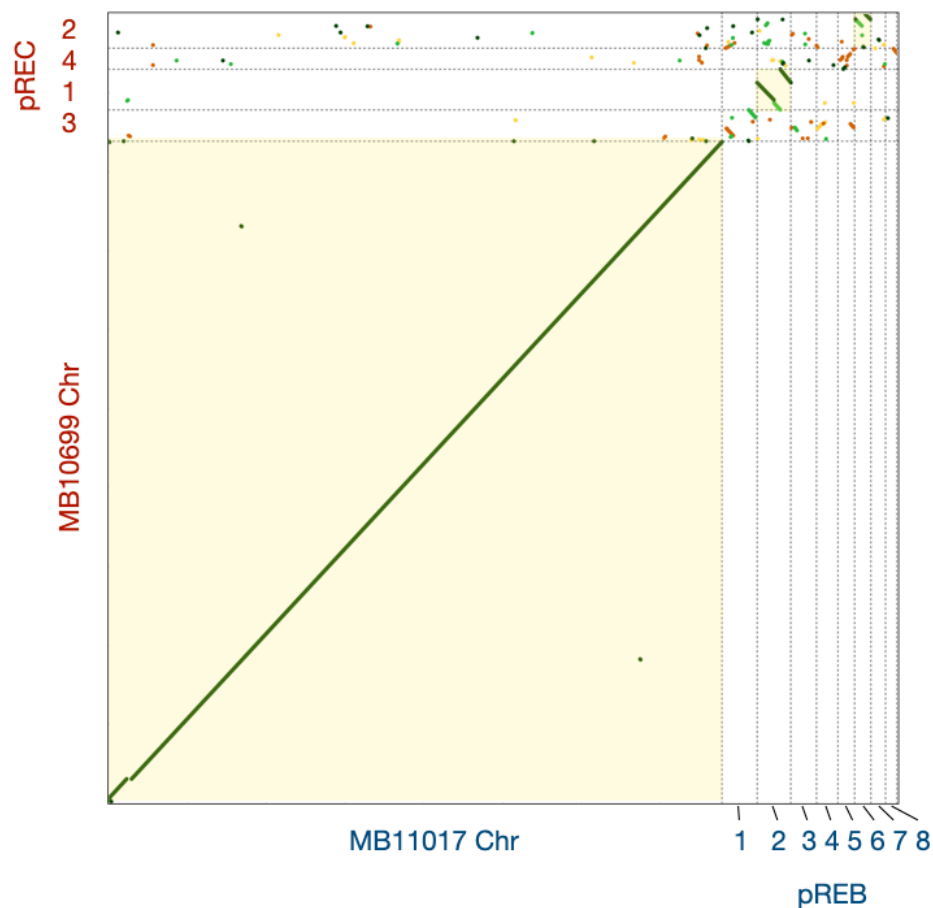
Figure 4. Dot plot analysis between *A. marina* MBIC10699 and MBIC11017. Dot plot alignment of *A. marina* MBIC10699 (vertical axis) versus *A. marina* MBIC11017 (horizontal axis). The color of each dot indicates the degree of identity, with 75% or more represented by green, 75–50% by light green, 50–25% by brown, and 25% or less by yellow, respectively. Matrices of combinations in which significant homology was detected are highlighted in light yellow.
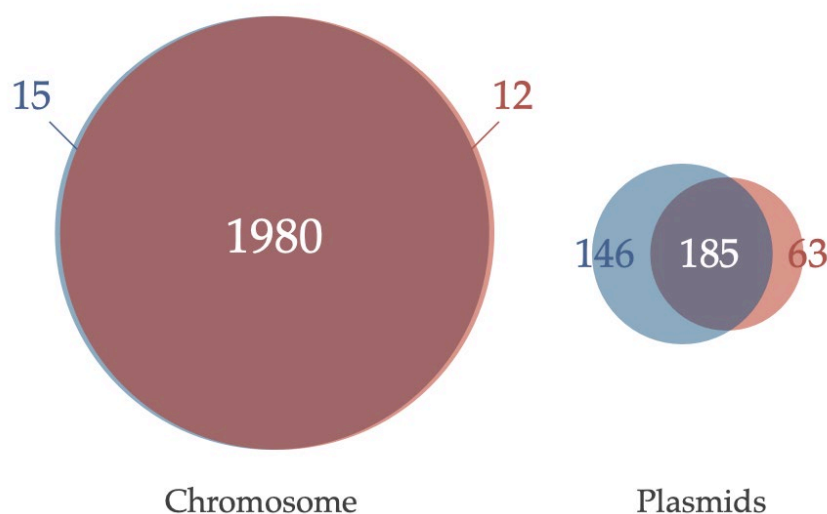
Figure 5. Common and unique genes in the two *Acaryochloris* strains. Euler diagram showing the common genes between *A. marina* MBIC10699 (red) and *A. marina* MBIC11017 (blue) for each chromosome and plasmids, respectively. Genes annotated by KEGG with the same ID were considered common. The numbers indicate the number of genes in the fraction.
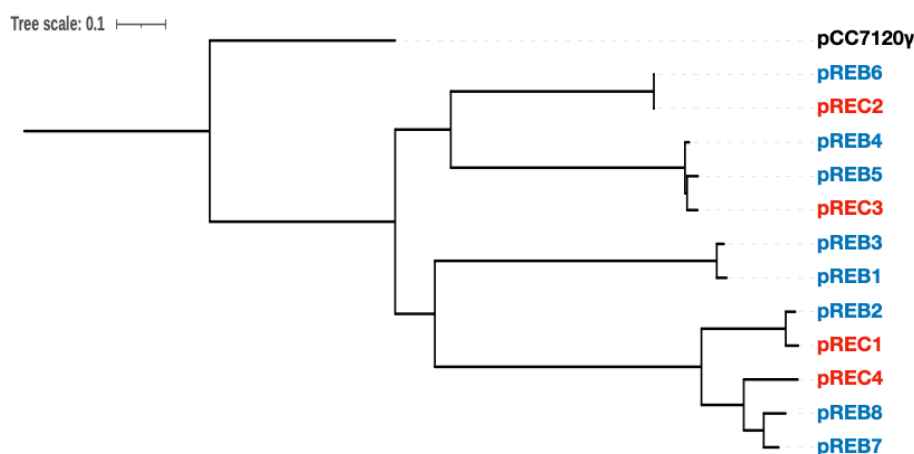
Figure 6. Phylogenetic relationship of each plasmid based on relaxase gene homology. The phylogenetic relationship of each plasmid was classified on the basis of the amino acid sequence of the relaxase (MobF) protein encoded in each plasmid. The plasmids shown in red are from *A. marina* MBIC10699, and the plasmids shown in blue are from *A. marina* MBIC11017. MobF encoded by pCC7120γ (black), the plasmid from *Anabaena* sp. strain PCC7120, was used as the outgroup.

The high degree of conservation of Chr sequences in these two strains and the high diversity among the plasmids suggest that various mutations in the plasmids have played a major role in the divergence of these two strains. Some plasmids that appear to be derived from the same origin show high homology while others do not, suggesting that the loss and gain of genes on the plasmids have not occurred evenly and that mutation frequency is different among the plasmids. Indeed, the presence or absence of PBP, the most characteristic phenotypic difference between MB10699 and MB11017, is determined by a group of genes accumulated on pREB3. Thus, even among closely related *Acaryochloris* strains, there is a high variety of gene contents and high flexibility of genetic variation on the plasmids, suggesting that *Acaryochloris* has taken advantage of this plasmid gene extensibility to adapt to various environments.

3.3. Comparative Analysis of Genes Conserved in Plasmids between Two Acaryochloris Species

To look at plasmid divergence in detail, we annotated the genes encoded in the plasmids of the two *Acaryochloris* strains with KEGG and selected genes unique for the respective species (Tables S4 and S5). A total of 63 specific genes were found in the MB10699 plasmids, 27 of which were not present in MB11017, including its Chr. Among the 63 genes, genes involved in sulfur metabolism (phosphoadenosine phosphosulfate reductase, sulfate adenyltransferase, and sulfur dioxygenase), metal ion transporters, and glycolysis (α-amylase, α-glucosidase, glycoside/pentoside: cation symporter) were found. In contrast, 146 genes unique to MB11017 plasmids were found, of which 60 were unique to MB11017 that were not present in MB10699 including its Chr. Genes specific to MB11017 plasmids include PBP-related genes, another set of ATP synthase, and bidirectional hydrogenase-related genes, as well as pyruvate-ferredoxin/flavodoxin oxidoreductase, acetyl-CoA synthetase, NAD(P)H-quinone oxidoreductase, fatty acid CoA ligase, and other carbon metabolism genes. We also found genes involved in the biosynthesis of porphyrins, such as divinyl chlorophyllide reductase and *hemL* (GSA) in MB11017 plasmids. To determine how these plasmid-specific genes were acquired, homologous regions were aligned for pREC1 and pREB2, considered having a common plasmid origin (Figure S1). The ORFs characteristic of pREC1 were found to form several clusters on the plasmid, with transposase genes located in the vicinity. This suggests that gene gain or loss on plasmids occur in clusters containing multiple genes. Alternatively, there may be 'hot-spot' loci where gene gain or loss is prone to occur by transposases, and gene gain or loss events may have been concentrated around these positions.

3.4. PBP-Related Genes

In MB11017, *cpcA/B* (α and β subunits of phycocyanin), *cpcC/D/G* (linker polypeptides that connect discs composed of phycocyanin heterodimers), *hemH* (ferrochelatase), *ho* (heme oxygenase), *pcyA* (phycocyanobilin oxidoreductase in the biosynthesis of phycocyanobilin), and *cpcE/F* (phycocyanobilin lyases that bind phycocyanobilin to the CpcA/B apo-protein) are encoded in the plasmid pREB3. Four paralogs of *cpcA/B*, three paralogs of *cpcG*, and two each of *cpcC/D*, *hemH*, and *ho* form multiple operons within pREB3. Cellular spectra of MB10699 indicated that it does not have PBPs as an antenna complex (Figure 1). As mentioned above, MB10699 does not harbor a plasmid corresponding to pREB3, and many of the PBP-related genes were not found in the entire genome. In MB11017, *hemH*, *ho*, and *pcyA*, which are involved in phycocyanobilin biosynthesis, are present as four, three, and two paralogous genes, respectively, and are encoded in both Chr and pREB3 (Figure 7). Although no paralogs of these three genes were found in the plasmids of MB10699, we found an equal number of paralogs of each gene in the Chr as in MB11017 Chr. Phycocyanobilin functions as a chromophore of the photoreceptor cyanobacteriochromes, in addition to PBPs [27,28]. Indeed, two paralogs of phycocyanobilin:ferredoxin oxidoreductases (PcyA) encoded by Chr and pREB3 show different biochemical properties, suggesting that these PcyAs are specialized in providing chromophores to apo-proteins of cyanobacteriochrome and phycocyanin, respectively [29]. PcyA conserved in Chr of MB10699 shows an identical amino acid sequence to that conserved in Chr of MB11017, suggesting that Chr-encoding PcyA is specialized to provide chromophores to cyanobacteriochrome. In MB10699, only *apcB*, which encodes β-subunit of allophycocyanin, is conserved in Chr, and this gene is similarly conserved in Chr in MB11017, while it has not yet been examined whether this *apcB* is actually expressed in MB11017. Since *apcB* is conserved even in MB10699, this *apcB* may serve some different functions rather than the allophycocyanin subunit of PBPs.
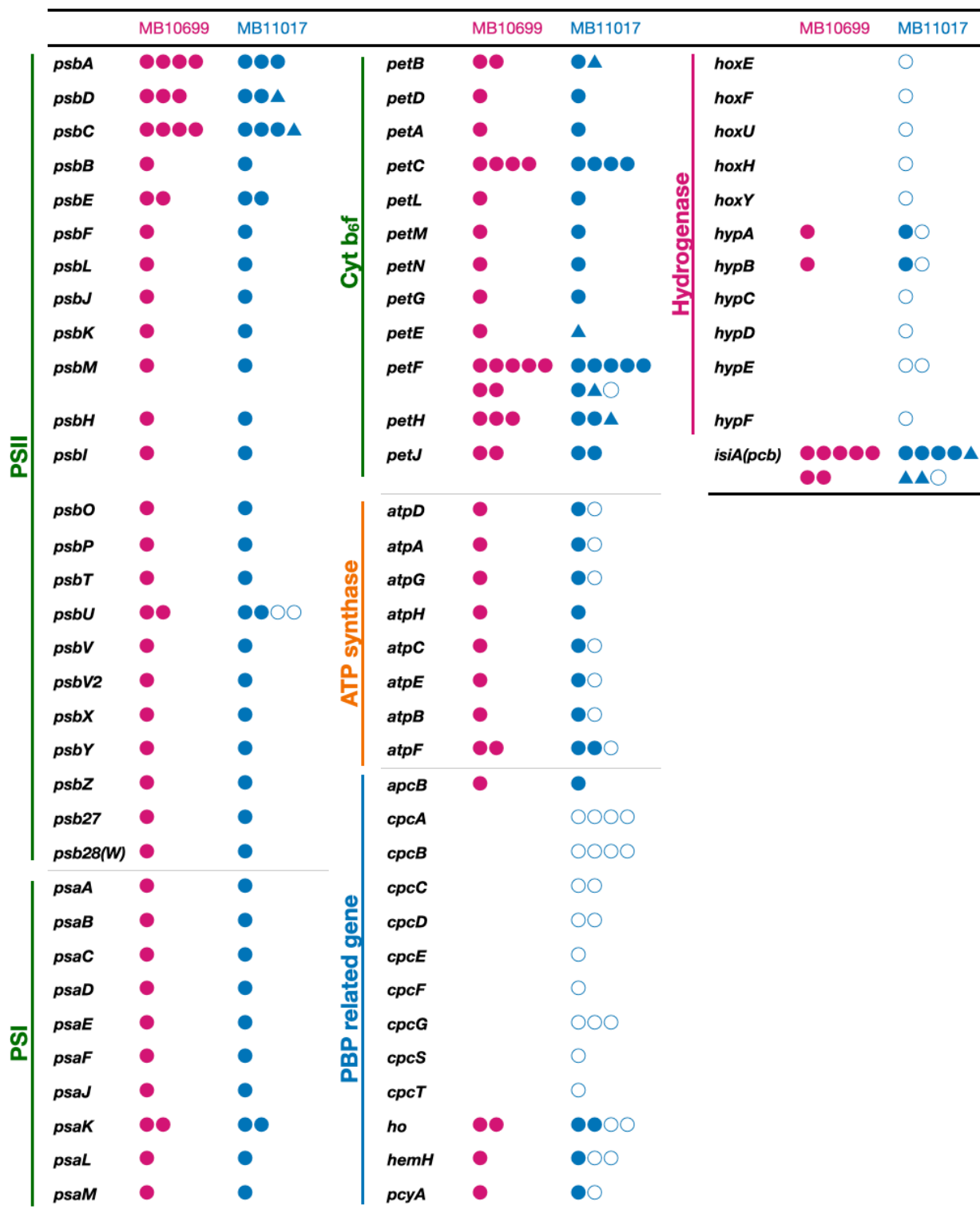
Figure 7. Comparison of the number of gene paralogs related to photosynthesis and electron transfer systems. The number of paralogs of genes related to photosystem I (*psa*), photosystem II (*psb*), cytochrome b_6f and electron transfer (*pet*), ATP synthase (*atp*), PBP, and hydrogenase are indicated by the number of red (*A. marina* MBIC10699) and blue (*A. marina* MBIC11017) symbols. Circles indicate completely identical amino acid sequences between the two strains and triangles indicate that the homology is not 100% (shown in *A. marina* MBIC11017). Filled circles indicate genes encoded by chromosome and empty circles indicate plasmid-encoded genes.

3.5. Genes Related to Alternative ATP Synthase and Hydrogenase

In MB11017, eight genes for *atpABCDEFGH*, subunits of F-type ATP synthase, are present in Chr, and these paralogs of the seven genes except *atpH* are also conserved in pREB4 (Figure 7). It is not known whether these two sets of ATP synthase share the Chr-coded δ subunit (AtpH) or whether the pREB4-encoded ATP synthase functions in a form that has lost its δ subunit (N-type ATP synthase). When functioning as an N-type ATP synthase, it is thought to transport Na^+ instead of H^+. Other marine cyanobacteria such as *Cyanothece* also conserved two sets of ATP synthase genes lacking one δ subunit, suggesting that this is a conserved form of the gene with some significance. In addition to ATP synthase, *hoxEFUYH* genes, which encode hydrogenase subunits, and *hypABCDEF* genes, which are involved in hydrogenase maturation, are conserved in pREB4 of MB11017 (only *hypAB* has their paralogs in Chr) (Figure 7). *hoxEFUYH* genes encode a bidirectional hydrogenase, which catalyzes the conversion of protons and molecular hydrogen in both directions. These genes for plasmid-encoded ATP synthase (alternative ATP synthase) and *hox*-hydrogenase are not conserved in MB10699. A total of 26 *Acaryochloris* strains reported to date were tested for the presence of the alternative ATP synthase and *hox*-hydrogenase genes (Figure 3). Since these gene sets were also found in several other *Acaryochloris* strains, no correlation was found between the acquisition of PBP and the conservation of this alternative ATP synthase and *hox*-hydrogenase. Interestingly, the conservation of both gene sets for the alternative ATP synthases and *hox*-hydrogenase was perfectly consistent in *Acaryochloris* species without any exceptions. Given the high correlation between the alternative ATP synthase and *hox*-hydrogenase conservation, the gain or loss of these genes is expected to occur synchronously. The conservation of these genes is inconsistent with the phylogenetic relationship of *Acaryochloris* (Figure 3), and species that conserved the alternative ATP synthase and *hox*-hydrogenase genes appear mosaic in all clades. This suggests that both gene sets for the alternative ATP synthase and *hox*-hydrogenase were retained at origin before these *Acaryochloris* species diverged, but that some species lost both genes after divergence. The complete coexistence of both of these genes suggests that they are functionally related. It has been suggested that bidirectional hydrogenase can release excess reducing power in fermentation by proton reduction [30]. Then, the energy stored as molecular hydrogen is extracted by the reverse reaction to produce reducing power (NADPH) when needed, and the resulting proton gradient drives ATP synthase to generate ATP. Since this combination of bidirectional hydrogenase and ATP synthase enables the generation of NADPH and ATP using molecular hydrogen, it may play an important role in conditions where both photosynthesis and respiratory electron transfer systems do not fully function, such as dark and anaerobic conditions.

3.6. Genes Related to Photosynthetic Reaction centers, Electron Transfer, and Chl-Binding Antenna Complex

Comparative analysis was performed between MB10699 and MB11017 for genes involved in photosystems, antenna complexes, and photosynthetic electron transfer systems that are potentially affected by the presence or absence of PBP (Figure 7). All genes encoding subunits of PSI, PSII, and the cytochrome b_6f complex in MB11017 were also conserved in MB10699 and nearly all of their amino acid sequences were identical in these two *Acaryochloris* strains, although some of them were found with a different number of paralogs. While three *psbA* paralogs (AM1_0488, 2166, 2889) are found in MB11017, four *psbA* paralogs (AM10699_32560, 54700, 47980, 47970) are found in MB10699, of which AM10699_32560 and AM1_0488 are identical, and the remaining three *psbA* paralogs were identical and perfectly matched with AM1_2166/2889. Four paralogous genes of *psbU* gene are present in MB11017, two of which are conserved in its Chr and the other two in pREB4 and pREB7. However, in MB10699, only two *psbUs* (AM10699_07930, 17380) were found in Chr and no *psbU* paralogs in plasmids. Although PsbU plays a role in stabilizing energy transfer between PSII and PBPs [31], *psbU* gene was also conserved in MB10699. The amino acid sequences of two plasmid-encoded PsbUs (AM1_D0138 and G0114) are

significantly different from those of the Chr-encoded PsbUs (Figure S2), suggesting that these plasmid-encoded PsbUs are involved in the complex formation of PBP and PSII in MB11017. In MB10699, seven *petF* paralogs encoding ferredoxin are conserved in its Chr, and in MB10117, another *petF* paralog is found in the plasmid in addition to the seven *petF* paralogs in its Chr. The *petF* lost in MB10699 is conserved in pREB3 of MB11017, which is most likely the ferredoxin involved in PBP biosynthesis (PcyA). *Acaryochloris* retains a unique type of antenna complex Pcb that binds Chl *a/d*, encoded by *isiA* that shows significant homology to *psbC* (CP43) [21]. MB10699 has seven *isiA* paralogs, and eight in MB11017. One extra paralog in MB11017, AM1_C0105, is also encoded by pREB3.

Most of these photosynthesis-related genes were identical in the amino acid sequence between the two *Acaryochloris* strains; however, significant differences were found in the following genes. *petE*, which encodes plastocyanin, showed 95% identity between the two strains, and other genes such as *petF* (AM1_5442) and FNR (*petH*) (AM1_2942) had 99% identity with their counterparts in MB10699. Three *isiA* genes, AM1_3362, 3364, and 3366 of MB11017 are consecutively located on Chr and form an operon, and the corresponding, AM10699_02480, 02490, and 02500 of MB10699 form a similar operon on Chr. AM10699_02480 and the corresponding paralog AM1_3362 showed 88% identity, which is significantly lower than other paralog pairs. AM10699_02490 and 02450, corresponding to AM1_3364 and 3366 in the same operon, showed 99% identity with significant differences. MB10699 should employ Pcbs (IsiA) as the main photosynthetic antenna complex because of the absence of PBPs. Compared to MB11017, the number of *isiA* paralogs was rather reduced by one in MB10699, and there was significantly less sequence homology in the operon with three consecutive *isiA*s between the two species. These three *isiA*s may represent examples of unique evolution influenced by the presence or absence of PBP. Further studies are required to determine how the amino acid differences in these IsiAs have altered their function.

3.7. Availability of Carbon Sources in Two Acaryochloris Strains

Recently, MB10699 and MB11017 were reported to differ in their ability to use exogenous carbon sources [10]. For example, MB10699 was capable of mixotrophic growth with sucrose and glucose as carbon sources, whereas MB11017 was lethal in cultures with these carbon sources. Conversely, MB11017 tolerates glycerol as a carbon source, but MB10699 was lethal in cultures containing glycerol. We discuss these phenotypic differences through a comparison of their genomic information. Comparison of sucrose metabolism genes in these two *Acaryochloris* strains showed that there are equal numbers of paralogs for sucrose phosphate, sucrose synthase, invertase, amylosucrase, and α-glucosidase genes encoded by Chr, but another paralog of α-glucosidase is conserved in pREC2 of MB10699 (AM10699_61640). Furthermore, α-amylase and glycoside/pentoside: cation symporter, are also specifically conserved in the plasmids of MB10699, suggesting that these plasmid-encoded genes are responsible for the superior sucrose/glucose metabolism in MB10699 compared to MB11017. The effect of glycerol addition on cyanobacteria was reported in *Synechococcus* sp. PCC7002, where glycerol had a negative effect on growth in mutants with loss of plastoquinone pool regulation [32]. In MB11017, the genes for bidirectional hydrogenase and NAD(P)H-quinone oxidoreductase are specifically conserved in the plasmids, and these genes seem to be involved in the redox regulation of plastoquinone pool. Thus, the different genetic composition of their plasmids would be expected to contribute to phenotypic differences in the availability of carbon sources, supporting that the differences in plasmid-constitutive genes determine the phenotypic differences between these two *Acaryochloris* strains. It is necessary to consider the effects of these plasmid-coded genes in combination with transcriptome analysis in the future.

4. Conclusions

In this report, we sequenced the complete genome of *Acaryochloris marina* MBIC10699, a PBP-less *Acaryochloris*, and found that it is closely related to type-strain *A. marina*

MBIC11017. Chr sequence homology between these two *Acaryochloris* strains was extremely high, with no major inversions or duplications, and their genetic composition was largely identical. However, the plasmid composition and the genes encoded in the plasmids were significantly different, and the presence or absence of PBP was also explained by the genes encoded in the plasmids. Although amino acid sequences of most photosynthesis-related genes were identical between the two *Acaryochloris* strains, we found genes with significantly lower homology and genes that differed in the number of their paralogs, suggesting that these differences resulted from the presence or absence of PBPs. Furthermore, the differences in available carbon sources in the two *Acaryochloris* strains also appear to result from differences in the genes encoded in their plasmids, suggesting that most of the phenotypic differences of the two *Acaryochloris* strains are due to differences in the genes constitutive in the plasmids. Comparison of genomic information between these two closely related *Acaryochloris* strains reveals that *Acaryochloris* has acquired and altered traits through gene acquisition using a highly extensible plasmid. This finding provides the basis for a detailed analysis of how horizontal gene transfer of large gene clusters occurs in *Acaryochloris*.

References

1. Chen, M. Chlorophyll modifications and their spectral extension in oxygenic photosynthesis. *Annu. Rev. Biochem.* **2014**, *83*, 317–340. [CrossRef]
2. Dekker, J.P.; Boekema, E.J. Supramolecular organization of thylakoid membrane proteins in green plants. *Biochim. Biophys. Acta* **2005**, *1706*, 12–39. [CrossRef]
3. Miyashita, H.; Ikemoto, H.; Kurano, N.; Adachi, K.; Chihara, M.; Miyachi, S. Chlorophyll *d* as a major pigment. *Nature* **1996**, *383*, 402. [CrossRef]

4. Hu, Q.; Miyashita, H.; Iwasaki, I.; Kurano, N.; Miyachi, S.; Iwaki, M.; Itoh, S. A photosystem I reaction center driven by chlorophyll *d* in oxygenic photosynthesis. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 13319–13323. [CrossRef]

5. Tomo, T.; Okubo, T.; Akimoto, S.; Yokono, M.; Miyashita, H.; Tsuchiya, T.; Noguchi, T.; Mimuro, M. Identification of the special pair of photosystem II in a chlorophyll *d*-dominated cyanobacterium. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 7283–7288. [CrossRef] [PubMed]

6. Miller, S.R.; Abresch, H.E.; Baroch, J.J.; Fishman Miller, C.K.; Garber, A.I.; Oman, A.R.; Ulrich, N.J. Genomic and Functional Variation of the Chlorophyll *d*-Producing Cyanobacterium *Acaryochloris marina*. *Microorganisms* **2022**, *10*, 569. [CrossRef]

7. Swingley, W.D.; Chen, M.; Cheung, P.C.; Conrad, A.L.; Dejesa, L.C.; Hao, J.; Honchak, B.M.; Karbach, L.E.; Kurdoglu, A.; Lahiri, S.; et al. Niche adaptation and genome expansion in the chlorophyll *d*-producing cyanobacterium *Acaryochloris marina*. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 2005–2010. [CrossRef] [PubMed]

8. Ulrich, N.J.; Uchida, H.; Kanesaki, Y.; Hirose, E.; Murakami, A.; Miller, S.R. Reacquisition of light-harvesting genes in a marine cyanobacterium confers a broader solar niche. *Curr. Biol.* **2021**, *31*, 1539–1546.e1534. [CrossRef] [PubMed]

9. Pfreundt, U.; Stal, L.J.; Voß, B.; Hess, W.R. Dinitrogen fixation in a unicellular chlorophyll *d*-containing cyanobacterium. *ISME J.* **2012**, *6*, 1367–1377. [CrossRef] [PubMed]

10. Tsuzuki, Y.; Tsukatani, Y.; Yamakawa, H.; Itoh, S.; Fujita, Y.; Yamamoto, H. Effects of Light and Oxygen on Chlorophyll *d* Biosynthesis in a Marine Cyanobacterium *Acaryochloris marina*. *Plants* **2022**, *11*, 915. [CrossRef]

11. Zapata, M.; Rodríguez, F.; Garrido, J.L. Separation of chlorophylls and carotenoids from marine phytoplankton: A new HPLC method using a reversed phase C8 column and pyridine-containing mobile phases. *Mar. Ecol. Prog. Ser.* **2000**, *195*, 29–45. [CrossRef]

12. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [CrossRef] [PubMed]

13. De Coster, W.; D'Hert, S.; Schultz, D.T.; Cruts, M.; Van Broeckhoven, C. NanoPack: Visualizing and processing long-read sequencing data. *Bioinformatics* **2018**, *34*, 2666–2669. [CrossRef] [PubMed]

14. Kolmogorov, M.; Yuan, J.; Lin, Y.; Pevzner, P.A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **2019**, *37*, 540–546. [CrossRef]

15. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K.; et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **2014**, *9*, e112963. [CrossRef] [PubMed]

16. Tanizawa, Y.; Fujisawa, T.; Nakamura, Y. DFAST: A flexible prokaryotic genome annotation pipeline for faster genome publication. *Bioinformatics* **2018**, *34*, 1037–1039. [CrossRef] [PubMed]

17. Lee, I.; Ouk Kim, Y.; Park, S.C.; Chun, J. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. *Int. J. Syst. Evol. Microbiol.* **2016**, *66*, 1100–1103. [CrossRef] [PubMed]

18. Cabanettes, F.; Klopp, C. D-GENIES: Dot plot large genomes in an interactive, efficient and simple way. *PEERJ* **2018**, *6*, e4958. [CrossRef]

19. Page, A.J.; Cummins, C.A.; Hunt, M.; Wong, V.K.; Reuter, S.; Holden, M.T.; Fookes, M.; Falush, D.; Keane, J.A.; Parkhill, J. Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **2015**, *31*, 3691–3693. [CrossRef]

20. Miyashita, H.; Adachi, K.; Kurano, N.; Ikemot, H.; Chihara, M.; Miyach, S. Pigment Composition of a Novel Oxygenic Photosynthetic Prokaryote Containing Chlorophyll *d* as the Major Chlorophyll. *Plant Cell Physiol.* **1997**, *38*, 274–281. [CrossRef]

21. Chen, M.; Quinnell, R.G.; Larkum, A.W. The major light-harvesting pigment protein of *Acaryochloris marina*. *FEBS Lett.* **2002**, *514*, 149–152. [CrossRef]

22. Li, Z.K.; Yin, Y.C.; Zhang, L.D.; Zhang, Z.C.; Dai, G.Z.; Chen, M.; Qiu, B.S. The identification of IsiA proteins binding chlorophyll *d* in the cyanobacterium *Acaryochloris marina*. *Photosynth. Res.* **2018**, *135*, 165–175. [CrossRef] [PubMed]

23. Jain, C.; Rodriguez, R.L.; Phillippy, A.M.; Konstantinidis, K.T.; Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **2018**, *9*, 5114. [CrossRef] [PubMed]

24. Murray, C.S.; Gao, Y.; Wu, M. Re-evaluating the evidence for a universal genetic boundary among microbial species. *Nat. Commun.* **2021**, *12*, 4059. [CrossRef] [PubMed]

25. Garcillán-Barcia, M.P.; Francia, M.V.; de la Cruz, F. The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiol. Rev.* **2009**, *33*, 657–687. [CrossRef]

26. Smillie, C.; Garcillán-Barcia, M.P.; Francia, M.V.; Rocha, E.P.; de la Cruz, F. Mobility of plasmids. *Microbiol. Mol. Biol. Rev.* **2010**, *74*, 434–452. [CrossRef] [PubMed]

27. Ikeuchi, M.; Ishizuka, T. Cyanobacteriochromes: A new superfamily of tetrapyrrole-binding photoreceptors in cyanobacteria. *Photochem. Photobiol. Sci.* **2008**, *7*, 1159–1167. [CrossRef] [PubMed]

28. Narikawa, R.; Fushimi, K.; Ni Ni, W.; Ikeuchi, M. Red-shifted red/green-type cyanobacteriochrome AM1_1870g3 from the chlorophyll *d*-bearing cyanobacterium *Acaryochloris marina*. *Biochem. Biophys. Res. Commun.* **2015**, *461*, 390–395. [CrossRef]

29. Miyake, K.; Fushimi, K.; Kashimoto, T.; Maeda, K.; Ni Ni, W.; Kimura, H.; Sugishima, M.; Ikeuchi, M.; Narikawa, R. Functional diversification of two bilin reductases for light perception and harvesting in unique cyanobacterium *Acaryochloris marina* MBIC 11017. *FEBS J.* **2020**, *287*, 4016–4031. [CrossRef]

30. Carrieri, D.; Wawrousek, K.; Eckert, C.; Yu, J.; Maness, P.C. The role of the bidirectional hydrogenase in cyanobacteria. *Bioresour. Technol.* **2011**, *102*, 8368–8377. [CrossRef]

31. Veerman, J.; Bentley, F.K.; Eaton-Rye, J.J.; Mullineaux, C.W.; Vasil'ev, S.; Bruce, D. The PsbU subunit of photosystem II stabilizes energy transfer and primary photochemistry in the phycobilisome-photosystem II assembly of *Synechocystis* sp. PCC 6803. *Biochemistry* **2005**, *44*, 16939–16948. [CrossRef] [PubMed]
32. Xu, D.; Liu, X.; Zhao, J. FesM, a membrane iron-sulfur protein, is required for cyclic electron flow around photosystem I and photoheterotrophic growth of the cyanobacterium *Synechococcus* sp. PCC 7002. *Plant Physiol.* **2005**, *138*, 1586–1595. [CrossRef] [PubMed]

Article

Genomic and Functional Variation of the Chlorophyll *d*-Producing Cyanobacterium *Acaryochloris marina*

Scott R. Miller *[ORCID], Heidi E. Abresch, Jacob J. Baroch, Caleb K. Fishman Miller, Arkadiy I. Garber, Andrew R. Oman and Nikea J. Ulrich

Division of Biological Sciences, University of Montana, Missoula, MT 59812, USA;
heidi.abresch@umconnect.umt.edu (H.E.A.); jake.baroch@umconnect.umt.edu (J.J.B.);
ckfmmt@gmail.com (C.K.F.M.); rkdgarber@gmail.com (A.I.G.); andy.oman@umconnect.umt.edu (A.R.O.);
nikea.ulrich@umconnect.umt.edu (N.J.U.)
* Correspondence: scott.miller@umontana.edu

Abstract: The Chlorophyll *d*-producing cyanobacterium *Acaryochloris marina* is widely distributed in marine environments enriched in far-red light, but our understanding of its genomic and functional diversity is limited. Here, we take an integrative approach to investigate *A. marina* diversity for 37 strains, which includes twelve newly isolated strains from previously unsampled locations in Europe and the Pacific Northwest of North America. A genome-wide phylogeny revealed both that closely related *A. marina* have migrated within geographic regions and that distantly related *A. marina* lineages can co-occur. The distribution of traits mapped onto the phylogeny provided evidence of a dynamic evolutionary history of gene gain and loss during *A. marina* diversification. Ancestral genes that were differentially retained or lost by strains include plasmid-encoded sodium-transporting ATPase and bidirectional NiFe-hydrogenase genes that may be involved in salt tolerance and redox balance under fermentative conditions, respectively. The acquisition of genes by horizontal transfer has also played an important role in the evolution of new functions, such as nitrogen fixation. Together, our results resolve examples in which genome content and ecotypic variation for nutrient metabolism and environmental tolerance have diversified during the evolutionary history of this unusual photosynthetic bacterium.

Keywords: *Acaryochloris*; chlorophyll; genomics; cyanobacteria; far-red photosynthesis; plasmid; horizontal gene transfer

1. Introduction

The marine cyanobacterium *Acaryochloris marina* is unique in its use of Chlorophyll *d* (Chl *d*) as its major photosynthetic pigment [1]. Although Chl *d* is nearly identical in structure to the ubiquitous Chl *a* of other cyanobacteria, algae, and plants, this pigment absorbs strongly at far-red wavelengths that are inaccessible to most oxygenic phototrophs. Because these wavelengths attenuate rapidly in water, *A. marina* appears to be restricted to shallow environments enriched in filtered far-red light, and its abundance has been observed to decrease with depth [2]. Nonetheless, Chl *d* and its derivatives can be abundant in marine coastal ecosystems [3], and *A. marina* is widely distributed in temperate and tropical saline environments, often in association with red algae or animals [1,4–8]. These cyanobacteria have also been confirmed, through either laboratory cultivation or environmental DNA sequencing, to occur in marine stromatolites [9], a saline lake epilithic biofilm [10,11], a microbial mat from a high-elevation brackish lake [12], and the aquatic plant rhizosphere of a heavily canopied freshwater stream [13].

Despite its broad distribution, the genomic diversity of *A. marina* remains poorly understood. This is because, until recently [8], genome data were available for only two laboratory strains [14,15]: the type strain MBIC110017 and strain CCMEE 5410, which was

isolated from the Salton Sea, California. Similarly, we are largely ignorant regarding the extent of functional variation for traits that are expected to be important for *A. marina* life history, including nitrogen metabolism and salinity tolerance. Improving our understanding of this variation is essential for inferring the different ways that this bacterium may contribute to ecosystem processes in far-red enriched environments. We investigated this issue for our collection of 37 *A. marina* laboratory strains, including twelve newly isolated strains (Table S1). The collection consists of epiphytes of red algae, strains associated with tunicates (MBIC11017, MU11, MU12, and MU13), and an epilithic strain (CCMEE 5410). We identified roles for both plasmid maintenance and loss, along with horizontal gene transfer (HGT), in the origins of metabolic diversity during *A. marina* evolution. We also report ecotypic variation for phenotypic traits, including salinity tolerance and nitrogen sources, particularly nitrogen fixation. Finally, we identified differences among strains with respect to the presence of an intact CRISPR-Cas system, which may be selectively maintained when foreign DNA-induced mortality is high but otherwise lost.

2. Materials and Methods

2.1. Laboratory Strain Isolation

Twelve new laboratory strains of *A. marina* were isolated as part of this study from the following sample collections (Table S1): red algae on intertidal rocks at Hug Point State Park, Oregon (28 June 2017); red algae on a low tide sand flat at Wreck Beach, Vancouver, British Columbia (7 August 2018); red algae on intertidal rocks at Praia de Carcavelos, Portugal (24 October 2018); red algae from dock tires at Friday Harbor Laboratories, Washington (5 January 2019); and a red alga on an intertidal rock at Queroianella, Italy (30 December 2019). Strains were isolated from these environmental samples as previously described [8] and are available upon request. Briefly, algal samples were incubated at 20 °C in screw-cap vials containing 15 mL of IOBG-11 medium under low far-red light (~1.25 μmol photons m^{-2} s^{-1} from LED diodes with 710 nm peak emission). Clonal laboratory isolates were obtained from successful enrichments through repeated streaks and transfers on IOBG-11 agar plates.

2.2. Genomics

DNA was extracted using the DNeasy PowerBiofilm DNA extraction kit (Qiagen) following manufacturer instructions. Sample libraries were prepared and sequenced on an Illumina NextSeq 550 platform (150-bp paired-end) at the University of Pittsburgh Microbial Genome Sequencing Center. Sequence reads were trimmed of trailing low-quality bases and filtered based on read length and sequence quality with Trimmomatic version 0.36 [16], and draft genome assemblies were obtained with SPAdes version 3.12.0 [17] using manually optimized parameter settings to maximize the N50. Following the removal of contaminant contigs with Bandage version 0.8.1 [18] and Kraken 2 [19], assemblies were annotated by RAST [20]. Genome completeness was estimated with CheckM version 1.0.18 [21] and BUSCO version 4.1.2 [22]. Genome sequence data for strains described in this study are available at NCBI BioProject PRJNA649288.

2.3. Phylogenetics

A genome-wide phylogeny was reconstructed for the *A. marina* strains *A. thomasi* RCC1774 (NCBI accession NZ_PQWO00000000) and *Cyanothece* sp. PCC 7425 (NCBI accession GCA_000022045.1). A total of 1369 single copy groups of orthologous CDS were identified by OrthoFinder v2.2.7 [23] to create a concatenated alignment of protein sequences. We constructed a maximum likelihood tree with 1000 ultrafast bootstrap replicates [24] using IQ-TREE version 2.0 [25] according to the JTT+F+R5 model of sequence evolution selected by AIC in ModelFinder [26]. Additionally, a 16S rRNA gene phylogeny was reconstructed by maximum likelihood according to the TIM3+F+R2 model. For the *nif* dataset, maximum likelihood trees were reconstructed for *nifHDK* and 16S rRNA genes

with a GTR+I+G model and 1000 bootstrap replicates, and a SH test [27] was used to test for topological congruence of trees.

2.4. Genomics and Bioinformatics

To identify *A. marina* core genes, we performed two local BLAST [28] approaches for each genome: (1) a blastn search of each CDS (protein coding sequence) in the *A. marina* reference strain MBIC11017 against each target genome assembly; and (2) a tblastn search of each protein sequence in a genome against the reference strain MBIC11017 genome. For (2), hits with <50% overlap between query and target and/or <50% sequence identity were removed. We used an E-value cut-off of 0.01 for both approaches. For each strain's genome, BLAST outputs were merged, and a nonredundant gene set was then obtained with a custom Python script. Gene sets for all genomes were then combined, and a custom Python script was used to count how many genomes possessed a given gene.

KEGG analysis was performed with eggNOG [29]. Of CDS sequences in the inferred *A. marina* core genome, 2642 out of 2757 met the criteria of a minimum hit E-value of 0.001, a minimum sequence identity of 40%, and minimum query and subject coverages of 20%. FeGenie was used with default settings to identify iron metabolism genes [30], and a dendrogram was generated with Pvlcust [31], which hierarchically clustered normalized gene counts by Euclidian distance. Putative siderophore synthesis gene clusters were confirmed using AntiSMASH [32]. CRISPR arrays were obtained from RAST, and CrisprFinder [33] was used to extract individual spacers, which were queried by BLAST against plasmid and virus databases with CRISPRTarget [34].

2.5. Nitrogen Fixation and Salt Tolerance Experiments

Stock cultures were grown in 125 mL flasks containing 75 mL of ASN-III medium (https://www.atcc.org, accessed on 20 July 2020), which contains 0.43 M NaCl. Flasks were incubated at 25 °C under a 12:12 h light:dark cycle with a mean light intensity of ~25 μmol photons $m^{-2} s^{-1}$ of cool white fluorescent light. For the salt tolerance experiment, each stock of growing cells was inoculated to an optical density at 750 nm (OD_{750}) value of 0.01 in duplicate 125 mL flasks for each experimental salinity treatment (75 mL of ASN-III medium modified to contain final NaCl concentrations of 0, 0.2, 0.43, 0.5, 0.62, 0.8, or 1.6 M). A similar approach was taken for the nitrogen fixation experiment. For the +N control treatment, stock cells were inoculated into fresh ASN-III; for the -N treatment, cells were inoculated into ASN-III-N medium (ASN-III without sodium nitrate added) after rinsing with ASN-III -N. For the -N rinse, cells were first pelleted in a 2 mL microfuge tube and the supernatant removed; then, 1 mL of ASN III -N media was added to the tube, the tube was vortexed, and the process repeated until cells had been rinsed three times. Experimental flasks were incubated at randomized positions in a growth chamber under the same environmental conditions as above. Every 48 h, the OD_{750} of a 2 mL homogenized culture sub-sample was measured for each flask with a Beckman Coulter DU 530 spectrophotometer (Indianapolis, IN, USA), and flasks were randomly re-positioned in the chamber. Growth was measured as the increase OD_{750} over time. Exponential growth rates were estimated for cultures that sustained growth for at least three population doublings. All statistical models were analyzed with JMP version 14.2 (SAS Institute Inc., Cary, NC, USA).

3. Results and Discussion

3.1. A. marina Phylogeny and Core Genome

We took a phylogenomics approach to develop a better understanding of *A. marina* genome variation and functional diversity. Our data set consisted of genome data for 37 *A. marina* laboratory strains from diverse geographic locations (Figure 1; Tables S1 and S2). Together with previously available data [8,14,15], this included genome assemblies for twelve newly isolated strains. Among these are the first *A. marina* strains from the Salish Sea (San Juan Island, WA, USA and Vancouver, BC, Canada), the eastern Atlantic Ocean

(Carcavelos, Portugal), and the Mediterranean Sea (Queroianella, Italy). Most are draft assemblies constructed from Illumina sequence data but are virtually complete (Table S2).
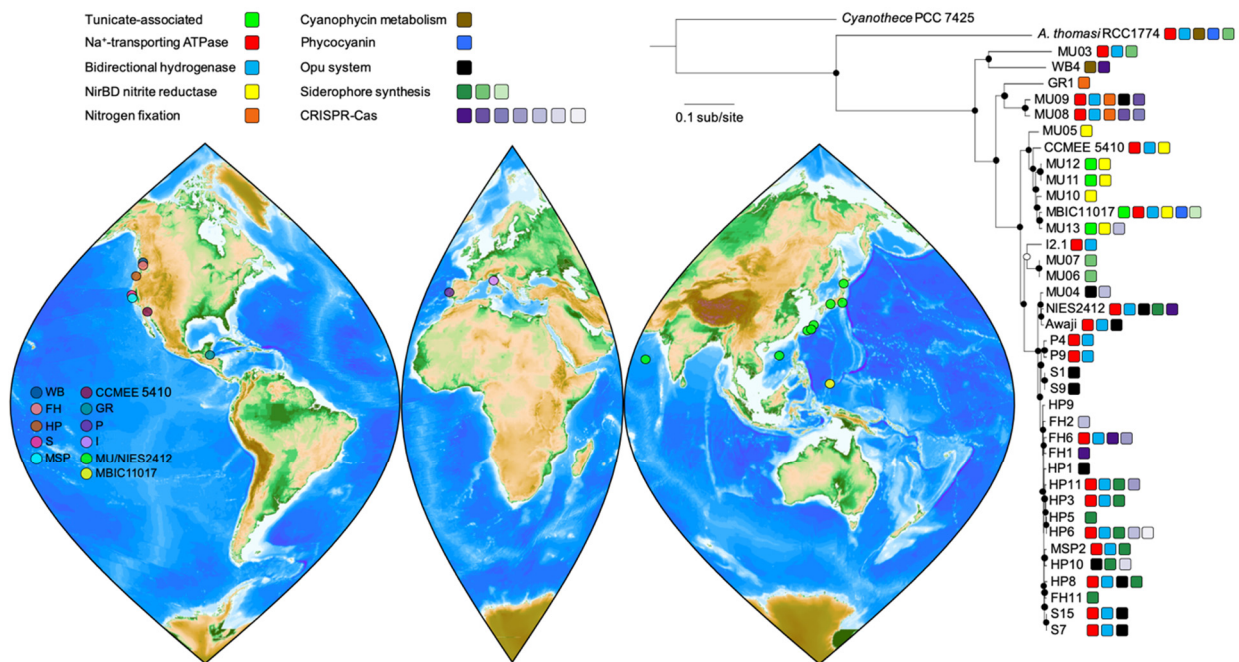

Figure 1. Sampling locations for *A. marina* strains and their evolutionary relationships. The genome-wide maximum likelihood phylogeny was reconstructed for a concatenation of 1369 protein sequences from single-copy orthologs according to the JTT+F+R5 model of sequence evolution and outgroup-rooted with *Cyanothece* sp. PCC 7425. Branch lengths are in units of the expected number of amino acid substitutions per site. Bootstrap support values of 100% (closed circles) and >90% (open circle) are shown for 1000 ultrafast bootstrap replicates. Selected traits discussed in the text are color coded as indicated. For the map of collection sites, note that the markers for the S. China Sea and Arabian Sea strains are approximate, as the exact sampling locations are unknown.

Our phylogeny provides fresh insights on *A. marina* diversification (Figure 1). As we previously reported [8], *A. thomasi* strain RCC1774, which produces Chl *b* instead of Chl *d*, was sister to the *A. marina* clade. *A. marina* strains isolated from the west coast of North America (sites FH, HP, S, MSP) were generally very closely related but did not cluster exclusively by geographic location (Figure 1), which implies a history of migration along the coast. New strains from Portugal and Italy provided a first look at European *A. marina* diversity, which revealed close connections with strains from both the western and eastern Pacific. Notably, Portugal strains P4 and P9 belonged to the Pacific West Coast clade (and were identical in 16S rRNA gene sequence to these strains). By contrast, the Italy strain I2.1 was sister to strains isolated from Awaji Island, Japan. Finally, although more data will be required to better resolve *A. marina* global biogeography, it is already clear that distantly related lineages can co-occur (for example, MU03 and MU04 from Hokkaido, Japan; Figure 1).

Tunicate-associated strains (Figure 1) formed part of a clade isolated from Palau (MBIC11017) and Okinawa (MU11, MU12, MU13). All were derived from the surface of didemnid tunicates that have a symbiotic association with the cyanobacterium *Prochloron*, although members of this clade can also be algal epiphytes (e.g., MU10). *A. marina* DNA has also been detected from a sponge sample collected in Palau [35]. Whether animal-associated *A. marina* are restricted to subtropical/tropical environments is not currently known.

Our data set also informs our understanding of the *A. marina* core genome. We considered a gene to belong to the core if it was detected in at least 36 of the 37 genomes. 2757 protein-coding genes belonged to the core (Figure 2; Table S3; $N = 2258$ genes for

a strict core of all 37 genomes). Few plasmid genes ($N = 6$) in the *A. marina* MBIC11017 reference genome are in this core, although they make up ~22% of its genome [14]. The majority of these were annotated as hypothetical proteins, with the exception of an RNA-binding protein and a phage integrase. This observation corroborates the earlier proposal that gene content of *A. marina* plasmids is highly dynamic and diverges rapidly between genomes [15]. 1502 core genes were assigned a KEGG Orthology ID, with 48 complete KEGG modules (Table S4; 35 additional modules are missing only one block).
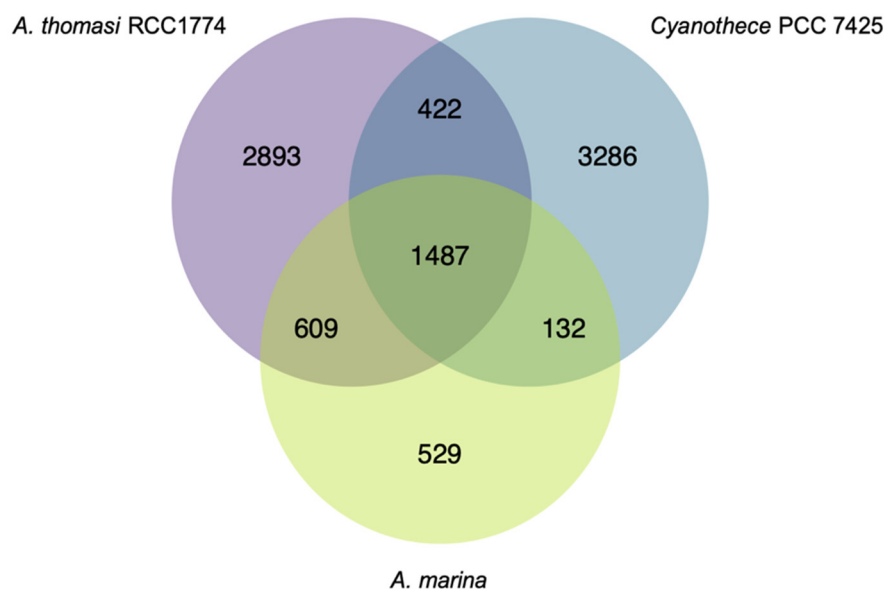
Figure 2. Venn diagram of shared and idiosyncratic protein-coding gene content among the *A. marina* core genome, the *Acaryochloris thomasi* RCC1774 genome, and *Cyanothece* PCC 7425 genomes. The diagram was generated with the R package VennDiagram.

529 genes in the *A. marina* core were not present in the outgroup strains (Figure 2; Table S5). This set of "unique core" genes potentially includes the unidentified gene(s) responsible for Chl *d* synthesis. Although Chl *d* is produced from Chl *a* and requires molecular oxygen [36], the identity of this enzyme(s) remains a longstanding puzzle in *Acaryochloris* biology. Yoneda et al. [37] identified 23 candidates from three oxygenase families, many of which were not co-expressed with other chlorophyll biosynthesis genes. These include the Rieske oxidase (RO; [38]), antibiotic biosynthesis monooxygenase (ABM) [39], and cytochrome P450 [40] families. Our core analysis excludes several of these, which do not belong to the *A. marina* core: AMB genes AM1_3889 and AM1_4175 (strain MBIC11017 gene designations), RO genes AM1_0159 and AM1_A0067, and P450 genes AM1_3563, AM1_4161, and AM1_5780. In fact, only one of the 23 genes belongs to the *A. marina* unique core: RO gene AM1_2905, annotated as a pheophorbide *a* oxygenase (PAO). PAO is involved in Chl *a* degradation through the oxygenolytic opening of the porphyrin ring [41]; the PAO enzyme in *A. marina* MBIC11017 is likely AM1_0031, which is the ortholog of PAO in *A. thomasi* RCC1774 (61% amino acid identity). AM1_2905 appears to have been acquired by an ancient HGT event prior to the diversification of extant *A. marina*. Whether this PAO homolog has evolved a new function in chlorophyll metabolism remains to be tested.

The estimated total number of genes (and genome size; Table S2) varies greatly among *A. marina* genomes (~5800–9600) but in all cases greatly exceeds the inferred core. Below, we focus on several examples of variation in gene content among members of the *A. marina* radiation.

3.2. Differential Retention of Ancestral Plasmid Gene Content Contributes to A. marina Functional Variation

Several notable genes that are observed in *A. thomasi* RCC1774 and were likely present in the *A. marina* common ancestor appear to have been differentially retained or lost during *A. marina* diversification (Figure 1). For example, many *A. marina* strains possess a second set of ATP synthase (ATPase) genes first reported for strain MBIC11017 [14]. These annotated sodium-transporting ATPase (Na^+-ATPase) genes are located on a conserved ~100 kbp block of plasmid sequence in the closed (or nearly closed) genomes of strains MBIC11017 (plasmid pREB4), CCMEE 5410, and S15 (Figure 3A,B). The genes are homologous to and share conserved gene order with the Na^+-ATPase operon of the halotolerant cyanobacterium *Aphanothece halophytica* [42]. The *A. halophytica* Na^+-ATPase is a sodium pump with increased activity at increasing NaCl concentrations [43], and heterologous expression conferred enhanced salt tolerance in the freshwater cyanobacterium *Synechococcus* PCC 7942 [42]. ATPase selectivity for sodium ions versus protons is controlled by the protein environment in the vicinity of the ion-binding site of the ATPase c-ring (AtpH) [44,45]. In *A. marina*, the amino acids of this binding site are conserved with *A. halophytica* [42], which suggests that these genes indeed encode a Na^+-ATPase that may be involved in salt tolerance (see below).
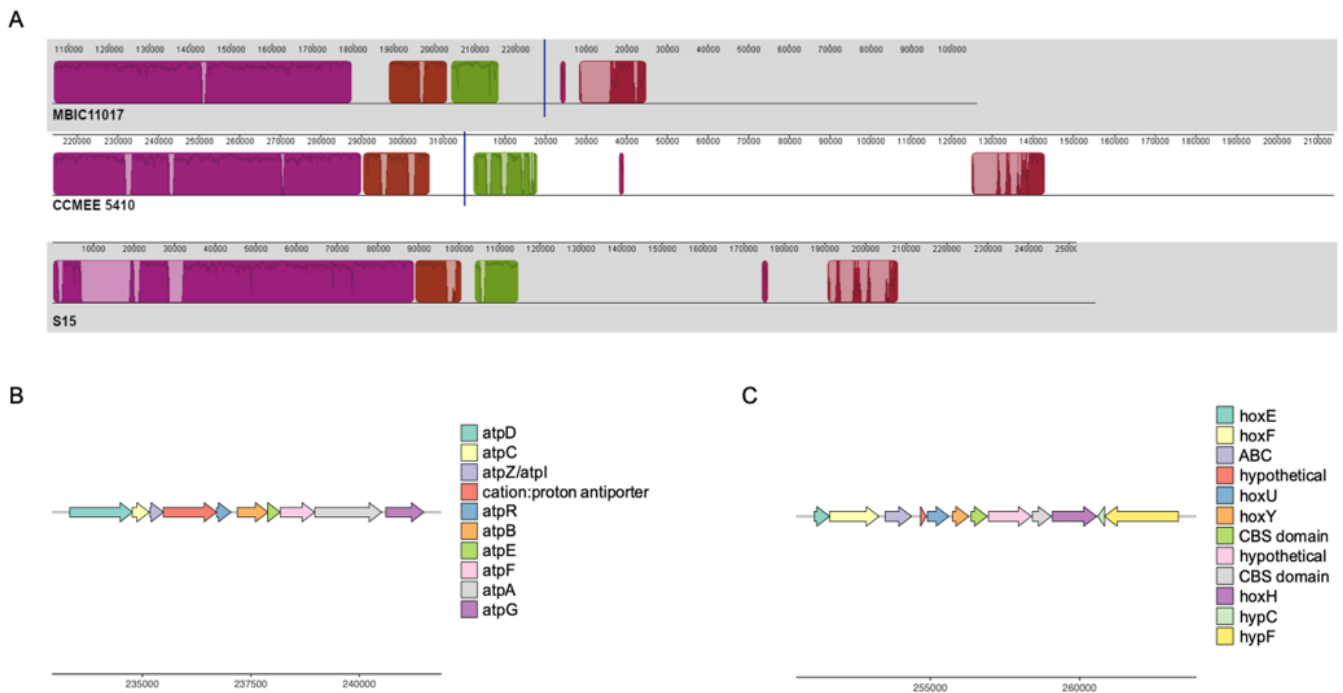
Figure 3. (**A**). Multiple alignment of plasmids from *A. marina* strains MBIC11017, CCMEE 5410, and S15 encoding sodium-transporting ATPase and bidirectional hydrogenase genes. Homologous blocks of aligned sequence share the same color, with missing DNA transparent. Traces within blocks indicate sequence similarity. (**B**). Gene map of the sodium-transporting ATPase region for *A. marina* CCMEE 5410. (**C**). Gene map of the bidirectional hydrogenase region for *A. marina* CCMEE 5410.

The putative Na^+-ATPase completely co-varies with genes encoding a bidirectional NiFe-hydrogenase (*hoxEFUYH*) and its associated maturation proteins that are found on the same plasmid as the *atp* genes (Figures 1 and 3A,C). Many cyanobacteria possess a bidirectional hydrogenase, but it is most commonly observed in strains isolated from environments that periodically experience microoxic or anoxic conditions, including microbial mats, salt marshes, and intertidal habitats [46]. The physiological role of this enzyme in cyanobacteria is still debated. Activity is biased toward hydrogen production in *Synechocystis* PCC 6803 [47], and proposed functions include as an electron sink in fermentation [46]

and/or photosynthesis [48], or as a general regulator of the cellular oxidation state [49]. In *Synechocystis* PCC 6803, electrons for hydrogen production are provided by reduced ferredoxin from photosystem I and pyruvate ferredoxin oxidoreductase (PFOR) activity, respectively, and a functional hydrogenase is essential for growth under anaerobic, nitrate-limiting conditions [50].

Conserved plasmid content also includes a glycogen phosphorylase (MBIC11017 nucleotide positions 164,576–167,203 in Figure 3) and PFOR genes (MBIC11017 positions 209,710–213,333). We propose that the plasmid may therefore provide a module for maintaining redox balance under fermentative conditions similar to the model of Khanna and Lindblad [51], with (1) glycogen catabolism to pyruvate, followed by (2) the oxidation of pyruvate to acetyl-CoA and reduction of ferredoxin by PFOR, and (3) hydrogen production via electron donation from ferredoxin to hydrogenase. This is in accord with the observation that *hox* expression in *A. marina* MBIC11017 is induced under anoxic, low light, and far-red light conditions [52,53]. Expression may potentially be controlled by conserved sensory and regulatory genes on the plasmid, which include a RpoD family sigma factor and a PAS domain S-box protein with histidine kinase and response regulator domains. PAS family proteins act as oxygen, redox status, or light sensors [54], and the gene includes a GAF domain with homology to FhlA ($E = 4 \times 10^{-11}$). FhlA plays a role in the transcription of the formate hydrogenlyase complex, which is responsible for fermentative hydrogen production in *E. coli* [55]. Whether there is a metabolic connection between the hydrogenase and the Na^+-ATPase remains to be investigated. In addition to a possible halotolerance function, nitrogen-starved cells of *Synechocystis* PCC 6803 use ATPase to produce a sodium motive force that maintains sufficient levels of ATP for survival during metabolic dormancy [56]. Both sodium bioenergetics and glycogen catabolism are important during the resumption of growth in *Synechocystis* PCC 6803 following nitrogen starvation [56], and future work could address if hydrogenase activity also contributes.

The distribution of *atp* and *hox* genes (Figure 1) indicates a dynamic history of retention and loss of the entire plasmid that bears these genes over millions of years of *A. marina* diversification [11]. The persistence of plasmids is a "paradox" [57] that has long puzzled biologists. Since they often impose a fitness cost on the host (e.g., via an expression burden), plasmids are generally expected to be lost over evolutionary time through either segregational instability or by the integration of beneficial plasmid genes into the chromosome [58]. Yet, it is possible for plasmids to be stably maintained not only in selectively favored environments but also in the absence of selection [59,60]. This can occur through compensatory evolution that reduces the fitness cost to the host or by plasmid mechanisms that promote persistence, such as a partitioning system to prevent segregational loss, conjugative transfer, or an addiction module such as a toxin-antitoxin system [61]. While the conserved core of the *atp*/*hox* plasmid lacks conjugation machinery and addiction modules, it does include plasmid partition genes *parAB*.

The role of selection in *atp*/*hox* plasmid persistence remains to be determined, as does whether plasmid loss is neutral or favored in particular environments. However, it is clear that these ecological and evolutionary dynamics can operate at a local scale. For example, strains with or without the plasmid co-occur at Hug Point, Oregon (Figure 1), where samples were collected at low tide from exposed rock and tide pools. One avenue for future research would be to investigate whether cells that have retained the plasmid are more likely to experience periods of anoxia, salt stress, and/or metabolic dormancy, perhaps due to microenvironmental differences in the probability or duration of tidal exposure.

3.3. HGT and the Evolution of Nitrogen Metabolism

A. marina strains have the coding potential to assimilate a variety of nitrogen (N) sources, including ammonium, nitrate, nitrite, urea, and cyanate. In addition, some *A. marina* genomes contain idiosyncratic N metabolism genes beyond these core capacities. For example, in addition to an assimilatory ferredoxin nitrite-reductase, one *A. marina* clade possesses transposase-flanked *nirBD* genes for nitrite reductase (Figure 1) that were likely

acquired by HGT from another cyanobacterium based on sequence similarity. In enterobacteria, NirBD is a dissimilatory enzyme that detoxifies nitrite produced by respiratory nitrate reduction under anoxic conditions [62]. Its role in *A. marina* is not clear; however, in strain CCMEE 5410, expression of these genes is similar to assimilatory ferredoxin nitrite-reductase in N deficient versus N replete conditions [37], which suggests that it may be an alternative assimilatory nitrite reductase.

N fixation of atmospheric N_2 to ammonium is an important source of biologically available N. The ability to fix N has been previously reported for one strain of *A. marina* (HICR111A) [63]. Like many unicellular, N-fixing cyanobacteria, strain HICR111A fixes N during the dark period [63], when nitrogenase-inhibiting oxygen is not produced by photosynthesis. N fixation genes are absent from the genomes of strains MBIC11017 and CCMEE 5410, and N fixation ability in HICR111A has been proposed to have been acquired by HGT [63]. However, the evolutionary history of this trait, including whether the common ancestor of *A. marina* possessed *nif* genes, has not been fully resolved due to limited taxon sampling.

Our new phylogeny provides fresh insights into this history. We have identified a clade of epiphytic *A. marina* near the base of the phylogeny that also have N fixation genes (Figure 1). Members of this clade can grow with N_2 as the sole N source at the same rate as on nitrate (Figure 4A). Although genome data are not publicly available for HICR111A, we believe that this strain likely belongs to this clade as well, based on a 16S rRNA gene tree (Figure S1), the *nif* gene sequence similarities of these strains, and the conservation of genes flanking the *nif* region among genomes [63].
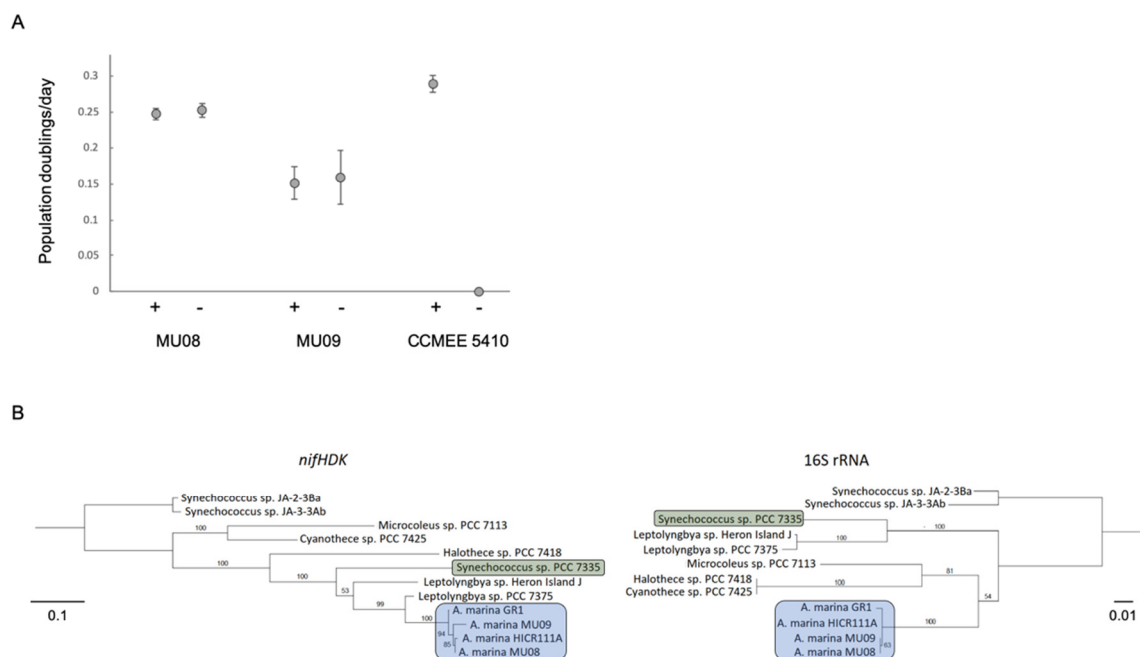
Figure 4. (**A**). Population growth rate in the presence (+) or absence (−) of combined nitrogen for *nif*-containing strains MU08 and MU09 and for *nif*-lacking strain CCMEE 5410. Error bars are standard errors. (**B**). Maximum likelihood phylogenies for *nifHDK* and 16S rRNA genes reconstructed with a GTR+I+G model. Bootstrap support values greater than 50% are shown for 1000 bootstrap replicates. Branch lengths are in units of expected number of nucleotide substitutions per site. The two topologies were significantly different by an SH test ($p < 0.0001$).

There are two alternative hypotheses for how these strains obtained *nif* genes. The first is that they were vertically inherited from the *A. marina* common ancestor, with subsequent gene loss in other descendant strains. If this were the case, then we predict that the *nif* phylogeny will resemble the species tree inferred with 16S rRNA gene data.

The second is that the most recent common ancestor of this clade acquired the *nif* genes more recently by HGT, in which case the *nif* and species trees will differ in topology. To test these hypotheses, we inferred maximum likelihood trees for both *nifHDK* (encoding the nitrogenase complex proteins) and 16S rRNA genes. In the species tree, the sister group of *A. marina* included *Cyanothece* PCC 7425 (Figure 4B), as has been observed in genome-wide phylogenies (Figure 1) [8,64]. By contrast, the most closely related sequences of *A. marina nifHDK* genes were from *Leptolyngbya* PCC 7375, with *Cyanothece* PCC 7425 genes more distantly related (Figure 4B). The topologies of the two trees were significantly different from each other, as determined by an SH test ($-\Delta\ln$ L = 310.4, $p < 0.0001$) [27]. This indicates that the *A. marina* common ancestor did not possess *nif* genes; rather, these genes were obtained later by the ancestor of the GR1/MU08/MU09 clade (Figure 1) via HGT from another cyanobacterium.

Finally, N storage pools vary among strains. The genes required to produce and degrade cyanophycin, the primary N storage compound of many cyanobacteria [65], is absent in all *A. marina* strains except strain WB4 (Figure 1). The amino acid sequences of cyanophycin synthetase and cyanophycinase in WB4 have high similarity (83% and 79%) with *A. thomasi* RCC1774, indicating that these genes were likely present in the *A. marina* common ancestor but lost by other strains. Phycobilisomes are the primary light-harvesting complexes of most cyanobacteria and can account for approximately 50% of cell soluble protein [66,67]; however, they are quickly degraded during N limitation to provide N for other functions [68]. We have previously reported that the *A. marina* common ancestor had lost the genes for phycobiliprotein production and degradation but were uniquely reacquired by strain MBIC11017 [8]. Even with the addition here of twelve new *A. marina* genomes, MBIC11017 is still the only known strain to produce phycobiliproteins (Figure 1).

3.4. Ecotypic Variation in Salt Tolerance

All of the *A. marina* strains in our collection were isolated from different zones along tidal gradients, from the upper intertidal to brackish or saline subtidal environments (Table S1). Consequently, they are expected to have historically experienced different frequencies and durations of stress associated with tidal activity, including the degree of exposure to both hyper- and hyposaline conditions. The salt tolerance range of marine algae (on which *A. marina* often grows as epiphytes) is generally broader for upper intertidal taxa than for subtidal species [69,70]. We therefore predicted that *A. marina* strains isolated from the upper intertidal zone would exhibit broader salinity tolerance than other strains.

To test this, we assayed growth at a range of NaCl concentrations for six *A. marina* strains. Strains S1 and S15 were isolated from upper intertidal and subtidal habitat, respectively, along a tidal gradient on the northern California coast (Shelter Cove; Table S1).

Like S1, HP9 was isolated from the upper intertidal zone (Hug Point, OR). Strain FH11, which is very closely related to the above strains (Figure 1), was isolated from a submerged estuarine environment (San Juan Island, WA) that has a salinity that is approximately 85% of that of seawater (30 ppt; https://www.bco-dmo.org/dataset/775732, accessed on 26 May 2021). Strains MBIC11017 and CCMEE 5410 are from submerged habitats in the tropical western Pacific [1] and the Salton Sea, CA (34 ppt salinity at time of collection; [11]), respectively.

As predicted, upper intertidal strains S1 and HP9 exhibited the broadest range of NaCl concentration that could support sustained growth (0.2–0.8 M; Figure 5; all six strains bleached at 0 M and 1.6 M NaCl). S1 and subtidal strain S15 did not differ significantly in growth rate at 0.2 M and 0.43 M NaCl; however, S1 grew faster than S15 at higher NaCl concentrations (FDR-adjusted $p = 0.08$ at 0.5 M, $p < 0.01$ at 0.62 M, and only S1 could grow at 0.8 M). The lower salt tolerance of S15 was expected given its subtidal origin. Estuarine strain FH11 was even less halotolerant, with optimal growth observed at lower NaCl concentrations ($p = 0.38$ for the comparison between 0.2 M and 0.43 M) and particularly low growth rates at NaCl concentrations greater than 0.43 M (including no sustained growth at 0.8 M). Other strains from less variable salinity environments grew well at intermediate

NaCl concentrations but did not grow at either the lowest salinity (MBIC11017) or highest salinity (CCMEE 5410) treatments. We conclude that *A. marina* halotolerance generally matches local environmental conditions.
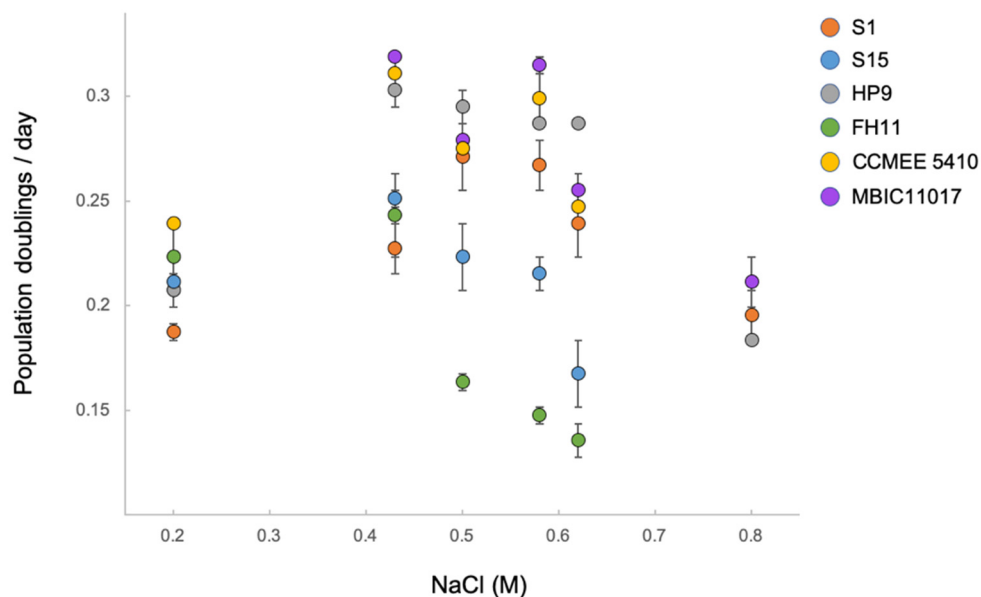
Figure 5. Population growth rates at different NaCl concentrations for upper intertidal *A. marina* strains S1 and HP9, subtidal strains S15 and MBIC11017, estuarine strain FH11, and saline lake strain CCMEE 5410. Error bars are standard errors. Note that 0.43 M is the NaCl concentration of standard marine ASN-III medium. Rates were estimated if growth was sustained for at least three generations. Pigments of strains for which growth was not observed at 0.2 M or 0.8 M NaCl did not bleach, whereas all strains bleached at 0 M and 1.6 M.

In cyanobacteria, halotolerance involves both the export of sodium ions and the production of compatible solutes to maintain osmotic balance under conditions of high salinity; however, other key aspects of tolerance, including how cells recognize salt stress signals and regulate salt acclimation, is not well understood [71]. Understanding the observed differences in halotolerance among *A. marina* strains is therefore not possible from a consideration of gene content alone. Nonetheless, it is instructive to briefly discuss how differences in gene content may contribute to these phenotypic differences.

A. marina strains vary in terms of the types and number of systems for removing sodium ions. For example, strain S1 has three CPA1 family Na^+/H^+ antiporter genes, compared with one each in the close relatives S15, FH11, and HP9. Most strains also have one copy of the *mrpCDEFGAB* gene cluster, a multi-subunit CPA3 family Na^+/H^+ antiporter. However, the *mrp* cluster is absent in the FH1 genome, whereas strains S15, S7, HP8, and HP11 have two copies. Recent duplicates in *A. marina* are often found on different genetic elements [15], and we confirmed that duplicated Mrp genes are on a plasmid (contig 10, nucleotide positions 263,524–267,402) for strain S15. The homologous *A. halophytica* cluster complements a salt-sensitive mutant of *E. coli* [72]; in *Anabaena* PCC 7120, *mrpA* is upregulated in response to increasing NaCl, and its inactivation by transposon mutagenesis results in a salt-sensitive phenotype [73]. Finally, as discussed above, many strains have plasmid-encoded genes for a sodium-transporting ATPase (Figure 1).

Strains also exhibit variation in genes involved in the production or import of compatible solutes. All possess *ggpP* and *ggpS* genes required to synthesize glucosylglycerol, a common compatible solute of marine cyanobacteria [71]. In addition, many strains also encode an ABC transporter in the osmoprotectant uptake (Opu) family of compatible solute transporters [74] with unknown substrate specificity (Figure 1).

3.5. Iron Metabolism

Iron is an important nutrient for most organisms but is often limiting in oxic environments, where it is insoluble [75]. Consequently, bacteria have evolved mechanisms for scavenging and storing iron when it is rare. This includes the ability to synthesize and/or import siderophores, low molecular weight compounds that chelate iron with high affinity [76]. The ability to acquire iron may vary among organisms depending on iron availability in the environment. We used FeGenie [30] to identify genes related to iron metabolism in *A. marina* genomes. Strains clustered into two groups based on iron gene content that do not closely reflect the *A. marina* phylogeny (Figure 6 and Figure S2). Strains with more genes related to iron metabolism tend to be enriched in genes involved in the regulation of iron assimilation, siderophore transport, and, often, siderophore synthesis (Figure 6; Table S6).
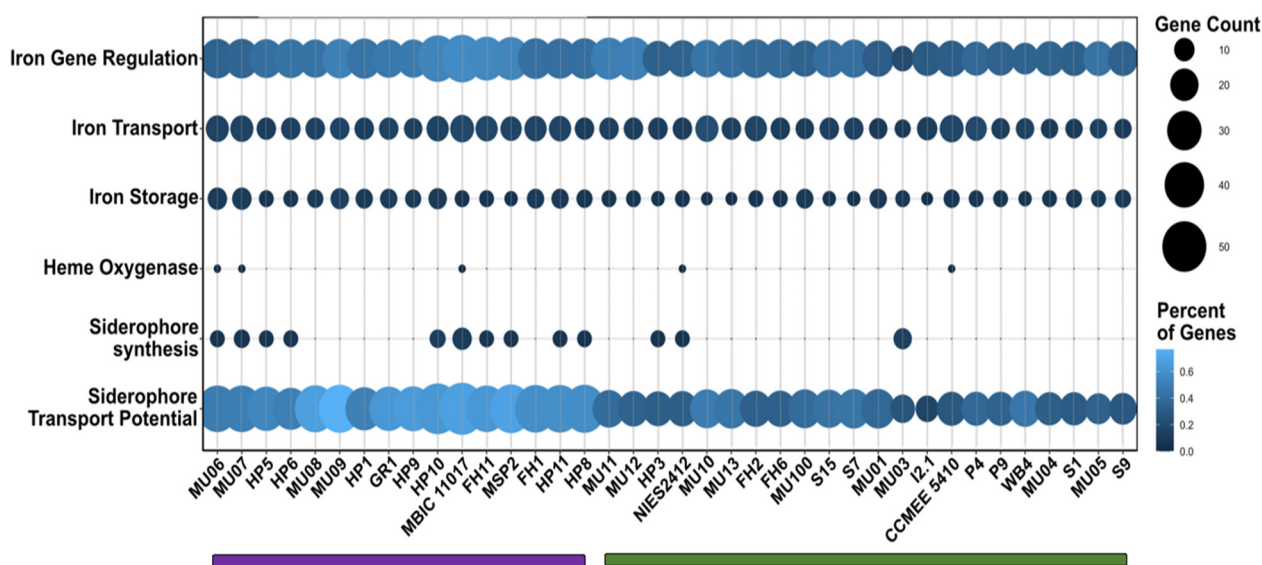
Figure 6. Distribution of genes involved in the regulation, acquisition, and storage of iron among *A. marina* strains based on both gene count and percent of genes in the genome. A dendrogram clustered strains into two groups based on high (purple) and low (green) total iron gene content. The approximately unbiased *p*-value of each cluster was 99%. See Figure S2 for the hierarchical clustering dendrogram.

All strains possess annotated siderophore import genes (though they vary in number). By contrast, genes involved in the synthesis of siderophores are scattered throughout the phylogeny and appear to have been independently acquired at least three times during *A. marina* diversification (Figure 1). In addition to the MBIC11017 cluster, the most common siderophore gene cluster was observed in strains isolated from the U.S. West Coast (FH, HP, and MSP sites) and in the Japanese strain NIES2412. The closest matches in the antiSMASH database of microbial secondary metabolites [32] were paenibactin and myxochelin. Homologs of these genes were detected by BLAST in an assembly of the cyanobacterium *Leptolyngbya* from metagenome data obtained for a South African peritidal stromatolite (NCBI BioProject PRJNA612530, assembly SM1_1_3) [77]. It is not clear whether this cluster was inherited from a common ancestor or independently acquired. A third cluster in strains MU06, MU07, and MU03 exhibited the closest similarity to cupriachelin in the antiSMASH database and exhibit homology to genes in *A. thomasi* RCC1774 (48–68% amino acid identity between the latter and MU06).

Enhanced iron gene content may be selectively favored in low iron environments. For example, the genome of strain MBIC11017, which was isolated from an iron-poor region of the Pacific (1), contains both plasmid-encoded iron transport gene duplicates and a

transposase-flanked cluster of genes [78] that is homologous to characterized siderophore synthesis genes in the cyanobacterium *Anabaena* PCC 7120 [79]. These genes are associated with higher rates of iron acquisition and growth under conditions of low iron availability compared with strain CCMEE 5410 [78], which is a member of the "low iron" gene content cluster (Figure 6).

3.6. Complex History of CRISPR-Cas Systems

A. marina genomes also vary with respect to the presence of CRISPR-Cas systems (Figure 1), all of which are either type I or type III class 1 systems (Table 1; Figure 7). The number of CRISPR arrays varied among these genomes (1–8 arrays), as did the total number of spacers (17–140 spacers; Table S7). The majority of spacers were derived from uncharacterized viral sequences, although plasmid sequences were also detected (Table S8).

Table 1. CRISPR-Cas systems detected in *A. marina* genomes.

Strains	Type	Repeat Consensus Sequence
MU08, MU09	I-D	GTTGCAAAAACGCTAAAACCCTCcAAGGGATTGAAAC
WB4, FH6, FH1, NIES2412	I-E	GTTGTCCCCACGCCTGTGGGGGTGGTCCG
HP10	I-MYXAN	AGCGGTGATTTAAGGTTTCCGGCCTGAAGCTTTGATGGACTT
MU08	III-A	GTTTCATCACTCATTCCCCGCAAGGGGACGGAAAC
HP6	III-B	GTTTCCAATAATTCCGATTGAAGTCAATCGGTAAAG
MU04, HP6, MU13, FH2	III-B	GTTTCCAATAATTCCGATTGAAGTCAATCGGTAAAG
HP11, FH6	III-B	GTTTTCATTTATTCGCCTTCCTACTGAATAGGAAG
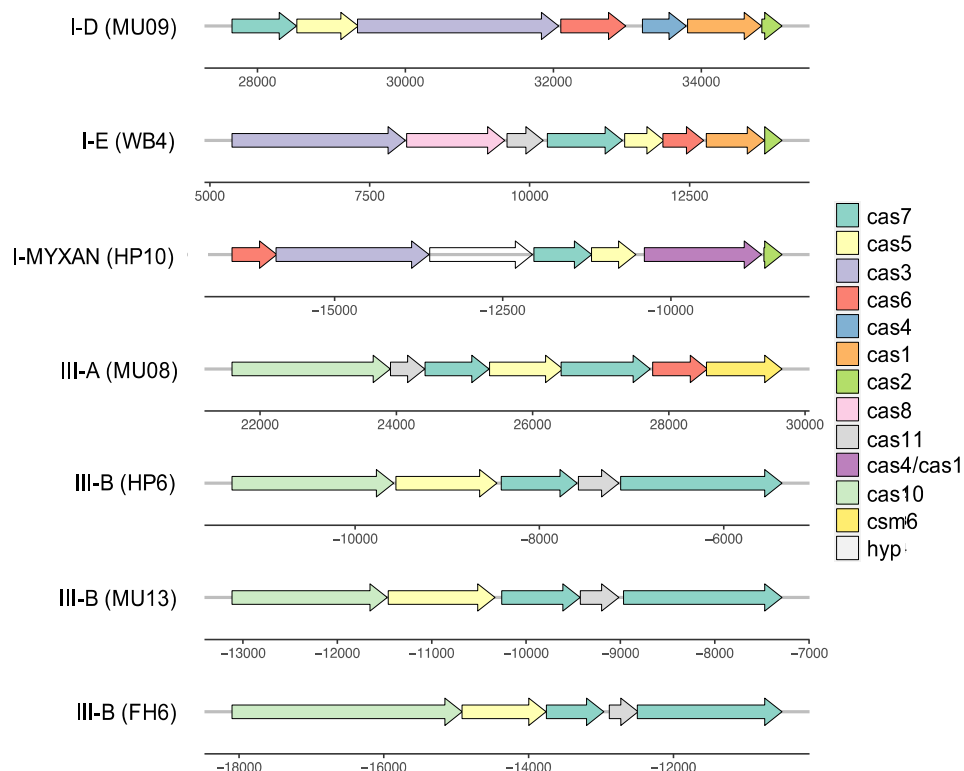
Figure 7. Representative gene maps for CRISPR-Cas systems detected in *A. marina* genomes (strain is indicated in parentheses).

The scattered distribution of CRISPR-Cas among extant *A. marina* strains (Figure 1) suggests a complicated history of gains, retentions, and losses. Two of the systems (the I-E and one of the III-B systems) are observed in strains from geographically distant locations (both cases involve western North America and Japan) among distantly related strains

(Figure 1). Both exhibit a highest sequence identity with distantly related cyanobacteria (*Romeria gracilis* (I-E) and *Aphanocapsa montana* (III-B), respectively) and therefore clearly have been acquired by HGT.

There is both theoretical and empirical support for the idea that organisms with functional CRISPR-Cas systems obtain protection from phages and plasmids at the cost of a reduction in the ability to acquire potentially beneficial genes by HGT [80,81]. We speculate that such a trade-off may have played out during *A. marina* diversification. An intact CRISPR-Cas may be selectively favored under conditions of high foreign DNA-induced mortality but otherwise lost; its absence may then result in a greater influx of new genes by HGT.

4. Conclusions

Laboratory strain collections and genomic resources are essential for developing an understanding of the genetic and phenotypic variation of recently discovered microorganisms. Our study illustrates several ways by which genome content and ecotypic variation for nutrient metabolism and environmental tolerance have evolved during *A. marina* diversification. It further identifies phenotypic variation of interest for future comparative physiology and omics investigations.

References

1. Miyashita, H.; Ikemoto, H.; Kurano, N.; Adachi, K.; Chihara, M.; Miyachi, S. Chlorophyll *d* as a major pigment. *Nature* **1996**, *383*, 402. [CrossRef]
2. Behrendt, L.; Larkum, A.W.D.; Trampe, E.; Norman, A.; Sørensen, S.J.; Kühl, M. Microbial diversity of biofilm communities in microniches associated with the didemnid ascidian *Lissoclinum patella*. *ISME J.* **2012**, *6*, 1222–1237. [CrossRef] [PubMed]
3. Kashiyama, Y.; Miyashita, H.; Ohkubo, S.; Ogawa, N.O.; Chikaraishi, Y.; Takano, Y.; Suga, H.; Toyofuku, T.; Nomaki, H.; Kitazato, H.; et al. Evidence of global chlorophyll *d*. *Science* **2008**, *321*, 658. [CrossRef] [PubMed]
4. Murakami, A.; Miyashita, H.; Iseki, M.; Adachi, K.; Mimuro, M. Chlorophyll *d* in an epiphytic cyanobacterium of red algae. *Science* **2004**, *303*, 1633. [CrossRef] [PubMed]
5. Mohr, R.; Voß, B.; Schliep, M.; Kurz, T.; Maldener, I.; Adams, D.G.; Larkum, A.D.W.; Chen, M.; Hess, W.R. A new chlorophyll *d*-containing cyanobacterium: Evidence for niche adaptation in the genus *Acaryochloris*. *ISME J.* **2010**, *4*, 1456–1469. [CrossRef] [PubMed]
6. Larkum, A.W.; Chen, M.; Li, Y.; Schliep, M.; Trampe, E.; West, J.; Salih, A.; Kühl, M. A novel epiphytic Chlorophyll *d*-containing cyanobacterium isolated from a mangrove-associated red alga. *J. Phycol.* **2012**, *48*, 1320–1327. [CrossRef]

7. Behrendt, L.; Nielsen, J.L.; Sørensen, S.J.; Larkum, A.W.D.; Winther, J.R.; Kühl, M. Rapid TaqMan-based quantification of Chlorophyll *d*-containing cyanobacteria in the genus *Acaryochloris*. *Appl. Environ. Microbiol.* **2014**, *80*, 3244–3249. [CrossRef]

8. Ulrich, N.J.; Uchida, H.; Kanesaki, Y.; Hirose, E.; Murakami, A.; Miller, S.R. Reacquisition of light-harvesting genes in a marine cyanobacterium confers a broader solar niche. *Curr. Biol.* **2021**, *31*, 1539–1546. [CrossRef]

9. Goh, F.; Allen, M.A.; Leuko, S.; Kawaguchi, T.; Decho, A.W.; Burns, B.P.; Neilan, B. Determining the specific microbial populations and their spatial distribution within the stromatolite ecosystem of Shark Bay. *ISME J.* **2009**, *3*, 383–396. [CrossRef]

10. Wood, A.M.; Miller, S.R.; Li, W.K.W.; Castenholz, R.W. Preliminary studies of cyanobacteria, picoplankton, and virioplankton in the Salton Sea with special attention to phylogenetic diversity among eight strains of filamentous cyanobacteria. *Hydrobiologia* **2002**, *473*, 77–92. [CrossRef]

11. Miller, S.R.; Augustine, S.; Le Olson, T.; Blankenship, R.E.; Selker, J.; Wood, A.M. Discovery of a free-living chlorophyll *d*-producing cyanobacterium with a hybrid proteobacterial/cyanobacterial small-subunit rRNA gene. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 850–855. [CrossRef] [PubMed]

12. Fleming, E.D.; Prufert-Bebout, L. Characterization of cyanobacterial communities from high-elevation lakes in the Bolivian Andes. *J. Geophys. Res.* **2010**, *115*. [CrossRef]

13. Zhang, Z.-C.; Li, Z.-K.; Yin, Y.-C.; Li, Y.; Jia, Y.; Chen, M.; Qiu, B.-S. Widespread occurrence and unexpected diversity of red-shifted chlorophyll producing cyanobacteria in humid subtropical forest ecosystems. *Environ. Microbiol.* **2019**, *21*, 1497–1510. [CrossRef] [PubMed]

14. Swingley, W.D.; Chen, M.; Cheung, P.C.; Conrad, A.L.; Dejesa, L.C.; Hao, J.; Honchak, B.M.; Karbach, L.E.; Kurdoglu, A.; Lahiri, S.; et al. Niche adaptation and genome expansion in the chlorophyll *d*-producing cyanobacterium *Acaryochloris marina*. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 2005–2010. [CrossRef] [PubMed]

15. Miller, S.R.; Wood, A.M.; Blankenship, R.E.; Kim, M.; Ferriera, S. Dynamics of gene duplication in the genomes of chlorophyll *d*-producing cyanobacteria: Implications for the ecological niche. *Genome Biol. Evol.* **2011**, *3*, 601–613. [CrossRef]

16. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]

17. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comp. Biol.* **2012**, *19*, 455–477. [CrossRef]

18. Wick, R.R.; Schultz, M.B.; Zobel, J.; Holt, K.E. Bandage: Interactive visualization of de novo genome assemblies. *Bioinformatics* **2015**, *31*, 3350–3352. [CrossRef]

19. Wood, D.E.; Lu, J.; Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **2019**, *20*, 257. [CrossRef]

20. Brettin, T.; Davis, J.J.; Disz, T.; Edwards, R.A.; Gerdes, S.; Olsen, G.J.; Olson, R.; Overbeek, R.; Parrello, B.; Pusch, G.D.; et al. RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* **2015**, *5*, 8365. [CrossRef]

21. Parks, D.H.; Imelfort, M.; Skennerton, C.T.; Hugenholtz, P.; Tyson, G.W. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **2015**, *25*, 1043–1055. [CrossRef] [PubMed]

22. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [CrossRef] [PubMed]

23. Emms, D.M.; Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **2019**, *20*, 1–14. [CrossRef] [PubMed]

24. Hoang, D.T.; Chernomor, O.; von Haeseler, A.; Minh, B.Q.; Vinh, L.S. UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **2018**, *35*, 518–522. [CrossRef]

25. Nguyen, L.-T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2014**, *32*, 268–274. [CrossRef] [PubMed]

26. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; von Haeseler, A.; Jermiin, L.S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **2017**, *14*, 587–589. [CrossRef]

27. Shimodaira, H.; Hasegawa, M. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **1999**, *16*, 1114–1116. [CrossRef]

28. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef]

29. Huerta-Cepas, J.; Szklarczyk, D.; Heller, D.; Hernández-Plaza, A.; Forslund, S.K.; Cook, H.V.; Mende, D.R.; Letunic, I.; Rattei, T.; Jensen, L.J.; et al. eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **2018**, *47*, D309–D314. [CrossRef]

30. Garber, A.I.; Nealson, K.H.; Okamoto, A.; McAllister, S.M.; Chan, C.S.; Barco, R.A.; Merino, N. FeGenie: A comprehensive tool for the identification of iron genes and iron gene neighborhoods in genome and metagenome assemblies. *Front Microbiol.* **2020**, *11*, 37. [CrossRef]

31. Suzuki, R.; Shimodaira, H. Pvclust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **2006**, *22*, 1540–1542. [CrossRef] [PubMed]

32. Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S.Y.; Medema, M.H.; Weber, T. antiSMASH 5.0: Updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **2019**, *47*, W81–W87. [CrossRef] [PubMed]

33. Grissa, I.; Vergnaud, G.; Pourcel, C. CRISPRFinder: A web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* **2007**, *35*, W52–W57. [CrossRef] [PubMed]

34. Biswas, A.; Gagnon, J.N.; Brouns, S.J.J.; Fineran, P.C.; Brown, C.M. CRISPRTarget: Bioinformatic prediction and analysis of crRNA targets. *RNA Biol.* **2013**, *10*, 817–827. [CrossRef] [PubMed]

35. Ohkubo, S.; Miyashita, H. Selective detection and phylogenetic diversity of *Acaryochloris* spp. that exist in association with didemnid ascidians and sponge. *Microbes Environ.* **2012**, *27*, 217–225. [CrossRef] [PubMed]

36. Schliep, M.; Crossett, B.; Willows, R.D.; Chen, M. ^{18}O labeling of chlorophyll *d* in *Acaryochloris marina* reveals that chlorophyll *a* and molecular oxygen are precursors. *J. Biol. Chem.* **2010**, *285*, 28450–28456. [CrossRef]

37. Yoneda, A.; Wittmann, B.J.; King, J.D.; Blankenship, R.E.; Dantas, G. Transcriptomic analysis illuminates genes involved in chlorophyll synthesis after nitrogen starvation in *Acaryochloris* sp. CCMEE 5410. *Photosynth. Res.* **2016**, *129*, 171–182. [CrossRef]

38. Barry, S.M.; Challis, G.L. Mechanism and catalytic diversity of Rieske non-heme iron-dependent oxygenases. *ACS Catal.* **2013**, *3*, 2362–2370. [CrossRef]

39. Sciara, G.; Kendrew, S.G.; Miele, A.E.; Marsh, N.G.; Federici, L.; Malatesta, F. The structure of ActVA-Orf6, a novel type of monooxygenase involved in actinorhodin biosynthesis. *EMBO J.* **2003**, *22*, 205–215. [CrossRef]

40. Denisov, I.G.; Makris, T.M.; Sligar, S.G.; Schlichting, I. Structure and chemistry of cytochrome P450. *Chem. Rev.* **2005**, *105*, 2253–2277. [CrossRef]

41. Pružinská, A.; Tanner, G.; Anders, I.; Roca, M.; Hörtensteiner, S. Chlorophyll breakdown: Pheophorbide a oxygenase is a Rieske-type iron–sulfur protein, encoded by the accelerated cell death 1 gene. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 15259–15264. [CrossRef] [PubMed]

42. Soontharapirakkul, K.; Promden, W.; Yamada, N.; Kageyama, H.; Incharoensakdi, A.; Iwamoto-Kihara, A.; Takabe, T. Halotolerant cyanobacterium *Aphanothece halophytica* contains an Na$^+$-dependent F$_1$F$_0$-ATP synthase with a potential role in salt-stress tolerance. *J. Biol. Chem.* **2011**, *286*, 10169–10176. [CrossRef] [PubMed]

43. Soontharapirakkul, K.; Incharoensakdi, A. Na$^+$-stimulated ATPase of alkaliphilic halotolerant cyanobacterium *Aphanothece halophytica* translocates Na$^+$ into proteoliposomes via Na$^+$ uniport mechanism. *BMC Biochem.* **2010**, *11*, 30. [CrossRef] [PubMed]

44. Schulz, S.; Iglesias-Cans, M.; Krah, A.; Yildiz, Ö.; Leone, V.; Matthies, D.; Cook, G.M.; Faraldo-Gómez, J.D.; Meier, T. A New Type of Na$^+$-driven ATP synthase membrane rotor with a two-carboxylate ion-coupling motif. *PLoS Biol.* **2013**, *11*, e1001596. [CrossRef] [PubMed]

45. Leone, V.; Pogoryelov, D.; Meier, T.; Faraldo-Gómez, J.D. On the principle of ion selectivity in Na$^+$/H$^+$-coupled membrane proteins: Experimental and theoretical studies of an ATP synthase rotor. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E1057–E1066. [CrossRef]

46. Barz, M.; Beimgraben, C.; Staller, T.; Germer, F.; Opitz, F.; Marquardt, C.; Schwarz, C.; Gutekunst, K.; Vanselow, K.H.; Schmitz, R.; et al. Distribution analysis of hydrogenases in surface waters of marine and freshwater environments. *PLoS ONE* **2010**, *5*, e13846. [CrossRef]

47. McIntosh, C.L.; Germer, F.; Schulz, R.; Appel, J.; Jones, A.K. The [NiFe]-hydrogenase of the cyanobacterium *Synechocystis* sp. PCC 6803 works bidirectionally with a bias to H$_2$ production. *J. Am. Chem. Soc.* **2011**, *133*, 11308–11319. [CrossRef]

48. Appel, J.; Phunpruch, S.; Steinmüller, K.; Schulz, R. The bidirectional hydrogenase of *Synechocystis* sp. PCC 6803 works as an electron valve during photosynthesis. *Arch. Microbiol.* **2000**, *173*, 333–338. [CrossRef]

49. Carrieri, D.; Wawrousek, K.; Eckert, C.; Yu, J.; Maness, P.-C. The role of the bidirectional hydrogenase in cyanobacteria. *Bioresour. Tech.* **2011**, *102*, 8368–8377. [CrossRef]

50. Gutekunst, K. The Bidirectional NiFe-hydrogenase in *Synechocystis* sp. PCC 6803 is reduced by flavodoxin and ferredoxin and is essential under mixotrophic, nitrate-limiting conditions. *J. Biol. Chem.* **2014**, *289*, 1930–1937. [CrossRef]

51. Khanna, N.; Lindblad, P. Cyanobacterial hydrogenases and hydrogen metabolism revisited: Recent progress and future prospects. *Int. J. Mol. Sci.* **2015**, *16*, 10537–10561. [CrossRef] [PubMed]

52. Kiss, É.; Kós, P.B.; Chen, M.; Vass, I. Functioning of the bidirectional hydrogenase in different unicellular cyanobacteria. In *Photosynthesis Research for Food, Fuel and the Future*; Kuang, T., Lu, C., Zhang, L., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 733–736.

53. Loughlin, P.; Lin, Y.; Chen, M. Chlorophyll *d* and *Acaryochloris marina*: Current status. *Photosynth. Res.* **2013**, *116*, 277–293. [CrossRef] [PubMed]

54. Taylor, B.L.; Zhulin, I.B. PAS domains: Internal sensors of oxygen, redox potential, and light. *Microbiol. Mol. Biol. Rev.* **1999**, *63*, 479–506. [CrossRef] [PubMed]

55. Sankar, P.; Lee, J.H.; Shanmugam, K.T. Gene-product relationships of *fhlA* and *fdv* genes of *Escherichia coli*. *J. Bacteriol.* **1988**, *170*, 5440–5445. [CrossRef] [PubMed]

56. Doello, S.; Burkhardt, M.; Forchhammer, K. The essential role of sodium bioenergetics and ATP homeostasis in the developmental transitions of a cyanobacterium. *Curr. Biol.* **2021**, *31*, 1606–1615. [CrossRef] [PubMed]

57. Harrison, E.; Brockhurst, M.A. Plasmid-mediated horizontal gene transfer is a coevolutionary process. *Trends Microbiol.* **2012**, *20*, 262–267. [CrossRef] [PubMed]

58. San Millan, A.; MacLean, R.C. Fitness costs of plasmids: A limit to plasmid transmission. *Microbiol. Spectr.* **2017**, *5*, 65–79. [CrossRef]

59. Harrison, E.; Guymer, D.; Spiers, A.J.; Paterson, S.; Brockhurst, M.A. Parallel compensatory evolution stabilizes plasmids across the parasitism-mutualism continuum. *Curr. Biol.* **2015**, *25*, 2034–2039. [CrossRef]

60. Wein, T.; Hülter, N.F.; Mizrahi, I.; Dagan, T. Emergence of plasmid stability under non-selective conditions maintains antibiotic resistance. *Nature Comm.* **2019**, *10*, 2595. [CrossRef]

61. Carroll, A.C.; Wong, A. Plasmid persistence: Costs, benefits, and the plasmid paradox. *Can. J. Microbiol.* **2018**, *64*, 293–304. [CrossRef]

62. Stewart, V. Regulation of nitrate and nitrite reductase synthesis in enterobacteria. *Antonie Leeuwenhoek* **1994**, *66*, 37–45. [CrossRef] [PubMed]

63. Pfreundt, U.; Stal, L.J.; Voß, B.; Hess, W.R. Dinitrogen fixation in a unicellular chlorophyll *d*-containing cyanobacterium. *ISME J.* **2012**, *6*, 1367–1377. [CrossRef] [PubMed]

64. Shih, P.M.; Wu, D.; Latifi, A.; Axen, S.; Fewer, D.; Talla, E.; Calteau, A.; Cai, F.; de Marsac, N.T.; Rippka, R.; et al. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 1053–1058. [CrossRef] [PubMed]

65. Watzer, B.; Forchhammer, K. Cyanophycin synthesis optimizes nitrogen utilization in the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6803. *Appl. Environ. Microbiol.* **2018**, *84*, e01298-18. [CrossRef]

66. Boussiba, S.; Richmond, A.E. C-phycocyanin as a storage protein in the blue-green alga *Spirulina platensis*. *Arch. Microbiol.* **1980**, *125*, 143–147. [CrossRef]

67. Glazer, A.N. Phycobiliproteins. *Methods Enzymol.* **1988**, *167*, 291–303.

68. Allen, M.M.; Smith, A.J. Nitrogen chlorosis in blue-green algae. *Arch. Microbiol.* **1969**, *69*, 114–120. [CrossRef]

69. Russell, G. Salinity and seaweed vegetation. In *Plant Life in Aquatic and Amphibious Habitats*; Crawford, R.M.M., Ed.; Blackwell: Oxford, UK, 1987; pp. 35–52.

70. Karsten, U. Seaweed acclimation to salinity and desiccation stress. In *Seaweed Biology*; Wiencke, C., Bischof, K., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 87–107.

71. Hagemann, M. Molecular biology of cyanobacterial salt acclimation. *FEMS Microbiol. Rev.* **2011**, *35*, 87–123. [CrossRef]

72. Fukaya, F.; Promden, W.; Hibino, T.; Tanaka, Y.; Nakamura, T.; Takabe, T. An Mrp-like cluster in the halotolerant cyanobacterium *Aphanothece halophytica* functions as a Na^+/H^+ antiporter. *Appl. Environ. Microbiol.* **2009**, *75*, 6626–6629. [CrossRef]

73. Blanco-Rivero, A.; Leganes, F.; Fernandez-Valiente, E.; Calle, P.; Fernandez-Piñas, F. *mrpA*, a gene with roles in resistance to Na^+ and adaptation to alkaline pH in the cyanobacterium *Anabaena* sp. PCC7120. *Microbiology* **2005**, *151*, 1671–1682. [CrossRef]

74. Hoffmann, T.; Bremer, E. Guardians in a stressful world: The Opu family of compatible solute transporters from *Bacillus subtilis*. *Biol. Chem.* **2017**, *398*, 193–214. [CrossRef] [PubMed]

75. Emerson, D. The irony of iron–biogenic iron oxides as an iron source to the ocean. *Front. Microbiol.* **2016**, *6*, 1502. [CrossRef] [PubMed]

76. Kranzler, C.; Rudolf, M.; Keren, N.; Schleiff, E. Iron in cyanobacteria. *Adv. Bot. Res.* **2013**, *65*, 57–105.

77. Waterworth, S.C.; Isemonger, E.W.; Rees, E.R.; Dorrington, R.A.; Kwan, J.C. Conserved bacterial genomes from two geographically isolated peritidal stromatolite formations shed light on potential functional guilds. *Environ. Microbiol. Rep.* **2021**, *13*, 126–137. [CrossRef]

78. Gallagher, A.L.; Miller, S.R. Expression of novel gene content drives adaptation to low iron in the cyanobacterium *Acaryochloris*. *Genome Biol. Evol.* **2018**, *10*, 1484–1492. [CrossRef]

79. Jeanjean, R.; Talla, E.; Latifi, A.; Havaux, M.; Janicki, A.; Zhang, C.C. A large gene cluster encoding peptide synthetases and polyketide synthases is involved in production of siderophores and oxidative stress response in the cyanobacterium *Anabaena* sp. strain PCC 7120. *Environ. Microbiol.* **2008**, *10*, 2574–2585. [CrossRef]

80. Wheatley, R.M.; MacLean, R.C. CRISPR-Cas systems restrict horizontal gene transfer in *Pseudomonas aeruginosa*. *ISME J.* **2021**, *15*, 1420–1433. [CrossRef]

81. Jiang, W.; Maniv, I.; Arain, F.; Wang, Y.; Levin, B.R.; Marraffini, L.A. Dealing with the evolutionary downside of CRISPR immunity: Bacteria and beneficial plasmids. *PLoS Genet.* **2013**, *9*, e1003844. [CrossRef]

Article

Discovery of Chlorophyll *d*: Isolation and Characterization of a Far-Red Cyanobacterium from the Original Site of Manning and Strain (1943) at Moss Beach, California

Nancy Y. Kiang [1,2,*], Wesley D. Swingley [2,3,*], Dikshyant Gautam [3], Jared T. Broddrick [4], Daniel J. Repeta [5], John F. Stolz [6,7], Robert E. Blankenship [8], Benjamin M. Wolf [8,†], Angela M. Detweiler [4,9,‡], Kathy Ann Miller [10], Jacob J. Schladweiler [3], Ron Lindeman [11] and Mary N. Parenteau [2,4,*]

1   NASA Goddard Institute for Space Studies, New York, NY 10025, USA
2   Virtual Planetary Laboratory, Nexus for Exoplanet System Science, NASA Astrobiology Program, USA
3   Department of Biological Sciences, Northern Illinois University, DeKalb, IL 60115, USA; gautam.dikshyant@gmail.com (D.G.); jjschlad1408@gmail.com (J.J.S.)
4   NASA Ames Research Center, Moffett Field, CA 94035, USA; jared.t.broddrick@nasa.gov (J.T.B.); angeladetweiler@gmail.com (A.M.D.)
5   Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA; drepeta@whoi.edu
6   Department of Biological Sciences, Duquesne University, Pittsburgh, PA 15282, USA; stolz@duq.edu
7   Center for Environmental Research and Education, Duquesne University, Pittsburgh, PA 15282, USA
8   Departments of Biology and Chemistry, Washington University in St. Louis, St. Louis, MO 63130, USA; reblankenship@gmail.com (R.E.B.); wolfbenjamin25@gmail.com (B.M.W.)
9   Bay Area Environmental Research Institute, Moffett Field, CA 94035, USA
10 University Herbarium, University of California, Berkeley, CA 94720, USA; kathyannmiller@berkeley.edu
11 Citizen Scientist, Alameda, CA 94501, USA; ron.lindeman123@gmail.com
*   Correspondence: nancy.y.kiang@nasa.gov (N.Y.K.); wswingley@niu.edu (W.D.S.); mary.n.parenteau@nasa.gov (M.N.P.)
†   Current address: Pluton Biosciences, Inc., 11754 Westline Industrial Dr., St. Louis, MO 63146, USA.
‡   Current address: Chan Zuckerberg Biohub, San Francisco, CA 94158, USA.

Abstract: We have isolated a chlorophyll-*d*-containing cyanobacterium from the intertidal field site at Moss Beach, on the coast of Central California, USA, where Manning and Strain (1943) originally discovered this far-red chlorophyll. Here, we present the cyanobacterium's environmental description, culturing procedure, pigment composition, ultrastructure, and full genome sequence. Among cultures of far-red cyanobacteria obtained from red algae from the same site, this strain was an epiphyte on a brown macroalgae. Its Q_y *in vivo* absorbance peak is centered at 704–705 nm, the shortest wavelength observed thus far among the various known *Acaryochloris* strains. Its Chl *a*/Chl *d* ratio was 0.01, with Chl *d* accounting for 99% of the total Chl *d* and Chl *a* mass. TEM imagery indicates the absence of phycobilisomes, corroborated by both pigment spectra and genome analysis. The Moss Beach strain codes for only a single set of genes for producing allophycocyanin. Genomic sequencing yielded a 7.25 Mbp circular chromosome and 10 circular plasmids ranging from 16 kbp to 394 kbp. We have determined that this strain shares high similarity with strain S15, an epiphyte of red algae, while its distinct gene complement and ecological niche suggest that this strain could be the closest known relative to the original Chl *d* source of Manning and Strain (1943). The Moss Beach strain is designated *Acaryochloris* sp. (*marina*) strain Moss Beach.

Keywords: chlorophyll *d*; *Acaryochloris*; Moss Beach; cyanobacteria; far-red photosynthesis; photosynthetic pigments; absorbance spectra; genome sequence

1. Introduction

Chlorophyll *d* (Chl *d*) is a far-red-absorbing chlorophyll, which Manning and Strain [1] first discovered off the coast of Central California, and assumed to be a pigment of the red macroalgae growing in the intertidal zone. It was not until more than 50 years later that

Miyashita, et al. [2] found Chl *d* to be not an accessory pigment of algae, but a primary photopigment of a cyanobacterium, which they named *Acaryochloris marina* (type-strain MBIC11017). They discovered that Chl *d* replaces nearly all chlorophyll *a* (Chl *a*) in these organisms, enabling them to perform oxygenic photosynthesis at long wavelengths in the far-red/near-infrared [3]. This discovery overturned long-held wisdom about the primacy of chlorophyll *a* and the photon energy limits for oxygenic photosynthesis.

Li and Chen [4] and Allakhverdiev, et al. [5] provided detailed reviews of the molecular structures of the different chlorophylls, and of research to characterize the position and function of Chl *d* in Photosystems I and II (PSI and PSII). Whereas Chl *a* possesses a vinyl group at the C3 position of the chlorin macrocycle, with its Soret and Q_y bands peaking in absorbance at 435 and 666 nm, respectively, in methanol, Chl *d* instead possesses a formyl group, such that its Soret band is broadened with two peaks at 401 and 460 nm, and the Q_y peak is at 697–698 nm. PSI in *A. marina* turns out to have a Chl *d* homodimer for the special pair [6], while in Photosystem II (PSII) it remains to be confirmed whether the special pair is a Chl *a*/Chl *d* heterodimer [7] or a Chl *d*/Chl *d* homodimer [8]. Mielke, et al. [9] resolved questions as to whether *A. marina* could have inefficiencies in the use of long-wavelength photons due to recombinations, showing that the photon-energy-storage efficiency in *A. marina* is comparable to or higher than that of Chl *a* organisms. They found *A. marina* MBIC 11017 had $40 \pm 1\%$ thermal storage efficiency compared to that measured in Chl *a* organisms (ranging from $34.1 \pm 1\%$ in *Synechococcus leopoliensis* to $43 \pm 2\%$ in an alga, *Chlorella vulgaris*, Cha and Mauzerall [10]). It has been estimated through photoacoustics that the PSII trap wavelength is 723 ± 3 nm [11], and through laser flash absorption that the PSI trap is at 740 nm [12]; such that they are 40–46 nm and 40 nm to the red of their Chl *a* counterparts, respectively.

The discovery of *Acaryochloris* has motivated a hunt for more far-red oxygenic phototrophs and led to a proliferation of discoveries, including: additional strains of *Acaryochloris* [13–16]; numerous cyanobacteria that utilize Chl *d* as well as chlorophyll *f* (Chl *f*) [17] at yet longer wavelengths for uphill energy transfer to Chl *a* in Far-Red Light Photoacclimation (FaRLiP) [18–20]; and even a far-red Chl *a* as a photopigment in an alga [21]. All of these organisms live in niche environments depleted in visible light and relatively enriched in the far-red/near-infrared, receiving light filtered by Chl-*a*-containing organisms. *A. marina* str. MBIC11017 [2] was found off the coast of the Palau Islands under an ascidian that contains a Chl-*a*-utilizing cyanobacterial symbiont [22]. Murakami, et al. [15] found *A. marina* sp. strain Awaji to be an epiphyte of red algae at Awaji Island, Japan; this solved the original puzzle about whether str. MBIC11017 was a symbiont or epiphyte of its ascidian [22]. Larkum, et al. [13] later found another *Acaryochloris* epiphyte of red algae in a mangrove at Salt Pan Creek, Georges River NSW, Australia, *Acaryochloris* sp. MPGRS1. Miller, et al. [16] found *A. marina* str. CCMEE 5410 from a microbial mat rock biofilm in the Salton Sea, California, USA. Mohr, et al. [14] obtained *Acaryochloris* sp. HICR111A from scrapings off of a dead coral skeleton overgrown with other algae and cyanobacteria at Heron Island, Great Barrier Reef, Australia. Mehda, et al. [23] found cf. *Acaryochloris* in Sahara Desert biocrusts through phylogenetic analysis. Most recently, Ulrich, et al. [24] isolated strain S15 from a red alga off the coast of Northern California, USA. Numerous additional 16S rRNA entries of unpublished *Acaryochloris* discoveries can now be found in the databases of the Ribosomal Database Project (RDP) (https://rdp.cme.msu.edu) and the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov).

Despite these recent discoveries, no one, to our knowledge, has returned to Manning and Strain's site of the original discovery of Chl *d*, until now. Thus far, full genome sequences have only been obtained for *A. marina* MBIC11017 [25], *Acaryochloris* sp. CCMEE 5410 [16], and *A. marina* S15 [24]. Here, we present a strain of *Acaryochloris* from Moss Beach, CA, with a description of its isolation and culture methods, its pigment composition, ultrastructure, and a full genome sequence. We compare these results to other strains and species in the literature, and analyze the Moss Beach strain's genomic characteristics and its place in the *Acaryochloris* phylogenetic tree.

2. Materials and Methods

The original site of the Manning and Strain [1] discovery of chlorophyll *d* is a rocky intertidal zone at Montara State Marine Reserve, Moss Beach, CA, USA (122°30′55.7″ W 37°31′15.3″ N). Water temperatures range from 9–15 °C seasonally, and algal growth peaks in summer, with a particular abundance of red macroalgae (Figure 1).


Figure 1. Rocky intertidal zone field site at low tide on 2 August 2015, J.V. Fitzgerald Marine Reserve, Montara State Marine Reserve. Moss Beach, CA, USA. Shows depth layering of green and red macroalgae. (Pictured: Ron Lindeman).

2.1. Establishment of Cultures

On 2 August 2015, a variety of green, red, and brown macroalgae was collected at low tide at the J.V. Fitzgerald Marine Reserve protected area of Montara State Marine Reserve, Moss Beach, California, during the peak growing season for the red algae (Figure 1). Clippings of each sample were used to inoculate enrichment cultures in iron-enriched marine BG-11 growth medium, "FeMBG-11" [26] consisting of 10.4 mg/L EDTA and iron(III) monosodium salt (Fe-EDTA) added to 3.5% Instant Ocean artificial sea salts (Aquarium Systems, Mentor, OH, USA), and with trace minerals, ferric ammonium citrate, bicarbonate, potassium phosphate, as published previously [27]. However, instead of N-[Tris(hydroxymethyl)methyl]-2-aminoethanesulfonic acid (TES) buffer, we used Trizma HCL buffer adjusted to pH 8.2 with NaOH pellets prior to autoclaving; after autoclaving, the pH of the medium was typically above 8.7 to 9.0. In addition, enrichment cultures were augmented with "f/2" vitamins modified from the "f-1" recipe of Guillard and Ryther [28], consisting of 0.25 µg/L cyanocobalamin vitamin B12, 0.25 µg/L biotin vitamin H, and 0.2 mg/L Thiamin vitamin B-1.

The cultures were incubated at room temperature under a ~730 nm light-emitting diode (LED) light (LED Light Bar 730 nm Far Red, BML Horticulture, Austin, TX, USA), with irradiance over 400–740 nm ranging from 12–22 µmol m^{-2} s^{-1}. with a 12 h/12 h light/dark cycle. The spectrum for the LED, as measured with a JAZ-A-Irrad absolute irradiance spectroradiometer (Ocean Optics, Inc., Dunedin, FL, USA) at an irradiance location central to the growth chamber, is shown in Figure 2. The exact peak of the LED light bar is at 733.7 nm, and dashed lines demark wavelengths at 680, 700, 733.7, and 740 nm. This shows there is non-zero photon flux below 680 and 700 nm, such that photosynthesis with Chl *a*-based P680 and P700 could still receive a small amount of photosynthetically active radiation (PAR, 400–700 nm), but otherwise 93% of the 400–740 nm irradiance in photon flux is in the far red, beyond the range of P700. The intensity of the light bar was adjusted such that irradiance over 400–740 nm incident to the samples ranged between 11–22 µmol m^{-2} s^{-1}.
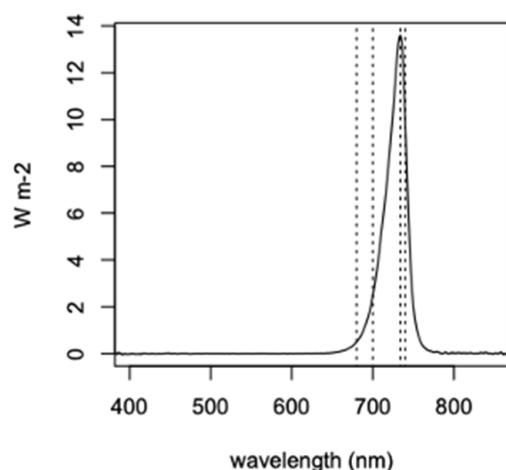
Figure 2. ~730 nm LED light bar irradiance spectrum to a central position in the *Acaryochloris* growth chamber. Dashed lines mark wavelengths at 680 (P680), 700 (P700), 733.7 (LED peak), and 740 nm (P740). Over 400–700 nm, the integrated flux is 1.74 μmol m^{-2} s^{-1}. Over 400–740 nm, the integrated flux is 21.87 μmol m^{-2} s^{-1} and 3.63 W m^{-2}.

We unsuccessfully attempted to establish axenic cultures from the enrichments using serial dilutions to extinction using inoculum that was sonicated to break up clumps of cells. However, we were successful in establishing unialgal cultures by selecting single-pigmented colonies on Petri plates containing FeMBG-11 medium hardened with 0.4% agarose. To generate more biomass for subsequent pigment and physiological characterizations, pigmented colonies from a plate were used to inoculate Erlenmeyer flasks ranging from 25 mL to 1 L, in the latter maintaining a medium depth of no more than 1.5 inches to allow for greater surface area of the medium for CO_2 exchange and diffusion to the bottom of the flask. Subsequent to genome sequencing, further maintenance of the cultures included antibiotic treatment to remove contaminating chemotrophic organisms. A plate colony was treated with an antibiotic mix of 25 μg/mL tetracycline, 25 μg/mL kanamycin, 100 μg/mL ampicillin, incubated in the dark for 24 h, and then the cells were transferred to regular FeMBG-11 growth medium. Biomass was preserved in -80 °C in the spent growth medium, and the maintained cultures descend from this antibiotic treatment.

2.2. In Vivo Spectral Absorbance

In vivo spectral absorbance was obtained as follows: 200 μL, 500 μL, and 800 μL aliquots of cultures were diluted into 10 mL of 10 mM Tris 50 mL NaCl pH 8.0 and gently vacuum-filtered onto 2.4 cm diameter GF/C filters. The filters were placed on top of a 96-well microtiter plate and their absorbance spectrum was measured on a Molecular Devices SpectraMax M5 spectrometer, at a spectral resolution of 1 nm over 350–900 nm. As a blank, 10 mL of buffer was filtered on GF/C filters and analyzed in the same manner, and was used to correct the baseline of the *in vivo* spectral absorbance. For each filter, 2 spots were analyzed as technical replicates and averaged together. The absorbance readings of the 200 μL and 500 μL were corrected to a total volume of 800 μL, the optical density (OD) values were averaged, and the standard deviation was determined.

2.3. Pigment Characterization

Pigment analysis was performed on samples of whole, intact cells of the far-red cyanobacteria cultures enriched from the fronds of the brown macroalgae *Stephanocystis osmundacea* (sample A-19), which were centrifuged into pellets in growth medium, and stored at -80 °C in growth medium prior to thawing for pigment extraction.

Pigment Extraction. The pigments in the cell pellets were recovered by sequential sonic extraction alternating between 0.5 mL of 100% methanol and 0.5 mL of 100% acetone (10 min each) until the final extract was not visibly colored. Cells were resuspended before

each sonication by vortex mixing. The extracts were combined and reduced to a small volume (~1 mL) by rotary vacuum evaporation at room temperature.

High-performance liquid chromatography (HPLC). Culture extracts were analyzed for pigments on an Agilent 1200 HPLC system fitted with a diode-array detector. Pigments were separated on an Ascentis C-18 column (150 × 2.1 mm; 3 μm) eluted at 300 μL min^{-1} using a linear gradient from 100% solvent A (80%/20% *v/v* methanol/aqueous 50 mM ammonium acetate) to 100% solvent B (80/20 *v/v* methanol/acetone) over 25 min followed by isocratic elution at 100% B for an additional 30 min. Pigments were detected between 400–750 nm.

Biomass. The extracted biomass used for pigment analyses was then air dried and weighed to obtain dry biomass. The dry biomass was then used to calculate the moles Chl *d* per grams biomass. We note that these estimated dry weights represent a minimum in biomass dry weight, and therefore a slight overestimate of pigment per gram of cell dry biomass, as they do not account for the organic matter that was extracted from the sample by the methanol and acetone. Acetone and methanol will extract lipids, which are ~10–20% of the cell dry weight, as a rule of thumb [29]. Therefore, the Chl/biomass presented here is a maximum and may be an overestimate by 10–20% since it derives from an underestimate of total biomass. However, this will not affect the estimated Chl *d*/Chl *a* ratio.

Calibration for Chl *d* extinction coefficient. Chl *d* and Chl *a* were purified by HPLC using the same column and flow rate detailed above, but with isocratic elution with 100% solvent B. The Chl *d* fraction (3.8–4.2 min) was collected and dried by rotary vacuum evaporation at room temperature. Spectral absorbance of the purified Chl *d* in 100% methanol was measured on an Evolution 300 spectrophotometer over 350–800 nm with a spectral resolution of 1 nm. The mass extinction coefficient α_λ = 71.11 (L g^{-1} cm^{-1}) for Chl *d* obtained by Li, et al. [30] and the 697 nm absorbance were used to obtain the mass concentration C (g L^{-1}) of Chl *d* in the solution. This solution was used to calibrate the HPLC detector response factor. For Chl *a*, the mass extinction coefficient of 88.15 (L g^{-1} cm^{-1}) [31] and prior calibration of the HPLC were used to obtain the mass concentration of Chl *a*.

2.4. Transmission Electron Microscopy (TEM)

Cells were prepared for TEM using a modification of the method in Switzer Blum, et al. [32]. The cells were initially fixed with glutaraldehyde (2.5% final concentration) in their culture medium for 2 h. They were then rinsed 3 times in Phosphate Buffered Saline (PBS) with 6% sucrose, then post fixed for 1 h with 2% osmium tetroxide (0.5 M sodium acetate). Following a triple rinse with 0.5 M sodium acetate buffer, the cells were incubated overnight with 0.5% uranyl acetate. The cells were then dehydrated in an ethanol series (50%, 70%, 90%, 95%, and 100%), followed by propylene oxide treatment (first straight, then 1:1 with Spurr's low-viscosity embedding medium) and embedded in Spurr's. Ultrathin sections were observed on a JEOL 1210 TEM (JEOL, Peabody, MA, USA) at 80 kV equipped with an ORCA HR digital camera (Hamamatsu, Bridgewater, NJ, USA).

2.5. DNA Extraction

Approximately 10 mL of cells in mid-exponential growth phase were pelleted by centrifugation. Cells were washed with 1 mL of 50 mM Tris HCl, pH 7.5, and pelleted a second time by centrifugation. The resulting pellet was reconstituted in 370 μL of 50 mM Tris HCl, 100 mM NaCl, 20 mM EDTA, 300 μg RNase A, and 400 μg lysozyme, and incubated in a water bath at 58 °C for 1 h. To this mixture, 200 μg proteinase K and 20 μL of 20% SDS (final concentration 1%) were added and incubated an additional 2 h at 58 °C, inverting several times during the incubation. The mixture was then cooled for 5 min in an ice bath before pelleting residual cell debris by centrifugation at 10,000× *g* for 10 min. To the supernatant, a solution of 150 mM Tris HCl pH 7.5, 40 mM EDTA, 4% cetyl trimethyl ammonium bromide (CTAB) and 2.4 M NaCl was added in a 1:1 ratio and incubated at 58 °C for 15 min. The mixture was cooled on ice for 5 min, combined with

600 μL of chloroform and mixed by inversion for 3 min. Phase separation was achieved by centrifugation at 10,000× *g* for 10 min and the upper aqueous phase was transferred to a new 1.5 mL microcentrifuge tube.

An additional RNA removal step was accomplished by adding 200 μg of RNase A and incubating at room temperature for 15 min before performing a second chloroform addition (500 μL). After mixing and centrifugation, the aqueous phase was transferred to a new 1.5 mL microcentrifuge tube. DNA was precipitated by adding 2 volumes of a solution containing 100 mM Tris HCl pH 7.5, 20 mM EDTA and 2% CTAB. DNA was pelleted by centrifugation and washed with 1 mL 70% ethanol (EtOH). This genomic DNA (gDNA) pellet was washed again with 1 mL 70% EtOH for 30 min on a rotary mixer. The gDNA pellet was then washed a final time with 1 mL 70% EtOH before being reconstituted in 15 μL 10 mM Tris HCl, 50 mM NaCl pH 7.5 in nuclease-free water. DNA quality was assessed by fluorometric quantification (Qubit, Invitrogen, Waltham, MA, USA) and UV-Vis microvolume spectrophotometry (Nanodrop, Thermo Fischer Scientific, Waltham, MA, USA). All gDNA manipulation was performed using wide-bore pipet tips to prevent gDNA shearing.

2.6. DNA Sequencing

Native genomic DNA (gDNA) was sequenced on a MinION R9.4 flowcell (Oxford Nanopore Technologies (ONT), Oxford, UK). The sequencing library was prepared using the ONT Rapid Barcoding Sequencing kit (SQK-RBK004) according to the manufacturer's protocol, with the following modifications: two separate 0.2 mL PCR tubes, 1 and 0.5 μg gDNA, were diluted to 9 μL in ONT EB (10 mM Tris, 50 mM NaCl, pH 7.5). The barcoded fragmentation mix was added in a ratio of 3:1 and 1:1 (μg gDNA:μL fragmentation mix) to the 1 μg and 0.5 μg samples, respectively. Half the library (~0.75 μg) was loaded onto the MinION flowcell. ONT EB was used to bring the total library volume to 75 μL prior to loading. Sequencing was performed for 5 h on a flowcell with approximately 1700 active pores. Sequencing Fast5 files were base-called using the ONT Guppy base-caller (v3.2.2) with GPU acceleration on a laptop with an Intel i7-6550U processor and 8 GB RAM connected to external GPU housing with an Nvidia GTX1070 (1920 CUDA cores, 8 GB VRAM) via a Thunderbolt 3 connection. Quality filtering was enabled with default settings using the high-accuracy base-calling algorithm.

An Illumina library was made using the Nextera kit (Illumina, San Diego, CA, USA) following the protocol recommended by the manufacturer and then pair-end sequenced using shotgun sequencing on the HiSeq 2500 platform at the University of Illinois at Chicago (Chicago, IL, USA) Genome Research Core DNA services facility.

2.7. Genome Sequence Assembly and Analysis

Nanopore reads were assembled using Flye v2.6 [33] with the arguments "–plasmids –meta–asm-coverage 50–min-overlap 2500". As the original culture was unialgal but not axenic, the resulting contigs were filtered to identify those likely to originate from members of the genus Acaryochloris by comparing predicted proteins identified using PRODIGAL v2.6.3 [34] to the NCBI nr database using BLASTP in the Basic Local Alignment Search Tool (BLAST) 2.5.0+ [35]. The top five hits for each protein, based on E-value, were used to filter out any contigs that coded for a majority of proteins (>50%) with no matches to Acaryochloris strains in their top five hits. Filtered contigs were cleaned twice with Pilon v1.23 [36] using paired Illumina reads to create a hybrid assembly. Contigs were manually verified to be circular by identifying overlapping regions at the start and end of linear FASTA sequences using BLASTN. Redundant sequence information present at the beginning and end of any contigs was removed. Coverage from Nanopore reads was extracted from the output file of the Flye assembler, while the coverage from Illumina reads were extracted using bwa 0.7.5a-r405 [37] followed by Samtools v1.9 [38] to map reads back to assembled contigs.

Final published annotation was generated through NCBI using their NCBI Prokaryotic Genome Annotation Pipeline (PGAP). The resulting proteins obtained from PGAP were used to calculate the amino acid identity (AAI) shared with other publicly available species of *Acaryochloris* (accession numbers in Table S1; access date 1 December 2021) using CompareM v0.1.1 (https://github.com/dparks1134/CompareM). Insertion elements were calculated using isescan 1.7.2. [39] using its default settings. CRISPR elements were identified using CRISPRCASTyper [40].

Visualization of the *Acaryochoris* sp. Moss Beach genome and its plasmids versus closely related strains was generated using the CGView_comparison tool [41]. The scripts cgview_comparison_tool.pl and redraw_maps.sh were used, and the configuration file project_settings.conf was modified to generate the comparative analysis. The gene content and synteny of *Acaryochloris* sp. Moss Beach, *A. marina* MBIC11017, and *A. marina* S15 [24,26] were visualized using Artemis Comparison Tool (ACT) v18.1.0 [42], using BLASTN results to display similarity, insertions, and rearrangements across the entire genomes. Finally, reciprocal BLAST hits were generated between strains Moss Beach and S15 using BLASTP and a custom python script to calculate and display orthologous proteins between these closely related strains.

3. Results

3.1. Cultures

The species of red and brown macroalgae that were collected on the field trip of 2 August 2015, are listed in Table 1. We note that some of these species can appear very green, but are classified as red or brown. *Neogastroclonium* and *Chondracanthus* are very dark green, and *Mazzaella flaccida* bright green, but they are red algae. *Desmarestia* is a brown alga that turns green when out of water, because H_2SO_4 in its cells is released and destroys its brown pigments. The table marks which samples contained Chl *d* by directly solvent-extracting from algal clippings (analyzed via HPLC) ("p"), and which samples yielded successful cyanobacterial cultures ("c"). The enrichment in far-red LED light and FeMBG-11 growth medium produced growth of far-red cyanobacteria for 16 of 26 inoculants from the August samples. While solvent extracts of clippings did not necessarily yield Chl *d* for some samples (likely due to low concentrations), the same algal sample sometimes yielded far-red enrichment cultures; likewise, algal clippings that had the presence of Chl *d* did not always yield viable cyanobacterial cultures.

The cultures exhibited several different modes of growth: benthic biofilm, planktonic suspensions, and clumped cells settled on the bottom of the flask. Due to slow growth and the difficulty of establishing pure cultures of each enrichment, we focused on characterizing one culture originating from A-19. We continued to maintain only cultures from a few lineages, as highlighted in Table 1. Cells from *A-19* as well as from cultures from other algal samples were unicellular, ovoid, 2–3 μm long, and 1.5–2 μm in diameter (Figure 3).
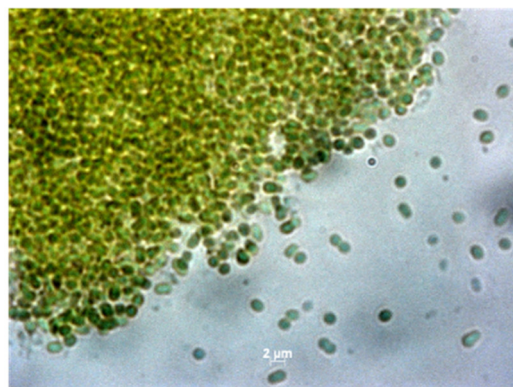
Figure 3. Phase contrast microscope images at 100X magnification showing cells of the *Acaryochloris* sp. culture from Moss Beach, CA, USA.

Table 1. Macroalgae species sampled from Montara State Marine Reserve, Moss Beach, California, USA on 2 August 2015. Column marked 'c' is for species from which enrichments produced far-red cyanobacterial cultures; 'p' for samples of algae clippings that were directly extracted and analyzed via HPLC. The species marked with an asterisk (*) are those for which cultures continue to be maintained. Further analyses in this paper were conducted on A-19, *Stephanocystis osmundacea*, a brown alga.

Sample No.	Species	Algae Color Class	c-Culture p-Pigment
A-1	*Plocamium pacificum*	red	
A-2	*Erythrophyllum delesserioides*	red	
A-3	*Microcladia coulteri*	red	
A-4	*Desmarestia herbacea*	brown	
A-5	*Costaria costata*	brown	
A-6	*Prionitis sternbergii*	red	C
A-7	*Farlowia compressa*	red	
A-8	*Cryptopleura lobulifera*	red	
A-9	*Neorhodomela larix*	red	
A-10	* *Osmundea spectabilis*	red	C
A-11	* *Neogastroclonium subarticulatum*	red	C
A-12	*Erythrophyllum delesseroides*	red	C
A-14	*Savoiea bipinnata*	red	C
A-15	*Sarcodiotheca gaudichaudii*	red	C
A-16	*Cryptopleura ruprechtiana*	red	C
A-17	*Chondracanthus canaliculatus*	red	C
A-18	* *Chondracanthus canaliculatus*	red	C
A-19	* *Stephanocystis osmundacea*	brown	p, c
A-20	*Ptilota densa*	red	C
A-21	*Cryptopleura ruprechtiana*	red	C
A-22	*Chondracanthus exasperatus*	red	
A-23	*Gelidium coulteri*	red	C
A-24	*Mazzaella flaccida*	red	
A-25	*Mazzaella splendens*	red	C
A-26	*Pikea californica*	red	C
A-27	*Mastocarpus jardinii*	red	C

TEM images, spectral characterizations, and genome sequencing were conducted on the lineage originating from the inoculation from a clipping of *Stephanocystis osmundacea* (A-19), a brown macroalga with the popular name "chainbladder kelp" (Figure 4).

Figure 4. Brown macroalgae *Stephanocystis osmundacea*, sample A-19 from the 2 August 2019, field trip and source of inoculum for the *Acaryochloris* culture analyzed in this study.

3.2. Pigment Characterization

***In vivo* spectral absorbance.** *In vivo* spectral absorbance of our *Acaryochloris* culture cell contents (Figure 5) exhibited a Q_y band peak absorbance at 705 nm. The two Soret band peaks at 405 and 454 nm may be of Chl *d in vivo*. There appears to be a slight shoulder at 740 nm.
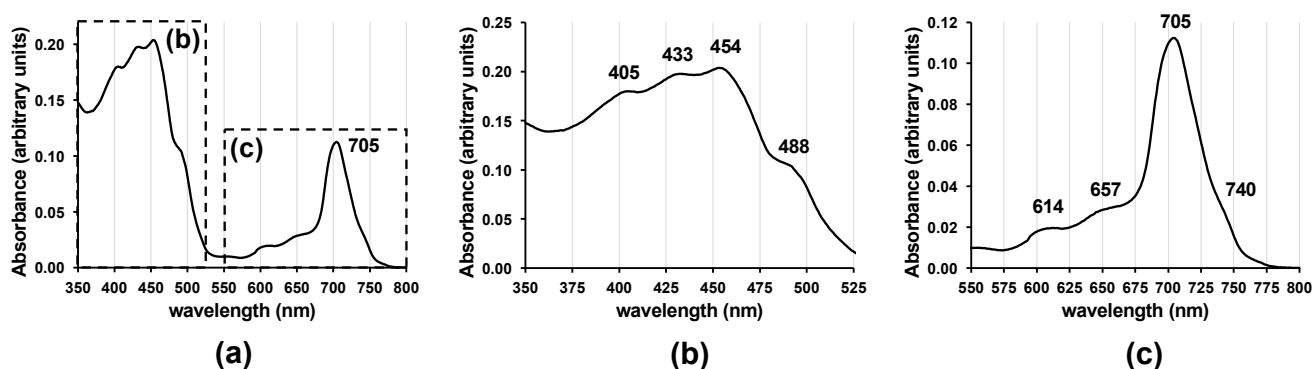
Figure 5. *In vivo* absorbance spectrum of whole cells of *Acaryochloris marina* strain Moss Beach on filter paper at a spectral resolution of 1 nm, (**a**) over 350–800 nm; (**b**) closeup over 350–525 nm; (**c**) closeup over 550–800 nm. From a culture originating from inoculation from a clipping of *Stephanocystis osmundacea*.

Pigment separation and quantification. The HPLC chromatograms are plotted in Figure 6 on the basis of Chl *a* Q_y absorption (at 660 nm) and Chl *a* Soret band absorption (at 440 nm). The chromatography allowed good separation of zeaxanthin (2.0 min), Chl *d* (~4.1 min), Chl *a* (~6.3 min), and ß-carotene (18.8 min). It is possible that the peak that we identified as ß-carotene could be or could include α-carotene, which has a very similar retention time and visible spectrum, but we did not have separate standards for each at the time. Since α-carotene is rarely seen in seawater samples, it is most likely the peak at 18.8 min is ß-carotene. A shoulder shortly after the peak for Chl *d* had a retention time less than a minute after, which could be consistent with the presence Chl *d'*, which is known to occur in PS I in MBIC11017 [5,43–46], or a degradation product. However, the absorbance spectrum of the shoulder shows no difference from Chl *d*.
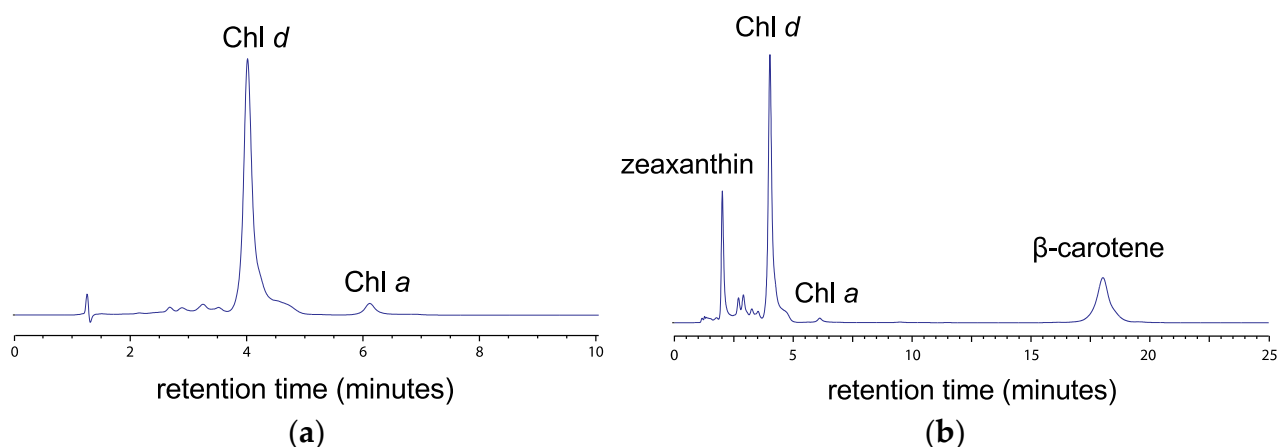
Figure 6. HPLC chromatograms for *Acaryochloris marina* strain Moss Beach (**a**) at 660 nm and (**b**) at 440 nm.

Chl *a*/Chl *d* ratio. The spectral absorbance of the purified Chl *d* in 100% methanol (Figure 7) matched those of Li, et al. [30]. The Chl *a*/Chl *d* ratio was 0.01, or Chl *d* comprised 99% of the total mass of Chl *a* and Chl *d*. The mass of Chl *d* per cell biomass obtained was 28.9–30.4 μg-Ch *d*/g biomass (2 samples), or 32.2–33.9 mol-Chl *d*/g biomass based on an average Chl *d* molecular weight of 895.5 [30], and these actual values may be 10–20% lower given the underestimate of total biomass from the extraction procedure. However, the estimated Chl *d*/Chl *a* ratio was not affected by these systematic biases in the total biomass estimate.
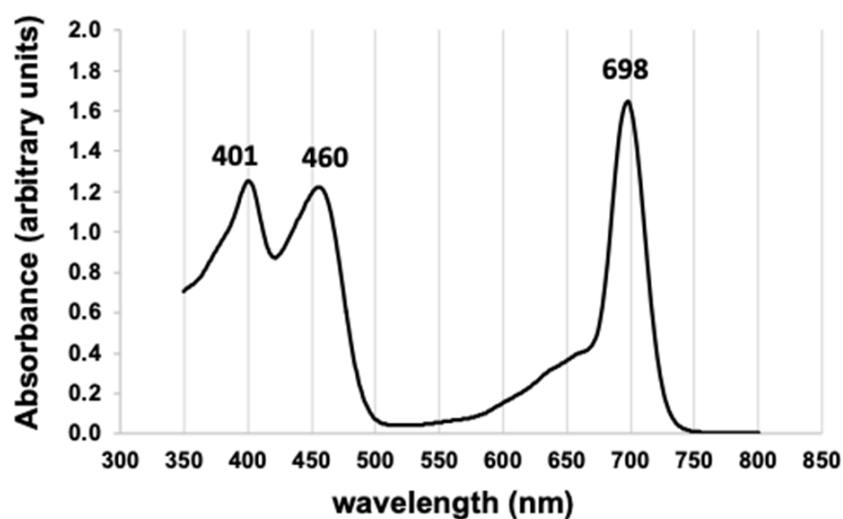
Figure 7. Absorbance spectrum of purified Chl *d* in 100% methanol, with wavelengths of local peak absorbance indicated.

3.3. TEM Images

TEM images (Figure 8) show the Moss Beach cells to be rod-shaped, ~1.3 μm in diameter, and ~2.6 nm in length. The cell sizes are similar to the other strains described [2,16]. Parietally arranged thylakoid membranes line the cell-wall interior [47]. Other notable features are the presence of carboxysomes (Figure 8b) in the cytoplasm and polyphosphate globules between the cytoplasmic membrane and the thylakoids.
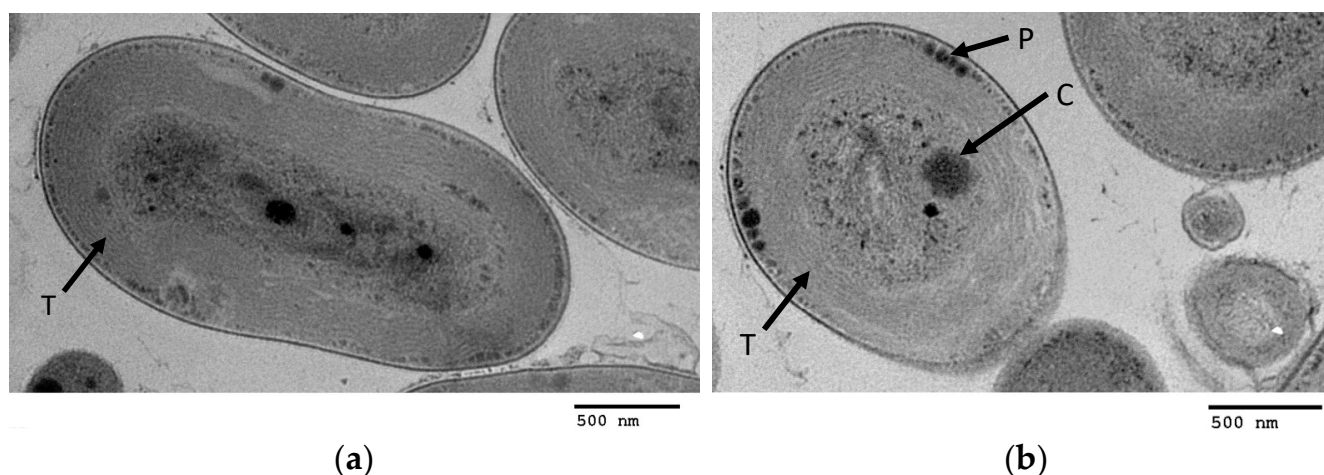
Figure 8. Oblong (**a**) and transverse (**b**) cross-sectional images of cells by transmission electron microscopy (TEM) showing the parietally arranged thylakoid membrane stacks (T) and polyphosphate granules (P) lined up along the cytoplasmic membrane of the cell wall, and carboxysome (C) in the center of the cell.

3.4. Genome Statistics

The complete genome assembly of *Acaryochloris* sp. strain Moss Beach yielded a circular chromosome of 7.25 Mbp and 10 circular plasmids ranging from 16 kbp to 394 kbp. This is consistent with the genome of type-strain *Acaryochloris marina* MBIC11017, at 8.36 Mbp and 9 plasmids, as summarized in Table 2 for genome features and Table S2 for guanine-cytosine (GC) content. The average coverage depth for nanopore sequence data was 174X and Illumina polishing coverage was estimated at 7X coverage. The GC average for the entire genome (the chromosome and all 10 plasmids) was 45%, with an average coding density of 78%. The genome codes for 5843 open reading frames, smaller than the 8528 for *A. marina* (Table S2). The genome size and statistics are consistent with other published strains [24,25,48,49], though the 10 plasmids in strain Moss Beach are the highest of any strain yet sequenced. CRISPR-Cas analysis on the Moss Beach strain showed that it lacks the *cas1* and *cas2* genes, consistent with previous studies in *Acaryochloris* [50], which showed that they lacked a *cas* operon and multiple spacer units.

Table 2. Genome features of strain Moss Beach (MB).

	Genome	pMB01	pMB02	pMB03	pMB04	pMB05	pMB06	pMB07	pMB08	pMB09	pMB10
Genome size	5,709,274	394,275	276,907	187,211	180,652	148,788	140,525	93,257	89,712	27,872	16,256
G+C content	47.01	45.67	46.16	46.61	46.32	45.52	45.52	44.53	44.34	42.33	42.14
Open reading frames	4589	311	212	155	143	128	114	73	81	25	12
Number conserved hypothetical	1075	143	63	75	81	71	77	41	53	17	5
Coding density	84.24	81.93	83.43	81.80	84.78	80.77	77.37	78.64	75.72	67.95	63.40
Average gene length	931	867	849	859	969	828	755	733	799	676	736
Ribosomal RNAs	6	0	0	0	0	0	0	0	0	0	0
Transfer RNAs	65	0	0	0	0	0	0	0	0	0	0
Pseudogenes	502	57	60	23	15	17	30	27	4	3	2
Insertion elements	150	40	39	9	6	12	12	23	3	1	1
Copy number (approximate)	1	1	1	1	1	1	1	1	1	1	1

3.5. Relationship of Strain Moss Beach with Other Acaryochloris Strains

Acaryochloris sp Moss Beach shares the highest average amino acid identity (AAI) with strain S15 [24], at ~97%, which is higher than the shared AAI with strains CCMEE 5410 and MBIC11017 at 89.1% and 88.9%, respectively (Table 3). The only sequenced *Acaryochloris* strain that does not produce the genus' signature Chl *d*, *Acaryochloris thomasi* RCC1774 [49], shares just ~65% AAI with the Moss Beach strain, reinforcing its position as a deeply
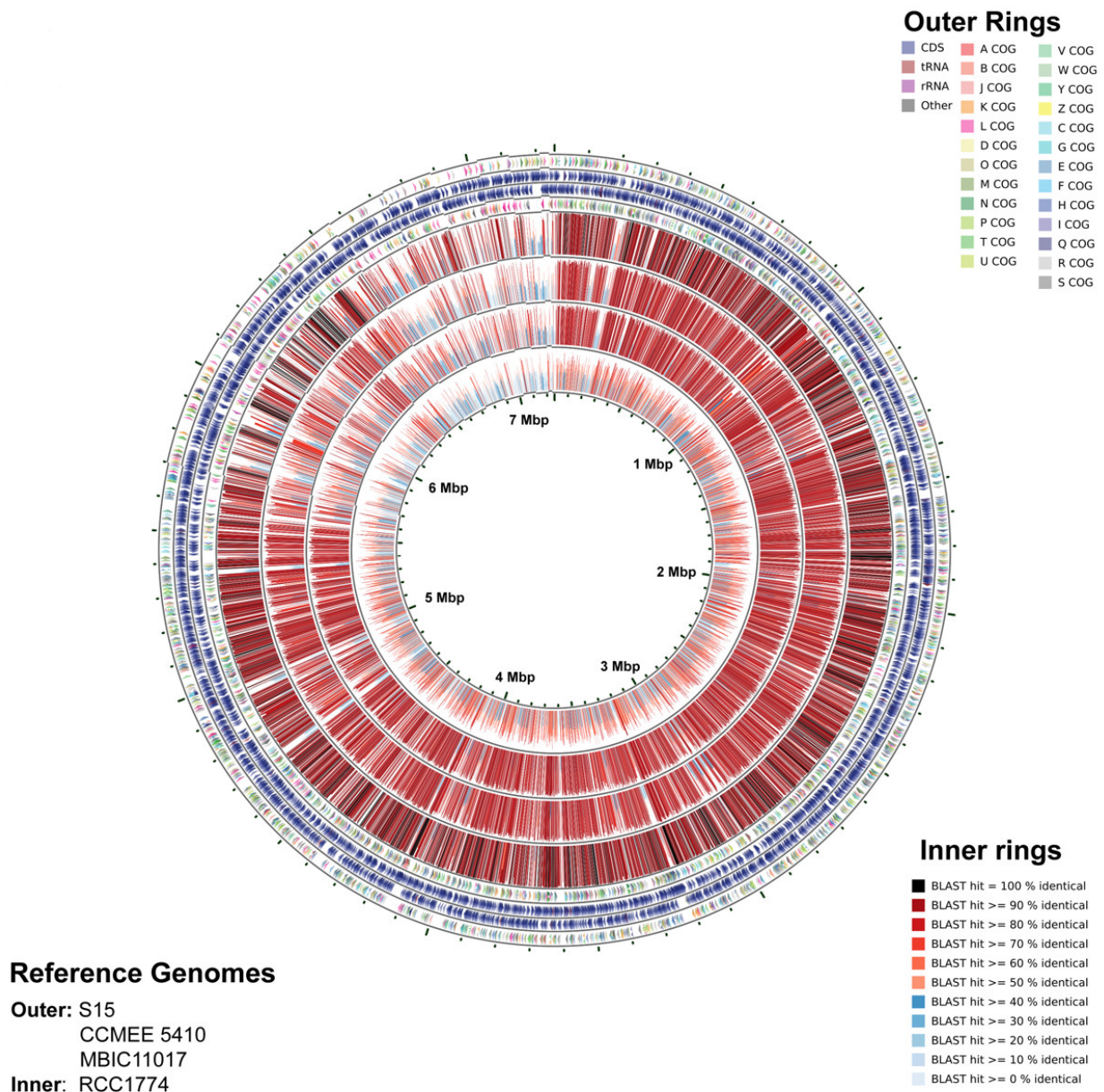
branching member of the genus, or perhaps suggesting that RCC1774 is not actually a true member of this genus. Finally, a set of draft (i.e., incomplete) metagenome-assembled genomes, SU_5_25, CRU_2_0, and RU_4_1 [51], share between 75–76% AAI with strain Moss Beach (Table 3).

Table 3. Amino acid identity (AAI) comparison of strain Moss Beach to other *Acaryochloris* strains.

Genome A	Genes in A	Genome B	Genes in B	Mean AAI (%)
Moss Beach	6583	S15	6307	97.6
Moss Beach	6583	CCMEE 5410	7493	89.1
Moss Beach	6583	MBIC11017	7495	88.9
Moss Beach	6583	SU_5_25	2261	76.4
Moss Beach	6583	CRU_2_0	4206	75.5
Moss Beach	6583	RU_4_1	4566	75.4
Moss Beach	6583	RCC1774	5531	65.8

Coding sequences (CDS) in strain Moss Beach are generally consistent with other Chl *d*-producing strains (Figure 9), with greater than 90% of CDS on the main chromosome shared with other strains. However, this continuity was much lower with RCC1774 (Figure 9, inner ring). While much of the Moss Beach chromosome shares the highest similarity to strain S15 (Figure S1), there are regions where the other two Chl *d* strains have higher similarity; for example, eight o'clock in Figure 9. This disruption in the close relationship between Moss Beach and S15 is more apparent in the plasmids, roughly 9 to 12 o'clock, whereas the Moss Beach strain shares either a lower similarity to S15 than other strains, or no similarity to any sequenced *Acaryochloris*. Further comparison of individual plasmids to S15 and MBIC11017 is shown in Figure S2.

Closely related strains Moss Beach and S15 shared approximately 5000 reciprocal best BLAST hits (Figure 10), as anticipated from their close AAI. Each genome codes for more than 1000 proteins that are not shared between the two strains, many of which were coded on plasmids. The comparison of orthologous proteins to other published *Acaryochloris* genomes is shown in Table S3 and the tree in Figure 11.

Outer Rings

CDS	A COG	V COG
tRNA	B COG	W COG
rRNA	J COG	Y COG
Other	K COG	Z COG
	L COG	C COG
	D COG	G COG
	O COG	E COG
	M COG	F COG
	N COG	H COG
	P COG	I COG
	T COG	Q COG
	U COG	R COG
		S COG

Inner rings

- BLAST hit = 100 % identical
- BLAST hit >= 90 % identical
- BLAST hit >= 80 % identical
- BLAST hit >= 70 % identical
- BLAST hit >= 60 % identical
- BLAST hit >= 50 % identical
- BLAST hit >= 40 % identical
- BLAST hit >= 30 % identical
- BLAST hit >= 20 % identical
- BLAST hit >= 10 % identical
- BLAST hit >= 0 % identical

Reference Genomes

Outer: S15
CCMEE 5410
MBIC11017
Inner: RCC1774

Figure 9. CGview Comparison Tool map representing the annotation of *Acaryochloris marina* strain Moss Beach (outer narrow rings) and BLAST comparison to *A. marina* S15 (outer wide band), *Acaryochloris* sp. CCMEE 5410 (next inward), *A. marina* MBIC11017 (next inward), and *Acaryochloris thomasi* RCC1774 (innermost wide band). In order from the outermost ring, each ring represents: Ring 1 and 4, COG functional categories for forward strand and reverse strand CDS; Ring 2 and 3, forwards and reverse strand sequence features; Rings 5–8, BLAST results (highest identity) compared to *Acaryochloris* genomes referenced above.

Figure 10. Reciprocal best BLAST hits for orthologous protein in *A. marina* strains Moss Beach and S15.
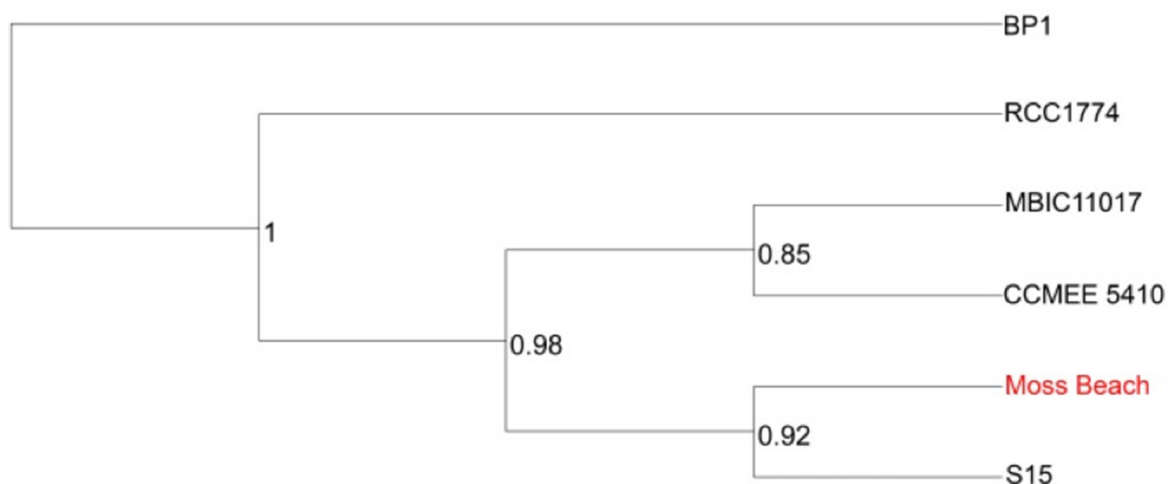
Figure 11. Phylogenetic tree of 6 concatenated OrthoFinder v2.5.4 [52] groups containing Species Tree inference from All Genes (STAG) support values at internal nodes. *Thermosynechococcus elongatus* BP-1 was used as an outgroup. The STAG support values represent the proportion of orthologous gene trees that were used to create the consensus tree that supports the particular branching pattern at that node.

3.6. Light-Harvesting Genes

Like the *Acaryochloris* type strain MBIC11017, the Moss Beach strain has two major light-harvesting and photoprotection systems along with accessory chlorophyll-binding proteins (CBPs) (Table S4). For the photosystem proteins and the CBP, a high percent identity suggests low divergence of these proteins between the two *Acaryochloris* strains. However, these two strains have a major difference in phycobiliprotein composition, as seen in the heat map in Figure 12. The divergence in phycobiliproteins in MBIC11017 is documented [24] and is a consistent feature in other Chl-*d*-containing *Acaryochoris* strains, though non-Chl *d*-producing strain RCC1774 does appear to have more similarity to MBIC11017 (Figure 12). Other than MBIC11017, all Chl-*d*-producing strains appear to produce only allophycocyanin, while, like MBIC11017, the non-Chl *d*-producing RCC1774 may also produce phycocyanin. As many of the large number of phycobiliproteins coded in MBIC11017 were found on plasmid pREB3 [25], the lack of plasmid similarity between strains likely accounts for this divergence in MBIC11017.
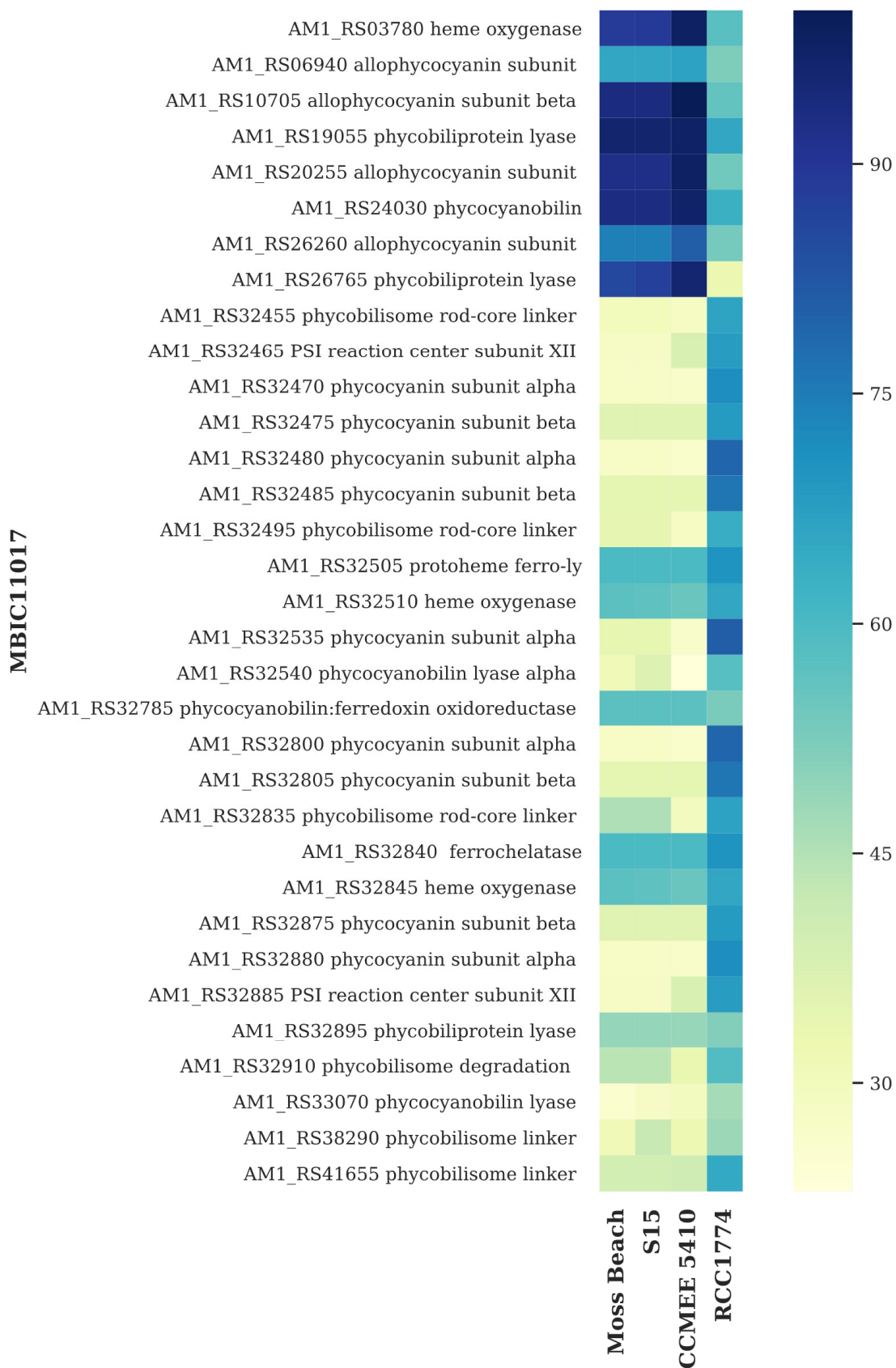
Figure 12. Heatmap displaying phycobiliprotein similarity of *Acaryochloris* species versus type strain MBIC11017 as the reference. Darker color indicates higher percent amino acid identity to the closest homolog in each of four strains (bottom).

4. Discussion

Algal host. *Acaryochloris* species or strains have been found as epiphytes on a variety of red algae: *A. marina* strain Awaji on the red alga *Ahnfeltiopsis flabelliformis* [15]; *Acaryochloris* sp. MPGRS1 on *Gelidium caulacantheum* [13]; *A. marina* S15 on *Pikea pinnata* [16,24], with three other *Acaryochloris* strains on red algae or red seaweed at the same site [24]. Ulrich, et al. [24] found at least ten more *Acaryochloris* strains on red algae from a wide sampling around Belize, Japan, the South China Sea, and the Arabian Sea. Manning and Strain [1] found Chl *d* on at least 18 species of red algae. From a preliminary diagnostic field trip we made to Moss Beach on 18 March 2015, we also found 16S rRNA evidence of *Acaryochloris* species on several algal hosts. A culture from an inoculation with a clipping of *Gelidium purpurascens* produced a 16S rRNA gene sequence with 99.65% sequence alignment with *Acaryochloris* sp. strain Awaji [15], a strain obtained from rinsing the surface of red macroalgae. A culture from inoculation with a clipping of *Cryptopleura ruprechtiana* yielded sequences with 100% coverage and 98% identity with *A. marina* CRS, *Acaryochloris* sp. CCMEE 5410, and *A. marina* MBIC 11017; 89% coverage and 99% identity with *Acaryochloris* sp. Awaji-1; and 99.6% identity with an unpublished sequence in the NCBI database (NIES 2412, Yamaguchi et al., submitted 6 October 2016). This 16S sequence also aligned perfectly with a culture from a clipping of *Polyneura latissima*.

From our 2 August 2015, field samples (Table 1), we used several algae species in common with those of Manning and Strain [1] as sources of inoculum, and additionally included brown algae. Far-red cyanobacterial cultures were successfully cultivated from both red and brown algae under the far-red light. As we only fully characterized one lineage of cultures from a brown alga, *Stephanocystis osmundacea,* we cannot assume that the Moss Beach site does not harbor multiple strains or species of far-red cyanobacteria, especially given that Ulrich, et al. [24] found four different *Acaryochloris* strains at the same Northern California site. The different modes of growth observed possibly suggest several different strains of *Acaryochloris* present or phenotypic plasticity within a strain. The broad occurrence of these far-red cyanobacteria across the several algae species also could indicate a cosmopolitan epiphyte growing on macroalgae in a variety of intertidal areas.

In vivo **absorbance spectrum.** The *in vivo* Chl *d* Q_y absorbance peak of the Moss Beach culture occurred at 705 nm (obtained by gently vacuum-filtering whole cells onto GF/C filters) (Figure 5). This is much shorter in wavelength than the Q_y peak at 714–718 nm measured by both Miyashita, et al. [2] and Li, et al. [53] for the type strain of *Acaryochloris marina* MBIC11017 (obtained by measuring whole cells suspended in growth medium). Murakami, et al. [15] found their *Acaryochloris* strain Awaji (from red algae) displayed a Q_y peak at 711 nm (using a photodiode array detector under a microscope on whole cells grown under fluorescent light). A spectrum of whole intact cells of MBIC11017 measured on filter paper [9] displayed the Q_y peak at 710 nm. Mohr, et al. [14] measured an *in vivo* absorbance Qy peak at 707 nm in *Acaryochloris* sp. HICR111A, in a culture obtained from dead coral at Heron Island, Australia, and grown in near-infrared light centered at 720 nm. This shorter wavelength peak was attributed by Mohr, et al. [14] to the greater difficulty of maintaining homogenous cell suspensions while measuring the spectra of whole cells in glycerol (Wolfgang Hess, personal communication). However, Duxbury, et al. [54] used a Taylor sphere to measure the *in vivo* absorbance spectra of suspended whole cells (clumps loosened through brief bead beating or sonication; Min Chen, personal communication), observing the Q_y peak centered at 709 nm in MBIC11017 and at 707 nm in CCMEE 5410.

Duxbury, et al. [54] discussed the differences among studies that found the Q_y peak at longer wavelengths for MBIC11017, and noted that one such study [55] had an uncorrected baseline, while differences between strains could be due to small differences in Chl *d* pigment-binding proteins or to different contents of phycobilins. Thus, the differences observed in the Q_y peak wavelength for the same strain, as well as among the variety of strains of *Acaryochloris*, may be due to intrinsic differences between strains, culture conditions, pigment ratios (Chl *a*, Chl *d*, phycobilins), PSII/PSI ratios, differences between measurement techniques, as well as data smoothing. The Moss Beach strain's *in vivo*

absorbance spectrum is plotted together with that of MBIC11017 and CCMEE 5410 from Duxbury et al. [1] in (Figure 13). Thus far, it exhibits the shortest measured wavelength for the Q_y peak among known strains. The slight *in vivo* absorbance shoulder at 740 nm is not normally apparent in absorbance spectra of other *Acaryochloris* strains. It could be due to P740, but as yet we do not have an explanation.
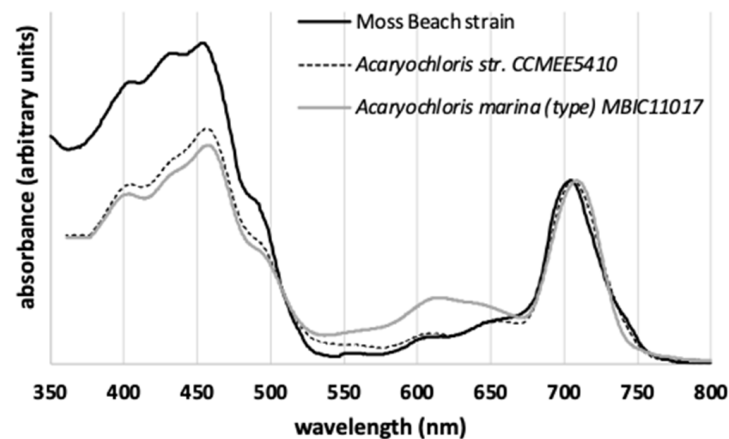
Figure 13. Comparison of the *in vivo* absorbance spectrum of the Moss Beach strain and those by Duxbury et al. [45] for *Acaryochloris marina* MBIC11017 and *Acaryochoris* str. CCMEE 5410. Only MBIC11017 shows the presence of phycobilins, manifested in the absorbance peak around 620 nm.

Chl *a*/Chl *d* ratio. Previously reported Chl *a/d* ratios and Chl *d* percent of total chlorophylls reported in *Acaryochloris* include 0.03–0.09 or 92–97% Chl *d* [3], 0.016 ± 0.003 Chl *a*/Chl *d* or 84% Chl *d* [16], and 0.016–0.039 Chl *a*/Chl *d* or 96–98% Chl *d* [54] in *A. marina Str.* MBIC11017; 0.025 ± 0.0013 [16]; and 0.022–0.032 Chl *a*/Chl *d* or 97% Chl *d* in *A. marina Str.* CCMEE 5410. While other studies grew cultures in white fluorescent light, the study by Duxbury, et al. [54] had culture treatments in white fluorescent light, and LED light at 625 and 720 nm, with the maximum Chl *d* percent occurring with the far-red LED light. Mohr, et al. [14] compared the Chl *a/d* ratios of MBIC11017 and *Acaryochloris* sp. HICR111A under low, medium, and high light regimes in LED light at 720 ± 20 nm and found MBIC11017 was able to adjust the Chl *a/d* from ~1% to as high as ~5.7%. The HICR111A strain was less able to acclimate, with Chl *a/d* ranging 2–3%. They measured an *in vivo* absorbance peak of 707 nm for HICR111A vs. 710 nm for MBIC11017. Larkum, et al. [13] found up to 90% Chl *d* of all cellular chlorophyll in *Acaryochloris* sp. MPGRS1 from red algae in Australia [13] and an *in vivo* absorbance peak at 710 nm. The high Chl *d* content of the Moss Beach strain, with Chl *a/d* at 99%, was a little higher than but comparable to that of MBIC11017 grown under both white light [16] and far-red LED light [54]. Our LED environment compared to other studies was at the longest wavelength in the far red at 730 nm.

The shoulder that occurs in the HPLC chromatogram about 1 s after the Chl *d* peak does not occur in HPLC chromatographs of pigments extracts reported in Miyashita, et al. [3] or Larkum, et al. [13]. Further study would be needed to explain if it is Chl *d* or a degradation product. *A. marina* strain MBIC11017 also contained trace amounts of phycobiliproteins, a Chl *c*-like pigments, and α-carotene but not β-carotene [3].

TEM. TEM micrographs of the ultrastructure of different *Acaryochloris* species have been published for *A. marina* MBIC1017 [2,3,14,49,53,56–60], sp. CCMEE 5410 [16], sp. CCNUM4 [61], and sp. RCC1774 [26,49]. Direct comparisons of the imaged ultrastructural features, however, are complicated by different growth conditions (e.g., iron-rich, iron-poor, far-red vs. white light) and the methods of fixation. TEM images exist in the literature that used the same fixation methods as ours for MBIC11017 [3,26,49,57,60]. CCMEE5410 was post-fixed with permanganate by Swingley, et al. [26] and with osmium tetroxide (the same as our method) by Miller, et al. [16]. In our TEM images, the Moss Beach strain showed

a parietal arrangement of the thylakoids, carboxysomes, and polyphosphate granules, features seen in other similarly prepared cells of *A. marina* strains MBIC11017 [2] and CCMEE 5410 [16]. MBIC11017 appears to have a tendency toward thylakoid membrane layers with wavy or "pinched" patterns (*sinsu* Swingley, et al. [26]), whereas CCMEE 5410 and the Moss Beach strain have the thylakoids parallel to the cells walls. However, we cannot conclude whether these are consistent differences or due to variation in growth conditions. CCMEE 5410 and the Moss Beach strain may be the most similar, possibly because MBIC11017 is the only strain to have phycocyanin. The Moss Beach strain does not show evidence of phycobilisomes, and we can also rule out their presence according to the absorbance spectra above and genome analysis below.

Genome. The high amino acid identity between strains Moss Beach and S15 suggests that both strains belong to the same species, identified as *Acaryochloris marina* in [24], which is in line with the shared 100% identity between Moss Beach and S15 16S rRNA genes and average nucleotide identity of 98%. While both strains were identified on the Pacific Coast of California, USA, the relatively high number of unique proteins in each strain (Figure 10), and lack of similarity in plasmid-encoded genes distinguishes these two as distinct strains. The S15 strain was isolated from Shelter Cove, Northern California, USA (40.022251° N, 124.0071373° W), while Moss Beach is located southward off the coast of Central California.

Phycobiliproteins aid in both light harvesting and photoprotection in most cyanobacteria. Phycobiliproteins were found to be associated with unique rod-shaped phycobilisomes in *A. marina* MBIC11017 from the Palau Islands, Japan [60]. *Acaryochloris* sp. MPGRS1 from red algae in a mangrove in Australia, was found to lack the gene for phycocyanin [13]. Miller, et al. [16] determined that *A. marina* strain CCMEE 5410 from the Salton Sea, California, does not produce phycobiliproteins. Mohr, et al. [14] found that *Acaryochloris* sp. HICR111A from Heron Island, Australia, lacks phycocyanin and phycoerythrin genes but has allophycocyanin genes. While type strain *A. marina* MBIC11017 produces a large amount of allophycocyanin and phycocyanin, as indicated by its dark color and signature "blue" phycobiliprotein absorbance peaks (see Figure 13 and Duxbury, et al. [54]), culture observation and spectra of strain Moss Beach suggest that this strain is significantly more deficient in these light-harvesting systems. Unlike MBIC11017, in which most of the phycobiliprotein-related genes are found on plasmid pREB3 [25], all of those in Moss Beach are coded on its main chromosome. Beyond the lack of phycocyanin-coding genes, Moss Beach only codes for a single set of genes for producing allophycocyanin, the most far-red light-absorbing phycobiliprotein, versus three copies of these genes in MBIC11017. This phycobiliprotein content was consistent with findings in other *Acaryochloris* strains [24], which suggested that phycocyanin-based light harvesting is exclusive to MBIC11017, allowing increased plasticity for acclimation to yellow light.

Given the high number of plasmids in the Moss Beach strain (and other *Acaryochloris* strains) compared to other bacteria, we checked whether any of these may be classified as chromids. Chromids are stable and essential extra-chromosomal replicons in bacteria [62], compared to the less stable plasmids, the latter being important agents for mediating lateral gene transfer [63,64]. We generated two dendrograms (not shown), by calculating distance based on tetranucleotide frequency between: (1) the chromosomes of MBIC11017, Moss Beach, and S15, and their respective plasmids; and (2) between the chromosome of Moss Beach and its plasmid. Both the dendrograms and a non-metric multidimensional scaling (NMDS) clustering analysis (not shown) showed no clear association between *Acaryochloris* chromosomes and any of the designated plasmids, suggesting they would not qualify as chromids. Harrison, et al. [62] found that about 10% of 897 fully sequenced bacteria possess chromids, among which were one genus of cyanobacteria, and that possession of chromids appears genus specific. Therefore, we conclude that the genus *Acaryochloris* does not have chromids.

5. Conclusions

We have isolated and identified an *Acaryochloris* from the site where chlorophyll *d* was originally discovered by Manning and Strain [1]. Whether this is the same organism from which these early researchers originally extracted Chl *d*, we cannot know, and it may be one of several Chl *d*-bearing strains or species coexisting at the site as epiphytes on multiple algae species.

Our strain is the first *Acaryochloris* to be isolated from a brown alga rather than red algae, although it is likely that it is also present on the red algae at the same site, from which far-red cyanobacteria cultures were also broadly obtained. Its *in vivo* spectral absorbance showed a Q_y peak centered at 704–705 nm, the shortest measured for *Acaryochloris* thus far. The range of values found in the literature both for the same strain as well as among strains has yet to be precisely explained. The Chl *d* content in the Moss Beach strain was very high, at 99% of total chlorophyll mass and a Chl *a*/Chl *d* ratio of 0.01, similar to that of MBIC11017. The TEM micrographs indicate the absence of phycobilisomes, which is further corroborated by genome analysis. The TEM imaging also suggests internal morphology closest to CCMEE 5410, as shown in Figure 2 from Miller, et al. [16].

Here, we contribute the full genome of the Moss Beach strain, which will add to the other three full genome sequences for *Acaryochloris*, the others being the type-strain MBIC11017, CCMEE 5410, and S15 [24,48]. The Moss Beach strain bears close 16S rRNA resemblance to the *Acaryochloris* sp. strain Awaji, discovered by Murakami, et al. [15] in Japan, for which there is no full genome sequence. However, it is closest in full genome comparison to strain S15 from the California coast [65], both epiphytes of red algae in the Pacific. Like strain S15, the Moss Beach strain also lacks complex phycobilisomes, as compared to type strain MBIC11017, perhaps indicating an adaptation to their similar red (or brown for strain Moss Beach) algal epiphyte spectral environment.

Given the multiple far-red cyanobacterial cultures obtained from red algae species at Moss Beach, in future work it remains to be seen whether they are the same strain as that isolated from the brown macroalga and characterized in this study, or whether there is differentiation by algae type. It also remains to be investigated whether the Q_y peak wavelength of the Moss Beach strain is sensitive to alternative growth conditions. There is an emerging wealth of *Acaryochloris* genome sequences that offer potential for additional comparative genomics, including data by Waterworth, et al. [51], other genomes linked to S15 [24], and an expected boom in more discoveries of far-red cyanobacteria. This new full genome sequence adds to the database for comparative studies to hopefully identify the elusive gene for Chl *d*.

Due to its close phylogenetic relationship to *Acaryochloris marina* strain S15, we designate this strain consistently as *Acaryochloris marina* strain Moss Beach. The genome is available under Genbank accession number GCA_021497025.1. Cultures may be obtained from the authors upon request.

References

1. Manning, W.M.; Strain, H.H. Chlorophyll d: A green pigment in red algae. *J. Biol. Chem.* **1943**, *151*, 1–19. [CrossRef]
2. Miyashita, H.; Ikemoto, H.; Kurano, N.; Adachi, K.; Chihara, M.; Miyachi, S. Chlorophyll d as a major pigment. *Nature* **1996**, *383*, 402. [CrossRef]
3. Miyashita, H.; Adachi, K.; Kurano, N.; Ikemoto, H.; Chihara, M.; Miyachi, S. Pigment composition of a novel oxygenic photosynthetic prokaryote containing chlorophyll d as the major chlorophyll. *Plant Cell Physiol.* **1997**, *38*, 274–281. [CrossRef]
4. Li, Y.Q.; Chen, M. Novel chlorophylls and new directions in photosynthesis research. *Funct. Plant Biol.* **2015**, *42*, 493–501. [CrossRef] [PubMed]
5. Allakhverdiev, S.I.; Kreslavski, V.D.; Zharmukhamedov, S.K.; Voloshin, R.A.; Korol'kova, D.V.; Tomo, T.; Shen, J.R. Chlorophylls d and f and their role in primary photosynthetic processes of cyanobacteria. *Biochem.-Mosc.* **2016**, *81*, 201–212. [CrossRef] [PubMed]
6. Hamaguchi, T.; Kawakami, K.; Shinzawa-Itoh, K.; Inoue-Kashino, N.; Itoh, S.; Ifuku, K.; Yamashita, E.; Maeda, K.; Yonekura, K.; Kashino, Y. Structure of the far-red light utilizing photosystem I of Acaryochloris marina. *Nat. Commun.* **2021**, *12*, 2333. [CrossRef]
7. Kobayashi, M.; Ohashi, S.; Iwamoto, K.; Shiraiwa, Y.; Kato, Y.; Watanabe, T. Redox potential of chlorophyll d in vitro. *Biochim. Biophys. Acta-Bioenerg.* **2007**, *1767*, 596–602. [CrossRef]
8. Tomo, T.; Okubo, T.; Akimoto, S.; Yokono, M.; Miyashita, H.; Tsuchiya, T.; Noguchi, T.; Mimuro, M. Identification of the special pair of photosystem II in a chlorophyll d-dominated cyanobacterium. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 7283–7288. [CrossRef]
9. Mielke, S.P.; Kiang, N.Y.; Blankenship, R.E.; Gunner, M.R.; Mauzerall, D. Efficiency of photosynthesis in a Chl d-utilizing cyanobacterium is comparable to or higher than that in Chl a-utilizing oxygenic species. *Biochim. Biophys. Acta-Bioenerg.* **2011**, *1807*, 1231–1236. [CrossRef]
10. Cha, Y.; Mauzerall, D.C. Energy-storage of linear and cyclic electron flows in photosynthesis. *Plant Physiol.* **1992**, *100*, 1869–1877. [CrossRef]
11. Mielke, S.P.; Kiang, N.Y.; Blankenship, R.E.; Mauzerall, D. Photosystem trap energies and spectrally-dependent energy-storage efficiencies in the Chl d-utilizing cyanobacterium, Acaryochloris marina. *Biochim. Biophys. Acta-Bioenerg.* **2013**, *1827*, 255–265. [CrossRef] [PubMed]
12. Hu, Q.; Miyashita, H.; Iwasaki, I.; Kurano, N.; Miyachi, S.; Iwaki, M.; Itoh, S. A photosystem I reaction center driven by chlorophyll d in oxygenic photosynthesis. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 13319–13323. [CrossRef] [PubMed]
13. Larkum, A.W.D.; Chen, M.; Li, Y.; Schliep, M.; Trampe, E.; West, J.; Salih, A.; Kühl, M. A novel epiphytic Chlorophyll d-containing cyanobacterium isolated from a mangrove-associated red alga. *J. Phycol.* **2012**, *48*, 1320–1327. [CrossRef]
14. Mohr, R.; Voss, B.; Schliep, M.; Kurz, T.; Maldener, I.; Adams, D.G.; Larkum, A.D.W.; Chen, M.; Hess, W.R. A new chlorophyll d-containing cyanobacterium: Evidence for niche adaptation in the genus Acaryochloris. *ISME J.* **2010**, *4*, 1456–1469. [CrossRef] [PubMed]
15. Murakami, A.; Miyashita, H.; Iseki, M.; Adachi, K.; Mimuro, M. Chlorophyll d in an epiphytic cyanobacterium of red algae. *Science* **2004**, *303*, 1633. [CrossRef]
16. Miller, S.R.; Augustine, S.; Le Olson, T.; Blankenship, R.E.; Selker, J.; Wood, A.M. Discovery of a free-living chlorophyll d-producing cyanobacterium with a hybrid proteobacterial/cyanobacterial small-subunit rRNA gene. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 850–855. [CrossRef]

17. Chen, M.; Schliep, M.; Willows, R.D.; Cai, Z.L.; Neilan, B.A.; Scheer, H. A Red-Shifted Chlorophyll. *Science* **2010**, *329*, 1318–1319. [CrossRef]

18. Gan, F.; Zhang, S.; Rockwell, N.C.; Martin, S.S.; Lagarias, J.C.; Bryant, D.A. Extensive remodeling of a cyanobacterial photosynthetic apparatus in far-red light. *Science* **2014**, *345*, 1312–1317. [CrossRef]

19. Ho, M.Y.; Gan, F.; Shen, G.Z.; Bryant, D.A. Far-red light photoacclimation (FaRLiP) in Synechococcus sp PCC 7335. II. Characterization of phycobiliproteins produced during acclimation to far-red light. *Photosynth. Res.* **2017**, *131*, 187–202. [CrossRef]

20. Ho, M.Y.; Gan, F.; Shen, G.Z.; Zhao, C.; Bryant, D.A. Far-red light photoacclimation (FaRLiP) in Synechococcus sp PCC 7335: I. Regulation of FaRLiP gene expression. *Photosynth. Res.* **2017**, *131*, 173–186. [CrossRef]

21. Wolf, B.M.; Niedzwiedzki, D.M.; Magdaong, N.C.M.; Roth, R.; Goodenough, U.; Blankenship, R.E. Characterization of a newly isolated freshwater Eustigmatophyte alga capable of utilizing far-red light as its sole light source. *Photosynth. Res.* **2018**, *135*, 177–189. [CrossRef] [PubMed]

22. Larkum, A.W.D.; Kuhl, M. Chlorophyll d: The puzzle resolved. *Trends Plant Sci.* **2005**, *10*, 355–357. [CrossRef] [PubMed]

23. Mehda, S.; Muñoz-Martín, M.Á.; Oustani, M.; Hamdi-Aïssa, B.; Perona, E.; Mateo, P. Microenvironmental conditions drive the differential cyanobacterial community composition of biocrusts from the sahara desert. *Microorganisms* **2021**, *9*, 487. [CrossRef] [PubMed]

24. Ulrich, N.J.; Uchida, H.; Kanesaki, Y.; Hirose, E.; Murakami, A.; Miller, S.R. Reacquisition of light-harvesting genes in a marine cyanobacterium confers a broader solar niche. *Curr. Biol.* **2021**, *31*, 1539–1546.e4. [CrossRef] [PubMed]

25. Swingley, W.D.; Chen, M.; Cheung, P.C.; Conrad, A.L.; Dejesa, L.C.; Hao, J.; Honchak, B.M.; Karbach, L.E.; Kurdoglu, A.; Lahiri, S.; et al. Niche adaptation and genome expansion in the chlorophyll d-producing cyanobacterium Acaryochloris marina. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 2005–2010. [CrossRef]

26. Swingley, W.D.; Hohmann-Marriott, M.F.; Le Olson, T.; Blankenship, R.E. Effect of iron on growth and ultrastructure of Acaryochloris marina. *Appl. Environ. Microbiol.* **2005**, *71*, 8606–8610. [CrossRef]

27. Chen, M.; Quinnell, R.G.; Larkum, A.W.D. Chlorophyll d as the major photopigment in Acaryochloris marina. *J. Porphyr. Phthalocyanines* **2002**, *6*, 763–773. [CrossRef]

28. Guillard, R.R.L.; Ryther, J.H. Studies on marine planktonic diatoms I. Cyclotella nana hustedt and detonula confervacea (Cleve) Gran. *Can. J. Microbiol.* **1962**, *8*, 229–239. [CrossRef]

29. Hedges, J.I.; Baldock, J.A.; Gélinas, Y.; Lee, C.; Peterson, M.L.; Wakeham, S.G. The biochemical and elemental compositions of marine plankton: A NMR perspective. *Mar. Chem.* **2002**, *78*, 47–63. [CrossRef]

30. Li, Y.; Scales, N.; Blankenship, R.E.; Willows, R.D.; Chen, M. Extinction coefficient for red-shifted chlorophylls: Chlorophyll d and chlorophyll f. *Biochim. Biophys. Acta (BBA)-Bioenerg.* **2012**, *1817*, 1292–1298. [CrossRef]

31. Jeffrey, S.W.; Humphrey, G.F. New spectrophotometric equations for determining chlorophylls A, B, C1 and C2 in higher-plants, algae and natural phytoplankton. *Biochem. Physiol. Pflanz.* **1975**, *167*, 191–194. [CrossRef]

32. Switzer Blum, J.; Burns Bindi, A.; Buzzelli, J.; Stolz, J.F.; Oremland, R.S. Bacillus arsenicoselenatis, sp. nov., and Bacillus selenitireducens, sp. nov.: Two haloalkaliphiles from Mono Lake, California that respire oxyanions of selenium and arsenic. *Arch. Microbiol.* **1998**, *171*, 19–30. [CrossRef] [PubMed]

33. Kolmogorov, M.; Yuan, J.; Lin, Y.; Pevzner, P.A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **2019**, *37*, 540–546. [CrossRef] [PubMed]

34. Hyatt, D.; Chen, G.-L.; LoCascio, P.F.; Land, M.L.; Larimer, F.W.; Hauser, L.J. Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **2010**, *11*, 119. [CrossRef]

35. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef]

36. Walker, B.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K.; et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **2014**, *9*, e112963. [CrossRef]

37. Li, H.W. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **2013**, arXiv:1303.3997.

38. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M.; et al. Twelve years of SAMtools and BCFtools. *GigaScience* **2021**, *10*. [CrossRef]

39. Xie, Z.; Tang, H. ISEScan: Automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics* **2017**, *33*, 3340–3347. [CrossRef]

40. Russel, J.; Pinilla-Redondo, R.; Mayo-Muñoz, D.; Shah, S.A.; Sørensen, S.J. CRISPRCasTyper: Automated identification, annotation, and classification of CRISPR-Cas loci. *CRISPR J.* **2020**, *3*, 462–469. [CrossRef]

41. Grant, J.R.; Arantes, A.S.; Stothard, P. Comparing thousands of circular genomes using the CGView comparison tool. *BMC Genom.* **2012**, *13*, 202. [CrossRef]

42. Carver, T.J.; Rutherford, K.M.; Berriman, M.; Rajandream, M.-A.; Barrell, B.G.; Parkhill, J. ACT: The Artemis comparison tool. *Bioinformatics* **2005**, *21*, 3422–3423. [CrossRef] [PubMed]

43. Sivakumar, V.; Wang, R.L.; Hastings, G. Photo-oxidation of P740, the primary electron donor in photosystem I from Acaryochloris marina. *Biophys. J.* **2003**, *85*, 3162–3172. [CrossRef]

44. Tomo, T.; Kato, Y.; Suzuki, T.; Akimoto, S.; Okubo, T.; Noguchi, T.; Hasegawa, K.; Tsuchiya, T.; Tanaka, K.; Fukuya, M.; et al. Characterization of highly purified photosystem I complexes from the chlorophyll d-dominated cyanobacterium Acaryochloris marina MBIC 11017. *J. Biol. Chem.* **2008**, *283*, 18198–18209. [CrossRef] [PubMed]

45. Akiyama, M.; Miyashita, H.; Kise, H.; Watanabe, T.; Mimuro, M.; Miyachi, S.; Kobayashi, M. Quest for minor but key chlorophyll molecules in photosynthetic reaction centers-unusual pigment composition in the reaction centers of the chlorophyll d-dominated cyanobacterium Acaryochloris marina. *Photosynth. Res.* **2002**, *74*, 97–107. [CrossRef]

46. Akiyama, M.; Miyashita, H.; Kise, H.; Watanabe, T.; Miyachi, S.; Kobayashi, M. Detection of chlorophyll d' and pheophytin a in a chlorophyll d-dominating oxygenic photosynthetic prokaryote Acaryochloris marina. *Anal. Sci.* **2001**, *17*, 205–208. [CrossRef]

47. Stolz, J. Bacterial Intracellular Membranes. In *eLS*; John Wiley & Sons, Inc: Hoboken, NJ, USA, 2020; pp. 1–7. [CrossRef]

48. Miller, S.R.; Wood, A.M.; Blankenship, R.E.; Kim, M.; Ferriera, S. Dynamics of gene duplication in the genomes of chlorophyll d-producing cyanobacteria: Implications for the ecological niche. *Genome Biol. Evol.* **2011**, *3*, 601–613. [CrossRef]

49. Partensky, F.; Six, C.; Ratin, M.; Garczarek, L.; Vaulot, D.; Probert, I.; Calteau, A.; Gourvil, P.; Marie, D.; Grebert, T.; et al. A novel species of the marine cyanobacterium Acaryochloris with a unique pigment content and lifestyle. *Sci. Rep.* **2018**, *8*, 9142. [CrossRef]

50. Cai, F.; Axen, S.D.; Kerfeld, C.A. Evidence for the widespread distribution of CRISPR-Cas system in the phylum cyanobacteria. *RNA Biol.* **2013**, *10*, 687–693. [CrossRef]

51. Waterworth, S.C.; Isemonger, E.W.; Rees, E.R.; Dorrington, R.A.; Kwan, J.C. Conserved bacterial genomes from two geographically isolated peritidal stromatolite formations shed light on potential functional guilds. *Environ. Microbiol. Rep.* **2021**, *13*, 126–137. [CrossRef]

52. Emms, D.M.; Kelly, S. OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **2015**, *16*, 157. [CrossRef] [PubMed]

53. Li, Z.K.; Yin, Y.C.; Zhang, L.D.; Zhang, Z.C.; Dai, G.Z.; Chen, M.; Qiu, B.S. The identification of IsiA proteins binding chlorophyll d in the cyanobacterium Acaryochloris marina. *Photosynth. Res.* **2018**, *135*, 165–175. [CrossRef] [PubMed]

54. Duxbury, Z.; Schliep, M.; Ritchie, R.J.; Larkum, A.W.D.; Chen, M. Chromatic photoacclimation extends utilisable photosynthetically active radiation in the chlorophyll d-containing cyanobacterium, Acaryochloris marina. *Photosynth. Res.* **2009**, *101*, 69–75. [CrossRef] [PubMed]

55. Chan, Y.W.; Nenninger, A.; Clokie, S.J.H.; Mann, N.H.; Scanlan, D.J.; Whitworth, A.L.; Clokie, M.R.J. Pigment composition and adaptation in free-living and symbiotic strains of Acaryochloris marina. *FEMS Microbiol. Ecol.* **2007**, *61*, 65–73. [CrossRef]

56. Chen, M.; Floetenmeyer, M.; Bibby, T.S. Supramolecular organization of phycobiliproteins in the chlorophyll d-containing cyanobacterium Acaryochloris marina. *FEBS Lett.* **2009**, *583*, 2535–2539. [CrossRef]

57. Hu, Q.; Marquardt, J.; Iwasaki, I.; Miyashita, H.; Kurano, N.; Morschel, E.; Miyachi, S. Molecular structure, localization and function of biliproteins in the chlorophyll a/d containing oxygenic photosynthetic prokaryote Acaryochloris marina. *Biochim. Biophys. Acta-Bioenerg.* **1999**, *1412*, 250–261. [CrossRef]

58. Kashimoto, T.; Miyake, K.; Sato, M.; Maeda, K.; Matsumoto, C.; Ikeuchi, M.; Toyooka, K.; Watanabe, S.; Kanesaki, Y.; Narikawa, R. Acclimation process of the chlorophyll d-bearing cyanobacterium Acaryochloris marina to an orange light environment revealed by transcriptomic analysis and electron microscopic observation. *J. Gen. Appl. Microbiol.* **2020**, *66*, 106–115. [CrossRef]

59. Loughlin, P.; Lin, Y.K.; Chen, M. Chlorophyll d and Acaryochloris marina: Current status. *Photosynth. Res.* **2013**, *116*, 277–293. [CrossRef]

60. Marquardt, J.; Mörschel, E.; Rhiel, E.; Westermann, M. Ultrastructure of Acaryochloris marina, an oxyphotobacterium containing mainly chlorophyll d. *Arch. Microbiol.* **2000**, *174*, 181–188. [CrossRef]

61. Zhang, Z.-C.; Li, Z.-K.; Yin, Y.-C.; Li, Y.; Jia, Y.; Chen, M.; Qiu, B.-S. Widespread occurrence and unexpected diversity of red-shifted chlorophyll producing cyanobacteria in humid subtropical forest ecosystems. *Environ. Microbiol.* **2019**, *21*, 1497–1510. [CrossRef]

62. Harrison, P.; Kim, N.; Young, J.P. Introducing the bacterial 'chromid': Not a chromosome, not a plasmid. *Trends Microbiol.* **2010**, *18*, 141–148. [CrossRef] [PubMed]

63. diCenzo, G.C.; Mengoni, A.; Perrin, E. Chromids aid genome expansion and functional diversification in the family burkholderiaceae. *Mol. Biol. Evol.* **2019**, *36*, 562–574. [CrossRef] [PubMed]

64. Hulter, N.; Ilhan, J.; Wein, T.; Kadibalban, A.S.; Hammerschmidt, K.; Dagan, T. An evolutionary perspective on plasmid lifestyle modes. *Curr. Opin. Microbiol.* **2017**, *38*, 74–80. [CrossRef]

65. Miller, S.R.; Abresch, H.E.; Ulrich, N.J.; Sano, E.B.; Demaree, A.H.; Oman, A.R.; Garber, A.I. Bacterial adaptation by a transposition burst of an invading IS element. *Genome Biol. Evol.* **2021**, *13*, evab245. [CrossRef] [PubMed]

Article

A Cyanobacteria Enriched Layer of Shark Bay Stromatolites Reveals a New *Acaryochloris* Strain Living in Near Infrared Light

Michael S. Johnson [1,2,*], Brendan P. Burns [3,4], Andrei Herdean [5], Alexander Angeloski [6], Peter Ralph [5], Therese Morris [4,7,8], Gareth Kindler [4], Hon Lun Wong [3,8], Unnikrishnan Kuzhiumparambil [5], Lisa M. Sedger [2,9] and Anthony W. D. Larkum [5]

1 School of Medical Sciences, University of New South Wales, Sydney 2052, Australia
2 School of Life Sciences, University of Technology Sydney, Sydney 2007, Australia; lisa.sedger@wimr.org.au
3 School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney 2052, Australia; brendan.burns@unsw.edu.au (B.P.B.); h.l.wong@unsw.edu.au (H.L.W.)
4 Australian Centre for Astrobiology, University of New South Wales, Sydney 2052, Australia; therese.morris@blueearthscience.com.au (T.M.); garethseankindler@gmail.com (G.K.)
5 Climate Change Cluster, University of Technology Sydney, Sydney 2007, Australia; andrei.herdean@uts.edu.au (A.H.); peter.ralph@uts.edu.au (P.R.); unnikrishnan.kuzhiumparambil@uts.edu.au (U.K.); a.larkum@sydney.edu.au (A.W.D.L.)
6 School of Mathematical and Physical Sciences, University of Technology Sydney, Sydney 2007, Australia; alexander.angeloski@uts.edu.au
7 School of Earth and Planetary Sciences, Curtin University, Perth 6102, Australia
8 Department of Aquatic Microbial Ecology, Institute of Hydrobiology, Biology Centre of the Academy of Sciences of the Czech Republic, 37005 České Budějovice, Czech Republic
9 Centre for Virus Research, Westmead Institute for Medical Research, Sydney 2145, Australia
* Correspondence: michael.s.johnson@unsw.edu.au; Tel.: +61-2-0413809670

Abstract: The genus *Acaryochloris* is unique among phototrophic organisms due to the dominance of chlorophyll *d* in its photosynthetic reaction centres and light-harvesting proteins. This allows *Acaryochloris* to capture light energy for photosynthesis over an extended spectrum of up to ~760 nm in the near infra-red (NIR) spectrum. *Acaryochloris* sp. has been reported in a variety of ecological niches, ranging from polar to tropical shallow aquatic sites. Here, we report a new *Acarychloris* strain isolated from an NIR-enriched stratified microbial layer 4–6 mm under the surface of stromatolite mats located in the Hamelin Pool of Shark Bay, Western Australia. Pigment analysis by spectrometry/fluorometry, flow cytometry and spectral confocal microscopy identifies unique patterns in pigment content that likely reflect niche adaption. For example, unlike the original *A. marina* species (type strain MBIC11017), this new strain, *Acarychloris LARK001*, shows little change in the chlorophyll *d/a* ratio in response to changes in light wavelength, displays a different Fv/Fm response and lacks detectable levels of phycocyanin. Indeed, 16S rRNA analysis supports the identity of the *A. marina* LARK001 strain as close to but distinct from from the *A. marina* HICR111A strain first isolated from Heron Island and previously found on the Great Barrier Reef under coral rubble on the reef flat. Taken together, *A. marina* LARK001 is a new cyanobacterial strain adapted to the stromatolite mats in Shark Bay.

Keywords: cyanobacteria; *Acaryochloris*; chlorophyll *d*; near infrared; stromatolite; shark bay

1. Introduction

Chlorophyll *d* (Chl *d*) was discovered in 1943 and was first attributed to marine macrophytic red algae ([1] and references therein). It was the fourth chlorophyll to be discovered, hence the name attribution. It differed from chlorophyll *a* (Chl *a*) in having a formyl group at C3 in ring A instead of a vinyl group. This small change meant that the Qy peak was red shifted (in vitro) from 665 nm to 696 nm. For the next fifty years, the existence of Chl *d* was debated and often relegated to a breakdown product resulting

from changes during extraction [1]. More than half a century later, the cyanobacterium *Acaryochloris marina* (strain *Acaryochloris marina* gen. et sp. Nov., MBIC11017 (cyanobacteria) was isolated [2] and later confirmed to be an oxygenic photosynthetic prokaryote containing Chl *d* as its major photopigment [3]. It was subsequently recognised that the Chl *d* found in red algal species originated from colonial *Acaryochloris* spp. living on the surface of the algae [4–6].

A second isolate was later obtained as a free-living epiphytic cyanobacterium present on the surface of a number of red algae [6]. Thus, *Acaryochloris* grows as a free-living organism, and a large part of the presence of this species in didemnid ascidians is due to an epiphytic growth [4,5]. A third isolate belonging to the *A. marina* clade (strain CCMEE5401) was characterized in 2005, isolated from a unique site at the Salton Sea saltwater lake in California, USA [7]. However, a comparison of the 16S rRNA sequences of both strains indicates firstly that they are highly similar (99.2% nucleotide sequence identity) and also that the *A. marina* clade contains a characteristic small-subunit 16S rRNA gene insertion likely originating from a b-proteobacterium; hence, *Acaryochloris* is a proteobacterial-cyanobacterial hybrid entity based on 16S rRNA analysis [7].

Several additional strains of *Acaryochloris marina* have been isolated and cultured: *A. marina* MBIC11017 isolated from the squeezed extract of didemnid ascidian in Palau Island [2,3], which is likely the same epiphytic organism also recovered elsewhere [4,5], *A. marina* AWAJI-1 isolated as an epiphyte from red algae in Japan [6], *A. marina* sp. CCMEE 5410 isolated from a eutrophic hypersaline lake in Salton Sea, California [7], *A. marina* sp HICR111A isolated from dead coral on the reef flat at Heron Island, Australia [8] and the isolated strains, *A. marina* MPGRS1—an epiphyte from the red alga *Gelidium caulacantheum* collected from Georges River, Australia [9] and another from a shaded calcified substrate on a coral reef [10]. Of note, *A. marina* has also been found in a number of habitats [11,12] including Antarctica [13], although 16S rRNA data are lacking from these more recently isolated strains.

Unlike most other photooxygenic organisms, *A. marina* uses Chl *d*, instead of Chl *a*, as a main pigment in the reaction centres and for photosynthetic light-harvesting. Chl *d*, differs from Chl *a* at longer wavelengths by absorbing in the near-infrared (NIR) region (approx. 710 nm in vivo) rather than the red region of the visible spectrum. Chl *d* was at first believed to be unique to the genus *Acaryochloris* [3,5]. However, Chl *d* has been found in a range of cyanobacteria [14–19], including some cyanobacteria that possess both Chl *d* and Chl *f* [16,19]. Investigation of the gene cluster FaRLiP (Far Red Light Photoacclimation) indicates that this is a genetic locus responsible for the synthesis of Chl *d* and Chl *f* in several cyanobacteria—correlating with the capability of growing in far-red light [16] (Op. Cit). In cyanobacterial species containing FaRLiP, Chl *d* forms only a minor part of the total chlorophyll, yet it has been shown to be present in the reaction centre of photosystem II in *Chroococcidiopsis* [20]. Nevertheless, all analyses of the 16S rRNA sequence data of *A. marina* cyanobacteria from environmental DNA samples that have been examined to date indicate that these photosynthetic bacteria are very widespread in nature [21–24]. Despite Chl *d* having been discovered in 2003/2004 in a range of other cyanobacterial genera (op. cit.), it is only recently that a second species, *Acaryochloris thomasi* [22], has been identified. This species differs from all other isolates of *Acaryochloris* in that it completely lacks Chl *d* but rather has monovinyl-Chl *a* and *b* content similar to that in *Prochloron* and *Prochlorothrix*.

The genome of *A. marina* MBIC11017 (type strain) was recently analysed and found to be one of the largest cyanobacterial genomes, comprising 8.3 million nucleotides distributed between one master chromosome of 6,503,724 base pairs and 9 plasmids [23]. This unusual genome organization, particularly the fact that approximately 25% of its genes reside in plasmids, gave rise to the idea of an increased plasmid-based lateral gene transfer within this genus [23]. Furthermore, the genes located within the plasmid components include those that code for small-molecule biosynthesis, central or intermediary metabolism, energy metabolism, signal transduction, DNA metabolism, transcription, protein synthesis/fate and surface-associated features [23]. Presumably, the plasmids have the capacity for lateral

gene transfer to confer an environmental adaptation advantage. Despite these intriguing features of the type strain, the *Acaryochloris* species as a whole is still poorly described genetically and phenotypically using currently available technologies.

Here, we have characterised a strain of *A. marina* isolated from the 4–6 mm layer of smooth stromatolites from Nilemah, Hamelin Pool of Shark Bay in Western Australia. This layer receives mainly NIR radiation [24] consistent with the need for specialized NIR absorbing chlorophylls, such as Chl *d* and/or Chl *f*. This study therefore characterises this unique *A. marina* strain from Australian stromatolites and compares it with the *A. marina* type strain (MBIC11017), using 16S phylogenetic typing and pigment analysis by spectrometry/fluorometry, flow cytometry and spectral confocal microscopy to reveal its unique pattern in pigment distribution that likely reflects stromatolite niche adaption.

2. Materials and Methods

2.1. Site Description, Sampling and Mat Sectioning

Sampling was undertaken at Nilemah, on the southern shore of the Hamelin Pool (26°27′336″ S, 114°05.762″ E), at 1200 h on 21 June 2015 (see Figure 1A). This is an area of high salinity due to the isolation of the Hamelin pool from the rest of Shark Bay as a result of a sill at the entrance to the Hamelin Pool (salinity 66–72 PSU). At the time of sampling, the water temperature was 20.1 °C, the salinity was 67.4 PSU (Practical Salinity Units), i.e., a hypersaline environment, and the pH was 8.13. The samples were collected under DBCA license FO25000229-2, and the research was conducted on Malgana country and in consultation with the traditional owners by AWDL. A sample was collected and transported to UTS for the various analyses. It was stored at 4 °C in the dark prior to analysis. The mat types analysed here were designated smooth mats, as they are extensively distributed in Shark Bay and have been examined in detail at the taxonomic and metagenomic levels [25–28]. A map showing the location of sampling is shown in Figure 1A. Sections of mat approx. 5 cm × 5 cm and 2 cm deep were obtained with a sharp knife and placed in waterproof plastic containers for air transport to the University of Technology, Sydney within 24 h; they were then sored at 4 °C in the dark until examination within 2 months. For confocal microscopy, thin sections (Figure 1B) of the mat material (approx. 2 mm thick) were obtained and placed on slides, which were examined with a Nikon A1 Confocal Microscope equipped with a spectral detection capability from 420–750 nm in 10 nm bandwidths. These mat samples are considered representative of stromatolitic mats in Shark Bay, as shown by previous studies. Cultures of *A. marina* MBIC11017 were obtained from the laboratory of Prof. Min Chen at the University of Sydney.
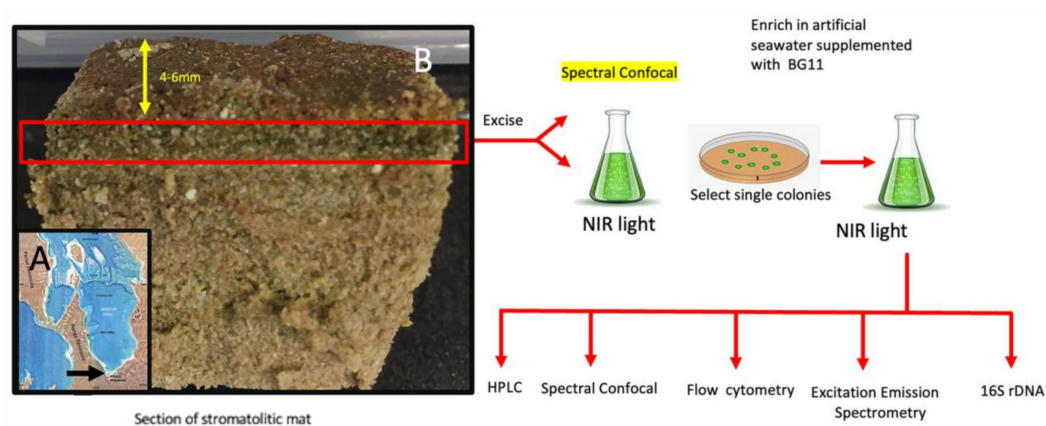
Figure 1. Sampling and enrichment: (**A**) Map showing the Hamelin Pool, with an arrow indicating the position of the sampling site at Nilemah. (**B**) Overview of sampling, culture enrichment and analysis methods.

2.2. Isolation and Culturing

The overall isolation and culturing strategy is summarised in Figure 1B. Briefly, a thin slice (2 mm thick) representing the green layer (4–6 mm) from the upper surface was excised with a scalpel blade. This layer was transferred to a sterile 35 mm glass bottom petri dish (imaging grade WPI dish) and imaged by confocal microscopy. After imaging the section, it was transferred to artificial seawater supplemented with BG11 liquid broth and grown in NIR light to enrich far red absorbing microorganisms. The enriched culture was transferred to agar plates containing artificial seawater supplemented with BG11 so as to obtain single colonies under NIR conditions. A single colony was then sub- cultured into artificial seawater supplemented with BG11 liquid broth and further grown in NIR light to an appropriate cell mass for analysis. Purity was checked by 16S rRNA sequencing. Once purity reached greater than 95%, flow cytometry, confocal microscopy, spectroscopy and HPLC were employed to analyse the culture material.

2.3. Liquid Culture Enrichment

Culture aggregates from the 4–6 mm layer of several stromatolite samples were ground up with a 10 mL *Potter-Elvehjem* (*glass-teflon*) hand homogeniser with 5 strokes and were resuspended in BG11 medium (Sigma 73816) dissolved in seawater (at 30 PSU) and incubated in NIR radiation (720 nm, 10 μJ m^{-2} s^{-1}) for up to 12 weeks. Further subculturing involved homogenising 10 mL of the existing culture and diluting the homogenate into new medium at a 1/10 ratio. Liquid cultures were then grown in NIR light (720 nm, 10 μmols m^{-2} s^{-1}), with shaking at 60 rpm at room temperature (23 °C).

2.4. Culturing on Agar Plates

A 500 μl of aliquot liquid culture homogenate was spread onto agar plates (artificial seawater supplanted with BG11 and 15% agar) and incubated as above in NIR or white light. Single green colonies were selected and re-inoculated with a sterile loop onto BG11 supplemented seawater agar plates and grown in white light (WL) (100 μmols m^{-2} s^{-1}) or NIR radiation (720 nm, 10 μmols m^{-2} s^{-1}) for up to 12 weeks (Figure 1B).

2.5. Fluorescence Microscopy

Hyperspectral confocal microscopy was used to determine emission spectra matched to cell morphologies. Mat samples sections, small aliquots of liquid culture or colonies from agar plates were laid onto WPI Fluoro-dishes (coverslip thickness 0.017 mm), and images were scanned on a Nikon A1 Confocal microscope using the 458 nm laser with emission collected through the spectral detector (up to 750 nm) in order to obtain a lambda stack. Cells of interest were highlighted by the Regions of Interest (ROIs) function, and the emission spectra were collected for those cells.

For the 3D imaging of isolates from liquid culture, cells were fixed in 4% Paraformaldehyde/PBS and mounted on a 0.017 mm coverslip in a prolong glass (NA1.520) mounting medium (ThermoFisher, Waltham, MA, USA) and imaged on a Leica Stellaris 8 Confocal microscope with a 63× oil object (NA1.4). The excitation and emission settings were set to detect phycocyanin fluorescence (excitation 620 nm, emission 675/10 nm) and Chl *d* fluorescence (excitation 708 nm, emission 720/10 nm). A step size of 100 nm was used to collect the whole volume of the cells. For morphometric analysis, the fixed and mounted cells (above) were imaged on an Olympus BX51 epifluorescence microscope fitted with a 100× oil immersion objective (NA1.3) and an Olympus DP73 camera. Both brightfield and fluorescence images (excitation 500 nm–530 nm emission 575 nm LP) were collected. Cell shape (aspect ratio) and size (diameter) were determined by using Image J—FIJI [29]. For each isolate, a threshold level was set for the red auto-fluorescence, where the fluorescence perimeter of the cells matched the brightfield perimeter. The segmented cells were made into a binary image, and any joining cells were separated by water-shedding. The "analyze particles . . . " function was then used to determine the shape (aspect ratio). Further segmentation based on circularity (where non-round/elongated cells, i.e., cells with circularity

less than 0.8, were discarded) was applied, and the diameter of these "round" cells was also determined using the "analyze particles ..." function.

2.6. Absorbance and Fluorescence Spectroscopy

Slides of cultured *Acaryochloris sp.* were placed on glass microscope slides with coverslips and placed on an upright microscope stage (Nikon Ti, Tokyo, Japan) with a $40\times$ (NA0.6) objective at room temperature. The cells were brought into the focal plane under white light. The absorbance of the light that passed through the cells was measured with a spectrometer (Avantes Avaspec with a spectral resolution of 0.4 nm) connected via an optical cable to one of the oculars of the microscope. An area devoid of cells was used to collect the background spectra that was later subtracted in the image analysis processing.

2.7. Excitation Emission Matrices

Liquid cultures were placed into the wells of a 96 welled (black walled) plate, and the fluorescence emissions at every wavelength from 620–760 nm were measured at every excitation wavelength from 300–750 nm in a fluorescence plate reader (Tecan Infinite M1000 Pro, Tecan, Männedorf, Switzerland). The raw data were exported to Microsoft Excel, and heatmaps were created as described by [30].

2.8. HPLC Pigment Analysis

High performance liquid chromatography (HPLC) was used to determine the concentrations of Chl *a* and *d* in the *Acaryochloris* cultures grown in either white light or NIR cultures. Each culture was pelleted at 5000 g for 10 min to collect the biomass for pigment signal detection. The extraction of samples were carried out following [31], with slight modifications. Chilled acetone was added to the pelleted biomass, probe sonicated (30 s at 50 W on ice), vortexed three times for 30 s each under cold, dark conditions to limit pigment degradation and then stored at 4 °C overnight. The pigment extracts were then filtered through a 0.2 μM PTFE 13 mm syringe filter and stored at −80 °C until analysis. An Agilent 1290 HPLC system equipped with a binary pump with an integrated vacuum degasser, thermostat-controlled column compartment modules, an Infinity 1290 autosampler and a PDA detector was used for the analysis. Column separation was performed using a Zorbax Eclipse C18 HPLC 4.6 mm $\times$ 150 mm column (Agilent, Santa Clara, CA, USA) using a gradient of ammonium acetate (0.01 M), methanol, acetonitrile and ethyl acetate. A sandwich injection approach was set using the auto injector program, where the mixture of MeOH: ACN (8:2) and the samples were drawn alternatively in the sequence, 100:60:100 (μL), and then mixed in the loop and injected. A complete pigment spectrum from 270 to 780 nm was recorded using a PDA detector with 3.4 nm bandwidth. The following wavelengths were used to monitor the chromatogram: 406, 440, 660, 696 and 706 nm.

2.9. 16SSU rRNA Analysis

DNA was extracted using the DNeasy PowerBiofilm Kit DNA extraction kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. Biological duplicates were extracted separately and then pooled prior to sequencing to minimise potential heterogeneity biases. Purified DNA samples were sequenced on an Illumina MiSeq for 16S V1–V3 amplicon (27F-519R) on a v3 2 $\times$ 300 bp run by the Ramaciotti Centre for Genomics (UNSW Sydney, Australia). Sequences were processed through the Mothur pipeline [32] with default parameters, with searches of representative sequences against the SILVA [33] database. For the most abundant sequence in the cultures, representative sequences were then aligned against Cyanobacteria 16S rRNA sequences retrieved from the SILVA and NCBI databases [34], with Chloroflexi as the outgroup. A maximum likelihood tree was constructed with IQ-TREE2 with 1000 bootstraps, and the best-fit model TIM3+F+I+G4 was chosen by ModelFinder. The tree was visualized using iTOL [35].

2.10. Flow Cytometry

Sample acquisition and instrument settings. Aliquots of *Acaryochloris*—grown cultures in BG11/seawater (grown to a purity of 95% by rRNA sequencing as described above in Section 2.2)—were prepared for flow cytometry analysis by executing five strokes of a *Potter-Elvehjem* (*glass-teflon*) hand homogenizer tube to break up cyanobacterial aggregates. Analysis was carried out on an LSRII flow cytometer (Becton Dickenson, BD Biosciences, Franklin Lakes, NJ, USA) with DIVA acquisition software (version 8.0.2). First, FSC-A/SSC-A two-parameter dotplots were viewed with logarithmic scale settings. Sterile normal saline was acquired for 1 min pre-cyanobacterial sample acquisition. This permitted the detection of electronic noise events and permitted subsequent gating on larger events of cyanobacteria (R1) from the liquid cultures. Next, cell events were collected for each culture sample before again running sterile saline, also for 1 min. Voltages were set such that electronic noise and any non-fluorescent cyanobacteria were located in the bottom left-hand panel of the two-parameter dot-plots of all fluorescence detection channels. Thus, natural fluorescence was evident as events with higher laser-excited fluorescent emission and without any staining. The data were saved and exported as FCS3 data files.

Cytometry Data Analysis. The samples' file events were set to the same minimum 80,000 number using the DownsizeSampleV3 plugin. The samples were first viewed as two-parameter FSC-A/SSC-A dot-plots. Uncompensated data were also concatenated for subsequent analysis for SSC-A distribution (or FSC-A distribution) versus sample ID to confirm cyanobacterial events (R1) as distinct from electronic noise events (*). Events in R1 were assessed as FSC-A data via histogram overlays—gating on very small cells and debris, small cells and large cells. Uncompensated two-parameter fluorescence data excited by the 640 nm red laser were chosen for the fluorescent analysis, as this represented the presence or abundance of phycocyanin (detected via the 660/20 nm bandpass filter in detector array position C) and Chl *d* (detected via the 730/45 nm bandpass filter in detector array position B). All cytometry analysis and data overlays were performed with FlowJo (version 10.8.1, Flowjo LLC, Ashland, OR, USA), and multi-component data figures were prepared in Microsoft Powerpoint for Mac (version 16.57, Microsoft, Redmond, WA, USA).

2.11. Fluorometry by Pulsed Amplitude Modulated (PAM) Fluorescence

The quantum yield of photosystem II (F_v/F_m) was determined by the MAXI PAM model IMAG-MAX/L (Walz, Effeltrich, Germany) fluorometer. The instrument was fitted with blue actinic light emitting diodes (LEDs) with a peak wavelength of 450 nm which were used for applying the saturation pulse for the determination of F_m and for the determination of F_o. The samples were dark adapted for 10 min prior to the measurement. After image acquisition, data processing was performed with the ImagingWin v2.56p instrument software (Walz, Effeltrich, Germany) [36].

2.12. Cell Counts

The growth rate was measured from liquid cultures using a Haemocytometer and an Olympus BX51 microscope fitted with a 20× (0.6 NA) objective. Prior to counting, the cell clumps were broken up by a potter Elvehjem-Potter homogeniser. The data were analysed in Graph Pad PRISM using the exponential growth curve function.

3. Results

3.1. Location

Cyanobacteria were isolated from stromatolitic mats located in the hypersaline environment at Nilemah in the Hamelin Pool (see Section 2.1 and Figure 1A).

3.2. Confocal Imaging of Mat Material

Hyperspectral confocal images were obtained of the mat at 4–6 mm from the upper surface. Both coccoid and filamentous cyanobacterial cells were visualised. Emission wavelengths evident from the filamentous cells exhibited a peak at 745–750 nm, indicating

Chl *f* (Figure 2A), and coccoid cells with a peak at approximately 735 nm, indicating Chl *d* (Figure 2B). The filamentous cells with the 745–750 nm peak emission also had a significant peak in the red at 680–685 nm, indicating the presence of significant amounts of Chl *a* (Figure 2A). There were also coccoid cells with a spectral emission profile indicative of mainly Chl *a* (Figure 2A—data not shown). Hyperspectral confocal imaging of an initial culture enrichment in NIR indicated the presence of coccoid cells with an emission peak of 735 nm, filamentous cells with an emission peak of 728 nm and diatoms with an emission peak of 710 nm (Figure 2C).

(A) **(B)** **(C)**
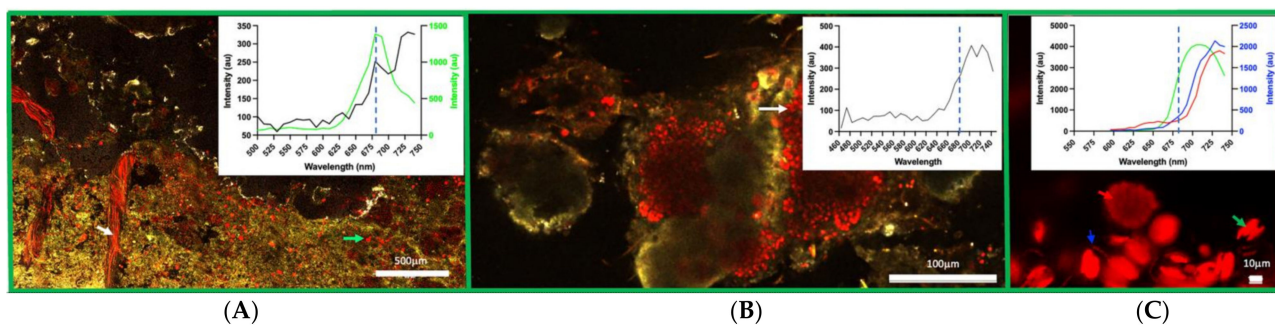
Figure 2. Confocal Imaging of the green layer. A side view of an excised segment randomly sampled from a smooth mat at Nilemah. For reference, the blue dotted line at 680 nm shows a typical Chl *a* emission. (**A**) A large section (1 cm × 0.6 cm) of the green layer imaged by hyperspectral confocal microscopy using a 458 nm laser with 10 nm continuous band pass emissions collected up to 750 nm. The red organisms consist of red and far red/NIR (700–750 nm) fluorescing pigments. Both filamentous and coccoid forms are observed. The filamentous forms (white arrow) contain a far-red emitting (750 nm peak) pigment that is most likely Chl *f*. The green arrow shows coccoid cells exhibiting a Chl *a* emission profile. Scale bar 500 μm. (**B**) Emission spectra of typical clusters of coccoid cells (white arrow) when excited by the 458 nm laser. Two peaks are evident at 704 nm and 724 nm, suggestive of the presence of Chl *d* (724 nm). Scale bar 100 μm. (**C**) The arrow shows the various microbial forms that were found in the initial NIR enrichment culture. The green arrow shows a diatom (emission peak 710 nm), the blue arrow shows a filamentous form (emission peak 730 nm) and the red arrow shows a cluster of coccoid cells (emission peak 730 nm). Scale bar 10 μm.

Present in the confocal microscopy analysis of the mat material were several other bacteria, some with near far-red shifted peaks in their chlorophylls (data not shown). The most abundant of these was what we identified as *Chroococcidiopsis sp.*, which possesses Chl *f* with a fluorescence emission peak at ~750 nm at RT. This *Chroococcidiopsis sp.* has not been previously reported to have been isolated to purity in in vitro culture. Similarly, another cyanobacterium with an apparent red-shifted chlorophyll is *Spirulina sp.*, which also has not been isolated to purity to date. Moreover, the 16S genotype sequence data confirm the presence of both of these microorganisms in stromatolite material [23]. Furthermore, the 16S data presented here indicate the presence of other phototrophic bacteria, *Phormidia*, *Halomicronema* and *Cyanobium*, all of which have been detected microscopically (data not shown). Microalgae (chloroplast) are also evident in the 16S data, and these were evident microscopically in our initial enriched cultures (Figure 2C). The presence of Caenarcaniphilales (Malainabacteria) is also of interest. The Melainabacteria are a newly described, non-photosynthetic sister phylum to the Cyanobacteria. They are often found in aphotic environments such as human and animal guts, grassland soil and wastewater treatments [37], and it is interesting that they have been detected in the attenuated light conditions associated with the 4–6 mm deep layer studied here.

3.3. Obtaining a Purified Isolate and Culture Morphology

Sub-sampling of the enriched culture resulted in the growth of the coccoid forms as colonies on agar plates in NIR. Further subculturing of agar plates in NIR of single colonies

led to the isolation of single colonies on agar plates (Figure S1, panel B). The MBIC11017 type strain had a propensity to form large clusters, compared with the LARK001 strain, when grown in liquid culture in NIR (Figure S1, panel A). The LARK001 strain exhibited a yellow green colour when grown on agar or when grown in liquid culture, whereas the MBIC11017 strain exhibited a darker blue-green colour in the same conditions.

3.4. HPLC Analysis

HPLC analysis revealed that both the LARK001 strain and MBIC11017 strain possess Chl *d* as their major photosynthetic pigment. The Chl *d* content in the LARK001 strain remained consistent in response to light conditions, with Chl *d* making up 93.4% of the chlorophyll content in white light and 92.9% in NIR (Table S1). However, the MBIC11017 strain showed a marked difference in chlorophyll content in response to light conditions, with 91% Chl *d* content evident in cells grown in white light, which increased to 96.4% Chl *d* in cells grown in NIR (Table S1).

3.5. Sequence Identity

The 16S DNA sequence identity of cultured *A. marina* LARK001 cyanobacteria from Shark Bay stromatolites indicated that it forms a clade closely linked to *A. marina* HICR111A, first isolated from Heron Island, as demonstrated by maximum likelihood phylogenetic tree analyses of representative species (Figure 3A); 98.67% similarity. Indeed, whilst the LARK001 isolate is clearly an *A. marina* cyanobacteria, it clusters away from the *A. marina* MBIC11017 type strain (Figure 3A). Moreover, although not entirely genetically pure, this LARK001 in vitro cultured isolate is over 95% *A. marina* and can sustain in vitro growth when isolated together in co-culture with up to three other species (Figure 3B). In the initial enrichment, 16S rRNA analysis indicated the presence *Phormidesmiales*, *Oxyphotobacteria* and *Halomicronema*, along with *Acaryochloris*. In the late enriched culture, a 16S rRNA gene analysis revealed the presence of *Halomicronema* and Chloroplasts (most likely diatoms), which were both also found by confocal microscopy in an enriched culture—see Figure 2C. It is interesting to see the presence Caenarcaniphilales in the late enriched culture. These are a newly described order of organisms, belonging to the phylum Melainabacteria, that share phylogenetic and structural similarities with cyanobacteria but lack the ability to photosynthesise. Nevertheless, this new Australian *Acaryochloris* strain constituted 95% of the late enriched culture and is distinctly phylogenetically distant from the clade containing the MBIC11017 type strain.

3.6. Spectroscopic Analysis

Next an absorbance spectral analysis was performed on cultures of the *A. marina* MBIC11017 strain and the *A. marina* LARK001 strain. The absorbance peak of *A. marina* MBIC11017 was 710 nm, whereas that of *A. marina* LARK0001 was at 704 nm (Figure 4B). However, whilst *A. marina* MBIC11017 additionally has a distinct absorbance peak at 610–640 nm (attributed to phycocyanin), this was not detectable in the new *A. marina* LARK001 (Figure 4A).

The fluorescence emission spectra of the two strains differed markedly (Figure 4C): the *A marina* MBIC11017 had a major emission peak at ~724 nm, whereas the *A. marina* LARK001 had a major peak at 735 nm (Figure 4C). Excitation emission scanning further highlighted the different profiles of these strains (Figure 4D). Notably, an additional fluorescence emission peak at 650 nm (attributed to phycocyanin) is present in the *A. marina* MBIC11017 strain, and there is no evidence of this in the LARK001 (Figure 4C). Finally, spectral confocal imaging analysis also confirmed that Chl *d* fluorescence emission peaks of 715 nm were evident for the *A. marina* MBIC11017, but in *A. marina* LARK001 (Figure 4; panels D,F), an additional fluorescence emission peak of 735 nm appears to predominate. Despite differences in the spectral characteristic at a cellular in vivo level, there were no such differences observed in the absorbance profiles in the HPLC Chl *d* fraction from each strain (Panel F).
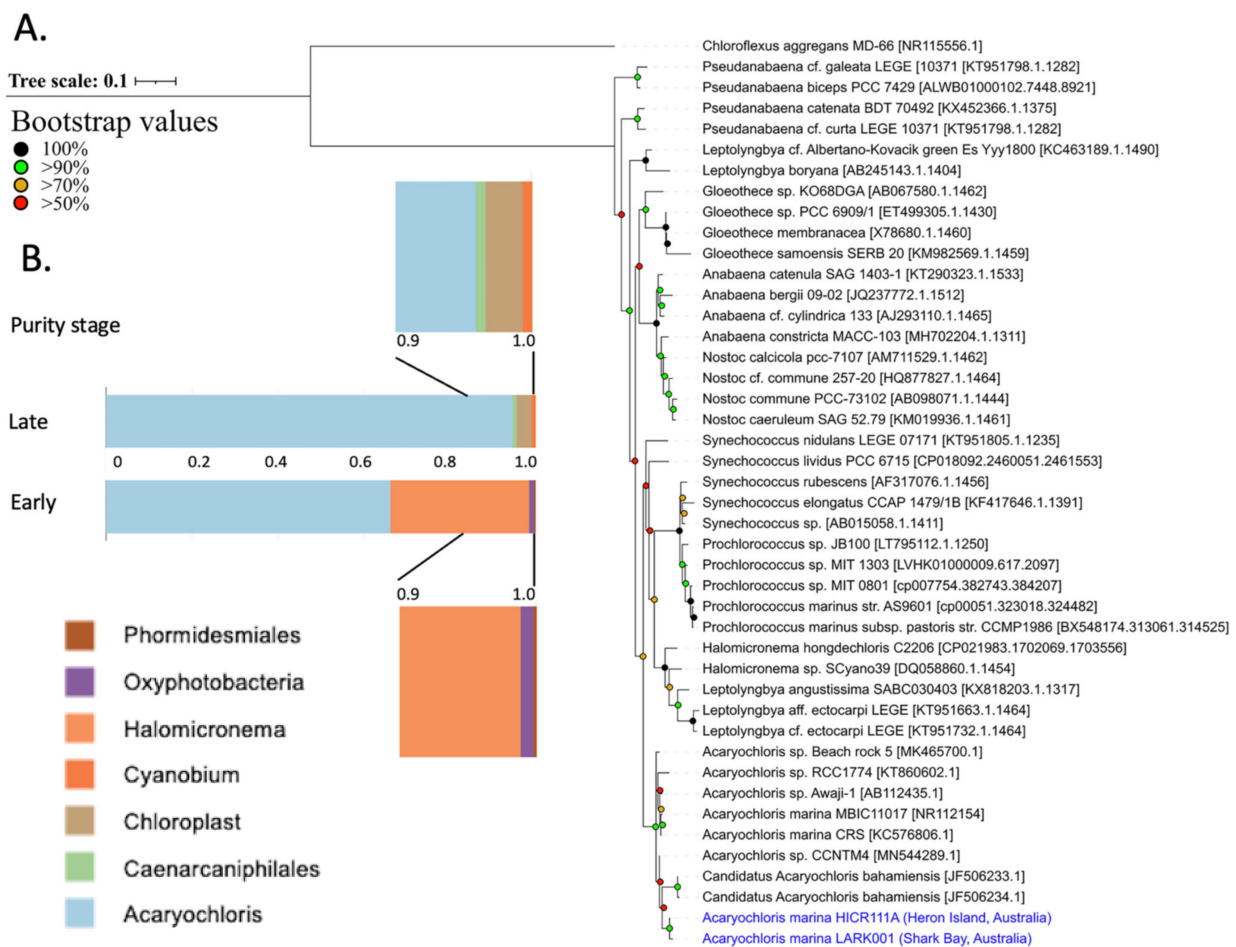
Figure 3. 16S rRNA sequencing. (**A**) 16S rRNA gene phylogeny of the LARK001 isolate from showing the relationship to other isolates. (**B**) 16S rRNA gene sequence representation of the LARK001 isolate at different stages of culture purification.

3.7. Flow Cytometry and Morphometric Analysis

In vitro cultures of *A. marina* MBIC11017 and LARK001 were examined by flow cytometry after homogenisation to enable single cell cytometry. The small size and diverse culture shapes of these cyanobacteria are evident in forward and side scatter dot-plots, and the preparation procedure appears to produce very small cells and debris, but these are easily made distinct from electronic noise by comparing the files generated by saline (without cyanobacteria) and gating on side-scatter or forward scatter profiles (Figure 5A). Thus, the in vitro cultured cyanobacteria cells were detected in three general size groupings: very small, small and large (Figure 5A). Simple two-parameter uncompensated fluorescent data overlays of 640 nm red-laser excited emission profiles can easily profile both phycocyanin (evident on the 660/20 nm bandpass/detector channel) and Chl *d* (detected at the 730/45 nm detector channel) where distinct populations are evident (Figure 5B). A clear feature of this data is that the largest cells contain the most abundant Chl *d*, and the presence of phycocyanin in the MBIC11017 type strain easily separates the two strains (Figure 5C). In comparison, the small cells both contain low fluorescence in 660/20 nm and 730/45 nm, and the very small cells and debris are either weakly fluorescent or non-fluorescent in these channels (Figure 5C). The flow cytometry profiles also closely mirrored the images of two strains obtained by confocal fluorescent microscopy (Figure 5D), using the 620 nm laser excitation and collecting emissions at 675/10 nm for phycocyanin fluorescence and using laser excitation at 708 nm with emissions collected at 720/10 nm for Chl *d* (Figure 5D). Here, 3D rendering indicated that, in the MBIC11017 strain, the phycocyanin was present

near the periphery of the cells, and Chl *d* was situated throughout the entire cell volume (Figure 5D), whereas no phycocyanin was present in the LARK001 strain.
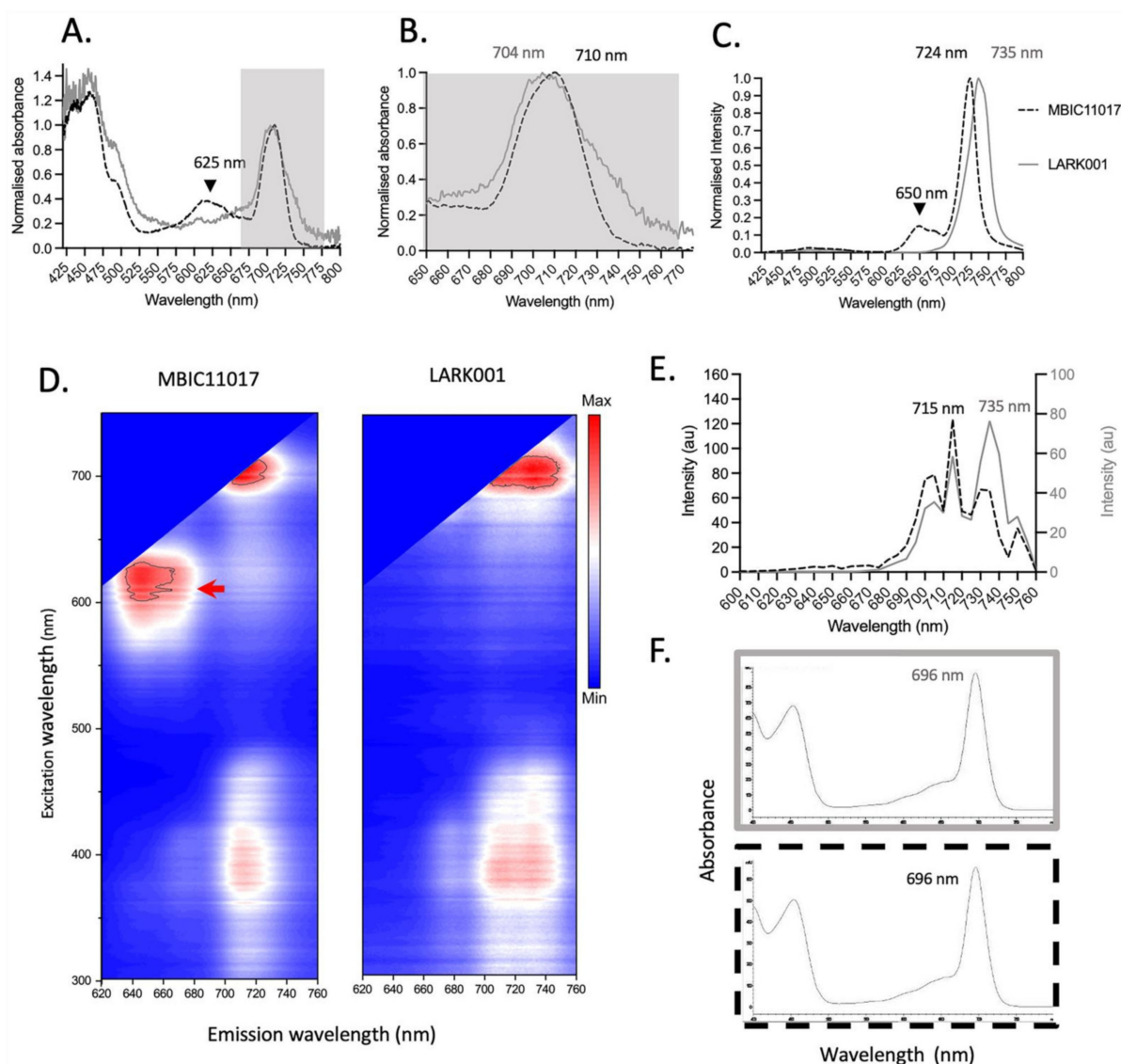
Figure 4. Absorbance spectra and fluorescence spectra of *A. marina* MBIC11017 vs. *A. marina* LARK001 grown in NIR. (**A**) Absorption spectra of both strains indicate differences between the two stains. At 625 nm (black arrow), a peak is present in the MBIC11017 strain that is not present in the LARK001 strain. Differences in the Chl *d* absorption region are also present (grey boxed region). (**B**) The grey boxed region from Panel A is magnified to further highlight the difference in peak Chl *d* absorption— 704 nm for LARK001 and 710 nm for MBIC11017. (**C**) The emission profiles for each strain reveal a peak at 650 nm (black arrow) that is unique in the MBIC11017 strain, whilst the Chl *d* emission peaks differ substantially, with the peaks at 724 nm and 735 nm for MBIC11017 and at 735 nm for LARK001 strain. (**D**) An excitation emission matrix shows a unique spectral component (excitation max. approx. 625 nm and emission max. approx. 640 nm) in the MBIC11017 strain. The LARK001 strain exhibits two distinct emission profiles for Chl *d* at 718 nm (approx.) and 735 nm (approx.). Only one emission profile at approx. 718 nm is evident for the MBIC11017 strain. (**E**) Hyperspectral confocal scan (top panel) confirms the evidence of two emission peaks in LARK001, with the 735 nm peak being the major contributor in this strain. The same two peaks are evident in the MBIC11017 strain, but the 715 nm peak is the major peak. (**F**) The grey and black dashed panels reveal no difference in the absorbance profiles in the Chl *d* fraction that was purified by HPLC for each strain.
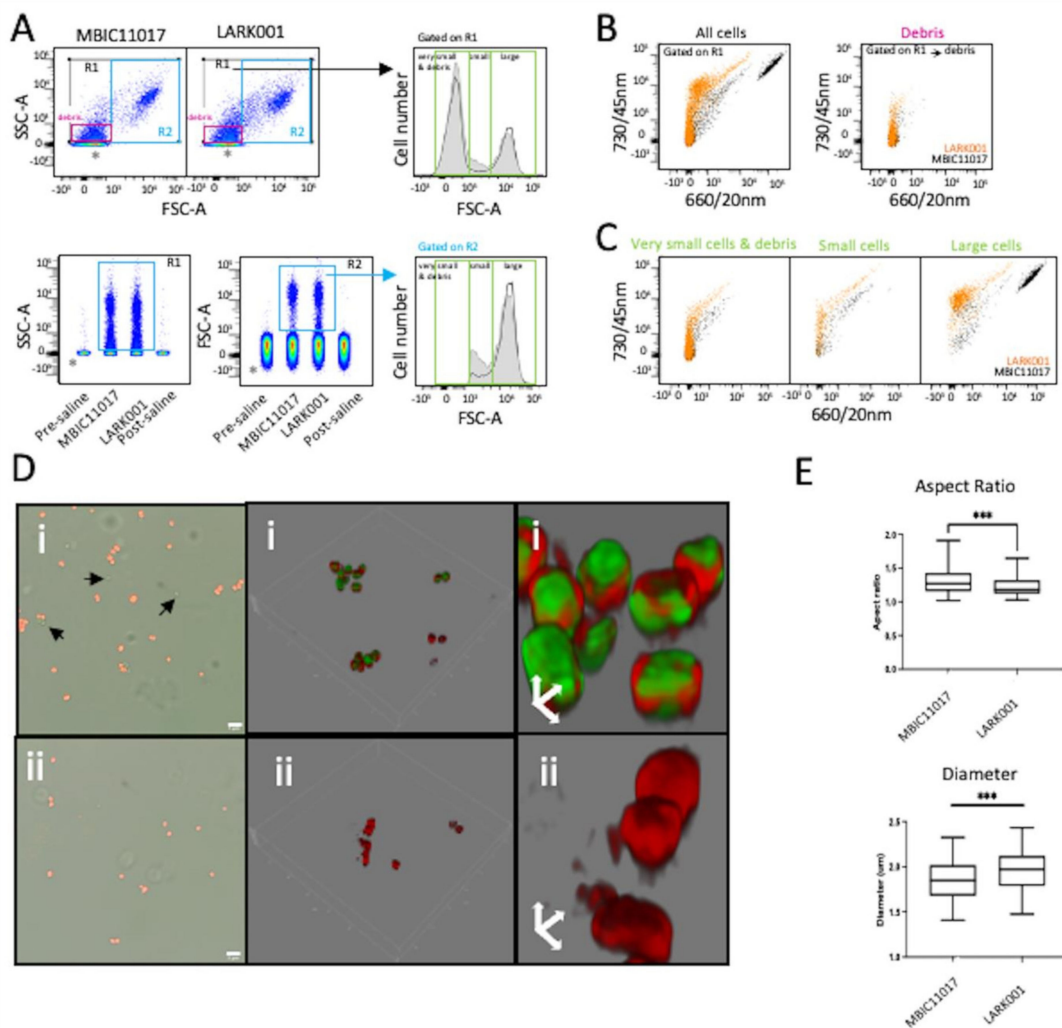
Figure 5. Cytometry and confocal microscopy analysis of A. *marina* type strains MBIC11017 and LARK001. (**A**) MBIC11017 and LARK001 cyanobacterial cells in FSC-A/SSC-A dot-plots indicating distinct cell populations. Also shown are concatenated SSC-A distribution data, confirming electronic noise events in pre- and post-saline samples (acquired for 1 min). Cell events in R1 exclude electronic noise (*) and demonstrate evidence of two sub-populations based on cell size: designated smaller (R2) and larger (R3) cyanobacteria cells in histogram overlays: MBIC11017 (black unfilled histogram) and LARK001 (grey filled histogram). Tiny debris events are also evident within the R1 gate, but size gating (R4) removes these events. The arrow indicates the shoulder of slightly larger cells within the R2 gated population. (**B**) Two-parameter dot-plots of natural fluorescence at 730/45 nm versus 660/20 nm in *A. marina* MBIC11017 (black) and LARK001 (orange) strains upon 640 nm red laser excitation. (**C**) Two-parameter dot-plots of fluorescence at 730/45 nm versus FSC-A for R2 and R3 gated events of *A. marina* MBIC11017 (black) and LARK001 (orange) strains, also upon 640 nm red laser excitation. (**D**) Widefield and confocal images of *A. marina* MBIC11017 (**i**) and *A. marina* LARK001 (**ii**) cells (scale bar is 5 mm); also shown are 3-D rendered and zoomed images of the same two strains. (**E**) Aspect-ratio ($n = 107$ for MBIC11017 and $n = 94$ for LARK001) and diameter analysis ($n = 65$ for MBIC11017 and $n = 73$ for LARK001) of cells from microscopy data acquired in D. *** $p < 0.005$ by a student *t*-test.

Finally, both the cytometry side scatter profiles and confocal imaging strongly suggested that the MBIC11017 strain and the LARK001 strain may exhibit morphometry differences. Therefore, the fluorescence microscopy image data were analysed with respect to aspect ratio and diameter, and here, the literal physical shape differences were confirmed

in that the LARK0001 strain was statistically more spherical and larger in diameter than the MBIC11017 strain (Figure 5E). However, it should be taken into account that the MBIC11017 strain divides at a faster rate than the LARK001 strain, potentially resulting in a greater representation of elongated (non-spherical) forms.

3.8. Fluorometry

Pulse amplitude modulated (PAM) fluorescence measurements facilitated the determination of the ratio of variable fluorescence (F_v) to maximum fluorescence (F_m) and hence the potential quantum efficiency of photosystem II for each strain under each light condition. The F_v/F_m values of the MBIC110117 strain did not appear to change significantly in response to the light condition. However, the LARK001 strain showed a significant drop in potential quantum efficiency in photosystem II when incubated in NIR as compared to the cells grown in white light (Figure 6 Panels A,B). The LARK0001 cells grown in NIR showed the lowest quantum efficiency (F_v/F_m) of the all strains/conditions tested, and this was shown to be significantly different from all other strain and condition combinations tested. Additionally, the LARK001 strain grown in NIR resulted in the slowest growth conditions, with a 53 day dividing time (Figure 6 Panel C). This is more than double the doubling time of any other combination of strain and light condition.
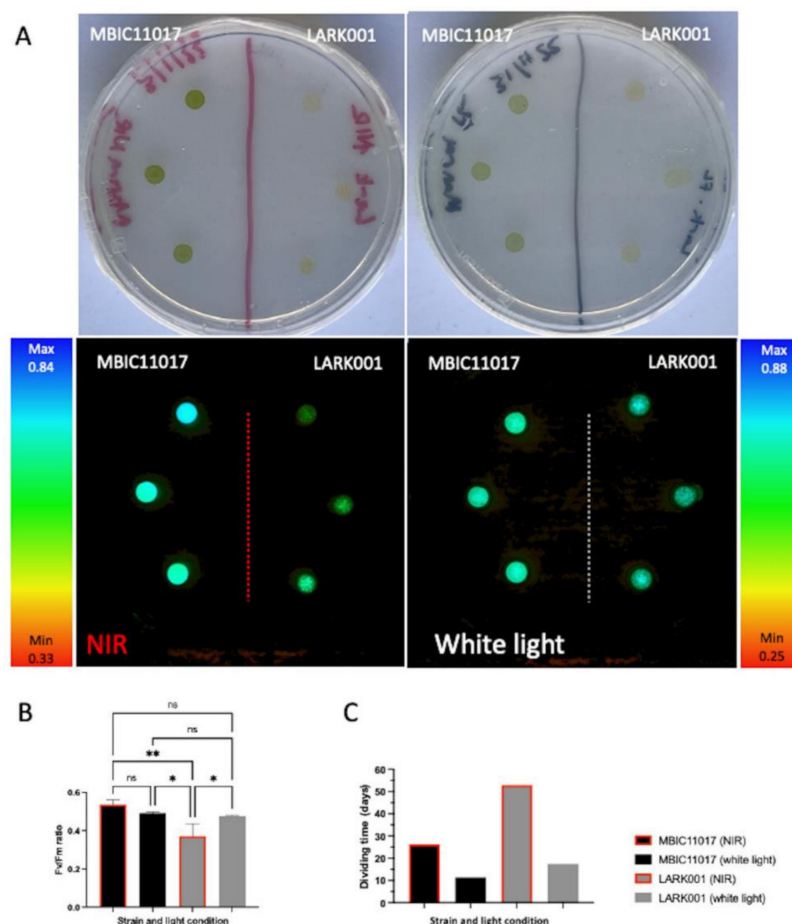
Figure 6. Fluorimetry measurement by PAM. (**A**) Clockwise from top left: Cultures grown for 3 weeks in NIR after 50,000/25 µL cells of each strain were added to the surface of the agar plate in triplicate. Cultures grown for 3 weeks in white light after 50,000/25 µL cells of each strain were added to the surface of the agar plate in triplicate. PAM imaging of the white light incubated cultures. PAM imaging of the NIR incubated cultures. (**B**) Fv/Fm measurements for the cultures and graph showing pairwise comparison * $p < 0.05$, ** $p < 0.01$. ns: not significant. (**C**) Doubling time of each strain in either white light or NIR conditions.

4. Discussion

We have isolated a new strain of *A. marina* from stromatolitic mats from the southern end of the Hamelin Pool, Shark Bay in Western Australia. This site is the one designated as available for research in the stromatolite area of the World Heritage Area. It is, in fact, the same site where all approved stromatolite research has been conducted over the last >20 years. All of the published and unpublished research, to date, indicates that, although there is some difference in calcium deposition, overall, these structures represent a typical photosynthetic community, as found on the columnar stromatolites 20 km East, further along the southern side of the Hamelin pool [38]. The new Australian stromatolite-derived *A. marina* strain described here (*A. marina* LARK001) differs in several ways from the MBIC11017 original strain. Firstly, in 16S rRNA phylogenetic analysis, the new strain is closest to the previously isolated strain *A. marina* HICR111A [8], from a shallow site on the reef flat at Heron Island, Southern Great Barrier Reef. Consistent with this genetic closeness, the *A. marina* LARK001 has a room-temperature (RT) fluorescence peak at 735 nm, indicating the abundance of Chl *d*. Indeed, both the HICR111A and LARK001 strains appear to possess Chl *d* that is conjugated with a specialized pigment protein that shifts the fluorescence emission to 735 nm and above. It has been shown that Chl *d* is bound by a light-harvesting pigment protein and is also present in the reaction centres of photosystem I and Photosystem II [20]. Instead, the MBIC11017 type strain has a shorter RT fluorescence emission of ~720 nm, indicating that *Chl d* is the major chlorophyll in both photosystem I and photosystem II and that its light-harvesting chlorophyll protein complex has a fluorescence emission at ~720 nm. Thus, it is likely that *A. marina* HICR111A and LARK001 contain different light-harvesting Chl *d*-binding proteins that cause a shifted emission at 735 nm and above at RT. Moreover, the light harvesting Chl pigment harvests NIR light efficiently and passes it on to photosystem II—and possibly to photosystem I as well, as evident in the PAM analysis. This interpretation is consistent with the HPLC analysis that confirmed only Chl *d* as the major pigment present within LARK001 cultures, with only a minor amount of Chl *a* (6–7%) being detected in this analysis (Table S1). Future studies will investigate the effect of environmental light conditions on Chl *d* content relative to Chl *a*. Another similarity between the HICR111A and LARK001 strains of cyanobacteria is that they both appear to contain very little phycocyanin. Interestingly, LARK001 divides slowly in NIR compared to in white light (Figure 6C). On face value, we conclude that the two strains of *A. marina* LARK001 and HICR111A have different light harvesting properties compared to the MBIC11017 type strain. This is in spite of HICR111A occurring on a coral reef in a shaded rubble zone on the reef flat in Eastern Australia and LARK001 existing several millimeters below the surface of a stromatolitic mat in the Hamelin Pool, Western Australia—some 3800 km distant. Both sites are subtropical with respect to latitude (Heron Island 22° S and Hamelin Pool, Shark Bay 26° S) but are differentiated in terms of the ecosystems involved. However, the *A. marina* MBIC11017 type strain was isolated from a tropical site (Palau 7.5 °N) in the western pacific, which has a very different ecological niche compared to the location in which we isolated the *A. marina* LARK001. Interestingly, the Hamelin Pool in Western Australia, Heron Island and Palau are equidistant from each other. However, the *A. marina* MBIC11017 type strain grows much better in NIR light compared to LARK001. In natural light, the MBIC11017 strain is augmented by some orange-red light (600–700 nm) absorbed by phycocyanin. In this regard, the MBIC11017 strain grows on the underside of didemnid ascidians, where it receives mainly radiation that has been highly filtered by the *Prochloron didemni* that lives in the atrial cavity of the ascidians. Thus, it may generally receive NIR light in the 720 nm range, with incidental flecks of light in the orange red spectral region. In contrast, both the *A. marina* LARK001 and HICR111A strains seem to rely on NIR at 704–706 nm but also have the ability to exist in white light, where Chl *d* can absorb light in the blue region augmented by absorption by the carotenoids (460–540 nm) without any reliance on phycocyanin in the orange-red region of the visible spectrum. It remains possible that all *Acaryochloris* strains might also rely on the absorption of light by Chl *a*, which is present in all strains, even if in smaller quantities (Table S1).

A subtle but important feature of our data is that both the flow cytometry and fluorescence microscopy (with morphometric analysis) demonstrated that the smaller cells within the in vitro *culture* of the *A. marina* LARK001 strain are slightly larger (1.48 μm–2.43 μm diameter) than the smaller cells in the MBIC111017 type strain cultures (1.41–2.32 μm diameter). This is in contrast to reports of the size of the HICR111A Heron Island strain, which, although also lacking in phycocyanin, is a much smaller at 0.75–1.00 μm [8]. Hence, the 16S genotyping similarities of the *A. marina* LARK001 strain to HICR111A also coincide with similarities in its morphometric features—despite the obvious and significant ecological niche differences described above. A deeper, full genome analysis of *A. marina* LARK001 is underway in our laboratory, which should better explain and define the similarities and differences between these three strains. This will provide a deeper understanding of the genetic limitations and molecular signatures of the adaptation capability—particularly the potential for serious sensitivities to habitat changes within the stromatolite microbial communities, which would be expected to occur with the ongoing climate change impacts, or insults from viral and other microbial predators.

5. Conclusions

We provide here the first physical and initial genetic characterisation of a unique strain of *A. marina* cyanobacteria, termed the LARK001, isolated from living Australian stromatolitic mats in Shark Bay, Western Australia. Using HPLC, spectroscopy, flow cytometry and microscopic techniques, we demonstrate that *A. marina* LARK001 contains plentiful amounts of Chl *d* and minor signatures of Chl *a* and lacks the phycocyanins that are present in the *A. marina* type strain MBIC11017. Despite the difference in ecological niche and the large location distance, *A. marina* LARK001's closest genetic relative is the *A. marina* HICR111A strain from Heron Island, Eastern Australia. Taken together, these data suggest that the *A. marina* LARK001 strain is a novel cyanobacterial strain and the first to be isolated from a section of stromatolite where incident light is attenuated and enriched in NIR.

References

1. Holt, A. Further evidence of the relation between 2-desvinyl-2-formyl-chlorophyll a and chlorophyll d. *Can. J. Bot.* **1961**, *39*, 327–331. [CrossRef]
2. Miyashita, H.; Ikemoto, H.; Kurano, N.; Adachi, K.; Chihara, M.; Miyachi, S. Chlorophyll d as a major pigment. *Nature* **1996**, *383*, 402. [CrossRef]
3. Miyashita, H.; Ikemoto, H.; Kurano, N.; Miyachi, S.; Chihara, M. *Acaryochloris marina* gen. et sp. Nov. (cyanobacteria), an oxygenic photosynthetic prokaryote containing Chl d as a major pigment. *J. Phycol.* **2003**, *39*, 1247–1253. [CrossRef]
4. Kühl, M.; Chen, M.; Ralph, P.J.; Schreiber, U.; Larkum, A.W. Ecology: A niche for cyanobacteria containing chlorophyll d. *Nature* **2005**, *433*, 820. [CrossRef] [PubMed]
5. Larkum, A.W.; Kühl, M. Chlorophyll d: The puzzle resolved. *Trends Plant Sci.* **2005**, *10*, 355–357. [CrossRef] [PubMed]
6. Murakami, A.; Miyashita, H.; Iseki, M.; Adachi, K.; Mimuro, M. Chlorophyll d in an epiphytic cyanobacterium of red algae. *Science* **2004**, *303*, 1633. [CrossRef] [PubMed]
7. Miller, S.R.; Augustine, S.; Olson, T.L.; Blankenship, R.E.; Selker, J.; Wood, A.M. Discovery of a free-living chlorophyll d-producing cyanobacterium with a hybrid proteobacterial/cyanobacterial small-subunit rRNA gene. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 850–855. [CrossRef]
8. Mohr, R.; Voss, B.; Schliep, M.; Kurz, T.; Maldener, I.; Adams, D.G.; Larkum, A.D.; Chen, M.; Hess, W.R. A new chlorophyll d-containing cyanobacterium: Evidence for niche adaptation in the genus *Acaryochloris*. *ISME J.* **2010**, *4*, 1456–1469. [CrossRef]
9. Larkum, A.W.; Chen, M.; Li, Y.; Schliep, M.; Trampe, E.; West, J.; Salih, A.; Kühl, M. A Novel Epiphytic Chlorophyll d-containing Cyanobacterium Isolated from a Mangrove-associated Red Alga. *J. Phycol.* **2012**, *48*, 1320–1327. [CrossRef]
10. Behrendt, L.; Larkum, A.W.; Norman, A.; Qvortrup, K.; Chen, M.; Ralph, P.; Sorensen, S.J.; Trampe, E.; Kühl, M. Endolithic chlorophyll d-containing phototrophs. *ISME J.* **2011**, *5*, 1072–1076. [CrossRef]
11. Goh, F.; Allen, M.A.; Leuko, S.; Kawaguchi, T.; Decho, A.W.; Burns, B.P.; Neilan, B.A. Determining the specific microbial populations and their spatial distribution within the stromatolite ecosystem of Shark Bay. *ISME J.* **2009**, *3*, 383–396. [CrossRef] [PubMed]
12. Loughlin, P.; Lin, Y.; Chen, M. Chlorophyll d and *Acaryochloris marina*: Current status. *Photosynth. Res.* **2013**, *116*, 277–293. [CrossRef] [PubMed]
13. De los Rios, A.; Grube, M.; Sancho, L.G.; Ascaso, C. Ultrastructural and genetic characteristics of endolithic cyanobacterial biofilms colonizing Antarctic granite rocks. *FEMS Microbiol. Ecol.* **2007**, *59*, 386–395. [CrossRef] [PubMed]
14. Airs, R.L.; Temperton, B.; Sambles, C.; Farnham, G.; Skill, S.C.; Llewellyn, C.A. Chlorophyll f and chlorophyll d are produced in the cyanobacterium *Chlorogloeopsis fritschii* when cultured under natural light and near-infrared radiation. *FEBS Lett.* **2014**, *588*, 3770–3777. [CrossRef]
15. Averina, S.; Velichko, N.; Senatskaya, E.; Pinevich, A. Far-red light photoadaptations in aquatic cyanobacteria. *Hydrobiologia* **2018**, *813*, 1–17. [CrossRef]
16. Ho, M.Y.; Bryant, D.A. Global Transcriptional Profiling of the Cyanobacterium *Chlorogloeopsis fritschii* PCC 9212 in Far-Red Light: Insights into the Regulation of Chlorophyll d Synthesis. *Front. Microbiol.* **2019**, *10*, 465. [CrossRef]
17. Ho, M.Y.; Shen, G.; Canniffe, D.P.; Zhao, C.; Bryant, D.A. Light-dependent chlorophyll f synthase is a highly divergent paralog of PsbA of photosystem II. *Science* **2016**, *353*, aaf9178. [CrossRef]
18. Ho, M.Y.; Soulier, N.T.; Canniffe, D.P.; Shen, G.; Bryant, D.A. Light regulation of pigment and photosystem biosynthesis in cyanobacteria. *Curr. Opin. Plant Biol.* **2017**, *37*, 24–33. [CrossRef]
19. Miyashita, H.; Ohkubo, S.; Komatsu, H.; Sorimachi, Y.; Fukayama, D.; Fujinuma, D.; Akutsu, S.; Kobayashi, M. Discovery of chlorophyll d in *Acaryochloris marina* and chlorophyll f in a unicellular cyanobacterium, strain KC1, isolated from Lake Biwa. *J. Phys. Chem. Biophys.* **2014**, *4*, 1. [CrossRef]
20. Nürnberg, D.J.; Morton, J.; Santabarbara, S.; Telfer, A.; Joliot, P.; Antonaru, L.A.; Ruban, A.V.; Cardona, T.; Krausz, E.; Boussac, A.; et al. Photochemistry beyond the red limit in chlorophyll f-containing photosystems. *Science* **2018**, *360*, 1210–1213. [CrossRef]
21. McNamara, C.J.; Perry, T.D.; Bearce, K.A.; Hernandez-Duque, G.; Mitchell, R. Epilithic and endolithic bacterial communities in limestone from a Maya archaeological site. *Microb. Ecol.* **2006**, *51*, 51–64. [CrossRef] [PubMed]
22. Partensky, F.; Six, C.; Ratin, M.; Garczarek, L.; Vaulot, D.; Probert, I.; Calteau, A.; Gourvil, P.; Marie, D.; Grebert, T.; et al. A novel species of the marine cyanobacterium *Acaryochloris* with a unique pigment content and lifestyle. *Sci. Rep.* **2018**, *8*, 9142. [CrossRef] [PubMed]
23. Swingley, W.D.; Chen, M.; Cheung, P.C.; Conrad, A.L.; Dejesa, L.C.; Hao, J.; Honchak, B.M.; Karbach, L.E.; Kurdoglu, A.; Lahiri, S.; et al. Niche adaptation and genome expansion in the chlorophyll d-producing cyanobacterium *Acaryochloris marina*. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 2005–2010. [CrossRef] [PubMed]
24. Fisher, A.; Wangpraseurt, D.; Larkum, A.W.D.; Johnson, M.; Kuhl, M.; Chen, M.; Wong, H.L.; Burns, B.P. Correlation of bio-optical properties with photosynthetic pigment and microorganism distribution in microbial mats from Hamelin Pool, Australia. *FEMS Microbiol. Ecol.* **2019**, *95*, fiy219. [CrossRef]
25. Ruvindy, R.; White III, R.A.; Neilan, B.A.; Burns, B.P. Unravelling core microbial metabolisms in the hypersaline microbial mats of Shark Bay using high-throughput metagenomics. *ISME J.* **2016**, *10*, 183–196. [CrossRef]
26. Wong, H.L.; Smith, D.-L.; Visscher, P.T.; Burns, B.P. Niche differentiation of bacterial communities at a millimeter scale in Shark Bay microbial mats. *Sci. Rep.* **2015**, *5*, 15607. [CrossRef]

27. Wong, H.L.; Visscher, P.T.; White III, R.A.; Smith, D.-L.; Patterson, M.M.; Burns, B.P. Dynamics of archaea at fine spatial scales in Shark Bay mat microbiomes. *Sci. Rep.* **2017**, *7*, 1–12. [CrossRef]

28. Wong, H.L.; White, R.A.; Visscher, P.T.; Charlesworth, J.C.; Vázquez-Campos, X.; Burns, B.P. Disentangling the drivers of functional complexity at the metagenomic level in Shark Bay microbial mat microbiomes. *ISME J.* **2018**, *12*, 2619–2639. [CrossRef]

29. Schindelin, J.; Arganda-Carreras, I.; Frise, E.; Kaynig, V.; Longair, M.; Pietzsch, T.; Preibisch, S.; Rueden, C.; Saalfeld, S.; Schmid, B.; et al. Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **2012**, *9*, 676–682. [CrossRef]

30. Herdean, A.; Hall, C.C.; Pham, L.L.; Macdonald Miller, S.; Pernice, M.; Ralph, P.J. Action Spectra and Excitation Emission Matrices reveal the broad range of usable photosynthetic active radiation for *Phaeodactylum tricornutum*. *Biochim. Biophys. Acta Bioenerg.* **2021**, *1862*, 148461. [CrossRef]

31. Van Heukelem, L.; Thomas, C.S. Computer-assisted high-performance liquid chromatography method development with applications to the isolation and analysis of phytoplankton pigments. *J. Chromatogr. A* **2001**, *910*, 31–49. [CrossRef]

32. Schloss, P.D.; Westcott, S.L.; Ryabin, T.; Hall, J.R.; Hartmann, M.; Hollister, E.B.; Lesniewski, R.A.; Oakley, B.B.; Parks, D.H.; Robinson, C.J.; et al. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **2009**, *75*, 7537–7541. [CrossRef] [PubMed]

33. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glockner, F.O. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **2013**, *41*, D590–D596. [CrossRef] [PubMed]

34. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef]

35. Letunic, I.; Bork, P. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **2019**, *47*, W256–W259. [CrossRef]

36. Hill, R.; Schreiber, U.; Gademann, R.; Larkum, A.; Kühl, M.; Ralph, P. Spatial heterogeneity of photosynthesis and the effect of temperature-induced bleaching conditions in three species of corals. *Mar. Biol.* **2004**, *144*, 633–640. [CrossRef]

37. Soo, R.M.; Skennerton, C.T.; Sekiguchi, Y.; Imelfort, M.; Paech, S.J.; Dennis, P.G.; Steen, J.A.; Parks, D.H.; Tyson, G.W.; Hugenholtz, P. An expanded genomic representation of the phylum cyanobacteria. *Genome Biol. Evol.* **2014**, *6*, 1031–1045. [CrossRef] [PubMed]

38. Suosaari, E.P.; Reid, R.P.; Playford, P.E.; Foster, J.S.; Stolz, J.F.; Casaburi, G.; Hagan, P.D.; Chirayath, V.; Macintyre, I.G.; Planavsky, N.J.; et al. New multi-scale perspectives on the stromatolites of Shark Bay, Western Australia. *Sci. Rep.* **2016**, *6*, 20557. [CrossRef]

Article

Chromatic Acclimation Processes and Their Relationships with Phycobiliprotein Complexes

Fanyue Wang [ORCID] and Min Chen *[ORCID]

School of Life and Environmental Sciences, University of Sydney, Sydney, NSW 2006, Australia;
fwan8133@uni.sydney.edu.au
* Correspondence: min.chen@sydney.edu.au; Tel.: +61-2-9036-5006

Abstract: Chromatic acclimation (CA) is a widespread mechanism for optimizing the composition of phycobiliprotein complexes to maximize the cyanobacterial light capture efficiency. There are seven CA types, CA1-CA7, classified according to various photoregulatory pathways. Here, we use sequence analyses and bioinformatics to predict the presence of CA types according to three GAF (cGMP phosphodiesterase/adenylyl cyclase/FhlA)-containing photoreceptors, CcaS (cyanobacterial chromatic acclimation sensor), RcaE (regulator of chromatic adaptation), and RfpA (regulator for far-red photoacclimation). These photoreceptors were classified into three different phylogenetic groups leading different CA types in a diverse range of cyanobacteria. Combining with genomic information of phycobilisome compositions, the CA capabilities of various cyanobacteria were conjectured. Screening 65 accessible cyanobacterial genomes, we defined 19 cyanobacteria that have the capability to perform far-red light photoacclimation (FaRLiP) under the control of RfpA. Forty out of sixty-five cyanobacteria have the capability to perform green/red light photoacclimation, although they use different photoreceptors (RcaE and/or CcaS) and photoregulatory pathways. The reversible response of photoreceptors in CA regulation pathways trigged by changed light conditions reflects the flexibility of photoregulatory mechanisms in cyanobacteria and the putative independent evolutionary origin of photoacclimation types.

Keywords: cyanobacterial photoreceptors; phycobiliproteins; chromatic acclimation; linker proteins; phylogenetic comparison

1. Introduction

Cyanobacteria are primary producers and contribute significantly to the global carbon and nitrogen flows [1]. They exist in diverse ecological niches ranging from desert to rock surface, ice sheet, saline soil, or forest soil, from fresh water to marine water bodies. As photoautotrophic organisms, light is an imperative energy source and an essential environmental indicator [2]. Since light can vary both in quality and quantity, sensing and responding to light are key mechanisms of cyanobacterial eco-physiological versatility, including increasing light-harvesting efficiency for photosynthesis.

The cyanobacterial photosynthetic apparatus comprises two major pigment-binding protein complexes: chlorophyll (Chl)-binding protein complexes localized inside the thylakoid membranes and bilin-binding protein complexes localized outside thylakoid membranes. Several different types of light-harvesting bilin-binding proteins are classified based on the bound chromophore types responsible for their different absorption maxima: phycocyanin (PC) bound phycocyanobilin (PCB) and gives the absorption maximum (λ_{max}) = ~620 nm [3]; phycoerythrin (PE) bound phycoerythrobilin (PEB) and gives the λ_{max} = ~565 nm; phycoerythrocyanin (PEC) bound PCB or PVB and gives the λ_{max} = ~575 nm [3,4]; and allophycocyanin (APC) bound PCB and gives the red-shifted λ_{max} = ~650 nm [5]. Multiple bilin-binding proteins (also named as chromophorylated phycobiliproteins) form an $\alpha_6\beta_6$ oligomers disc unit and those discs are

assembled by their associated non-chromophorylated linker proteins and are integral to the supramolecular phycobilisome (PBS) complexes. PBS is the main antenna system for cyanobacteria and captures light of wavelengths that are poorly absorbed by Chls and transfers the energy to Chls in photosystems [6]. The organization of PBS structure relates to the presence of PBS linker proteins, especially the linkers associated with APC core and membranes [7]. The linker proteins facilitate the assembly of PBS and modulate the absorption properties of PBS, promoting the unidirectional down-hill energy transfer pathways from outside short-wavelength absorption rod structures (PEC, PE, and PC) to the core structure of APC, and then to photosystems supporting the light-harvesting efficiency [8].

In many cyanobacteria, the composition of PBS can change to accommodate the prevalent wavelength of light in the environment. Chromatic acclimation (CA) is a phenomenon observed in cyanobacteria, in which cyanobacteria perceive the wavelength of light in the environment and transduce the light signals into a sequence of biochemical events resulting in changing the activities of genes encoding specific phycobiliproteins [9,10]. Cellular CA responses represent the immediate changes occurring by changes at the level of gene expression. In addition, light intensity and nutrients can change gene expression profiles and PBS structures. The current article will focus on the diversity of molecular events involved in CA by screening known genomic information. Most of the information concerning green/red and red/far-red transducing reactions are presented here.

Currently, seven CA variants have been classified and defined as CA1–CA7 [11,12]. The CA1, CA2, CA3, and CA7 represent the response to green and red light [13]; CA4 represents the response to green and blue light [14,15]; CA5 and CA6 govern photo-regulation triggered by red and far-red light. However, the photoreceptor regulating CA4 (Green/blue light acclimations) and CA5 (reported in *Acaryochloris marina* MBIC11017 only) are still to be defined [16,17]. Here, we studied the types of CA regulated by GAF (cGMP phosphodiesterase/adenylyl cyclase/FhlA)-containing photoreceptors. GAF domain-containing photoreceptors use bilins as their chromophore and, typically, can absorb light in two regions of the wavelengths to interchange between photoconvertible isomers [18]. The types of CA responding to green/red light are mainly regulated either by the chromatic acclimation sensor of CcaS reported in CA1, CA2 and CA7 or the chromatic acclimation sensor of RcaE governing CA3 processes [11,12,19]. The canonical CA3-capable cyanobacterium, *Fremyella diplosiphon*, has almost no PE in red light conditions and each rod contains three PC discs (hexamers). This organism will synthesize PE, and each rod contains two to three PE discs and one PC disc under green light conditions [20]. Both CcaS and RcaE photosensors contain the PAS/GAF core modules and defined as a green/red photosensor using different photo-regulatory pathways [21]. The GAF domain of CcaS has the maximal absorption at wavelengths of 536 nm (green) and 672 nm (red) [19]. CA1 shows no altered levels of red-light absorbing PC and green light-absorbing PE in the rod of PBS and a newly assembled CpcL-rod-shaped PBS alternatively associated with Photosystem I under prevalent green light conditions [22,23]. CcaS leads to green-activated phosphorylation of transcriptional regulator CcaR to regulate the expression of a gene operon that encodes the rod–membrane linker (CpcL) in CA1, PE-associated rod linkers and regulators in CA2, and PEC in CA7 [11]. CA2 is controlled by the CcaS photoregulatory pathway and exhibited that PE levels are photo-acclimated components in response to the changed light signals [19,24].

CA3 is also known as complementary chromatic acclimation (CCA), which showed that PE is accumulated under green light and PC differentially accumulates in the outer portion of PBS rods under red light [13]. RcaE transduces the light signals through two response regulators, RcaF and RcaC, to control the pigment content of PBS associated with CA3 responses. In red light, genes encoding PE proteins are repressed and in the green light, the repression of PE synthesis genes are removed resulting in PE accumulation [2,21].

The newly defined CA6 regulates the far-red-absorbing variants of PBSs to couple with far-red light photo-acclimation (FaRLiP) upregulation and Chl *f* production [25–28]. The FaRLiP gene cluster is controlled by a three-gene operon, *rpfB-rfpA-rpfC*, and protein RfpA is

a phytochrome photoreceptor containing a PAS, GAF and histidine kinase domains [27,29]. The null mutant of three genes, *rfpA, rfpB,* and *rfpc*, caused no CA6-capability in CA6-capable cyanobacteria [27]. Here, we analyzed publicly available cyanobacterial genomes and show that 19 cyanobacteria have the FaRLiP gene cluster containing a three-gene operon, *rpfB-rfpA-rpfC*.

CcaS, RcaE and RfpA have a common bilin-binding GAF domain and have a unique photocycle in response to the changed illumination [27,30]. However, the CcaSR, RcaEFC and RfpABC photosensory systems show distinct signal transduction pathways and target different PBS genes, suggesting that they originate from a common photosensory system and differentiated during the evolution of CA in cyanobacteria. PBS proteins are a major constituent in the cells that allows cells to optimize absorption of light for photosynthesis. Most CA-capable cyanobacteria are able to assemble two different PBSs based on their light wavelength-dependent accumulation of phycobiliproteins and their associated linkers. The cooperation of the variety of PBS becomes an essential element for multiple photoregulatory pathways to enhance the efficiency of photoacclimation and provide cell growth advantage. Here, we report that CA types controlled by CcaS, RcaE or RfpA are widespread throughout the cyanobacterial phylum. Comparing conserved C-terminal protein sequences of CpcG (PBS rod–core linker protein) and CpcL (PBS rod–membrane linker protein), we reannotated CpcL encoding genes and the possibility of the formation of CpcL-rod PBS induced by corresponding light signals was predicted, which may have direct implications on the regulation of the structure of PBS in vivo.

2. Materials and Methods

2.1. Selected Cyanobacteria Genomes and Predicated Photosensory Proteins

The representatives' species covering each defined CA type were selected after retrieving GAF-containing photoreceptors of CcaS, RcaE, and RfpA, and the characteristic PBS from the accessible genomic datasets. The selected sixty-five cyanobacterial genomic sequences were obtained from GeneBank (NCBI) and Integrated Microbial Genomes of the DOE joint Genome institute (Table S1).

16S rRNA sequences were deduced from the genome sequence and used to classify cyanobacteria into groups (Table S1). Further cyanobacterial taxonomy information was retrieved from JGI IMG database (Joint Genome Institute (JGI) Integrated Microbial Genomes (IMG); https://img.jgi.doe.gov/, accessed on November 2021–May 2022).

An array of GAF-containing protein homologs was defined after genomic annotation and re-defined by the BLASTP program at www.ncbi.nlm.nih.gov to screen for the presence of putative GAF-containing proteins. The conserved protein domains and domain structure of predicted GAF-containing proteins were compared and reannotated after the known photoreceptors reported from cyanobacteria, such as CcaS of *Synechocystis* sp. PCC 6803 (WP_014407164) and RcaE of *Fremyella diplosiphon* Fd33 (WP_045871568). RfpA homologs were determined after defined FaRLiP cluster and known RfpA protein module organization of *Halomicronema hongdechloris* C2206 (WP_080806407).

Domain organization of photoreceptor proteins was determined using the Conserved Domains Search program in NCBI connected with the Pfam (protein families) database. The CA types were predicted by combining the presence of photoreceptors and photoacclimated PBS components based on the definition of CA types [17].

2.2. Sequence Homology and Phylogenetic Analysis

The sequences of GAF domains (lengths of 149–174 amino acids) from representatives of CcaS, RcaE and RfpA homologs were extracted after defined functional domains. The multiple sequence alignments were generated through the MEGA-5 program using ClustalW program [31] and refined after published GAF-domain alignment in Hirose et al. [30]. The phylogenetic analyses were performed using the Neighbour-Joining method and Jones–Thornton–Taylor (JTT)-based model in MEGA-5 [31]. The equal Aa

substitution rates were used, and the consensus trees were generated and supported by 5000 bootstrap replicates.

Protein sequences of phycobiliproteins and linker proteins were obtained using BLASTP searches through the selected cyanobacterial genomic sequences to collect any possible mis-annotated linker proteins. The PBS rod–core linker proteins (CpcGs, containing PBS linker domain Pfam00427) were verified and separated CpcL (also containing PBS linker domain Pfam00427) from CpcG linkers by comparing the predicted secondary structure if there was a hydrophobic helix at the C-terminal region (https://services.healthtech.dtu.dk/service.php?TMHMM-2.0, accessed on March–June 2022). The phylogenetic analyses of linker proteins were performed using the Neighbour-Joining method and Jones–Thornton–Taylor (JTT)-based model in MEGA-5 [31]. The equal Aa substitution rates were used, and the consensus trees were generated and supported by 500 bootstrap replicates.

The genes encoding 15,16-dihydrobiliverdin:ferredoxin oxidoreductase (PebA) and phycoerythrobilin:ferredoxin oxidoreductase (PebB) or their homologs are used to verify the presence of PEB-associated protein subunits, PE-rods.

3. Results and Discussion

3.1. Photoreceptors and CA-Capable Cyanobacteria

3.1.1. Predication of Photoreceptors

The GAF-containing protein homologs were retrieved after screening 65 selected cyanobacterial genomes. All retrieved GAF-containing proteins were classified by comparing the protein domain architectural structure to the characterized CcaS (WP_014407164), RcaE (WP_045871568), and RfpA (WP_080806407) (Figure S1). The constructed phylogenetic tree demonstrated three photoreceptor groups, verifying the prediction of photoreceptors (Figure 1). The chromophore binding sites (Cys) are conserved in three photoreceptors. Different amino acids involved in the function of the protochromic triad are noticed in RfpA. Conserved histidine (His) next to Cys is consistent in the GAF of RfpA and replaced leucine (Leu) in the CcaS and RcaE (Figure 1). The imidazole group of His represses the thiol adduct formation and plays an important role for the cleaving of the thioether linkage in the 15E state of bilins in RfpA, supporting red/far-red conversion [32].

3.1.2. Characterization of CA-Capabilities

After screening the 65 cyanobacterial genomes, 16 out of 65 cyanobacteria contain CcaS homologs of *Synechocystis* sp. PCC 6803, including 3 of them that also contain the RcaE photoreceptor (Figure 2). The co-existing CcaS and RcaE in *Calothrix* sp. NIES-2100, Calothrix brevissima NIES-22, and *Calothrix* sp. 336/3 allow them to conduct multiple types of green-/red-light-induced CA, including CA1, CA2, and CA3 (Figure 2; Table 1).

Twenty-seven out of 65 cyanobacteria contain RcaE homologs of *F. diplosiphon*, and all of them have the capability to synthesize PEB (Table S1; Figure 2). The occurrence of the RcaE photoreceptor is directly related to the capability of assembling PE-containing PBS and it is independent of cyanobacterial families (Figure 3).

Initial searches for RfpA homologs (WP_080806407, *H. hongdechloris*) were performed via BLASTP and returned many hits of GAF domain-containing proteins. We combined the results of searching for the FaRLiP cluster near the putative RfpA homologs or operon of *rfpB-rfpA-rfpC* and defined 19 out of 65 cyanobacteria containing the FaRLiP gene cluster including 5 cyanobacteria their FaRLiP cluster placed in different genomic scaffolds (Figure 4 and Appendix S1). As demonstrated in Figure 4, all 14 FaRLiP clusters contain core subunits of PSI, PSII and PBS except *Fischerella* sp. NIES-4106. *Fischerella* sp. NIES-4106 contain only one copy of *psaA* but two copies of *psaB*. The PsaB encoding gene in the FaRLiP cluster of *Fischerella* sp. NIES-4106 is shorter, ~602 Aa, due to missing ~150 Aa at the N-terminal of typical PsaB. A different organisation of photosynthetic apparatus induced by CA6 could be expected in *Fischerella* sp. NIES-4106 due to dysfunctional PSI core encoding genes in FaRLiP cluster. The presences *rfpBAC* operon and FaRLiP cluster confirmed they are CA6-capable cyanobacteria.
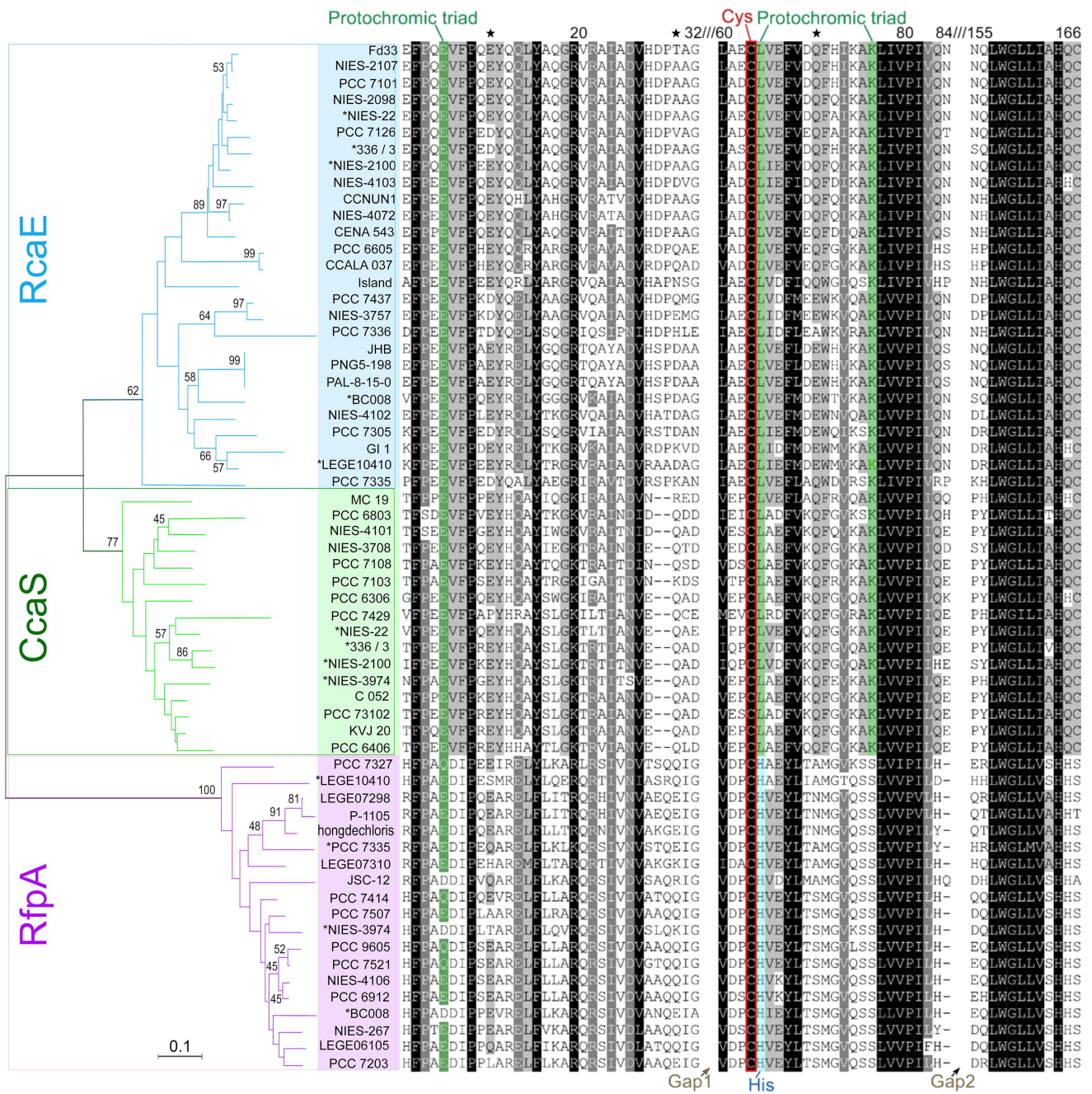
Figure 1. Phylogenetic analysis and GAF domain alignments of photoreceptors. The groups of RcaE (Blue), CcaS (green), and RfpA (purple) are supported by the alignment fragments of GAF domains. The conserved Cys sites are highlighted in a red framed box. The protochromic triad sites are highlighted in green and conserved His sites of RfpA are highlighted in blue. The asterisk represents that the strain uses two different photoreceptors. The branch supports are calculated by 5000 replicates. The five-pointed stars represent the positions of Aa alignment at 10, 30, and 70, respectively. The strain culture numbers are used in the figure and the photoreceptor sequence accessions are listed in Table S1.
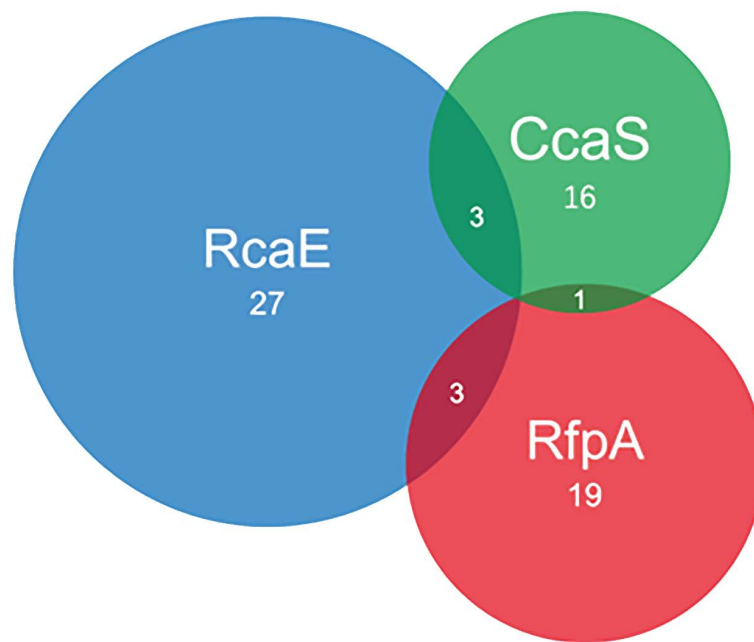
Figure 2. Venn diagram summarized the numbers of photoreceptors from 65 selected cyanobacterial genomes. The numbers in the overlayed regions represent the cases of co-existing photoreceptors. Green, CcaS; Blue, RcaE; Red, RfpA.

Table 1. Co-existing photoreceptors and chromatic acclimation types.

Cyanobacteria	Photoreceptors		CA Types				Phycobiliproteins	Linker Proteins
Calothrix sp. NIES-3974	CcaS/RfpA	CA1	CA2			CA6	APC/PC/PE	ApcE×2/CpcG×3/CpcL×2
Calothrix sp. NIES-2100	CcaS/RcaE	CA1	CA2	CA3			APC/PC/PE	ApcE×1/CpcG×1/CpcL×2
Calothrix brevissima NIES-22	CcaS/RcaE	CA1	CA2	CA3			APC/PC/PE	ApcE×1/CpcG×2/CpcL×2
Calothrix sp. 336/3	CcaS/RcaE	CA1	CA2	CA3			APC/PC/PE	ApcE×1/CpcG×2/CpcL×2
Calothrix sp. 336/3	CcaS/RcaE	CA1	CA2	CA3			APC/PC/PE	ApcE×1/CpcG×2/CpcL×2
Mastigocoleus testarum BC008	RcaE/RfpA			CA3	CA6		APC/PC/PE	ApcE×3/CpcG×1/CpcL×2
Pleurocapsales cyanobacterium LEGE 10410	RcaE/RfpA			CA3	CA6		APC/PC/PE	ApcE×2/CpcG×1
Synechococcus sp. PCC 7335	RcaE/RfpA			CA3	CA6		APC/PC/PE	ApcE×2/CpcG×1/CpcL×1
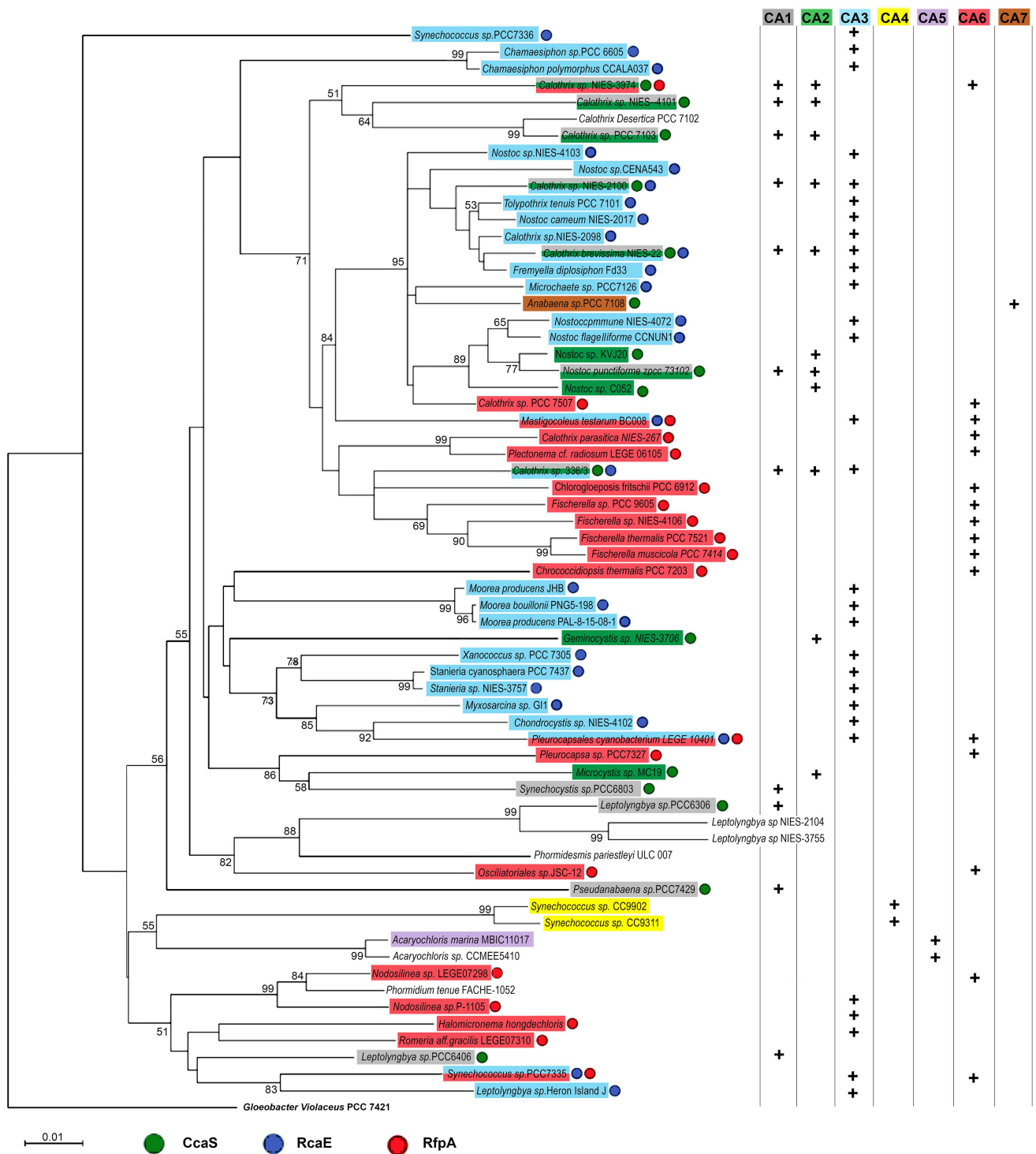
Figure 3. Neighbor-joining phylogeny of selected 65 cyanobacteria based on 16S rRNA sequences and defined chromatic acclimation (CA) types. The CA types are color-coded as CA1, gray; CA2, green; CA3, blue; CA4, yellow; CA5, purple; CA6, red; and CA7, brown. The photoreceptors are illustrated as circles next to the cyanobacterial names and are color-coded as CcaS (Green), RcaE (Blue), and RfpA (Red). The branched groups are supported with a bootstrap of ≥50%. The scale bar represents a phylogenetic distance of 0.01 nucleotide substitutions per site. All sequence accessions are listed in Table S1.
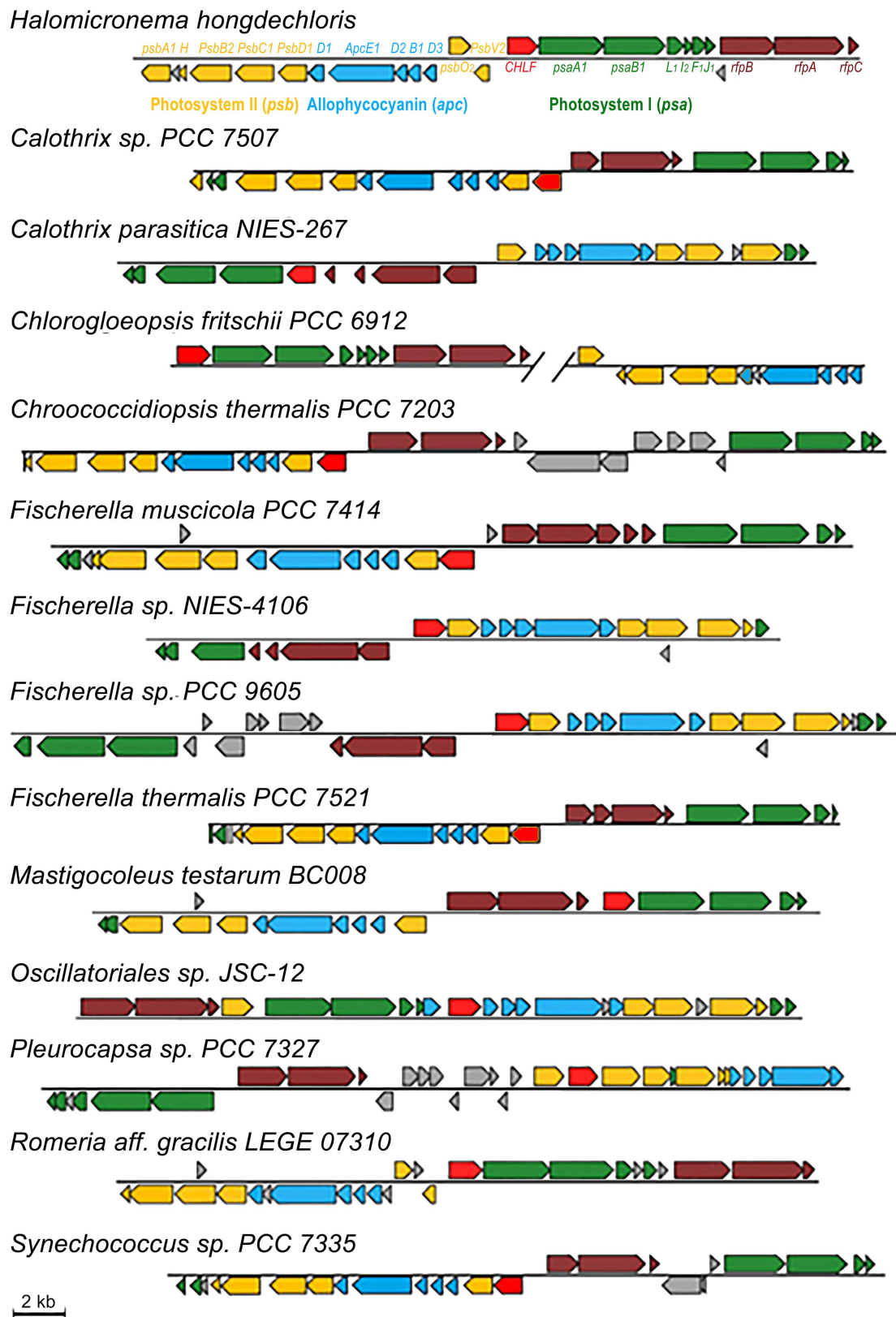
Figure 4. FaRLiP gene clusters from 14 CA6-capable cyanobacteria. The identities of the genes from *H. hongdechloris* are used as references. Blue represents the Allophycocyanin (Apc) subunits; green represents genes encoding photosystem I (*psa*); yellow represents genes encoding photosystem II (*psb*); red represents Chl *f* synthase (CHLF) homolog; dark-red represents the Rfp operons (*rfp*ABC); gray represents hypothetical proteins.

There are RfpA-like homologs detected in *Calothrix desertica* PCC 7102 and *Nostoc sp.* C052, but no detectable FaRLiP gene cluster in the genomes (data not shown). Those RfpA-like photoreceptors are not involved in CA6-type photoacclimation and their regulatory functions are unknown. Interestingly, screening the genomic draft of *Calothrix desertica* PCC 7102, we could not find homologs of CcaS and RcaE, although there are 27 GAF-containing hypothetical proteins (Table S1). We proposed that *Calothrix desertica* PCC 7102 does not have known CA capability. Additionally, another five out of 65 cyanobacteria demonstrated no known CA1-CA3 and CA6 capability due to lacking known CA-regulators of CcaS, RcaE, or RfpA including the outgroup cyanobacterium *Gloeobacter violaceus* PCC 7421. One common feature of those non-CA capable cyanobacteria, excluding the outgroup of *Gloeobacter violaceus* PCC 7421, is that APC and PC are dominant subunits of PBS and no PE and PE-associated encoding genes are found in their genomes (Table S1). However, *Acaryochloris* sp. CCMEE 5410 is assigned as non-CA capable cyanobacterium due to lacking PBS rod encoding genes.

As described above, the altered ratio of PE:PC in PBS rods reflects the spectral distribution of green/red light and is governed mainly by photoreceptors of CcaS and RcaE. In this report, 26 out of 65 cyanobacteria do not have PE proteins, including 1 CA7 representative, 2 CA5-capable representatives, and 4 CA1-capable cyanobacteria (Table S1).

Most CA6-capable cyanobacteria possess PBS without involvement of PE and only 5 out of 19 CA6-capable cyanobacteria contain PE encoding genes. The far-red light photosensor of RfpA co-exists with RcaE in *Mastigocoleus testarum* BC008, *Pleurocapsales cyanobacterium* LEGE 10410, and *Synechococcus* sp. PCC 7335, resulting in their complicated photoacclimation response of green/red (CA3) and red/far-red (CA6). The three CA3 and CA6 enabling cyanobacteria belong to three different orders of cyanobacteria and have different morphological features, suggesting the CA development is independent of the morphological properties and their family orders (Table 1). In *Synechococcus* sp. PCC 7335, a null mutant of *rfpA* affected the ratios of PE:PC, suggesting the interaction between CA3 and CA6 regulatory pathways [33].

3.2. Cyanobacterial PBS Structure and CA Responses

Tuning PBS composition in response to distinct wavelengths has been recognised, such as the PBS having extended PE rods will prefer to capture green light observed in CA2-capable cyanobacteria. Knowing the genes encoding phycobiliproteins and linkers by screening genomic information could help us to predict the involvement of PBS in CA processes. In this section, we identified the possible types of PBS according to the PBS encoding gene profiles.

3.2.1. Changeable PE Rods

The presence of PE encoding genes was learned from published genomic sequences and confirmed by capability of PEB synthesis, i.e., if there are PebA and PebB encoding genes annotated in the genome. In total, 39 out of 65 cyanobacteria can synthesize PE and only two cyanobacteria, *Calothrix desertica* PCC 7102 and *Gloeobacter violaceus* PCC 7421, demonstrate constitutively assembling PE-containing PBS without predicted photoacclimation changes. The assigned 27 CA3-capable cyanobacteria will change the ratio of PE:PC in the rod under control of the RcaE mechanism (Table S1). The altered ratio of PE:PC results in visible cellular color changes. No PE component can be detected in the CA3-capable cyanobacteria grown under red light conditions, resulting in blue-green colored appearance [13].

The assigned 11 CA2-capable cyanobacteria contain the CcaS regulatory pathway. The upregulating product of PE in green light extended the rod length without changes of PC amount at the base part of the rod [12]. PE-associated genes are constitutively active and PE subunits are an important composition of PBS independent of culture conditions, but PE components will be increased in green light conditions.

PE is the main acclimating rod component controlled by green light in CA2- or CA3-capable cyanobacteria. However, not all PE-containing PBS photoacclimate. Approximately 27% of PE-containing cyanobacteria were not capable of altering either PE or PC amount under changing light colors [13]. The photoacclimating morphological changes are only observed in CA2- and CA3-capable cyanobacteria (Table S1). For example, *Gloeobacter violaceus* PCC 7421 demonstrated purple cellular color due to constitutive expression of PE encoded genes, which are not governed by photoacclimation.

3.2.2. Assembling CpcL-Rod PBS Structure

CA1-capable cyanobacteria share the same photoreceptor of CcaS, although the gene organization controlled by CcaS might be differentiated to regulate products of PE or CpcL. The co-presence of the rod–membrane linker (CpcL) and CcaS regulatory pathways are the essential features of CA1-capable cyanobacteria [11]. In the process of CA1, CcaS mediate the product of *cpcL* gene and assemble phycobiliprotein rod interacting with PSI via CpcL to enhance energy distribution into PSI under green light conditions [34]. Both CpcL and CpcG contain a conserved linker domain of pfam00427 but have different hydrophilic/hydrophobic properties (Figure S2). CpcL serves as a rod–membrane linker and the hydrophobic helix at the C-terminal region enables it to connect the rod directly to photosynthetic reaction centers inside membranes [23]. In this study, we screened all annotated *cpcG* genes in the selective cyanobacterial genomes and predicted the presence of hydrophobic helices using the TMHMM program. In total, 42 out of 65 cyanobacteria contain CpcL encoding genes, while only 11 of them contain photoreceptor CcaS. Reading through the locus of *cpcL*, we noticed that most *cpcL* genes are localized in one or another phycobiliprotein associated operons (Figure 5). The formation of CpcL-rod PBS could be regulated together with other phycobiliproteins by sharing photoreceptors and photoregulatory pathways.
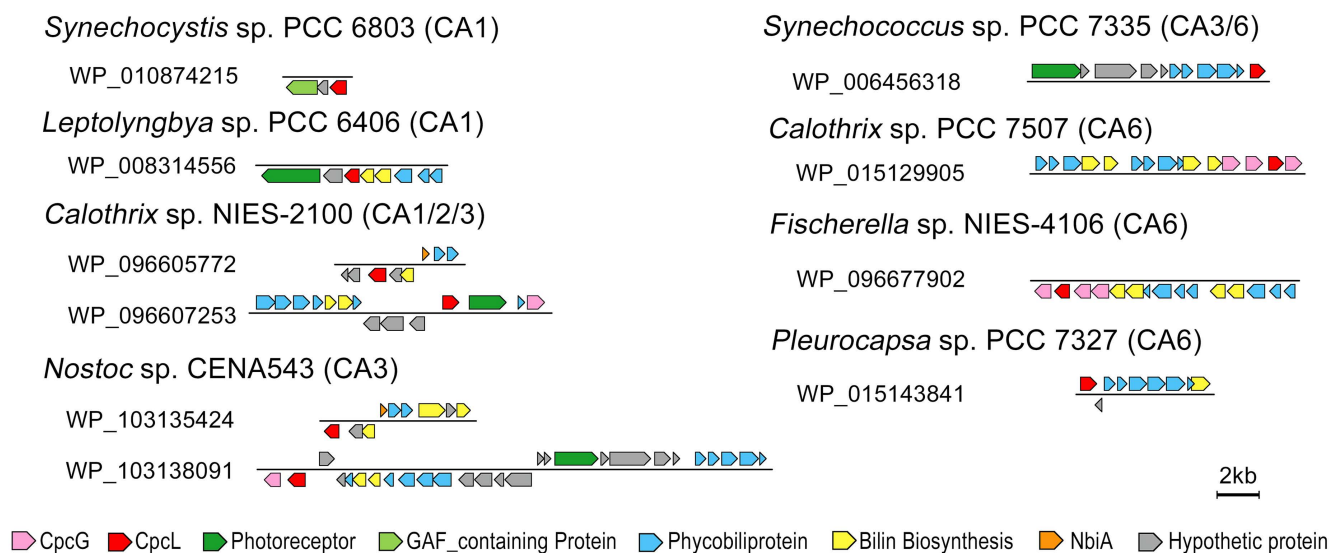
Figure 5. Associated gene locus containing *cpcL* from representative cyanobacteria.

As shown in Figure 5, *cpcL* is localized in the operon containing photoreceptors or close to the gene encoding phycobilin lyase (such as *cpcS*, *cpcE*, and *cpcF*) and phycobiliproteins. Assembling simple CpcL-rod PBS appears to be a commonly used mechanism to balance the energy flow between PSI and PSII, especially under certain light and nutrient conditions [35]. The CpcL-rod PBS was only isolated from green light cultured CA1-capable cyanobacterium *Leptolyngbya* sp. PCC 6406, indicating that the expression level of *cpcL* was regulated by green light [11].

Previous reports also indicated that CpcL-rod PBS in some cyanobacteria are consistently present. In CA5-capable cyanobacterium *A. marina* MBIC11017, CpcL-PC-rod-shaped PBS is the basic PBS structure associated with PSI and PSII [36]. In *Anabaena* sp. PCC 7120, isolated PBS-CpcL-PSI complexes indicated the constant expression of the CpcL encoding gene without involvement of photoregulatory mechanisms [35].

Within the 19 identified CA6-capable cyanobacteria, only three of them including *H. hongdechloris* do not have CpcL encoding genes (Table S1). Most *cpcL* genes from CA6-capable cyanobacteria are localized within a phycobiliprotein operon containing phycobiliprotein subunits and phycobilin assembling/biosynthesis enzymes (Figure 5). Interestingly, phylogenetic analysis of CpcL revealed a separate group of CpcL from some CA6-capable cyanobacteria, although there are CpcLs from other CA6-capable cyanobacteria that are placed in different phylogenetic groups. Independent evolutionary/development pathways of CpcL-rod PBS and photoacclimation are proposed (Figure S3).

There are 14 out of 27 CA3-capable cyanobacteria having CpcL encoding genes, and most *cpcL* genes are in the gene operon of phycobiliproteins modulated by RcaE or close to phycobilin biosynthesis enzymes, which might be modulated by different photoreceptors. Further investigation on the formation of CpcL-rod in CA3-capable cyanobacteria could reveal updated knowledge of photoacclimation and adaptive photosynthetic apparatus.

3.2.3. PBS Having Red-Shifted Absorption Characteristics

There are two different PBS structures supported by two copies of core–membrane linker encoding genes (*apcE*) in CA6-capable cyanobacteria (Table S1) [28]. The conventional PBS are assembled under white light conditions, and red-shifted PBS are formed under far-red light conditions. All CA6-capable cyanobacteria have five additional genes encoding APC subunits in the FaRLiP gene cluster, which are essential APC subunits assembling the PBS core with red-shifted absorption [28]. The five APC encoding genes include three ApcD, one ApcB and one ApcE (Figure 4).

ApcE is a high molecular mass core–membrane linker and has a homology to phycobiliproteins at the N-terminal region and several repeats of linker polypeptides at the C-terminal region [6]. The phycobiliprotein-like N-terminal domain can bind a PCB and serve as a PBS terminal energy acceptor [6]. The number of repeats of linker domains of ApcE are different according to the size of assembled PBS [28]. In CA6-capable cyanobacteria, the product of *apcE* in the FaRLiP gene cluster has two repeats of linker domain, a shorter version of ApcE, assembling a smaller size of PBS. The smallest PBS isolated from far-red light grown *H. hongdechloris* culture provided experimental evidence that ApcE from FaRLiP is involved in assembling small PBS and modulates the absorption properties [25]. Figure 4 demonstrates that five gene clusters of *apcD-E-D-B-D* are present in all CA6-capable cyanobacteria and only *Oscillatoriales* sp. JSC-12 has an additional ApcD encoded gene in the cluster (Figure 4). These APC subunits are responsible for the red-shifted absorption of PBS isolated from *H. hongdechloris* [25,28]. The disorganized structure of PBS due to newly synthesized APC subunits induced by CA6 contributed to the changed spectral properties of PBS in *Synechococcus* sp. PCC 7335 [37].

ApcE proteins from FaRLiP not only have fewer repeats of linker domain, but also lack chromophore-binding sites (conserved cysteine, Cys) in the phycobiliprotein-like domain (Figure 6) [25,33]. Using the N-terminal domain of ApcE, we can separate canonical ApcE from the ApcE in the FaRLiP cluster (Figure 6). Additionally, one ApcD from FaRLiP also lacks chromophore binding sites [38]. The missed chromophore binding sites of phycobiliproteins/domains might play a vital role in modulating the spectral properties of PBS induced by CA6 [37,39].
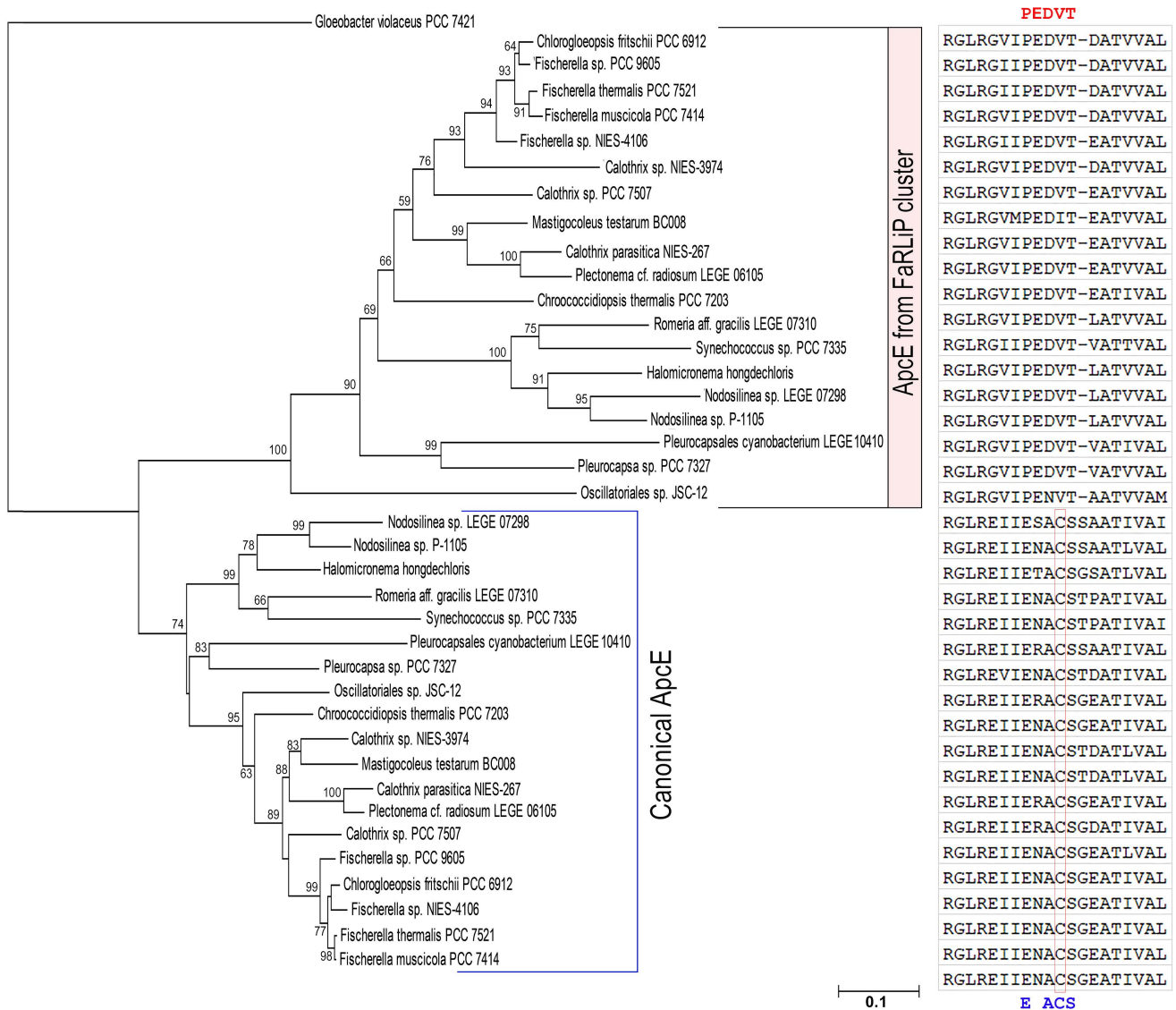
Figure 6. Neighbor-Joining phylogeny and alignment of conserved phycobiliprotein region of ApcEs from CA6-capable cyanobacteria. The conserved Cys sites from canonical ApcE proteins are highlighted in a red framed box. The conserved motif PEDVT in ApcE from FaRLiP cluster are highlighted in red. The conserved motif ExACS in canonical ApcE are highlighted in blue. The ApcE sequence accessions are listed in Table S1.

3.3. Photoacclimation and Alternative PBS

The hemidiscoidal shape of PBS is the typical form, which consists of a central APC core and radiating either PC or PC and PE/PEC rods. With variability in the rod length across species (e.g., CA3-capable cyanobacteria), re-equilibrated energy distribution among photosystems (e.g., CA1-capable cyanobacteria), and inducible PBS consisting of only the core component (e.g., CA6-capable cyanobacteria), altered PBS morphologies reflect the multi-CA-type and photoregulatory pathways. Sixty-five cyanobacteria were selected covering variations of PBS and CA types. Our results offer new insights into how cyanobacteria can flexibly combine different types of CA to acclimate to different light environments. However, with extensively increased numbers of fully sequenced cyanobacteria, new studies on photoreceptors in caynobacteria demonstrated that CA are also modulated by multiple photoreceptors. For example, cyanobacteriochromes (CBCRs) can sense a wide range of wavelengths and modulate CA by refined wavelength resulting in more complicated dual or triple control of CA processes [18,40]. In 2016, Wiltbank and Kehoe

demonstrated a photoreceptor DpxA (decreased phycoerythrin expression A) that senses different wavelengths of light from RcaE and regulated the amount of PE in CA3-capable cyanobacteria [41]. The proposed dual control system in CA3-capable cyanobacteria enables the organism to sense a wide spectrum of light with 'fine-tuning' CA responsive processes. A wide range of photoreceptors provide cyanobacteria flexibility in tuning cellular responses to light. Beyond highlighted photoreceptors retrieved from genomes and their functional role in photosynthesis, a range of GAF-containing proteins are involved in important cellular processes including growth, phototaxis, and cell aggregation, most of them are awaiting characterization and beyond current report scopes (Table S1) [18,42]. Here, we used photoreceptors, CcaS, RcaE and RfpA, that are known to regulate at least one CA process and predicted the similar physiological responses across cyanobacteria. Further studies on photoacclimation in cyanobacteria represent new research frontiers in understanding 'switchable' photosynthetic apparatus in cyanobacteria, but also in building new biotechnological tools.

4. Conclusions

CA is a photoacclimation strategy widely distributed in cyanobacteria. In this study, 65 selected cyanobacterial genomes were explored for the relationship between CA photoregulatory pathways and PBS components. Since the PBS compositions are the main switchable parts in response to changed light conditions, CA capabilities of cyanobacteria rely on the genetic features of PBS. We define the CA types according to the presence of known photoreceptors and potential components for assembling functional PBS.

The phylogeny of conserved GAF domains from different photoreceptors verified the classifications of CcaS, RcaE, and RfpA. The different amino acid residues surrounding the bound chromophore sites indicate the different characterized active wavelength ranges of photoreceptors [26]. The conserved His next to Cys in RfpA play an important role for the active range of red/far-red wavelength [32]. Most CA-capable cyanobacteria can assemble two different types of PBS: (1) the different length of rod and ratio of PE:PC in CA2- and CA3-capable cyanobacteria; (2) co-existing canonical PBS structure serving for PSII and single rod-shaped PBS serving for PSI via CpcL linkers; (3) newly assembled, red-shifted PBS will replace the canonical PBS in CA6-capable cyanobacteria induced by far-red light. With detailed protein alignments, we predicted key motifs and conserved amino acid residues contributing to the spectral characteristics of photoreceptors and the shifted absorption of PBS. The different phylogenetic relationships based on 16S rRNA and photoreceptors suggest that cyanobacteria acquired the CA genes by horizontal gene transfer. However, an independent evolutionary origin of photoacclimation types does not directly reflect the adverse environments.

References

1. Whitton, B.A.; Potts, M. Introduction to the Cyanobacteria. In *The Ecology of Cyanobacteria*; Whitton, B.A., Potts, M., Eds.; Springer: Dordrecht, The Netherlands, 2000; pp. 1–11.

2. Gutu, A.; Kehoe, D.M. Emerging perspectives on the mechanisms, regulation, and distribution of light color acclimation in cyanobacteria. *Mol. Plant* **2012**, *5*, 1–13. [CrossRef] [PubMed]

3. MacColl, R. Cyanobacterial phycobilisomes. *J. Struct. Biol.* **1998**, *124*, 311–334. [CrossRef] [PubMed]

4. Six, C.; Thomas, J.C.; Thion, L.; Lemoine, Y.; Zal, F.; Partensky, F. Two novel phycoerythrin-associated linker proteins in the marine cyanobacterium *Synechococcus* sp. strain WH8102. *J. Bacteriol.* **2005**, *187*, 1685–1694. [CrossRef]

5. Peng, P.; Dong, L.; Sun, Y.; Zeng, X.; Ding, W.; Scheer, H.; Yang, X.; Zhao, K.H. The structure of allophycocyanin B from *Synechocystis* PCC 6803 reveals the structural basis for the extreme redshift of the terminal emitter in phycobilisomes. *Acta Crystallogr. D* **2014**, *70*, 2558–2569. [CrossRef] [PubMed]

6. Liu, H.; Zhang, H.; Niedzwiedzki, D.M.; Prado, M.; He, G.; Gross, M.L.; Blankenship, R.E. Phycobilisomes supply excitations to both photosystems in a megacomplex in cyanobacteria. *Science* **2013**, *342*, 1104–1107. [CrossRef]

7. Watanabe, M.; Ikeuchi, M. Phycobilisome: Architecture of a light-harvesting supercomplex. *Photosynth. Res.* **2013**, *116*, 265–276. [CrossRef]

8. Bryant, D.A.; Canniffe, D.P. How nature designs light-harvesting antenna systems: Design principles and functional realization in chlorophototrophic prokaryotes. *J. Phys. B At. Mol. Opt.* **2018**, *51*, 033001. [CrossRef]

9. Federspiel, N.A.; Grossman, A.R. Characterization of the light-regulated operon encoding the phycoerythrin-associated linker proteins from the cyanobacterium *Fremyella diplosiphon*. *J. Bacteriol.* **1990**, *172*, 4072–4081. [CrossRef]

10. Bhaya, D. In the limelight: Photoreceptors in cyanobacteria. *MBio* **2016**, *7*, e00741-16. [CrossRef]

11. Hirose, Y.; Chihong, S.; Watanabe, M.; Yonekawa, C.; Murata, K.; Ikeuchi, M.; Eki, T. Diverse chromatic acclimation processes regulating phycoerythrocyanin and rod-shaped phycobilisome in cyanobacteria. *Mol. Plant* **2019**, *12*, 715–725. [CrossRef] [PubMed]

12. Sanfilippo, J.E.; Garczarek, L.; Partensky, F.; Kehoe, D.M. Chromatic acclimation in cyanobacteria: A diverse and widespread process for optimizing photosynthesis. *Annu. Rev. Microbiol.* **2019**, *73*, 407–433. [CrossRef] [PubMed]

13. Kehoe, D.M.; Gutu, A. Responding to color: The regulation of complementary chromatic adaptation. *Annu. Rev. Plant. Biol.* **2006**, *57*, 127–150. [CrossRef] [PubMed]

14. Everroad, C.; Six, C.; Partensky, F.; Thomas, J.-C.; Holtzendorff, J.; Wood, A.M. Biochemical bases of type IV chromatic adaptation in marine *Synechococcus* spp. *J. Bacteriol.* **2006**, *188*, 3345–3356. [CrossRef] [PubMed]

15. Shukla, A.; Biswas, A.; Blot, N.; Partensky, F.; Karty, J.A.; Hammad, L.A.; Garczarek, L.; Gutu, A.; Schluchter, W.M.; Kehoe, D.M. Phycoerythrin-specific bilin lyase-isomerase controls blue-green chromatic acclimation in marine *Synechococcus*. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 20136–20141. [CrossRef] [PubMed]

16. Duxbury, Z.; Schliep, M.; Ritchie, R.J.; Larkum, A.W.D.; Chen, M. Chromatic photoacclimation extends utilisable photosynthetically active radiation in the chlorophyll *d*-containing cyanobacterium, *Acaryochloris marina*. *Photosynth. Res.* **2009**, *101*, 69–75. [CrossRef] [PubMed]

17. Sanfilippo, J.E.; Nguyen, A.A.; Karty, J.A.; Shukla, A.; Schluchter, W.M.; Garczarek, L.; Partensky, F.; Kehoe, D.M. Self-regulating genomic island encoding tandem regulators confers chromatic acclimation to marine *Synechococcus*. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 6077–6082. [CrossRef]

18. Wiltbank, L.B.; Kehoe, D.M. Diverse light responses of cyanobacteria mediated by phytochrome superfamily photoreceptors. *Nat. Rev. Microbiol.* **2019**, *17*, 37–50. [CrossRef]

19. Hirose, Y.; Narikawa, R.; Katayama, M.; Ikeuchi, M. Cyanobacteriochrome CcaS regulates phycoerythrin accumulation in *Nostoc punctiforme*, a group II chromatic adapter. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 8854–8859. [CrossRef]

20. Kehoe, D.M.; Grossman, A.R. New classes of mutants in complementary chromatic adaptation provide evidence for a novel four-step phosphorelay system. *J. Bacteriol.* **1997**, *179*, 3914–3921. [CrossRef]

21. Terauchi, K.; Montgomery, B.L.; Grossman, A.R.; Lagarias, J.C.; Kehoe, D.M. RcaE is a complementary chromatic adaptation photoreceptor required for green and red light responsiveness. *Mol. Microbiol.* **2004**, *51*, 567–577. [CrossRef]

22. Kondo, K.; Geng, X.X.; Katayama, M.; Ikeuchi, M. Distinct roles of CpcG1 and CpcG2 in phycobilisome assembly in the cyanobacterium *Synechocystis* sp. PCC 6803. *Photosynth. Res.* **2005**, *84*, 269–273. [CrossRef] [PubMed]

23. Kondo, K.; Ochiai, Y.; Katayama, M.; Ikeuchi, M. The Membrane-associated CpcG2-phycobilisome in *Synechocystis*: A new photosystem I antenna. *Plant Physiol.* **2007**, *144*, 1200–1210. [CrossRef] [PubMed]

24. De Marsac, N.T.; Cohen-bazire, G. Molecular composition of cyanobacterial phycobilisomes. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 1635–1639. [CrossRef] [PubMed]

25. Li, Y.; Lin, Y.; Garvey, C.J.; Birch, D.; Corkery, R.W.; Loughlin, P.C.; Scheer, H.; Willows, R.D.; Chen, M. Characterization of red-shifted phycobilisomes isolated from the chlorophyll *f*-containing cyanobacterium *Halomicronema hongdechloris*. *Biochim. Biophys. Acta Bioenerg.* **2016**, *1857*, 107–114. [CrossRef] [PubMed]
26. Gan, F.; Shen, G.; Bryant, D. Occurrence of Far-Red Light Photoacclimation (FaRLiP) in diverse cyanobacteria. *Life* **2014**, *5*, 4–24. [CrossRef] [PubMed]
27. Zhao, C.; Gan, F.; Shen, G.; Bryant, D.A. RfpA, RfpB, and RfpC are the master control elements of far-red light photoacclimation (FaRLiP). *Front. Microbiol.* **2015**, *6*, 1303. [CrossRef] [PubMed]
28. Chen, M.; Hernandez-Prieto, M.A.; Loughlin, P.C.; Li, Y.; Willows, R.D. Genome and proteome of the chlorophyll *f*-producing cyanobacterium *Halomicronema hongdechloris: Adaptative* proteomic shifts under different light conditions. *BMC Genom.* **2019**, *20*, 207. [CrossRef] [PubMed]
29. Gan, F.; Zhang, S.; Rockwell, N.; Martin, S.; Lagarias, J.; Bryant, D. Extensive remodeling of a cyanobacterial photosynthetic apparatus in far-red light. *Science* **2014**, *345*, 1312–1317. [CrossRef]
30. Hirose, Y.; Rockwell, N.C.; Nishiyama, K.; Narikawa, R.; Ukaji, Y.; Inomata, K.; Lagarias, J.C.; Ikeuchi, M. Green/red cyanobacteriochromes regulate complementary chromatic acclimation via a protochromic photocycle. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 4974–4979. [CrossRef]
31. Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **2011**, *28*, 2731–2739. [CrossRef]
32. Sato, T.; Kikukawa, T.; Miyoshi, R.; Kajimoto, K.; Yonekawa, C.; Fujisawa, T.; Unno, M.; Eki, T.; Hirose, Y. Protochromic absorption changes in the two-cysteine photocycle of a blue/orange cyanobacteriochrome. *J. Biol. Chem.* **2019**, *294*, 18909–18922. [CrossRef] [PubMed]
33. Ho, M.-Y.; Gan, F.; Shen, G.; Bryant, D.A. Far-red light photoacclimation (FaRLiP) in *Synechococcus* sp. PCC 7335. II. Characterization of phycobiliproteins produced during acclimation to far-red light. *Photosynth. Res.* **2017**, *131*, 187–202. [CrossRef]
34. Hirose, Y.; Shimada, T.; Narikawa, R.; Katayama, M.; Ikeuchi, M. Cyanobacteriochrome CcaS is the green light receptor that induces the expression of phycobilisome linker protein. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 9528–9533. [CrossRef]
35. Watanabe, M.; Semchonok, D.A.; Webber-Birungi, M.T.; Ehira, S.; Kondo, K.; Narikawa, R.; Ohmori, M.; Boekema, E.J.; Ikeuchi, M. Attachment of phycobilisomes in an antenna–photosystem I supercomplex of cyanobacteria. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 2512–2517. [CrossRef] [PubMed]
36. Chen, M.; Floetenmeyer, M.; Bibby, T.S. Supramolecular organization of phycobiliproteins in the chlorophyll *d*-containing cyanobacterium *Acaryochloris marina*. *FEBS Lett.* **2009**, *583*, 2535–2539. [CrossRef] [PubMed]
37. Ho, M.-Y.; Gan, F.; Shen, G.; Zhao, C.; Bryant, D.A. Far-red light photoacclimation (FaRLiP) in *Synechococcus* sp. PCC 7335: I. Regulation of FaRLiP gene expression. *Photosynth. Res.* **2017**, *131*, 173–186. [CrossRef] [PubMed]
38. Li, Y.; Chen, M. The specificity of the bilin lyase CpcS for chromophore attachment to allophycocyanin in the chlorophyll *f*-containing cyanobacterium *Halomicronima hongdechloris*. *Photosynth. Res.* **2021**, *151*, 213–223. [CrossRef] [PubMed]
39. Soulier, N.; Laremore, T.N.; Bryant, D.A. Characterization of cyanobacterial allophycocyanins absorbing far-red light. *Photosynth. Res.* **2020**, *145*, 189–207. [CrossRef]
40. Ikeuchi, M.; Ishizuka, T. Cyanobacteriochromes: A new superfamily of tetrapyrrole-binding photoreceptors in cyanobacteria. *Photochem. Photobiol. Sci.* **2008**, *7*, 1159. [CrossRef]
41. Wiltbank, L.B.; Kehoe, D.M. Two cyanobacterial photoreceptors regulate photosynthetic light harvesting by sensing teal, green, yellow, and red light. *MBio* **2016**, *7*, e02130-15. [CrossRef]
42. Franklin, K.A. Light signals, phytochromes and crosstalk with other environmental cues. *J. Exp. Bot.* **2003**, *55*, 271–276. [CrossRef] [PubMed]

Attachment of Ferredoxin: NADP+ Oxidoreductase to Phycobilisomes Is Required for Photoheterotrophic Growth of the Cyanobacterium *Synechococcus* sp. PCC 7002

Xiying Li [1,*,†], Chenhui Huang [1,2,†] , Peijun Wei [1], Kun Zhang [1], Chunxia Dong [1], Qing Lan [1], Zhenggao Zheng [1], Zhengdong Zhang [1] and Jindong Zhao [1,*]

1   State Key Laboratory of Protein and Plant Genetic Engineering, School of Life Science, Peking University, Beijing 100871, China; huangchh@shsmu.edu.cn (C.H.); wei_peijun@pku.edu.cn (P.W.); zk1898@pku.edu.cn (K.Z.); dongcx@pku.edu.cn (C.D.); a7lan9001@163.com (Q.L.); zhengzhenggao@pku.edu.cn (Z.Z.); zhangzhd2017@pku.edu.cn (Z.Z.)
2   School of Medicine, Shanghai Jiao Tong University, Shanghai 200011, China
*   Correspondence: cici0216@pku.edu.cn (X.L.); jzhao@pku.edu.cn (J.Z.); Tel.: +86-10-62756421 (J.Z.)
†   These authors contributed equally to this work.

Abstract: Two types of cyanobacterial phycobilisomes (PBS) are present: the hemidiscoidal PBS (CpcG-PBS) and the membrane-bound PBS (CpcL-PBS). Both types of PBS have ferredoxin:NADP+ oxidoreductase (FNR) attached to the termini of their rods through a CpcD domain. To date, the physiological significance of the attachment remains unknown. We constructed a mutant (dF338) which contains an FNR lacking the N-terminal CpcD domain in *Synechococcus* sp. PCC 7002. Isolated CpcG-PBS from dF338 did not contain FNR and the cell extracts of the mutant had a 35 kDa protein cross-reacting to anti-FNR antibodies. dF338 grows normally under photoautotrophic conditions, but little growth was observed under photoheterotrophic conditions. A *cpcL* (*cpcG2*) mutant grows extremely slowly under photoheterotrophic conditions while a *cpcG* (*cpcG1*) mutant, in which PBS rods could not attach to the cores of the CpcG-PBS, can grow photoheterotrophically, strongly suggesting that the attachment of FNR to CpcL-PBS is critical to photoheterotrophic growth. We show that electron transfer to the plastoquinone pool in dF338 and the *cpcL* mutant was impaired. We also provide evidence that trimeric photosystem I (PSI) and intact CpcL-PBS with a full-length FNR is critical to plastoquinone reduction. The presence of a NADPH-dehydrogenase (NDH)-CpcL-PBS-PSI trimer supercomplex and its roles are discussed.

Keywords: cyanobacteria; FNR; NDH; photoheterotrophic growth; phycobilisome

1. Introduction

Cyanobacteria are a large group of prokaryotes that perform oxygenic photosynthesis, and they play important roles in global carbon and nitrogen cycles. Phycobilisomes (PBS) are supramolecular pigment–protein complexes composed of phycobiliproteins and linker proteins and they are the major antenna complexes for light energy harvesting in cyanobacteria [1–5]. There are two types of PBS in most cyanobacteria: hemidiscoidal PBS and CpcL-PBS. The structures of several hemidiscoidal PBS have been determined recently [5] and they all consist of peripheral rods and a central core. A peripheral rod usually has several hexamers joined together through the linker protein CpcC, which contains a Pfam00427 domain and a Pfam01383 domain [5–7]. The attachment of a peripheral rod to the core is accomplished by the linker protein CpcG1 [8]. The CpcG1-containing PBS is herein called CpcG-PBS. CpcG1 has a Pfam00427 domain at its N-terminus and an extension with α-helices at its C-terminus [7,9]. The Pfam00427 domain of CpcG1 is located in the cavity of the proximal hexamer of the rods and the C-terminal helices interact with the core. At the distal end of the peripheral rod is another linker protein, CpcD, which contains a

Pfam01383 domain and is responsible for distal termination of PBS rod extension [10]. The CpcL-PBS was characterized more recently [11,12] and it consists of only rod that is attached to the thylakoid membrane by the linker protein CpcL. The N-terminal portion of CpcL is homologous to CpcG1 and the genes encoding CpcL were earlier annotated as *cpcG2* (in *Synechococcus* sp. PCC 7002 and *Synechocystis* sp. PCC 6803) or *cpcG3* (in *Anabaena* sp. PCC 7120) [13]. The C-terminal portion of CpcL contains a hydrophobic tail that anchors the CpcL-PBS to the thylakoid membranes [11,12]. The CpcL-PBS is preferentially associated with photosystem I (PSI) and the light energy absorbed by CpcL-PBS is transferred mainly to PSI [11,12,14–16]. Although CpcG-PBS is more closely associated with photosystem II (PSII) [9,17,18], the light energy absorbed by CpcG-PBS can be transferred to either PSII or PSI and distribution of the absorbed light energy between the two photosystems are regulated by a process called state transitions [19–21].

In linear photosynthetic electron transfer, water is oxidized by PSII and the electrons are transferred through electron transfer chain to the final electron acceptor $NADP^+$ and the last reaction is carried out by ferredoxin:$NADP^+$ oxidoreductase (FNR) [22]. FNR in plant chloroplasts has a two-domain structure: a FAD-binding domain and a NAD-binding domain, and these two domains are responsible for electron transfer from ferredoxin to $NADP^+$ [23,24]. The FNR from cyanobacteria also contains the two domains of plant's FNR and carries out the same biochemical reactions in photosynthetic electron transfer. However, it additionally contains a CpcD domain [25], which belongs to Pfam01383 protein family [26]. Due to the presence of the CpcD domain at its N-terminus, some cyanobacterial FNR is attached to rods of CpcL-PBS as well as of CpcG-PBS [27–29]. However, the physiological roles of the attachment of FNR to PBS remain unknown.

Even though there is only one gene, the *petH* gene, encoding FNR in cyanobacteria, two isoforms of FNR are observed: the full-length FNR (FNR_L) that has three domains and a short version of FNR (FNR_S) that lacks the CpcD domain [25]. Analysis of *petH* translation reveals that different translational start points are responsible for the generation of the two isoforms in cyanobacteria [30,31]. Based on enzymatic activity and mutant phenotypes, it was postulated that in cyanobacteria, the FNR_L mainly has a role in reduction of $NADP^+$, whereas the FNR_S is mainly involved in cyclic electron transfer (CET) [30–32]. Although the electron transfer route in CET in both cyanobacteria and plant chloroplasts is not well understood at the moment [33], it has been firmly established that NADPH dehydrogenase complex (NDH) is critical to CET in both cyanobacteria and higher plants [34–37] and the NDH-PSI supercomplex formation is key to its functions [38,39]. The formation of NDH-PSI supercomplex in cyanobacteria has also been demonstrated and it requires participation of the CpcL-PBS [40]. The presence of a cytochrome b_6f complex (cyt b_6f)-bound FNR in chloroplasts suggests a possible route of a direct electron transfer from FNR to cyt b_6f complex [41] although a cyt b_6f complex with a bound FNR has not been observed in cyanobacteria.

Here we report that a truncated FNR without the CpcD domain was unable to attach to PBS and the strain with the mutant gene was unable to grow photoheterotrophically. We provide evidence that the CET in the cyanobacteria is impaired when FNR is not associated with PBS. We provide evidence that the CET in the cyanobacteria is impaired when FNR is not associated with PBS and demonstrate that FNR_L plays an important role in CET in *Synechococcus* 7002.

2. Materials and Methods

2.1. Culture Conditions and Mutant Construction

The wild-type and mutant strains of *Synechococcus* sp. PCC 7002 were grown in A^+ medium [42] at 37 °C with illumination of white fluorescent light at a light intensity of approximately 200 µmol photons m^{-2} s^{-1}. Green light condition was provided with LED light with a wavelength at 520 nm at an intensity of 50 µmol photons m^{-2} s^{-1}. The cultures were bubbled with air plus 1% CO_2. When needed, kanamycin was added to the media at a concentration of 50 µg mL^{-1}, and erythromycin of 25 µg mL^{-1}. Glycerol

(10 mM) and DCMU (10 μM) were used when cells were grown photoheterotrophically. Iron deficiency cultures were grown according to Alcántara-Sánchez et al. [43]. Cell density was determined by measuring optical density at 730 nm.

The deletion mutant of *cpcG2* (ΔcpcG2) was constructed as described [14]. The strain ΔcpcC (deletion of the *cpcC* gene) was constructed according to Zheng et al. [5]. The strain ΔpsaL was constructed as described [44]. The strain ΔcpcG1 (deletion of the *cpcG1* gene) was constructed as described [45]. The strain ΔcpcBA, which lacks the *cpcBA* genes, was constructed according to Huang et al. [46]. The strain dF338, in which the *petH* gene was replaced by *petH2D* encoding FNR_S, was constructed as follows. A 1000 bp fragment from 643 bp upstream to 357 bp downstream of the first ATG codon of the *petH* gene was amplified by PCR with primer pairs of P1/P2. DNA fragments encoding kanamycin resistance cassette (Kan^r), the *cpcBA* promoter and *petH2D* gene (and truncated *petH* gene for SLF8) were amplified by PCR with primer pairs of P3/P4, P5/P6 and P7/P8. With two-way fusion PCR methods [47], the resultant PCR products were transformed to wild-type strain and the transformants were selected on 50 μg mL^{-1} kanamycin. The transformants were screened by PCR with primer pairs of P9/P10 for complete replacement of the *petH* gene by the *petH2D* gene. The same strategy was used for construction of strain SLF8, in which a mutant *petH* encodes a mutant FNR with a deletion from Glu75 to Pro95 in its amino acid sequence. The primers used in the construction of the mutant *petH* for SLF8 were P11/P12 and P8/P13. To generate pku-F2d-oe strain, DNA fragments encoding the *cpcBA* promoter, a mutant *petH* encoding FNRs and kanamycin resistance cassette (Kan^r) were amplified by PCR with primer pairs of P6/P14, P7/P15, and P16/P17. The PCR fragments were fused by two-way fusion PCR followed by insertion into a chromosome docking site between 2588743–2588744. Fragments upstream and downstream of this site were amplified with the primer pairs of P18/P19 and P20/P21 for homologous recombination. The primers used in this study are listed in Supplementary Table S1.

2.2. Protein Extraction and Immunoblotting

Thylakoid membranes were isolated as previously described [48]. After removal of unbroken cells, the supernatant containing thylakoid membranes and soluble proteins were separated after centrifuged at 20,000× *g* for 30 min at 4 °C. Soluble proteins were precipitated for sample preparation. PBS were prepared according to previous work [5,9]. PBS from sucrose density gradient centrifugation were desalted with ultrafiltration using Milipore Amicon Ultra centrifugal filters. PSI was isolated according to previous work [15,49] for EM analysis. Proteins were separated with 15% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Proteins in the gel were transferred to a nitrocellulose membrane for immunoblotting analysis and the antibodies against FNR from cyanobacteria were used as primary antibodies, which were detected by the secondary antibodies from Promega (Beijing, China).

N-terminal sequences of proteins were determined by sequencing with a protein sequencer ABI491.

2.3. Fluorescence Measurements

Fluorescence emission spectra at 77 K and room temperature fluorescence were obtained following the methods in the previous work [14,50]. Post-illumination rise of Chl fluorescence was measured as described previously using Dual-PAM-100 instrument [48].

2.4. P700 Measurement

P700 of PSI was measured using a Dual-PAM-100 instrument (Heinz Walz, Effeltrich, Germany) as described [21]. Cells were harvested and adjusted to OD_{730nm} of 3.0 and were dark-adapted for 5 min before turning on the actinic light. The wavelengths of the actinic lights at 635 nm and 440 nm were used. DCMU at a concentration of 10 μM was added to culture when needed.

3. Results

### 3.1. The CpcD Domain Is Responsible for Association of FNR with PBS

To study roles of FNR attachment to PBS, we constructed a mutant strain of *Synechococcus* 7002 which produces only FNR$_S$ (strain dF338). Because deletion of *petH* is lethal in cyanobacteria, the *petH* gene of *Synechococcus* 7002 was replaced by a modified version of *petH*, named *petH2D*, that had a deletion from residue 1 to residue 95. The correct mutation was confirmed (Figure S1). PBS isolated from the wild-type strain (WT), dF338 and ΔCpcG2, which is a mutant lacking *cpcG2* and its CpcL-PBS is missing while its CpcG-PBS remain unchanged [14], were analyzed with SDS-PAGE. Protein bands with a molecular mass of 45 kDa could clearly be visualized by Commassie brilliant blue staining in the lanes with the isolated CpcG-PBS from WT and ΔCpcG2 while it was not present in CpcG-PBS from dF338 strain. The bands from WT and ΔCpcG2 were cross-reacted with the anti-FNR antibodies in immunoblotting (Figure 1A). The 45 kDa bands were further analyzed by N-terminal sequencing and they had a sequence of MYGITSTANSTGNQSYAN, confirming that they were FNR. In total cellular extracts, FNR$_L$, the 45 kDa bands were found to cross-react with the anti-FNR antibodies in WT and ΔCpcG2. In dF338, only a 35 kDa protein was detected by the immunoblotting (Figure 1B), confirming that the CpcD domain of FNR was deleted in the mutant. The proteins from the cell extracts were separated into membrane and soluble fractions and they were again probed with immunoblotting after SDS-PAGE. As shown in Figure 1C,D, FNR$_L$ is present in the both membrane and soluble fractions from the WT and ΔCpcG2 but not in the dF338's fractions. It also shows that FNR$_S$ is present in WT, dF338 and ΔCpcG2. In the soluble fractions, there were more FNR$_S$ than the membrane fractions in WT and ΔCpcG2 and there were two bands detected by immunoblotting, indicating that two isoforms of FNR$_S$ could be present in soluble fractions in WT and ΔCpcG2. It should be noted that we used a *cpcBA* promoter for the expression of the *petH2D* gene in dF338 and the amount of FNR in dF338 is higher than that in the WT. We have not been able to generate a mutant strain that contains a truncated form of *petH* with its own promoter.

3.2. The Strain dF338 Is Able to Perform State Transitions

Since CpcG-PBS are mostly associated with PSII [17] while FNR plays a major role in electron transfer of PSI, it was first predicted that the attachment of FNR to PBS may have a role in state transitions. We first measured 77 K fluorescence emission spectra of WT and dF338 with a chlorophyll-absorbing (440 nm) excitation light (Figure 2A) and found that they were nearly identical, indicating the that PSII to PSI stoichiometry of dF338 was the same as that of WT. The 77 K fluorescence emission spectra were then measured with PBS-absorbing light under state I or state II conditions and the results suggested that the state transitions in dF338 strain were not impaired (Figure 2B,C). Room temperature fluorescence inductions in the presence of DCMU also showed that the state transitions of dF338 were not impaired (Figure S2). It has been shown that impairing state transitions leads to slower growth under green light that is preferentially absorbed by PBS [21]. We therefore measured the growth rates of dF338 under a green light and found that they were almost identical to that of the WT (Figure S3), suggesting that the major role of the attachment of FNR to PBS is unlikely related to state transitions. Room temperature fluorescence inductions and growth curve under green light conditions were also measured for ΔcpcG2 (Figures S2C and S3). The results for ΔcpcG2 were almost identical as that of wild-type, demonstrating that CpcL-PBS are not involved in state transition.
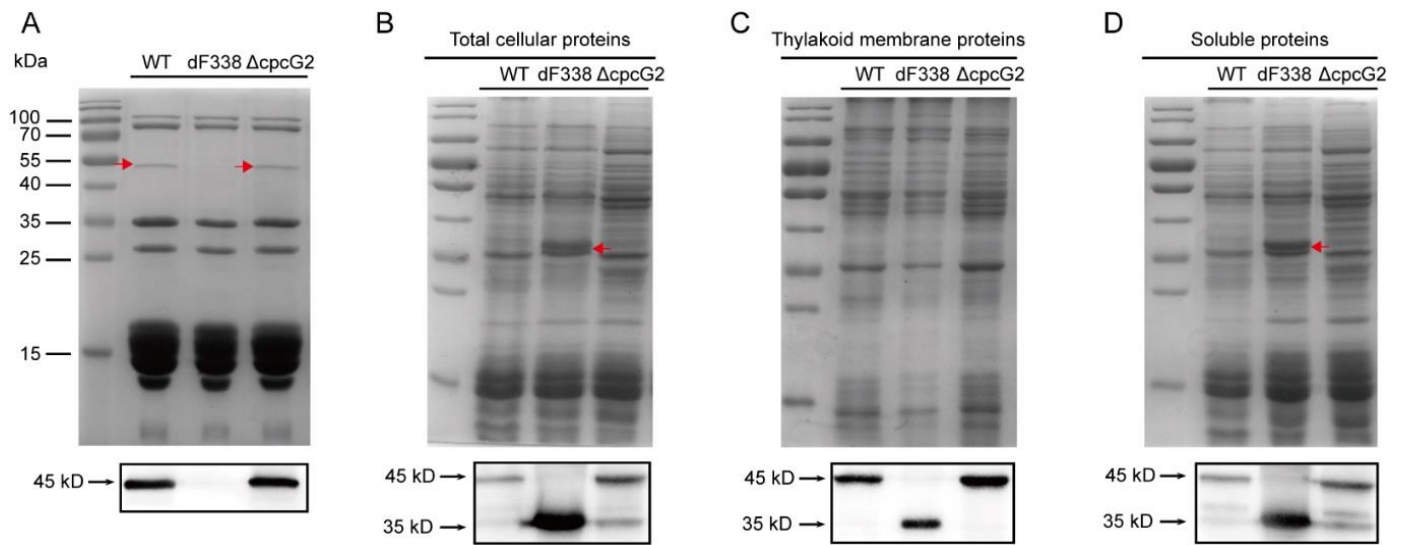
Figure 1. SDS-PAGE and immunoblotting analysis of ferredoxin:NADP⁺ oxidoreductase (FNR) from different strains. (**A**) SDS-PAGE separation of phycobilisome (PBS) proteins from the wild-type (WT), dF338 and ΔCpcG2. Red arrows indicate FNR protein, which is missing in the dF338 mutant. The molecular mass markers are shown as bars on the left side. Proteins separated by SDS-PAGE were transferred from gel to nitrocellulose membrane for immunoblotting and a portion of the blot is shown in the box below the gel image. (**B**–**D**), total cellular proteins, thylakoid membrane proteins and soluble proteins were extracted and separated by SDS-PAGE respectively, from the strains WT, dF338 and ΔCpcG2. The red arrows in panel (**B**,**D**) indicate an increased intensity of the protein bands at the position of approximately 35 kDa. Immunoblotting analysis of FNR proteins are shown in the boxes below each gel. The molecular masses of the immunoblotting bands are shown on the left of the boxes as indicated by black arrows.
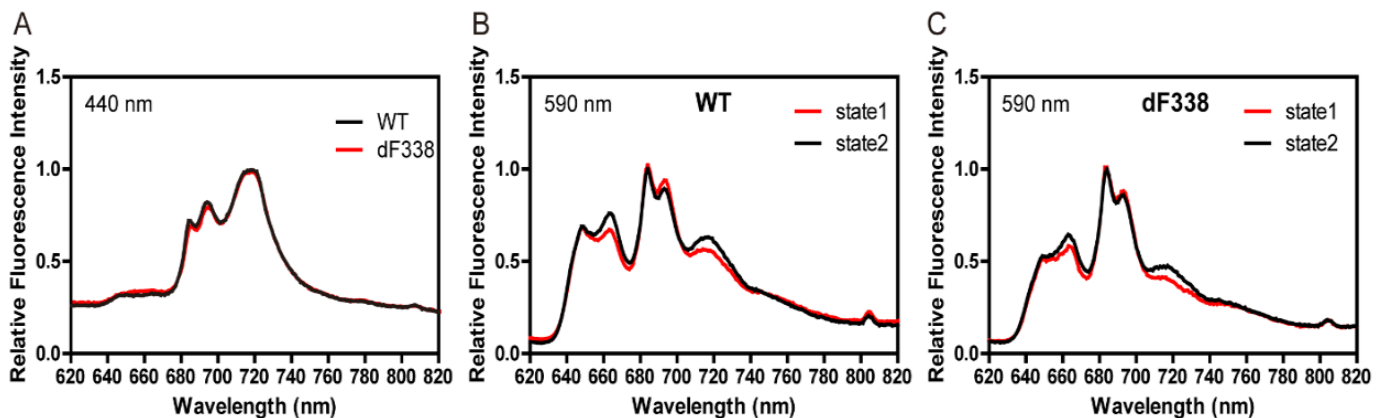

Figure 2. 77 K fluorescence spectra of the wild-type (WT) and dF338. (**A**) Emission spectra of WT (black line) and dF338 (red line) in state 2 conditions were collected with an excitation light at 440 nm. (**B**,**C**) 77 K fluorescence emission spectra of WT and dF338, respectively, were obtained with an excitation light at 590 nm. Cells were brought to state 1 by illuminating with blue light with DCMU (red line) and to state 2 with incubation in the dark (black line) before frozen in liquid nitrogen.

3.3. The Strain dF338 Is Impaired of Photoheterotrophic Growth

We measured growth of dF338 and several other strains under photoautotrophic and photoheterotrophic conditions. While the growth of the strains dF338 and ΔcpcG2 under photoautotrophic conditions with white light was very similar to that of WT (Figure 3A), their growth under photoheterotrophic conditions (10 mM glycerol and 10 μM DCMU) was much slower than that of WT (Figure 3B). The two other strains, ΔcpcBA and ΔcpcG1, were

also measured for their growth. ΔcpcBA lacks the genes encoding the α- and β-subunits of phycocyanin and no PBS rods can be synthesized. ΔcpcG1 lacks CpcG1 and the peripheral rods, which can be synthesized, could but cannot be attached to the cores (Figure S4A). Therefore, the mutated CpcG-PBS in ΔcpcBA and ΔcpcG1 were the similar in that they have no rods attached to the cores and therefore there would be no FNR attached to their CpcG-PBS. Both ΔcpcBA and ΔcpcG1 grew much slower than WT under photoautotrophic conditions (Figure 3A) because of insufficient light absorption due to a reduced light-absorbing cross-section sizes. However, their growth under photoheterotrophic conditions was different: the growth rates of ΔcpcG1 was slightly faster than that of WT while ΔcpcBA could hardly grow (Figure 3B). Since the CpcL-PBS in ΔcpcBA could not be synthesized and it is not affected by the deletion of *cpcG1* in ΔcpcG1, these results were suggestive that CpcL-PBS is more critical to the normal growth under photoheterotrophic conditions. We used another mutant ΔcpcC to further test this possibility. The strain ΔcpcC lacks the gene *cpcC* encoding the rod linker and the rods of both CpcG-PBS and CpcL-PBS contain only one hexamer [5] (Figure S4A). As shown in Figure 3A, photoautotrophic growth of ΔcpcC was slower than that of WT but significantly faster than that ΔcpcG1 and ΔcpcBA. Under photoheterotrophic conditions, little growth of ΔcpcC was observed (Figure 3B). Importantly, the attachment of FNR to the CpcC-less CpcG-PBS was confirmed by SDS-PAGE analysis (Figure S4B).
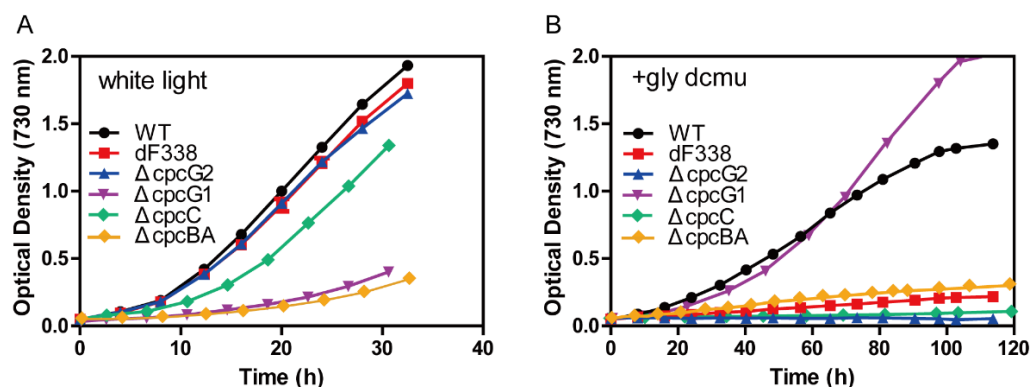
Figure 3. Growth curves of the wild-type (WT) and the mutant strains. (**A**) Photoautotrophic growth curves of WT and the mutant strains were obtained under white light condition provided with cool white fluorescent light at 200 μmol photons $m^{-2} s^{-2}$. (**B**) Photoheterotrophic growth curves of WT and the mutant strains in the presence of glycerol (10 mM) and DCMU (10 μM) under a light intensity of 50 μmol photons $m^{-2} s^{-2}$. Colors of the curves are as follows: black for WT; red for dF338; blue for ΔcpcG2; purple for ΔcpcG1; green for ΔcpcC; and yellow for ΔcpcBA.

3.4. Photosynthetic Electron Transfers

To understand why attachment of FNR_L to CpcL-PBS is required for growth under photoheterotrophic conditions, post-illumination rise of Chl fluorescence (PIRF), which is attributable to the reduction of plastoquinone (PQ) [38,40,41,51], was measured. WT exhibited significant increases of PIRF when the actinic light was turned off (Figure 4A). However, PIRF was not observed in dF338, ΔcpcG2 and ΔcpcC (Figure 4B–D), suggesting that these mutants were impaired of electron donation to PQ pool. When WT and dF338 cultures were exposed to a strong light (1300 μmol $m^{-2} s^{-1}$), dF338 showed a faster decrease of oxygen evolution rate than WT while its recovery from photoinhibition was slower than that of WT (Figure S5).
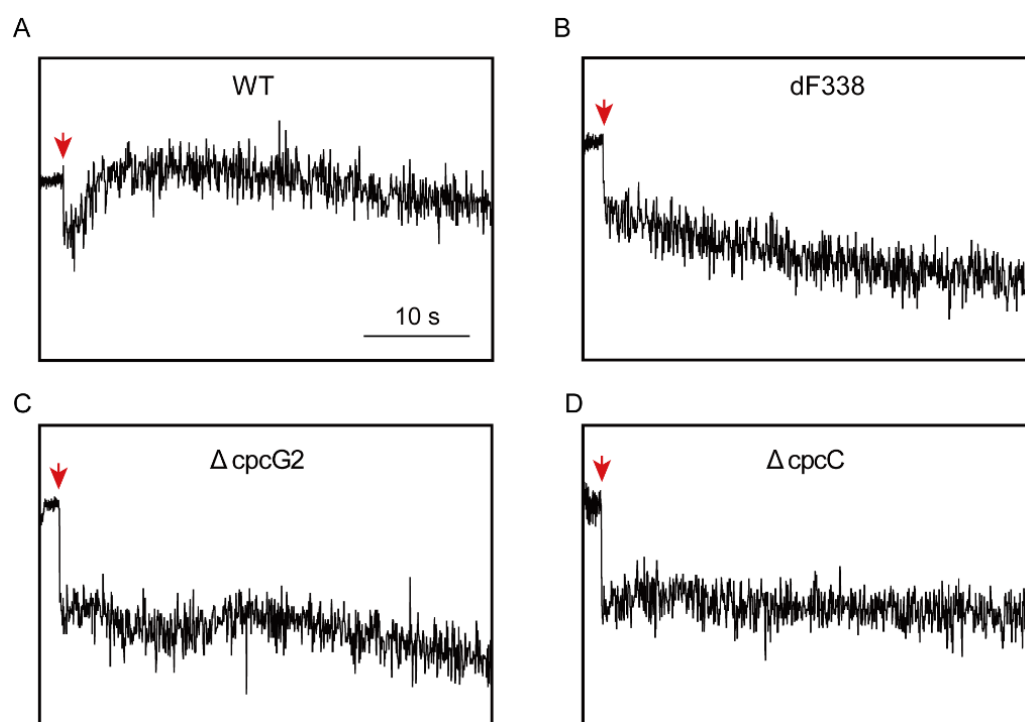
Figure 4. Post-illumination rise of Chl fluorescence (PIRF) analysis. PIRF analysis of the wild-type (WT) (**A**), dF338 (**B**), ΔcpcG2 (**C**) and ΔcpcC (**D**). Cells were illuminated with an actinic light for 30 s before the actinic light was turned off (downward arrows) and the transient increase of Chl fluorescence at the wavelength of 685 nm was monitored. Bar indicates 10 s.

The P700 redox kinetics of WT and dF338 were measured and it was found that the rates of P700$^+$ reduction in dF338 were comparable to the rates of P700$^+$ reduction in WT under the conditions tested (Figure 5). The P700 redox change induced with a PBS-absorbing (635 nm) light in WT cells are shown in Figure 5A. When the light was turned on, a spike of P700 signal was immediately observed, indicating a fast oxidation and a reduction, followed by a slower oxidation to reach an oxidized steady state. When the light was turned off, the P700$^+$ was rapidly reduced with a half time of 10–20 ms. In the presence of DCMU, which blocks electron transfer from PSII to PQ pool, the initial spike of P700 signal disappeared and the reduction of P700$^+$ was much slower with a half time of approximately 200 ms when the light was turned off (Figure 5B). The initial spike of the P700 signal observed in Figure 5A was also absent if the actinic light was a Chl-absorbing light at 440 nm (Figure 5C), which is preferentially used by PSI. For the strain dF338, while the initial spike of P700 signal is observed, the following oxidation of P700 was much slower (Figure 5E). The slower oxidation of P700 is abolished when DCMU was added (Figure 5F) or the actinic light was Chl-absorbing light (440 nm) (Figure 5G). The very slow oxidation of P700 observed in Figure 5E was likely a result from overexpression of the *petH2D* gene because overexpression of the *petH2D* gene in a WT background (strain pku-F2d-oe) also led to a similarly slow oxidation of P700 (Figure 5I). It is also possible that the strain with overproduced FNRs could have a faster electron donation to the PQ pool as observed in *Synechocystis* [52,53], leading to a slower oxidation of P700 (Figure 5E).
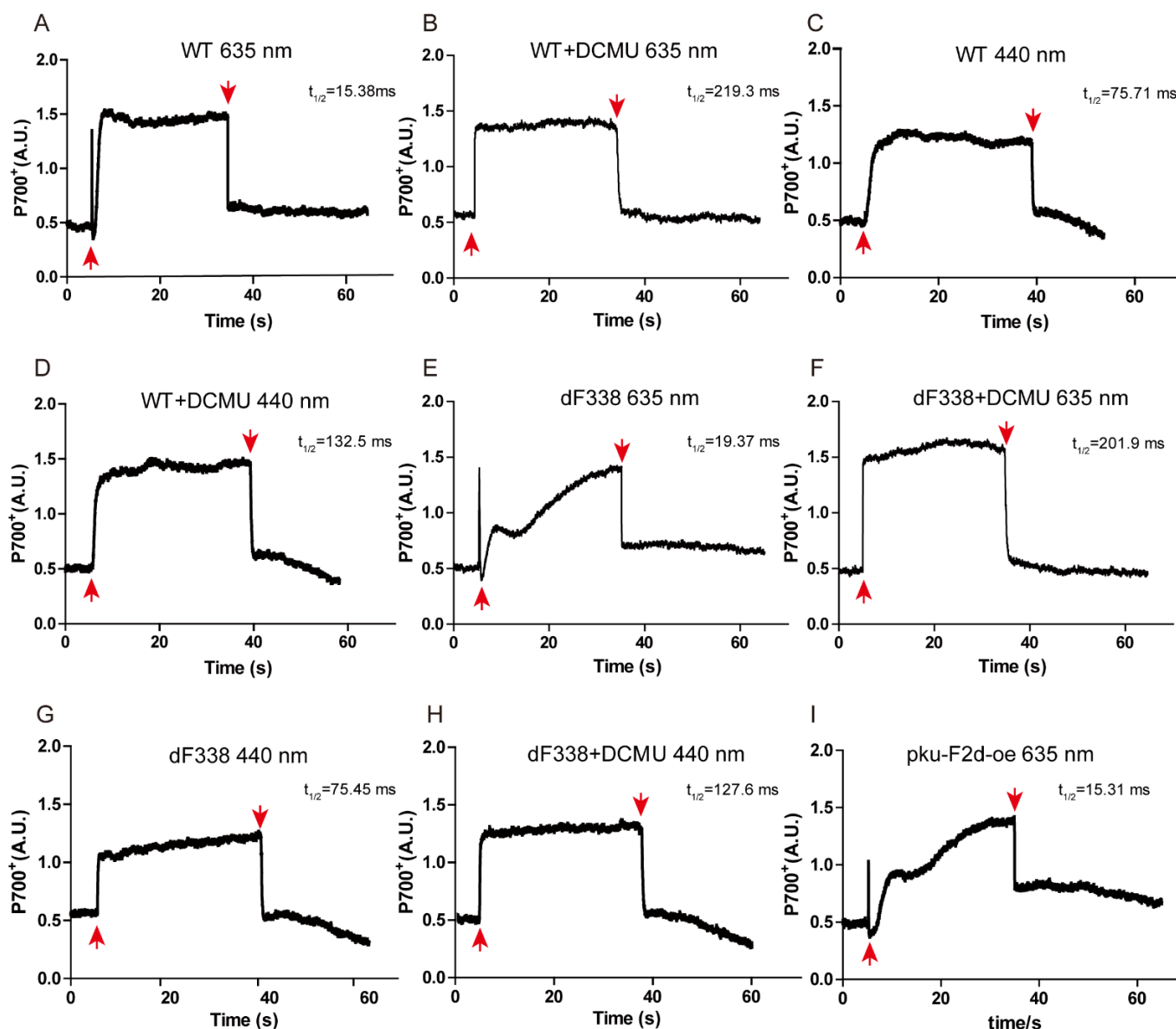
Figure 5. Kinetic analyses of P700 photooxidation and reduction of P700$^+$. Strains of WT (panels **A**–**D**), dF338 (panels **E**–**H**) and pku-F2d-oe (the strain that overexpresses the *petH2D* gene, panel **I**) were incubated in the dark for 5 min before the actinic lights at 635 nm (panels **A**,**B**,**E**,**F**,**I**) or at 440 nm (panels **C**,**D**,**G**,**H**) were switched on. When needed (panels **B**,**D**,**F**,**H**), DCMU was added at a concentration 10 μM before the dark incubation. The upward and downward red arrows in the panels indicate turning on or turning off of the actinic lights, respectively.

3.5. Association of CpcL-PBS with PSI

Based on an earlier study that CpcL-PBS was associated with PSI in the filamentous cyanobacterium *Anabaena* 7120 [15], we searched conditions for isolating CpcL-PBS-PSI complexes in *Synechococcus* 7002. We were unable to obtain CpcL-PBS-PSI complexes from WT and other strains under normal growth conditions. However, under the conditions that increase CET, such as low iron in growth media [43], we were able to isolate CpcL-PBS-PSI complexes with ultracentrifugation in high phosphate-sucrose buffer. As shown in Figure 6A(i), the isolated particles of PSI trimer and CpcL-PBS-PSI trimer complexes can be observed by EM with negative staining. Enlarged images of some particles are shown in Figure 6A(ii–v). The interactive site for CpcL-PBS association with PSI trimer

seemed to be the interface between PSI monomers (Figure 6A(ii,iii)). From the side view, the CpcL-PBS rods stood upwards, but the angles between PSI timer and the rods were variable (Figure 6A(iv,v)). We did not observe CpcL-PBS-PSI monomer complexes with negatively stained electron microscopy. Measurement of PIRF revealed that ΔpsaL, a *psaL* mutant that cannot form PSI trimer [44], had a much lower signal in PIRF (Figure 6B), suggestive of an impairment of electron transfer to PQ pool. Under the photoheterotrophic growth conditions, the *psaL* mutant grew more slowly than WT (Figure 6C) but faster than dF338. Under photoautotrophic conditions, the growth rate of the *psaL* mutant was nearly identical to that of WT (Figure S6).
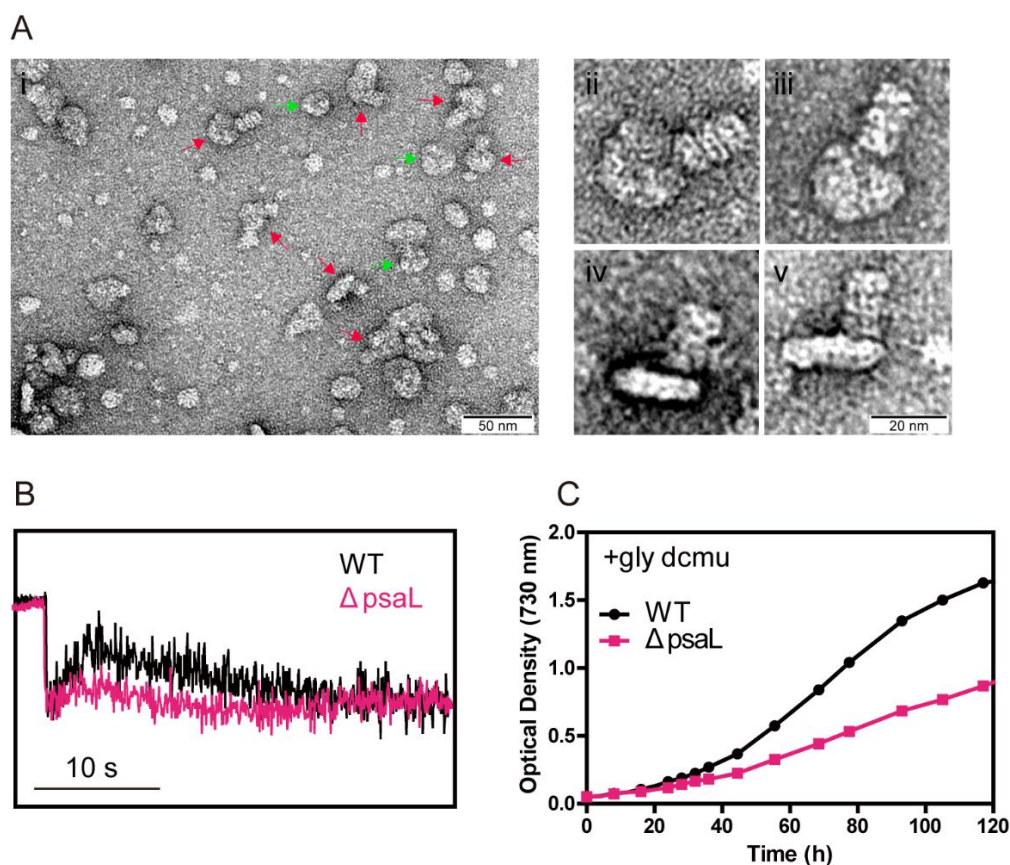
Figure 6. EM analysis of CpcL-PBS-photosystem I (PSI) complexes from *Synechococcus* 7002 and phenotypic characterization of ΔpsaL mutant. (**Ai**) Electron microscopic images of the negatively stained particles of the trimeric PSI (green arrows) and CpcL-PBS-PSI trimer complexes (red arrows). Bar indicates 50 nm. Panels (**Aii**) through (**Av**) show enlarged images of top view (**Aii,Aiii**) and side view (**Aiv,Av**) of individual complexes of CpcL-PBS-PSI trimer complexes. Bar indicates 20 nm. (**B**) PIRF analysis of the wild-type (WT) (black trace) and ΔpsaL (pink trace). (**C**) Photoheterotrophic growth of WT (black) and ΔpsaL (pink) in the presence of 10 mM glycerol and 10 μM DCMU under light intensity of 50 μmol photons $m^{-2} s^{-2}$.

As mentioned earlier, the CpcD domain of FNR_L is linked to the other two domains through a flexible linker region. To gain more insight into the mechanism of FNR_L in photoheterotrophic growth, we constructed a strain (SLF8) which has a mutant FNR with a shortened linker region. The region from 75 to 95 residues within the flexible region of FNR_L was deleted (Figure S7) and this mutant FNR (FNR_{Ls}) has a molecular mass of 43 kDa. It should be noted that the flexible region of FNR_L is not conserved among the cyanobacterial FNR_L [30] and a previous work did not find post-translational modification in residues 75–95 in this region of FNR_L in *Synechococcus* 7002 [54]. Analysis of isolated PBS from the strain SLF8 with SDS-PAGE showed that the band at 45 kDa position in WT was missing in the lane of SLF8 and a new band was present at the position of 43 kDa, which

cross-reacted with the anti-FNR antibodies (Figure 7A), demonstrating FNR_{Ls} was able to attach to PBS. Immunoblotting of the total cellular extracts (Figure 7B) showed there were two forms of FNR in SLF8, a 43 kDa protein and a 35 kDa protein, suggesting the mutant SLF8 is able to synthesize FNR_S as WT does. SLF8 was able to grow normally under photoautotrophic conditions (Figure S6) but could not grow under photoheterotrophic conditions (Figure 7C), and little PIRF was observed (Figure 7D). It is worthwhile to note that even though we used the *cpcBA* promoter for the expression of shortened *petH* gene in SLF8, the mutant FNR_{Ls} was less abundant than in dF338.
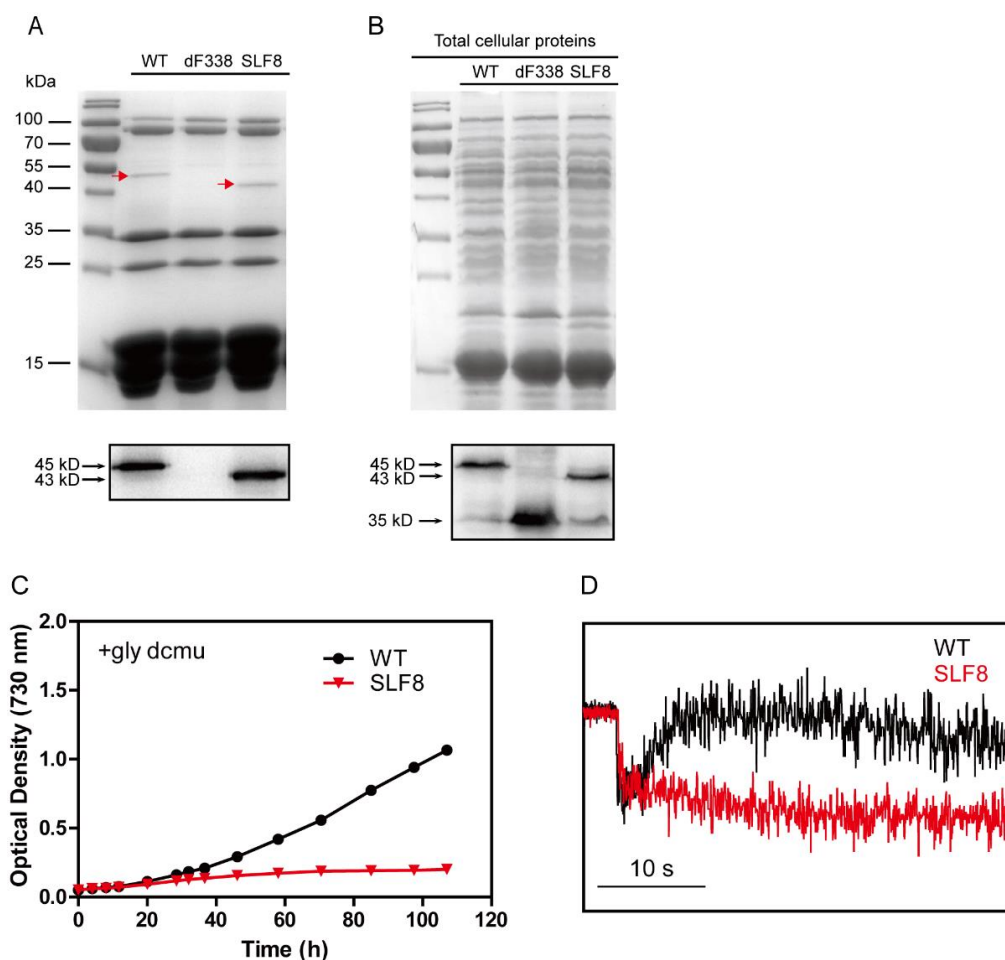
Figure 7. Characterization of the mutant strain SLF8. (**A**) PBS of the wild-type (WT), dF338 and SLF8 strains were isolated and the components were separated with SDS-PAGE. Red arrows indicate FNR protein bands, which was smaller in the lane from SLF8. Immunoblotting with FNR antibodies was performed and result is shown in the lower box with molecular masses on the left side. (**B**) Total cellular proteins from the same strains in panel A were separated by SDS-PAGE and stained with Coomassie blue. Immunoblotting with FNR antibodies was performed and result is shown in the lower box with molecular masses on the left side. In both gels, the left lanes contain molecular mass markers. (**C**) Photoheterotrophic growth of WT (black) and SLF8 (red) in the presence of 10 mM glycerol and 10 μM DCMU. (**D**) Analysis of electron donation to the PQ pool in WT (black) and SLF8 (red) by PIRF. Bar indicates 10 s.

4. Discussion

The cyanobacterial FNR contains an additional CpcD domain compared to the plant FNR and association of the cyanobacterial FNR$_L$ with PBS was discovered three decades ago [25]. The cyanobacterial *petH* gene encoding FNR$_L$ is also responsible for the formation of the plant-type FNR, the two domain FNR or FNR$_S$, which is generated by translation from the second translation start position [30,31] and it is located mainly in cytoplasm [43]. There are two types of PBS in most of the cyanobacteria, CpcG-PBS and CpcL-PBS, and they both harbor FNR$_L$ [29]. Even though the association of FNR$_L$ with PBS is expected to play important physiological roles as the FNR$_L$-PBS association is conserved in nearly the entire phylum of cyanobacteria, the functions of the attachment of FNR$_L$ to PBS remained mysterious. In this study, we report the construction of the mutant strain dF338, in which the gene *petH* encoding FNR$_L$ was replaced by a mutant *petH* gene, the *petH2D*, which encodes FNR$_S$. The inability for FNR$_S$ to attach to PBS (Figure 1) does not affect dF338's growth rate under normal growth conditions (Figure 3) and dF338 was able to perform state transitions as compared with WT (Figures 2 and S2). However, it did affect dF338's photoheterotrophic growth (Figure 3) when PSII is inhibited by the PSII inhibitor DCMU and the electron source is NADPH generated by oxidation of externally added glycerol. Therefore, the growth experiments strongly suggested that electron donation from NADPH to PQ pool is impaired in dF338.

A set of mutant strains that change PBS structures besides dF338 (Figure S4A) provided clues for understanding why attachment of FNR$_L$ to PBS is needed for photoheterotrophic growth. Among these mutants, dF338 is the only mutant that has no FNR$_L$ and the other mutants affect the attachment of FNR$_L$ to either CpcG-PBS or CpcL-PBS in different ways. In ΔcpcG1, no FNR$_L$ could be attached to the rod-less CpcG-PBS while its CpcL-PBS is unchanged. Its normal photoheterotrophic growth (Figure 3B) suggests that the attachment of FNR$_L$ to CpcG-PBS is not required for PQ pool reduction by NADPH. The evidence provided by the photoheterotrophic growth (Figure 3B) and PIRF (Figure 4), which is widely used in studies of the reduction of PQ pool by NDH [40,55], from the strains ΔcpcG2 (no CpcL-PBS), ΔcpcBA (no rods of any PBS) and dF338 collectively shows that the attachment of FNR$_L$ to CpcL-PBS is critical to the electron donation to PQ pool from NADPH and its function could not be replaced by CpcG-PBS.

It has been shown previously that CpcL-PBS is important for energy transfer from PBS to PSI [11,12,14,16] and it forms the CpcL-PBS-PSI complexes in *Anabaena* 7120 [15]. Here, we provide evidence that CpcL-PBS could interact with PSI trimer and forms a CpcL-PBS-PSI complexes in *Synechococcus* 7002 (Figure 6A). In plant chloroplasts, PSI could interact with NDH to form an NDH-PSI supercomplex and a recently determined structure of NDH-PSI supercomplex shows that LCH of PSI is key to the formation of supercomplex [38,39]. Although the NDH-PSI supercomplex has not been isolated from the cyanobacteria, its existence has been demonstrated in *Synechocystis* 6803 and its CpcL-PBS is required for the NDH-CpcL-PBS-PSI supercomplex formation [40]. The results in Figure 6 strongly suggest that PSI oligomer formation could be important to the supercomplex formation in *Synechococcus* 7002 because the photoheterotrophic growth of a PSI monomer mutant is significantly slower than that of WT and its PIRF is much smaller. Similar phenotypes have been observed in the *psaL* mutant of *Anabaena* 7120, which could not form PSI tetramer [48]. The roles of CpcL-PBS in photoheterotrophic growth are further studied with the strain ΔcpcC, in which both CpcG-PBS and CpcL-PBS have only one hexamer layer on their rods. Although no CpcL-PBS has been isolated from this strain, the CpcG-PBS was isolated from ΔcpcC and it has FNR$_L$ attached (Figure S4B). It is expected that the only hexamer layer CpcL-PBS from ΔcpcC should also have FNR$_L$ attached. Yet, ΔcpcC grows very poorly under photoheterotrophic conditions and has little PIRF (Figures 3B and 4D), strongly suggesting that the full-length CpcL-PBS is needed for proper electron donation to PQ pool from NADPH.

Based on the previous evidence and our current study, we propose a model for the cyanobacterial supercomplex of NDH-CpcL-PBS-PSI trimer (Figure 8). In this model, CpcL-

PBS is important not only in connecting PSI trimer with NDH, but also in delivering the attached FNR_L to the proper position of NDH so that more efficient electron transfer could be performed from NADPH to PQ pool. The position of FNR_L is critical to the supercomplex's function because a shortened flexible region between the CpcD domain and the other two domains in FNR_{Ls}, which could be attached to PBS, led to poor photoheterotrophic growth and low PIRF (Figure 7).
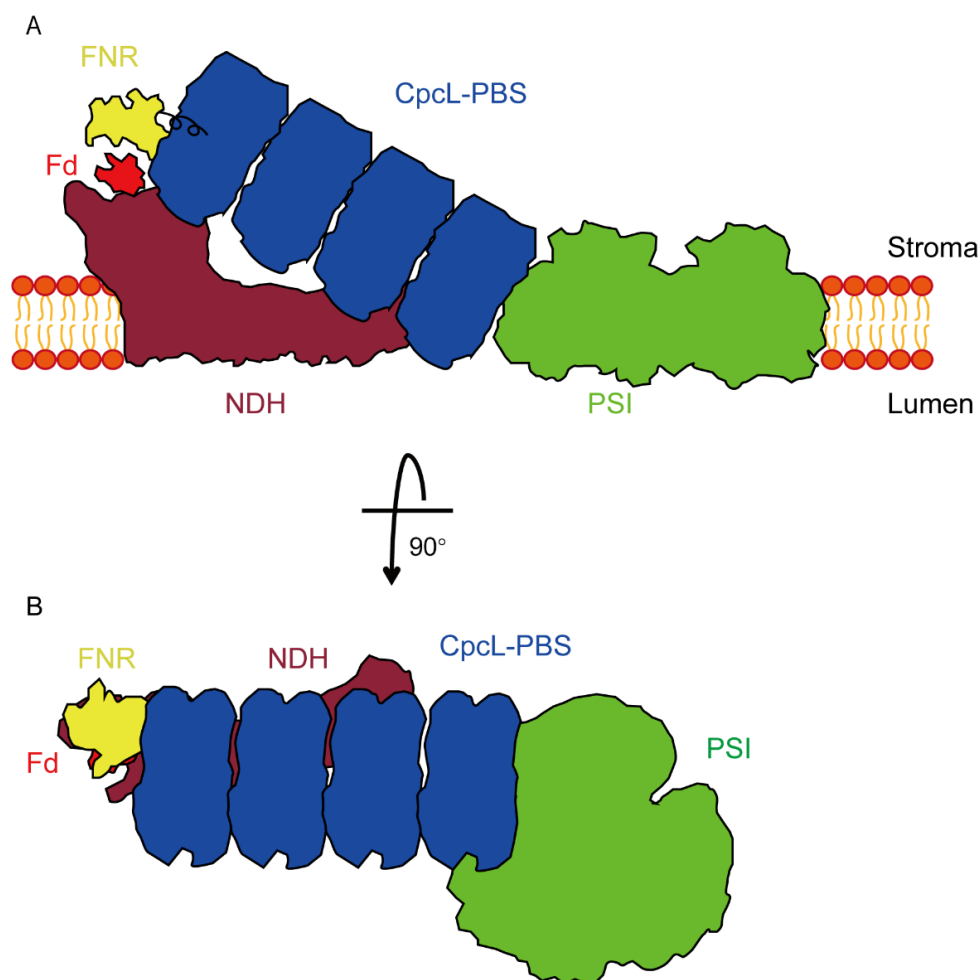
Figure 8. Schematic model of NADPH-dehydrogenase complex (NDH)-CpcL-PBS-PSI trimer supercomplex from sideview (**A**) and top-view (**B**). CpcL-PBS-PSI trimer complex interacts with NDH to form supercomplex. In the supercomplex, the FNR attached to CpcL-PBS could efficiently transfer electrons from NADPH to Fd that is in turn oxidized by NDH located on thylakoid membranes under photoheterotrophic growth conditions. The model is drawn in proportion according to the actual sizes of the supercomplex components.

Several types of NDH exist in the cyanobacteria and they play different roles in various cellular processes [34,35,56]. Although the mechanisms of CET in the cyanobacteria and chloroplasts are not completely understood at the moment [33], accumulated evidence has clearly demonstrated that NDH plays a major role in cyanobacterial CET [35,57–59]. Besides NDH, FNR_S could also be important to CET around PSI and it has been suggested that interaction of FNR_S with cyt b_6f complex [41] could play a critical role in this process. Under photoheterotrophic growth conditions, no electron from PSII could be transferred to the PQ pool and the ultimate reductant source for growth is sugar that generate NADPH for ATP generation through electron transfer chain shared by respiration and photosynthesis. The roles of NDH under these conditions are therefore two folds: (i) it could directly participate CET by receiving electrons from reduced ferredoxins from PSI directly [36,37,60,61] and

(ii) it uses reduced ferredoxin generated from NADPH through FNR to reduce PQ pool. It is evident that CET would eventually lose all its electrons without NADPH-generated ferredoxin as electron source under heterotrophic conditions because the electron transfer in CET is not 100% efficient and loss of electron is unavoidable. Our experimental results show that the attachment of FNR_L to CpcL-PBS is needed for optimal electron transfer activity from NADPH to PQ pool and this process is likely performed by the NDH-CpcL-PBS-PSI supercomplex. It should be noted that all mutants reported in our study remained alive for more than a week under photoheterotrophic conditions and their rates of $P700^+$ reduction were not significantly changed (Figure 5). Therefore, it is possible that FNR-dependent reduction of ferredoxin from NADPH could occur in cyanobacterial cytosol, which in turn reduce PQ pool through NDH and its capacity and efficiency are the key factors for photoheterotrophic growth. For *Synechococcus* 7002, they are not high enough to sustain a normal photoheterotrophic growth. Similarly, a *psaL* mutant of *Anabaena* 7120 showed little PIRF and its N_2-fixing ability was largely diminished [48]. On the other hand, the inability of FNR to attach to PBS in *Synechocystis* 6803 did not affect its reduction of PQ pool from NADPH significantly [30]. It was also reported that a *cpcG2* mutant lacking CpcL-PBS in *Synechocystis* 6803 could grow photoheterotrophically, but its PIRF was significantly reduced [40].

The roles of the attachment of FNR_L to CpcG-PBS remain unclear and more work is needed in the future. It is our speculation that attachment of FNR_L to CpcG-PBS might play a role in organization of PSII and PSI on thylakoid membranes. More work is needed to understand how each component of the supercomplex interacts with the others and how the electron transfers in the supercomplex are regulated.

5. Conclusions

Besides energy transfer from phycobiliproteins to PSI, CpcL-PBS are important to NDH-dependent electron transfer to the PQ pool in *Synechococcus* 7002 and this function could be carried out by a NDH-CpcL-PBS-PSI trimer supercomplex. The attachment of FNR_L with a functional flexible region to an intact CpcL-PBS is required for its optimal activity.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/microorganisms10071313/s1, Figure S1: PCR confirmation of the construction of the dF338 mutant. Figure S2: Room temperature fluorescence inductions of the wild-type (WT), dF338 and ΔCpcG2. Figure S3: Growth curves of the wild-type (WT) and mutant strains under green light condition. Figure S4: Schematic representations of CpcG-PBS and CpcL-PBS in different strains. Figure S5: Photoinhibition and recovery of wild-type (WT) and dF338. Figure S6: Growth curves of strains under photoautotrophic condition. Figure S7: Molecular characterization of wild-type (WT) and SLF8. Table S1: Sequences of the oligonucleotides primers.

References

1. Glazer, A.N. Light guides. Directional energy transfer in a photosynthetic antenna. *J. Biol. Chem.* **1989**, *264*, 1–4. [CrossRef]
2. MacColl, R. Cyanobacterial phycobilisomes. *J. Struct. Biol.* **1998**, *124*, 311–334. [CrossRef] [PubMed]
3. Adir, N.; Bar-Zvi, S.; Harris, D. The amazing phycobilisome. *Biochim. Biophys. Acta Bioenerg.* **2020**, *1861*, 148047. [CrossRef]
4. Sui, S.F. Structure of phycobilisomes. *Annu. Rev. Biophys.* **2021**, *50*, 53–72. [CrossRef]
5. Zheng, L.; Zheng, Z.; Li, X.; Wang, G.; Zhang, K.; Wei, P.; Zhao, J.; Gao, N. Structural insight into the mechanism of energy transfer in cyanobacterial phycobilisomes. *Nat. Commun.* **2021**, *12*, 5497. [CrossRef] [PubMed]
6. De Lorimier, R.; Guglielmi, G.; Bryant, D.A.; Stevens, S.E., Jr. Structure and mutation of a gene encoding a Mr 33.000 phycocyanin-associated linker polypeptide. *Arch. Microbiol.* **1990**, *153*, 541–549. [CrossRef]
7. Zhang, J.; Ma, J.; Liu, D.; Qin, S.; Sun, S.; Zhao, J.; Sui, S.F. Structure of phycobilisome from the red alga *Griffithsia pacifica*. *Nature* **2017**, *551*, 57–63. [CrossRef] [PubMed]
8. Yu, M.H.; Glazer, A.N. Cyanobacterial phycobilisomes role of the linker polypeptides in the assembly of phycocyanin. *J. Biol. Chem.* **1982**, *257*, 3429–3433. [CrossRef]
9. Chang, L.; Liu, X.; Li, Y.; Liu, C.C.; Yang, F.; Zhao, J.; Sui, S.F. Structural organization of an intact phycobilisome and its association with photosystem II. *Cell Res.* **2015**, *25*, 726–737. [CrossRef]
10. De Lorimier, R.; Bryant, D.A.; Stevens, S.E., Jr. Genetic analysis of a 9 kDa phycocyanin-associated linker polypeptide. *Biochim. Biophys. Acta* **1990**, *1019*, 29–41. [CrossRef]
11. Kondo, K.; Geng, X.X.; Katayama, M.; Ikeuchi, M. Distinct roles of CpcG1 and CpcG2 in phycobilisome assembly in the cyanobacterium *Synechocystis* sp. PCC 6803. *Photosynth. Res.* **2005**, *84*, 269–273. [CrossRef] [PubMed]
12. Kondo, K.; Ochiai, Y.; Katayama, M.; Ikeuchi, M. The membrane-associated CpcG2-phycobilisome in *Synechocystis*: A new photosystem I antenna. *Plant Physiol.* **2007**, *144*, 1200–1210. [CrossRef] [PubMed]
13. Bryant, D.A.; Stirewalt, V.L.; Glauser, M.; Frank, G.; Sidler, W.; Zuber, H. A small multigene family encodes the rod-core linker polypeptides of *Anabaena* sp. PCC 7120 phycobilisomes. *Gene* **1991**, *107*, 91–99. [CrossRef]
14. Deng, G.; Liu, F.; Liu, X.; Zhao, J. Significant energy transfer from CpcG2-phycobilisomes to photosystem I in the cyanobacterium *Synechococcus* sp. PCC 7002 in the absence of ApcD-dependent state transitions. *FEBS Lett.* **2012**, *586*, 2342–2345. [CrossRef] [PubMed]
15. Watanabe, M.; Semchonok, D.A.; Webber-Birungi, M.T.; Ehira, S.; Kondo, K.; Narikawa, R.; Ohmori, M.; Boekema, E.J.; Ikeuchi, M. Attachment of phycobilisomes in an antenna-photosystem I supercomplex of cyanobacteria. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 2512–2517. [CrossRef] [PubMed]
16. Niedzwiedzki, D.M.; Liu, H.; Blankenship, R.E. Excitation energy transfer in intact CpcL-phycobilisomes from *Synechocystis* sp. PCC 6803. *J. Phys. Chem. B* **2019**, *123*, 4695–4704. [CrossRef]
17. Chereskin, B.M.; Clement-Metral, J.D.; Gantt, E. Characterization of a purified photosystem II-phycobilisome particle preparation from *Porphyridium cruentum*. *Plant Physiol.* **1985**, *77*, 626–629. [CrossRef]
18. Wang, R.T.; Stevens, C.L.R.; Myers, J. Action spectra for photoreactions I and II of photosynthesis in the blue-green alga *Anacystis nidulans*. *Photochem. Photobiol.* **1977**, *25*, 103–108. [CrossRef]
19. Bonaventura, C.; Myers, J. Fluorescence and oxygen evolution from *Chlorella pyrenoidosa*. *Biochim. Biophys. Acta* **1969**, *189*, 366–383. [CrossRef]
20. Mullineaux, C.W. Excitation-energy transfer from phycobilisomes to photosystem-I in a cyanobacterium. *Biochim. Biophys. Acta* **1992**, *1100*, 285–292. [CrossRef]
21. Dong, C.; Tang, A.; Zhao, J.; Mullineaux, C.W.; Shen, G.; Bryant, D.A. ApcD is necessary for efficient energy transfer from phycobilisomes to photosystem I and helps to prevent photoinhibition in the cyanobacterium *Synechococcus* sp. PCC 7002. *Biochim. Biophys. Acta* **2009**, *1787*, 1122–1128. [CrossRef] [PubMed]
22. Aliverti, A.; Pandini, V.; Pennati, A.; de Rosa, M.; Zanetti, G. Structural and functional diversity of ferredoxin-NADP$^+$ reductases. *Arch. Biochem. Biophys.* **2008**, *474*, 283–291. [CrossRef] [PubMed]
23. Bruns, C.M.; Karplus, P.A. Refined crystal structure of spinach ferredoxin reductase at 1.7 A resolution: Oxidized, reduced and 2′-phospho-5′-AMP bound states. *J. Mol. Biol.* **1995**, *247*, 125–145. [CrossRef] [PubMed]
24. Hermoso, J.A.; Mayoral, T.; Faro, M.; Gómez-Moreno, C.; Sanz-Aparicio, J.; Medina, M. Mechanism of coenzyme recognition and binding revealed by crystal structure analysis of ferredoxin-NADP$^+$ reductase complexed with NADP$^+$. *J. Mol. Biol.* **2002**, *319*, 1133–1142. [CrossRef]
25. Schluchter, W.M.; Bryant, D.A. Molecular characterization of ferredoxin-NADP$^+$ oxidoreductase in cyanobacteria: Cloning and sequence of the *petH* gene of *Synechococcus* sp. PCC 7002 and studies on the gene product. *Biochemistry* **1992**, *31*, 3092–3102. [CrossRef]
26. Gao, X.; Zhang, N.; Wei, T.D.; Su, H.N.; Xie, B.B.; Dong, C.C.; Zhang, X.Y.; Chen, X.L.; Zhou, B.C.; Wang, Z.X.; et al. Crystal structure of the N-terminal domain of linker L(R) and the assembly of cyanobacterial phycobilisome rods. *J. Mol. Biol.* **2011**, *82*, 698–705. [CrossRef]
27. van Thor, J.J.; Gruters, O.W.; Matthijs, H.C.; Hellingwerf, K.J. Localization and function of ferredoxin:NADP$^+$ reductase bound to the phycobilisomes of *Synechocystis*. *EMBO J.* **1999**, *18*, 4128–4136. [CrossRef]

28. Gómez-Lojero, C.; Perez-Gómez, B.; Shen, G.; Schluchter, W.M.; Bryant, D.A. Interaction of ferredoxin:NADP$^+$ oxidoreductase with phycobilisomes and phycobilisome substructures of the cyanobacterium *Synechococcus* sp. strain PCC 7002. *Biochemistry* **2003**, *42*, 13800–13811. [CrossRef]

29. Liu, H.; Weisz, D.A.; Zhang, M.M.; Cheng, M.; Zhang, B.; Zhang, H.; Gerstenecker, G.S.; Pakrasi, H.B.; Gross, M.L.; Blankenship, R.E. Phycobilisomes Harbor FNR$_L$ in Cyanobacteria. *MBio* **2019**, *10*, e00669-19. [CrossRef]

30. Thomas, J.C.; Ughy, B.; Lagoutte, B.; Ajlani, G. A second isoform of the ferredoxin: NADP oxidoreductase generated by an in-frame initiation of translation. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 18368–18373. [CrossRef]

31. Omairi-Nasser, A.; de Gracia, A.G.; Ajlani, G. A larger transcript is required for the synthesis of the smaller isoform of ferredoxin: NADP oxidoreductase. *Mol. Microbiol.* **2011**, *81*, 1178–1189. [CrossRef]

32. Korn, A.; Ajlani, G.; Lagoutte, B.; Gall, A.; Sétif, P. Ferredoxin:NADP$^+$ oxidoreductase association with phycocyanin modulates its properties. *J. Biol. Chem.* **2009**, *284*, 31789–31797. [CrossRef]

33. Nawrocki, W.J.; Bailleul, B.; Picot, D.; Cardol, P.; Rappaport, F.; Wollman, F.A.; Joliot, P. The mechanism of cyclic electron flow. *Biochim. Biophys. Acta Bioenerg.* **2019**, *1860*, 433–438. [CrossRef]

34. Ogawa, T.; Mi, H. Cyanobacterial NADPH dehydrogenase complexes. *Photosynth. Res.* **2007**, *93*, 69–77. [CrossRef]

35. Peltier, G.; Aro, E.M.; Shikanai, T. NDH-1 and NDH-2 plastoquinone reductases in oxygenic photosynthesis. *Annu. Rev. Plant Biol.* **2016**, *67*, 55–80. [CrossRef]

36. Schuller, J.M.; Birrell, J.A.; Tanaka, H.; Konuma, T.; Wulfhorst, H.; Cox, N.; Schuller, S.K.; Thiemann, J.; Lubitz, W.; Sétif, P.; et al. Structural adaptations of photosynthetic complex I enable ferredoxin-dependent electron transfer. *Science* **2019**, *363*, 257–260. [CrossRef]

37. Laughlin, T.G.; Bayne, A.N.; Trempe, J.F.; Savage, D.F.; Davies, K.M. Structure of the complex I-like molecule NDH of oxygenic photosynthesis. *Nature* **2019**, *566*, 411–414. [CrossRef]

38. Iwai, M.; Takizawa, K.; Tokutsu, R.; Okamuro, A.; Takahashi, Y.; Minagawa, J. Isolation of the elusive supercomplex that drives cyclic electron flow in photosynthesis. *Nature* **2010**, *464*, 1210–1213. [CrossRef]

39. Shen, L.; Tang, K.; Wang, W.; Wang, C.; Wu, H.; Mao, Z.; An, S.; Chang, S.; Kuang, T.; Shen, J.R.; et al. Architecture of the chloroplast PSI-NDH supercomplex in *Hordeum vulgare*. *Nature* **2021**, *601*, 649–654. [CrossRef]

40. Gao, F.D.; Zhao, J.H.; Chen, L.P.; Battchikova, N.; Ran, Z.X.; Aro, E.M.; Ogawa, T.; Ma, W.M. The NDH-1L-PSI supercomplex is important for efficient cyclic electron transport in cyanobacteria. *Plant Physiol.* **2016**, *172*, 1451–1464. [CrossRef]

41. Zhang, H.M.; Whitelegge, J.P.; Cramer, W.A. Ferredoxin:NADP$^+$ oxidoreductase is a subunit of the chloroplast cytochrome b_6f complex. *J. Biol. Chem.* **2001**, *276*, 38159–38165. [CrossRef]

42. Stevens, S.E., Jr.; Patterson, C.O.P.; Myers, J. The production of hydrogen peroxide by blue-green algae: A survey. *J. Phycol.* **1973**, *9*, 427–430. [CrossRef]

43. Alcántara-Sánchez, F.; Leyva-Castillo, L.E.; Chagolla-López, A.; González de la Vara, L.; Gómez-Lojero, C. Distribution of isoforms of ferredoxin-NADP$^+$ reductase (FNR) in cyanobacteria in two growth conditions. *Int. J. Biochem. Cell Biol.* **2017**, *85*, 123–134. [CrossRef]

44. Schluchter, W.M.; Shen, G.; Zhao, J.; Bryant, D.A. Characterization of *psaI* and *psaL* mutants of *Synechococcus* sp. strain PCC 7002: A new model for state transitions in cyanobacteria. *Photochem. Photobiol.* **1996**, *64*, 53–66. [CrossRef]

45. Bryant, D.A. Genetic analysis of phycobilisome biosynthesis, assembly, structure and function in the cyanobacterium *Synechococcus* sp. PCC 7002. In *Light-Energy Transduction in Photosynthesis: Higher Plants and Bacterial Models.*; Stevens, S.E., Jr., Bryant, D.A., Eds.; American Society of Plant Physiologists: Rockville, MD, USA, 1988; pp. 62–90.

46. Huang, C.; Yuan, X.; Zhao, J.; Bryant, D.A. Kinetic analyses of state transitions of the cyanobacterium *Synechococcus* sp. PCC 7002 and its mutant strains impaired in electron transport. *Biochim. Biophys. Acta* **2003**, *1607*, 121–130. [CrossRef]

47. Berla, B.M.; Saha, R.; Immethun, C.M.; Maranas, C.D.; Moon, T.S.; Pakrasi, H.B. Synthetic biology of cyanobacteria: Unique challenges and opportunities. *Front. Microbiol.* **2013**, *4*, 246. [CrossRef]

48. Zheng, L.; Li, Y.; Li, X.; Zhong, Q.; Li, N.; Zhang, K.; Zhang, Y.; Chu, H.; Ma, C.; Li, G.; et al. Structural and functional insights into the tetrameric photosystem I from heterocyst-forming cyanobacteria. *Nat. Plants* **2019**, *5*, 1087–1097. [CrossRef]

49. Noji, T.; Watanabe, M.; Dewa, T.; Itoh, S.; Ikeuchi, M. Direct energy transfer from allophycocyanin-free rod-type CpcL-phycobilisome to photosystem I. *J. Phys. Chem. Lett.* **2021**, *12*, 6692–6697. [CrossRef]

50. Dong, C.; Zhao, J. ApcD is required for state transition but not involved in blue-light induced quenching in the cyanobacterium *Anabaena* sp. PCC 7120. *Sci. Bull.* **2008**, *53*, 3422–3424. [CrossRef]

51. Holland, S.C.; Kappell, A.D.; Burnap, R.L. Redox changes accompanying inorganic carbon limitation in *Synechocystis* sp. PCC 6803. *Biochim. Biophys. Acta* **2015**, *1847*, 355–363. [CrossRef]

52. Korn, A. Respective Roles of the Ferredoxin: NADP-Oxidoreductase Isoforms in the Cyanobacterium *Synechocystis* sp. PCC 6803. Ph.D. Thesis, Université Paris Sud, Paris, France, November 2010.

53. Miller, N.T.; Ajlani, G.; Burnap, R.L. Cyclic electron flow-coupled proton pumping in *Synechocystis* sp. PCC 6803 is dependent upon NADPH Oxidation by the soluble isoform of Ferredoxin:NADP-Oxidoreductase. *Microorganisms* **2022**, *10*, 855. [CrossRef] [PubMed]

54. Yang, M.; Yang, Y.; Chen, Z.; Zhang, J.; Lin, Y.; Wang, Y.; Xiong, Q.; Li, T.; Ge, F.; Bryant, D.A.; et al. Proteogenomic analysis and global discovery of posttranslational modifications in prokaryotes. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E5633–E5642. [CrossRef] [PubMed]

55. Mi, H.; Endo, T.; Ogawa, T.; Asada, K. Thylakoid membrane-bound, NADPH-specific pyridine nucleotide dehydrogenase complex mediates cyclic electron transport in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Cell Physiol.* **1995**, *36*, 661–668.
56. Laughlin, T.G.; Savage, D.F.; Davies, K.M. Recent advances on the structure and function of NDH-1: The complex I of oxygenic photosynthesis. *Biochim. Biophys. Acta Bioenerg.* **2020**, *1861*, 148254. [CrossRef]
57. Ogawa, T. A gene homologous to the subunit-2 gene of NADH dehydrogenase is essential to inorganic carbon transport of *Synechocystis* PCC 6803. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 4275–4279. [CrossRef]
58. Yu, L.; Zhao, J.; Muhlenhoff, U.; Bryant, D.A.; Golbeck, J.H. PsaE is required for in vivo cyclic electron flow around photosystem I in the cyanobacterium *Synechococcus* sp. PCC 7002. *Plant Physiol.* **1993**, *103*, 171–180. [CrossRef]
59. Miller, N.T.; Vaughn, M.D.; Burnap, R.L. Electron flow through NDH-1 complexes is the major driver of cyclic electron flow-dependent proton pumping in cyanobacteria. *Biochim. Biophys. Acta Bioenerg.* **2021**, *1862*, 148354. [CrossRef]
60. Zhang, C.; Shuai, J.; Ran, Z.; Zhao, J.; Wu, Z.; Liao, R.; Wu, J.; Ma, W.; Lei, M. Structural insights into NDH-1 mediated cyclic electron transfer. *Nat. Commun.* **2020**, *11*, 888. [CrossRef]
61. Pan, X.; Cao, D.; Xie, F.; Xu, F.; Su, X.; Mi, H.; Zhang, X.; Li, M. Structural basis for electron transport mechanism of complex I-like photosynthetic NAD(P)H dehydrogenase. *Nat. Commun.* **2020**, *11*, 610. [CrossRef]

Article

Evidence for Electron Transfer from the Bidirectional Hydrogenase to the Photosynthetic Complex I (NDH-1) in the Cyanobacterium *Synechocystis* sp. PCC 6803

Jens Appel [1,2] [ID], **Sean Craig** [3], **Marius Theune** [1,2], **Vanessa Hüren** [2], **Sven Künzel** [4], **Björn Forberich** [2], **Samantha Bryan** [3] and **Kirstin Gutekunst** [1,2,*]

1   Department of Molecular Plant Physiology, Bioenergetics in Photoautotrophs, University of Kassel, D-34132 Kassel, Germany
2   Department of Biology, Botanical Institute, Christian-Albrechts-University, D-24118 Kiel, Germany
3   BBSRC/EPSRC Synthetic Biology Research Centre, The Biodiscovery Institute, University of Nottingham, Nottingham NG7 2RD, UK
4   Max-Plank Institute for Evolutionary Biology, August-Thienemann-Straße 2, D-24306 Plön, Germany
*   Correspondence: kirstin.gutekunst@uni-kassel.de

Abstract: The cyanobacterial bidirectional [NiFe]-hydrogenase is a pentameric enzyme. Apart from the small and large hydrogenase subunits (HoxYH) it contains a diaphorase module (HoxEFU) that interacts with NAD(P)$^+$ and ferredoxin. HoxEFU shows strong similarity to the outermost subunits (NuoEFG) of canonical respiratory complexes I. Photosynthetic complex I (NDH-1) lacks these three subunits. This led to the idea that HoxEFU might interact with NDH-1 instead. HoxEFUYH utilizes excited electrons from PSI for photohydrogen production and it catalyzes the reverse reaction and feeds electrons into the photosynthetic electron transport. We analyzed hydrogenase activity, photohydrogen evolution and hydrogen uptake, the respiration and photosynthetic electron transport of Δ*hoxEFUYH*, and a knock-out strain with dysfunctional NDH-1 (Δ*ndhD1*/Δ*ndhD2*) of the cyanobacterium *Synechocystis* sp. PCC 6803. Photohydrogen production was prolonged in Δ*ndhD1*/Δ*ndhD2* due to diminished hydrogen uptake. Electrons from hydrogen oxidation must follow a different route into the photosynthetic electron transport in this mutant compared to wild type cells. Furthermore, respiration was reduced in Δ*hoxEFUYH* and the Δ*ndhD1*/Δ*ndhD2* localization of the hydrogenase to the membrane was impaired. These data indicate that electron transfer from the hydrogenase to the NDH-1 complex is either direct, by the binding of the hydrogenase to the complex, or indirect, via an additional mediator.

Keywords: hydrogen; photosynthesis; electron transport; photosystem I; respiration

1. Introduction

All respiratory complexes I that catalyze the oxidation of NADH and reduction of a quinone share 14 core subunits [1]. The first structural study on the cyanobacterial complex, now termed photosynthetic complex I, demonstrated that it receives electrons directly from ferredoxin [2]. In contrast to the canonical respiratory complexes, the photosynthetic complexes of cyanobacterial and plant chloroplasts do not contain the three outer subunits known to bind and oxidize NADH [2,3]. These subunits contain five additional FeS clusters that are absent from the photosynthetic structure. Instead, this kind of beheaded complex harbors only the three innermost FeS clusters with a binding site for ferredoxin on the top made up of three small additional subunits (NdhO, NdhV, and NdhS) only found in cyanobacteria and plant chloroplasts [3,4].

The absence of the outer subunits was already noted some time ago and the hypothesis was put forward that three of the subunits of the pentameric cyanobacterial bidirectional hydrogenase (HoxEFU) might take over. It was suggested that the hydrogenase forms a

large complex with the photosynthetic complex I [5,6]. In a number of other bacteria and archaea large hydrogenase complexes with strong homologies to the respiratory complex I exist [7,8].

Hydrogenases catalyze the formation of H_2 from two electrons and two protons or the reverse reaction. Due to its low redox potential H_2 is a versatile electron donor that can be used to reduce $NAD(P)^+$ or different quinones. In the reverse reaction protons serve as electron acceptors, e.g., under fermentative conditions to oxidize ferredoxin and form H_2. Since hydrogenases serve an important role in energy metabolism there is a plethora of different classes and subtypes in the microbial world differentiated based on their reaction partners [9]. Many microbial communities rely on the exchange of hydrogen and even though the composition of Earth's atmosphere changed dramatically and hydrogen is now a trace gas with 0.5 ppm, there is widespread use of hydrogenases in bacteria for sustained energy supply from the air [10].

Cyanobacteria are known to harbor two different types of [NiFe]-hydrogenases; the uptake, and the bidirectional hydrogenase. While the uptake hydrogenase is expressed in diazotrophic strains to regain the reducing power of the H_2 produced as a side product by the nitrogenase, the bidirectional enzyme is there to remove a surplus of electrons either during fermentation or when the cells restart photosynthesis from dark anaerobic conditions [11–13].

The cyanobacterial bidirectional hydrogenase accepts electrons from ferredoxin and flavodoxin while producing photohydrogen [14]. This is supported by the lower redox potentials of ferredoxin/flavodoxin compared to $NAD(P)H$ that prohibit H_2 production from the pyridine nucleotides at their normal physiological $NAD(P)H/NAD(P)^+$ ratio. Another study showed that the diaphorase module of this hydrogenase accepts electrons from ferredoxins to predominantly reduce NAD^+ to NADH; the K_M of NADPH was found to be about 25 times higher [15]. This is supported by previous investigations that also showed about six times higher rates of H_2 production with NADH compared to NADPH [16]. Thus, this type of hydrogenase is a central metabolic hub that mediates electron transfer between the three redox pools ferredoxin/flavodoxin, H_2 and predominantly NADH. In particular, NADH formation from reduced ferredoxin is peculiar for this enzyme and might open an additional channel aside the FNR or the transhydrogenases when the cells become strongly reduced.

It has long been known that this type of hydrogenase is responsible for hydrogen evolution in the light (so-called photoH_2) that is produced when dark anaerobic cells are suddenly exposed to strong light [17]. Recently, we showed that the enzyme is also working as an electron valve when the cells are shifted to high light intensities under aerobic conditions [18]. It remains unclear how the enzyme performs this task but NADH formation from reduced ferredoxin and its subsequent oxidation in respiration are a possible route.

In cyanobacteria not only one type of photosynthetic complex I exists. There are at least four different types of NDH-1 complexes in *Synechocystis* that contain different NdhD and NdhF subunits. Two of these complexes contain either NdhD1/NdhF1 or NdhD2/NdhF1 and are known to be involved in cyclic electron transport and respiration [19], while two additional complexes are known to be part of the carbon-concentrating mechanism (CCM, [20]) and contain either NdhD3/NdhF3 or NdhD4/NdhF4. Here we will name the complexes according to their NdhD subunits as NDH-1_1, NDH-1_2, NDH-1_3 and NDH-1_4.

Localization studies using GFP-labelled bidirectional hydrogenase found variable proportions of the enzyme in the cytoplasm and at the membrane [21]. No evidence for a direct interaction with the NDH-1 complex could be found in this study, although the hydrogenase was found to be clustered in so-called puncta at the membrane, which is reminiscent of localization studies of the cyanobacterial NDH-1 that found highly concentrated patches of the complex [22]. Thus, although the bidirectional hydrogenase is a soluble protein complex, its membrane attachment remains unclear and an interaction with NDH-1 is still an option.

This leaves us with two main unsolved issues for the cyanobacterial case. The first concerns the carbohydrate breakdown known for cyanobacteria. Except a possible route via the pyruvate:ferredoxin oxidoreductase [23], all others produce NAD(P)H but not reduced ferredoxin. In principle there are ferredoxins with higher redox potentials in the cyanobacterial cell that could be reduced by NAD(P)H, but it is not known if the FNR or any other enzyme could catalyze the required reactions and if the NDH-1 is able to oxidize them in return. So how do cyanobacteria submit electrons from glucose breakdown to their respiratory electron transport chain?

The second issue concerns the oxidation of hydrogen (also known as hydrogen uptake) that is catalyzed by the bidirectional hydrogenase after a short phase of photohydrogen evolution or when hydrogen is provided under anaerobic conditions in the light and photosystem II is blocked. Under these conditions, H_2 serves as a reductant for CO_2-fixation [11,12]. It still remains to be solved if the electrons that are delivered to the Calvin–Benson–Bassham (CBB) cycle are mediated either via ferredoxin (Fd) or NAD(P)H (H_2- > Fd- > PQ- > PSI- > CBB (Figure 1C) or H_2- > NAD(P)H- > CBB (Figure 1B)), or if they are fed into the PQ pool by a hydrogenase bound to the NDH-1 complex (H_2- > PQ- > PSI- > CBB, Figure 1D).
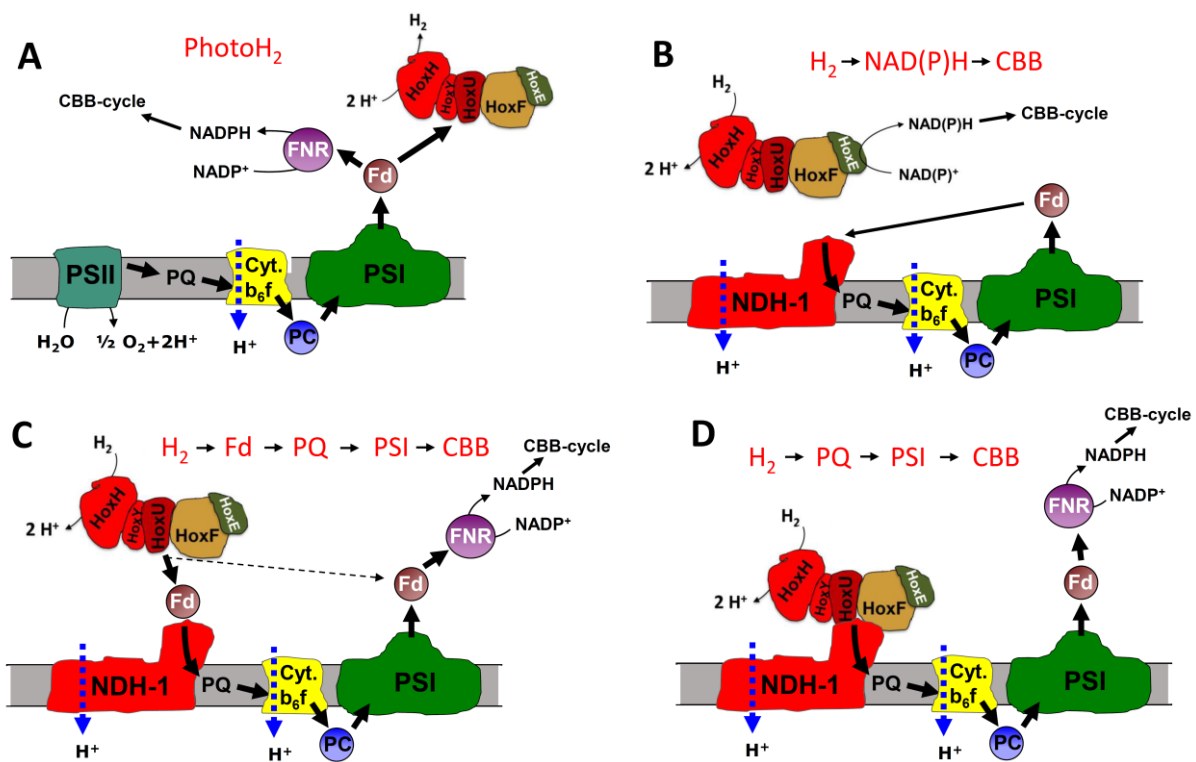
Figure 1. Schematic showing the different electron transport routes discussed. (**A**) Under anaerobic conditions when the CBB cycle is not activated the hydrogenase acts as an electron valve and produces photoH_2 when photosynthesis starts from darkness. Because of PSII forming O_2, conditions turn more oxidizing after a short time and H_2 evolution turns into hydrogen uptake (for an example see Figure 3). The electrons gained from H_2 oxidation could follow three different routes. (**B**) The hydrogenase could directly reduce NADPH, which would than serve the CBB cycle while cyclic electron transport provides the necessary ATP. (**C**) In an alternative pathway H_2 oxidation could result in ferredoxin reduction that is then re-oxidized by the photosynthetic NDH-1 complex. (**D**) In the third option, the hydrogenase could be part of the NDH-1 complex and deliver the electrons directly to the PQ pool. Abbreviations: Fd = ferredoxin, FNR = ferredoxin NADP reductase, PSII and PSI = photosystem II and I, PQ = plastoquinone, Cyt b6f = cytochrome b6f complex, NDH-1 = photosynthetic complex I, CBB = Calvin–Benson–Bassham cycle.

In this study, we undertook a number of investigations of the hydrogenase activity, hydrogen evolution in the light, hydrogen uptake, electron transport, and respiratory activity of the deletion strains of *Synechocystis* sp. PCC 6803 that are impaired in different NDH-1 complexes or the hydrogenase. The results indicate that the absence of the bidirectional hydrogenase causes a down-shift in respiratory activity and that the absence of specific NDH-1 complexes impairs hydrogen uptake. The colocalization of a GFP-labelled HoxF and a YFP-labelled NdhM show that the hydrogenase is found less in the membrane regions in a mutant without NDH-1$_1$ and NDH-1$_2$. Although we present physiological evidence only, this study provides strong evidence for electron transfer from the bidirectional NiFe-hydrogenase to the NDH-1 complex.

2. Materials and Methods

2.1. Growth Conditions

All strains were inoculated in 50 mL BG-11 and placed in 100 mL Erlenmeyer flasks on a rotary shaker at 28 °C, 50 µE m^{-2} s^{-1} and 100 rpm. After several days of growth, the cells were inoculated into 200 mL BG-11 at an OD$_{750}$ of 0.05 and placed into glass tubes bubbled with air at 120 µE m^{-2} s^{-1} at 28 °C. After four days of growth, cells were harvested by centrifuging at 4000× g for 5 min and then re-suspended in fresh BG-11 for measurement.

If not indicated otherwise we used OD$_{750}$ as a reference in all the measurements as it is a good proxy for cell dry weight [24] and because we did not find large differences between wild type and mutant strains concerning their chlorophyll content.

2.2. Construction of Mutants

Constructs were assembled in pBluescript SK- (Addgene, Watertown, MA, USA). The vector was cut with SacI and KpnI (Thermofisher Scientifc, Dreieich, Germany) and the PCR products including the resistance cassette were introduced by Gibson assembly [25] in a single step. All the primers used in this study are shown in Table S1. After sequencing, the respective constructs were transformed into glucose-tolerant *Synechocystis* sp. PCC 6803 wild type cells as described [26]. After re-streaking the resulting transformants at least four times, the clones were tested by PCR or Southern hybridization for complete segregation (Figure S1).

In this study we used a deletion strain without any of the three type 2 dehydrogenase genes (*ndbA* (*slr0851*), *ndbB* (*slr1743*), and *ndbC* (*sll1484*)) named Δ*ndh-2*, and a strain lacking only *ndbA* and *ndbC*, Δ*ndbA*Δ*ndbC*. Since there is evidence that NdbB is only used as an oxidoreductase to reduce prenyl naphthoquinones and prenyl benzoquinones during the biosynthesis of phylloquinone and plastoquinone and the respective deletion mutant Δ*ndbB* is lacking phylloquinone [27] we investigated parallel to the complete knock-out strain (Δ*ndh-2*) also the one still containing *ndbB* to make sure that the lack of phylloquinone biosynthesis does not result in additional effects on electron transport and hydrogen metabolism. In general, both strains behaved very similarly and showed only small deviations that are shown in the different figures.

2.3. Oxygen Measurements

To measure oxygen consumption the cell density was adjusted to 20 µg chlorophyll/mL and 1.6 mL was placed in the cuvette of the Dual-KLAS/NIR (Walz, Effeltrich, Germany). An oxygen microsensor OX-50 (Unisense, Aarhus, Denmark) was placed in a lab-made holder that allowed for the insertion of the sensor tip from above into the solution while simultaneously concealing the surface of the solution from gas exchange with the surroundings. The data were acquired with a time resolution of 0.2 s and the stirrer was on all the time. To eliminate photorespiration, 5 mM NaHCO$_3$ was added as well as 10 mM glucose when its effect on respiration or oxygen evolution was to be measured. Before measurements the oxygen sensor was calibrated as described by the manufacturer.

2.4. PhotoH$_2$ Measurements

Hydrogen was measured using a microsensor H2-50 (Unisense, Aarhus, Denmark) and the same holder as described was used. Before starting the measurements, the sensor was calibrated as described by the manufacturer. The cell density was adjusted to 20 μg chlorophyll/mL and 1.6 mL was placed in a cuvette of the Dual-KLAS/NIR or the Multicolor-PAM (Walz, Effeltrich, Germany). The total volume also contained 40 U/mL glucose oxidase, 50 U/mL catalase, and 10 mM glucose to induce anaerobiosis. Cells were kept in darkness for 5 min to monitor fermentative H$_2$ production and then a light with 750 μE/m^2/s was switched on. PhotoH$_2$ production and the subsequent H$_2$ oxidation were followed until completion.

2.5. Hydrogen Uptake

In this case 20 μM DCMU was added to the sample and it was kept anaerobic by including 10 mM glucose, 40 U/mL glucose oxidase, and 50 U/mL catalase. Before inserting the H$_2$-sensor, H$_2$-saturated BG-11 was used to dilute the sample to 20 μg chlorophyll/mL. Thereby H$_2$ concentrations between 400 and 500 μM were reached. In a 5 min dark period H$_2$ loss due to diffusion was followed and then a light intensity series from 16 to 420 μE m^{-2} s^{-1} was applied. Each intensity was measured for about 3 min and the resulting slope of H$_2$ uptake was corrected by the diffusion determined in the dark. Except for a brief mixing period at the start of the experiment, the stirrer was switched off to prevent the high diffusive loss of H$_2$.

2.6. Electron Flow Determination through Photosystem I

To determine electron flow through photosystem I we used dark-interval relaxation kinetics (DIRK) as described [28]. To this end the Dual-KLAS/NIR instrument was used. It was set to Fast Acquisition and after preincubation of the cells for 1 min at the respective light intensity to ensure a steady-state, DIRK measurements were started. During the following period the light was repeatedly shut off for 25 ms every 0.3 s for 400 times using a Fast-Kinetic Trigger File in combination with a Fast-Kinetic Multi-Run (4000 points, 0.5 ms resolution). The resulting traces were averaged to reduce noise. Based on the rates of re-reduction of P700 and plastocyanin in the dark intervals it is possible to calculate the electron throughput of PSI as μM e$^-$/s or e$^-$/PSI/s. For further details please refer to Reference [28].

2.7. State Change Measurements

Cell suspensions were diluted to 2.5 μg chlorophyll/mL and 1.25 mL of this suspension was placed in a stirred cuvette of the Multicolor-PAM (Walz, Effeltrich, Germany). The intensity of the measuring light and the gain were set to one. The actinic light and the measuring light were red with a maximum at 625 nm. After switching on the measuring light to determine F$_0$ one saturation pulse of 3500 μE/m^2/s was given. Subsequently, an actinic light of 110 μE/m^2/s was switched on and after 1 min another saturation pulse was given (F$_{m-glc}$). During a following dark phase of 30 s, glucose was added to 10 mM and after another 3 min of actinic light a second saturation pulse was applied (F$_{m+glc}$). To acquire a measure for the state change we used the formula (F$_{m-glc}$ − F$_{m+glc}$)/F$_0$.

2.8. Confocal Laser Scanning Microscopy and Colocalization

Cyanobacterial cultures were grown at 30 μmol photon m^{-2} s^{-1} in BG-11 at 150 rpm. Anoxia was achieved by supplementing BG11 with 10 mM glucose and 16 U mL^{-1} glucose oxidase and 20 U mL^{-1} of catalase. A 10 μL aliquot of each culture was spotted onto BG-11 agar, smeared into a patch of approximately 0.5 cm^2, and left to dry. Agar cubes of approximately 1 cm^2 were cut and placed on a 35 mm glass bottom dish (MatTek Life Sciences). All experiments were carried out in triplicate.

Cells were imaged using the ZEISS Elyra 7 (Carl Zeiss AG, Oberkochen, Germany) with a 63× water immersion objective lens, photosynthetic pigments were excited at 633 nm,

and the fluorescence was measured between 695–720 nm. The wavelengths for GFP and YFP excitation were 488 nm and 514 nm, respectively, and fluorescence was measured between 498–517 for GFP and 525–544 nm for YFP.

Photosynthetic pigments found in both Photosystem I and Photosystem II contain chromophores, which result in background autofluorescence. GFP and YFP both have strong absorption and emission spectrums, which bleach easily under excessive laser illumination at 488 nm and 514 nm [29]. To determine the fluorescence, which directly results from both the GFP and the YFP, images were recorded (pre-bleach), and then the laser intensity was increased for ten seconds to bleach both the GFP and the YFP (post bleach). Post bleach images were then recorded for both the GFP and the YFP. The post bleach images were subtracted from the pre bleached images, providing a true depiction of the distribution of GFP and YFP in the cell (Figures S9 and S10) [29]. Figure S11 depicts a wild type control (no GFP/YFP); images were captured, and then the cells were subject to increased illumination for 10 s. There was no impact on the background autofluorescence from the photosynthetic pigments and no contribution from these pigments to the GFP or YFP channels (Figure S11).

Images were processed using the Fiji software (version 2.4.0 available at https://imagej.net/software/fiji/). The wild type images were tabulated first, to establish the background fluorescence and set the thresholds for colocalization. The image channels were split; the post-bleached fluorescent protein images were subtracted from the pre-bleached images to generate a composite image. The contrast settings in the fluorescent protein channels were set to optimize visualization and the contrast settings were applied to all other images for that channel. Colocalization was quantified by assessing the fluorescence intensity of each pixel relative to the individual channels, using the colocalization function in Fiji. The Manders coefficient was used to determine the degrees of colocalization with the variance between conditions and strains being measured using the *t*-test, as previously described [30].

3. Results

3.1. Dark Respiration

The dark oxygen uptake of different deletion mutants of *Synechocystis* sp. PCC 6803 was measured (Figure 2). A mutant without any of the structural genes of the bidirectional hydrogenase (Δ*hoxEFUYH*) shows a lowered oxygen uptake. A similar low oxygen uptake is only visible in a strain without NDH-1$_1$ and NDH-1$_2$ (Δ*ndhD1*Δ*ndhD2*) and a strain without any of the terminal respiratory oxidases (Δ*ox*) while all other strains such as those without the type 2 NADH-dehydrogenase (Δ*ndh-2*, Δ*ndbA*Δ*ndbC*), and those lacking either NDH-1$_1$ or NDH-1$_2$, show an oxygen uptake close to wild type cells. The strain lacking only the large hydrogenase subunit (Δ*hoxH*) shows a reduced oxygen uptake but this reduction is not as strong as in Δ*hoxEFUYH* or Δ*ndhD1*Δ*ndhD2*. Surprisingly, the Δ*ndhD3*Δ*ndhD4* has a much higher respiratory oxygen consumption than all other strains.

The addition of glucose to cells in the dark induces an increase in oxygen consumption. This increase is also visible in Δ*hoxEFUYH* and Δ*hoxH* (Figure S2).

NDH-1$_1$ and NDH-1$_2$ are known to be the major NDH-1 complexes involved in respiration [19]. This is corroborated by our results since the Δ*ndhD1*Δ*ndhD2* shows an oxygen uptake similar to a mutant without terminal respiratory oxidases (Δ*ox*). We attribute the highly increased respiration in the Δ*ndhD3*Δ*ndhD4* to either a kind of compensatory upregulation of the respiratory complexes NDH-1$_1$ and NDH-1$_2$, or a need for higher respiration/ATP production due to the lack of the CCM. The reduction of respiration in the Δ*hoxEFUYH* in particular is surprising and indicates that the diaphorase (the NAD(P)H- and ferredoxin-handling part of the hydrogenase) also has a role in introducing electrons into the respiratory electron transport.
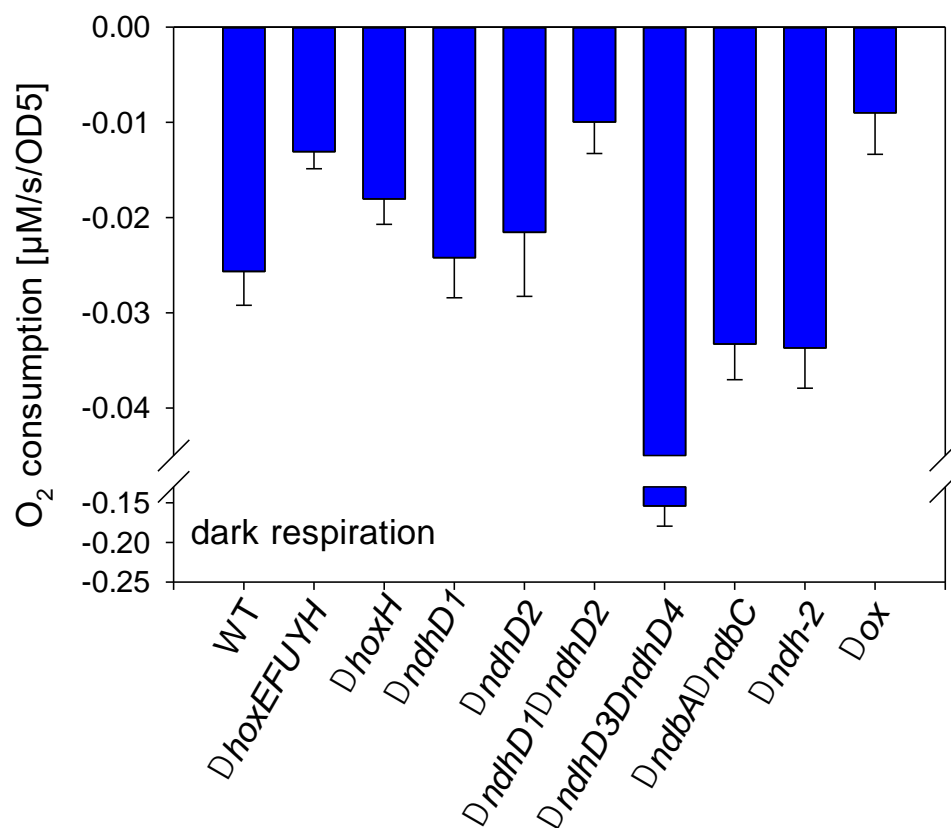
Figure 2. Dark respiration of different deletion strains of *Synechocystis* sp. PC 6803. Cell density was adjusted to the same OD_{750} and oxygen uptake measured for 5 min in the dark. The rate of oxygen uptake was taken from at least three different cultures of the same strain of the last 3 min. Error bars represent standard deviation.

The reduction of oxygen uptake in the Δ*hoxH* is also interesting, and is probably due to a destabilization of the whole enzyme due to the lack of the large subunit. We attribute the increase in dark respiration upon the addition of glucose to an increased oxidative pentose phosphate pathway under these conditions. Consequently, NADPH production would increase and could be degraded via FNR- > ferredoxin- > NDH-1. This is a pathway that could be pursued in the dark when the PQ pool is not reduced by PSII since FNR is known to work in the reverse direction [31].

3.2. PhotoH$_2$ Production

In Figure 3, original traces of measurements of photoH$_2$ production of the different strains are shown as recorded with a H$_2$ electrode. For a detailed analysis of the different phases of H$_2$ production and uptake including the standard deviation see Figure S3. To induce photoH$_2$ production, the cells were kept in darkness under anaerobic conditions for 5 min. After this dark phase the light was switched on with an intensity of about 700 μE/m^2/s.

The two strains that lack *ndhD1* produce higher amounts of H$_2$ in the light. While Δ*ndbA*Δ*ndbC* and Δ*ndhD2* are not different to wild type cells, the Δ*ndhD3ndhD4* shows diminished photoH$_2$ production.

When comparing the different strains, it is important to note the Δ*ndhD1*Δ*ndhD2* double mutant consistently shows a long phase of photoH$_2$ production and the lowest rates of H$_2$ uptake (Figures 3 and S3).

Direct measurements of the hydrogenase reveal that the activities of all the strains are in a small range around wild type activity (Figure S4). The Δ*ndhD1*Δ*ndhD2* mutant has

about 80%, and the Δ*ndhD3*Δ*ndhD4* and Δ*ndbA*Δ*ndbC* both have about 70% of the activity of wild type cells.
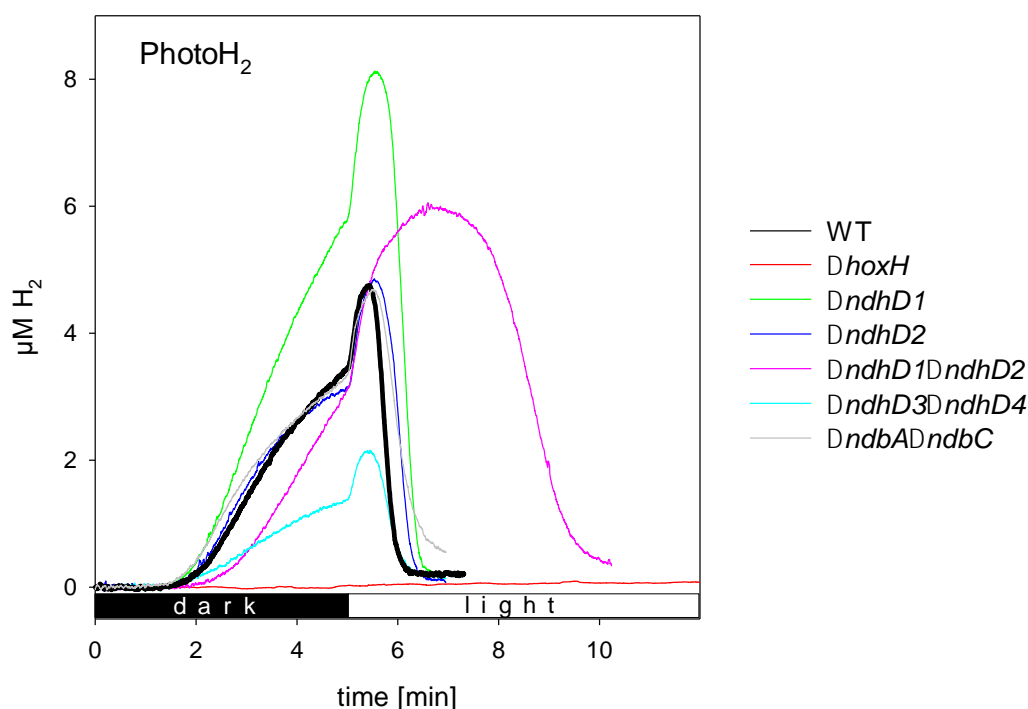
Figure 3. PhotoH$_2$ production of different deletion strains of *Synechocystis* sp. PC 6803 in comparison to wild type cells and cells without a hydrogenase (Δ*hoxH*). Cell density was adjusted to the same OD$_{750}$, anaerobiosis was induced by addition of 10 mM glucose, 40 U/mL glucose oxidase and 50 U/mL catalase. The original traces of wild type cells, Δ*hoxH* and the *ndhD* deletion strains are shown in comparison to a strain lacking the type 2 dehydrogenase (Δ*ndbA*Δ*ndbC*). For the variability between different cultures of the same strain please refer to Figure S3.

3.3. Electron Flow through PSI during H$_2$ Uptake

To follow H$_2$ uptake while measuring electron flow through PSI, anaerobiosis was induced by glucose, glucose oxidase, and catalase, as demonstrated above, and DCMU was added to block PSII activity. Under these conditions H$_2$ is the electron donor and photosynthesis including only PSI is running. H$_2$ consumption was measured by a H$_2$ electrode and electron flow was determined by DIRK (dark interval relaxation kinetics) measurements as described earlier [28].

Figure 4 on the one hand shows the increase of electron flow through PSI as calculated by subtracting the electron flow as measured without H$_2$ from the electron flow with H$_2$ (black symbols), and on the other hand the number of electrons generated by H$_2$ oxidation (red symbols). Please note that a strain without hydrogenase does not show an increased electron flow through PSI after the addition of H$_2$ [28].

While all the strains measured show a more or less parallel increase of PSI electron throughput compared to H$_2$ oxidation, the Δ*ndhD1*Δ*ndhD2* double mutant increases its H$_2$ uptake to only about 40% of the electron flow. To explain this deviating behavior, we investigate the electron flow of the different strains (Figure 4 and especially Figure S5) more closely. From the DIRK measurements made in parallel to the H$_2$ uptake measurements it is possible to calculate the NADPH/ATP ratio formed. When the cells fix CO$_2$ they have to produce a certain ratio of NADPH to ATP to fulfil the needs of the CBB cycle [32]. We found that the sum of linear and cyclic electron transport as determined by DIRK measurements when the cells perform oxygenic photosynthesis should yield a ratio of 2.2 [28]. This is very close to the value of 2.13 calculated on the basis of stoichiometric considerations for the biomass generated by *Synechocystis* [33]. Table 1 shows the NADPH/ATP ratio calculated

from the data in Figures 4 and S5. In this case we assumed that the major entry site for electrons into the electron transfer to PSI is the NDH-1 complex transferring a total of four protons per electron. Only in case of the $\Delta ndhD1 \Delta ndhD2$ mutant the entry site must be the type 2 dehydrogenases (NDH-2) that are not able to couple electron transfer to proton translocation. Thus, in this case, the transfer of two protons per electron was considered. These results confirm that the NDH-1 complex is not used in the $\Delta ndhD1 \Delta ndhD2$ but in all others.
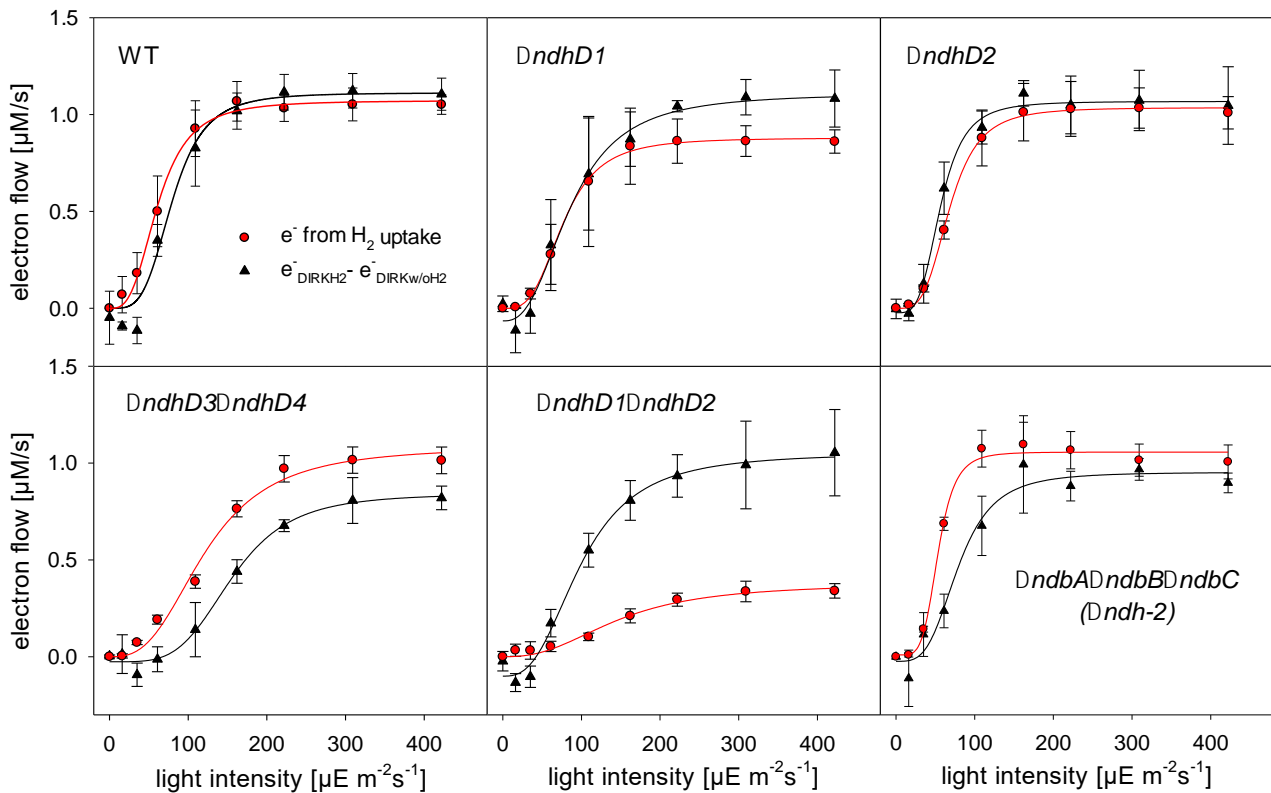
Figure 4. Parallel measurements of H_2 consumption (by a hydrogenase electrode) and electron flow through PSI by DIRK (dark interval relaxation kinetics) measurements of different *ndhD* deletion strains and the deletion strain of the type 2 dehydrogenase ($\Delta ndbA \Delta ndbB \Delta ndbC$). The black triangles and the black lines indicate the increase of electron flow through PSI due to the addition of H_2 and the red circles and red lines indicate the number of electrons generated by H_2 oxidation. The rates are given as μM electrons/s at a cell density of $OD_{750} = 5.7$. The error bars indicate the standard deviation of at least three independent cultures.

The calculated values are in a range from about 2.6 to 4.2 and probably differ from one mutant to the other because of the deviating use of different electron transport pathways and/or the inhibition of specific pathways due to the specific mutation. Under the conditions used here (pH 8 and 5 mM HCO_3^-) the $\Delta ndhD1 \Delta ndhD2$ and the $\Delta ndhD3 \Delta ndhD4d$ in particular—because of their impaired NDH-1 complexes—might be forced to use a higher proportion of ATP for the import of bicarbonate and thus need a higher ATP/NADPH ratio.

Based on these calculations the $\Delta ndhD1 \Delta ndhD2$ strain would generate an ATP/NADPH ratio of eight according to our measurements when using the NDH-1 complex. This is far outside the range of the other strains and corroborates that this strain, in contrast to the others, is not using this complex for electron transfer when using H_2 as electron donor for CO_2 fixation.

Table 1. ATP/NADPH ratio as calculated from electron flow through PSI ($V_{max} + H_2$) and electrons generated by H_2 oxidation ($V_{max}H_2$-uptake). If an electron is introduced into the intersystem chain by the NDH-1 complex it should translocate 4 H^+ into the lumen before passing PSI. In *Synechocystis* 4.66 H^+ are needed to produce one ATP resulting in the production of 10.6 ATP from 12.4 electrons. On the other hand, when H_2 is the only electron donor, two electrons from H_2 oxidation should yield one NADPH. In wild type cells three NADPH are formed from six e^- and an ATP/NADPH ratio of $10.6/3 = 3.5$ is produced. Standard deviations are given.

Strain	$V_{max} + H_2$ e^-/PSI/s	V_{max} H_2-Uptake e^-/PSI/s	ATP/NADPH Ratio [1]
WT	12.4 ± 0.5	6.0 ± 0.9	3.5 ± 0.5
$\Delta ndhD1$	10.2 ± 1.1	5.4 ± 0.9	3.3 ± 0.7
$\Delta ndhD2$	10.0 ± 0.4	6.4 ± 0.8	2.7 ± 0.4
$\Delta ndhD1\Delta ndhD2$	9.2 ± 1.6	1.9 ± 0.5	4.0 ± 1.2 [2]
$\Delta ndhD3\Delta ndhD4$	15.8 ± 1.7	6.4 ± 0.9	4.2 ± 0.8
Δndh-2	20.6 ± 0.7	14.1 ± 1.4	2.6 ± 0.3

[1] The ATP/NADPH ratio was calculated from the amount of ATP using $V_{max+H2} \times 4/4.66$ and the amount of NADPH by $V_{max\ H2\text{-uptake}}/2$, assuming that all the e^- passing PSI enter the electron transfer via the NDH-1. [2] In case of the $\Delta ndhD1\Delta ndhD2$ the entry site for electrons into the electron transfer must be predominantly NDH-2 thus only two protons are transferred per electron from the stromal to the luminal site. To calculate the amount of ATP the formula $V_{max+H2} \times 2/4.66$ was used.

When the cells take up H_2 under these conditions it is consumed completely. From the concentration curves measured against time it is possible to compute the rates of H_2 uptake for every concentration. We plotted the rates against the respective concentration (see Figure S6 for an example of such a plot). H_2 consumption by the bidirectional hydrogenase shows a cooperative behavior and when fitted with a Hill equation we acquired a K_M of 11.5 μM $\pm$ 1.5 and 9.0 μM $\pm$ 1.8 for wild type cells and the $\Delta ndhD1\Delta ndhD2$ double mutant, respectively. While these two values are very close to each other and inside the experimental error range, the V_{max} value of 6.9 nmol H_2/L/min/OD_{750} $\pm$ 0.7 for wild type cells is more than twice as large as 2.9 nmol H_2/L/min/OD_{750} $\pm$ 0.2 for the mutant. These results indicate that the hydrogenase itself shows the same affinity for H_2 in both strains but that the electron transfer to a possible acceptor is impaired or different in the mutant strain.

Taken together these results suggest that NDH-1_1 and NDH-1_2 specifically are needed for H_2 uptake. Since it is well known that NAD^+ is a good electron acceptor for electrons from the hydrogenase [11,14,16] its reduction to NADH and subsequent oxidation by the NDH-2 might be the alternative pathway in the $\Delta ndhD1\Delta ndhD2$ mutant. It seems highly improbable that ferredoxin 1 (Ssl0020), which is an electron donor to the hydrogenase [14], could be a mediator between the hydrogenase and the NDH-1 complex for H_2 uptake since it should be reduced to a certain extent by PSI under photosynthetic conditions. Using the Nernst equation, it is possible to calculate, which concentration of H_2 is in equilibrium with a certain ratio of reduced to oxidized ferredoxin (Fd_{red}/Fd_{ox}). If we assume that this ratio is in the range from 1.5 to 0.5 and if we take into account that the intracellular pH is about 7.3 [34], concentrations between 360 and 42 μM could be reached at the lowest, and even at a ratio of 0.1 (which seems unlikely during photosynthesis/CO_2-fixation) the H_2 concentration would still be 1.7 μM (see Figure S6). Since the H_2 level drops to about 0 μM during uptake (Figure 3) ferredoxin 1 is an unlikely electron mediator. Instead, it is much more likely that NAD^+ is the electron acceptor in case of the $\Delta ndhD1\Delta ndhD2$ strain and PQ via the NDH-1 complex in all other strains. This is supported by the normal H_2 uptake activity of the NDH-2 deletion strain, which is known to accept electrons from NADH [35] and that should not be possible if all strains were using NAD^+ as an electron acceptor.

The main conclusion we draw from these results is that the electron transfer from the hydrogenase to the NDH-1 during H_2 uptake is either direct and results in PQ reduction, or mediated by a carrier that is neither ferredoxin 1 nor NADH in wild type cells.

3.4. Glucose Utilization in the Light

If the hydrogenase would directly couple to the NDH-1 complex as suggested by the above measurements, its diaphorase should also funnel electrons into the intersystem chain. This is supported by our dark respiration measurements (Figure 2).

Here we made measurements of oxygen consumption in the light, without and after the addition of glucose, and of the electron flow through PSI. As shown in Figure 5, wild type cells and a number of deletion strains, including Δ*ox* without respiratory terminal oxidases; Δ*ndh-2* without the type 2 dehydrogenases; Δ*ndhD1*, without NDH-1$_1$; Δ*ndhD2* without NDH-1$_2$; Δ*ndhD1*Δ*ndhD2* without NDH-1$_1$ and NDH-1$_2$; Δ*ndhD3*Δ*ndhD4* without NDH-1$_3$ and NDH-1$_4$; Δ*hoxEFUYH* without any of the hydrogenase subunits; and Δ*hoxH* without the large hydrogenase subunit, were measured at different light intensities. In the experiment the cells were exposed to a specific light intensity for 1 min to reach a steady state. In the approximately 2 min following, changes in the rate of oxygen turnover were recorded and the electron transfer through PSI was determined by DIRK measurements. Subsequently, 10 mM glucose was added and after another 1.5 min under continuous light the oxygen and DIRK measurements were repeated.

After the addition of glucose, wild type cells decrease their oxygen evolution. Upon the addition of glucose the cells receive a surplus of carbon skeletons and reducing equivalents. To balance the consumption and further conversion of these metabolites the cells have three possibilities. They could either (1) increase respiration, (2) turndown the activity of PSII, (3) increase the consumption of low potential electrons by other acceptors, or undergo a combination of these options. In all three cases oxygen evolution should decrease in the presence of glucose. As is visible in Figure 5 increased respiration (option 1) cannot be the reason since the mutant of the respiratory terminal oxidases (Δ*ox*) and the one without the type 2 dehydrogenases (Δ*ndh-2*) show the same suppression of oxygen evolution as the wild type cells.

Concerning the bidirectional hydrogenase, it is striking that none of the hox mutants (Δ*hoxEFUYH* and Δ*hoxH*) show a decreased oxygen evolution. The most straight forward explanation is that under these conditions the diaphorase delivers electrons from glucose breakdown into the intersystem chain in wild type cells but not in the *hox* deletion strains. These electrons should be at the level of NADH or NADPH. A reduction of ferredoxin 1 (by HoxEFUYH) and its subsequent oxidation by the NDH-1 complex seems to be highly unlikely because of the more positive redox potential of the pyridine dinucleotides. Thus, a direct coupling of the diaphorase to the NDH-1 is much more probable.

All the mutants of NdhD subunits tested here also do not show the suppression of oxygen evolution, supporting the idea that the NDH-1 complex is involved in the oxidation of the additional reducing equivalents. The Δ*ndhD3*Δ*ndhD4* double mutant shows an increased oxygen consumption in the dark and at low light intensities. Therefore, the NDH-1$_1$ and NDH-1$_2$ complexes seem to deliver electrons into the intersystem chain especially in the dark and at low light, whereas the NDH-1$_3$ and NDH-1$_4$ complexes take over at higher light.

DIRK measurements (Figure S8) indicate that in all the strains showing the suppression of oxygen evolution (WT, Δ*ox*, Δ*ndbA*Δ*ndbC*) the electron transfer through PSI is slightly increased while in all other strains it is either the same or decreases.

If the breakdown of glucose results in a higher number of electrons introduced into the intersystem chain via the NDH-1 complex, we can expect that the plastoquinone pool might become more strongly reduced. Since the redox state of the PQ pool controls excitation energy distribution between the two photosystems [36] the surplus of electrons might induce state change. More specifically, we would expect that excitation energy/antenna are shifted away from PSII to decrease its activity (option 2).
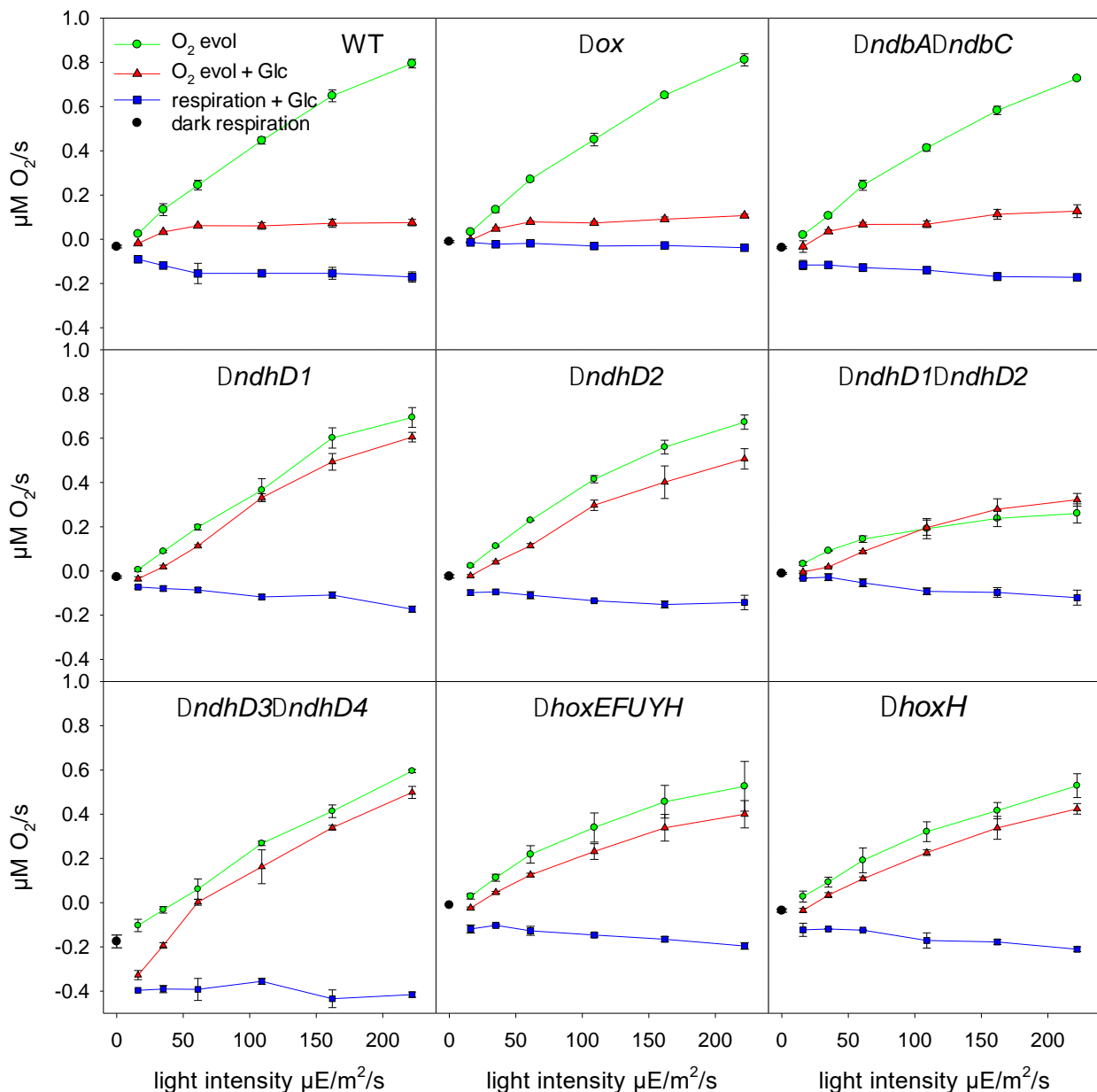
Figure 5. Oxygen evolution without (green lines and circles) and after glucose addition (red lines and triangles) and dark respiration (black circle) and respiration after glucose addition (blue lines and blue squares). The rates are given for a cell density of OD_{750} = 5.7. In all cases at least three different cultures of the same strain were measured and their standard deviation is indicated.

3.5. State Change Due to Glucose Addition

Parallel PAM-fluorescence measurements were performed for all the strains without and with glucose to monitor their putative state changes. By giving saturation pulses of high light, the excitation of PSII before and after glucose was measured. As visible in Figure 6, all the strains that show the increased oxygen consumption after the addition of glucose also show a state change and remove antenna from PSII (the F_m without glucose is higher), while all those that do not show this effect have a reduced state change or even have a negative value ($\Delta ndhD1\Delta ndhD2$) that indicates that more excitation energy is funneled into PSII.
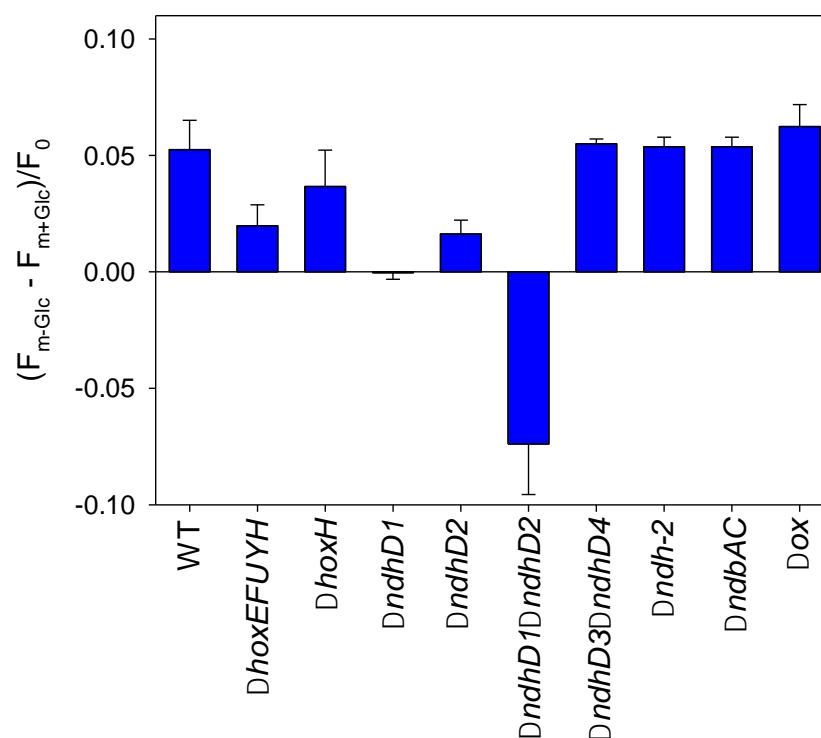
Figure 6. State change after glucose addition as measured by the PAM-fluorimeter of the different deletion strains. The cells were exposed to 120 μE/m^2/s at the same density as in Figure 4 for about 3 min and the Fm measured (F_{m-Glc}). After glucose addition the cells were continuously exposed to the same light and after 3 min another Fm was measured (F_{m+Glc}). The difference between both F_m should mirror the shift of antenna away from PSII because of the presence of glucose. At least three different cultures of the same strain were measured. The standard deviation is shown.

Taken together, these results indicate that glucose addition results in a stronger reduction of plastoquinone and suggest that wild type cells decrease the amount of excitation energy that is directed to PSII. This will result in a lowered oxygen evolution. Even the mutant without any of the respiratory terminal oxidases shows the same state change behavior, showing that the oxidases are not important in the withdrawal of surplus electrons as previously shown in Figure 5. However, since net oxygen evolution is strongly diminished in this strain upon glucose addition, it is likely that other players such as the flavodiiron proteins are also involved and increase oxygen consumption through a Mehler-like reaction (option 3).

The negative state change seen in the $\Delta ndhD1\Delta ndhD2$ is probably due to the higher amount of reducing equivalents generated by glucose breakdown that must be balanced by a higher ATP production. Since this strain is only able to increase ATP synthesis by increasing its linear electron transport driving the water–water cycle [28,37] it must shift a higher proportion of excitation energy to PSII.

Since state change is reduced especially in the $\Delta hoxEFUYH$ mutant this again indicates that electron flow into the PQ pool is impaired especially in the hox mutant without the diaphorase.

3.6. Effect of Glucose Addition on Electron Transport When PSII Is Inhibited by DCMU

Another possibility to measure the effect of glucose addition is to determine the electron flow in the absence and presence of glucose through PSI when PSII is blocked by DCMU. Under these conditions the PQ pool is strongly oxidized. This is obvious when comparing the oxidation level of PSI in the absence and presence of DCMU (Figure S8). In Figure 7 the results for all the deletion strains investigated here, including a strain lacking the succinate dehydrogenase ($\Delta sdhB1\Delta sdhB2$), are shown.
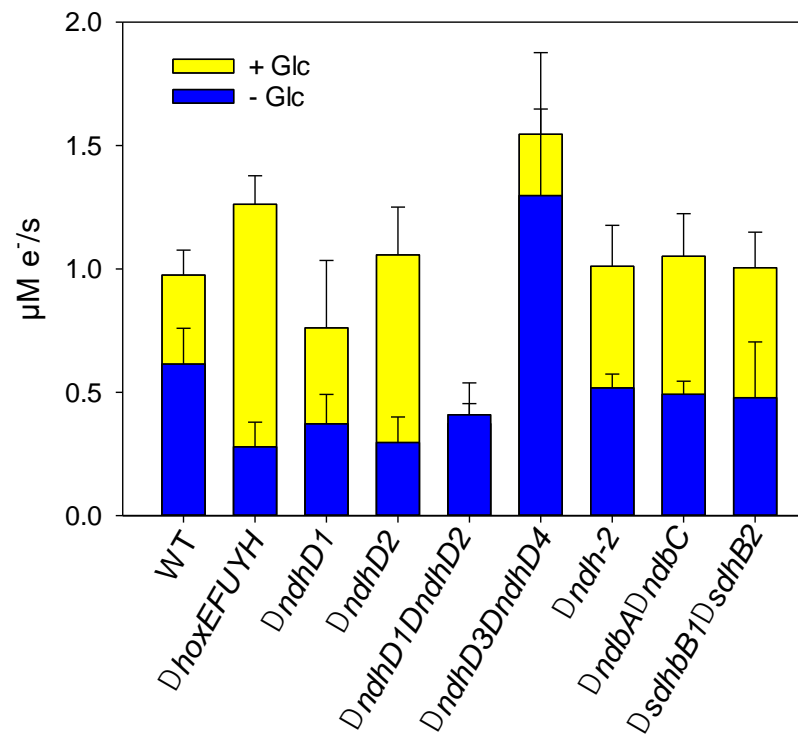
Figure 7. Electron flow with and without glucose in the presence of DCMU for the different deletion strains. Cell suspensions were diluted to $OD_{750} = 5.7$ and were illuminated with 222 $\mu E/m^2/s$ for 3 min before DIRK measurements were started. After the first round of these measurements 10 mM glucose was added and after 1.5 min another set of DIRK measurements was started. The measurements were conducted at least in triplicates. Their standard deviation is indicated.

Previously we already showed that the $\Delta ndhD1\Delta ndhD2$ mutant is unable to increase its electron flow upon glucose addition [23]. From Figure 7 it is obvious that all other deletion strains are able to increase their electron flow through PSI. Thus, specifically the complexes NDH-1_1 and NDH-1_2 are needed to re-oxidize the reducing equivalents and reduce PQ. It is not sufficient to delete one of these two complexes to block the electron flow triggered by glucose addition. This is different to normal photosynthetic conditions (Figure 5) where $\Delta ndhD1$ and $\Delta ndhD2$ already show reduced oxygen evolution which indicates that both are required to introduce the electrons from glucose breakdown.

Because the PQ pool is strongly oxidized under these conditions and the electron flow through PSI is much smaller than when PSII is active [28], ferredoxin 1 could be an electron donor to the NDH-1 complexes. This would also explain why the $\Delta hoxEFUYH$ mutant is able to increase its electron flow after the addition of glucose since the FNR could replace the diaphorase and work in reverse, as previously discussed concerning the respiration measurements.

3.7. Colocalization of HoxF and NdhM

To determine the colocalization of the bidirectional hydrogenase and NDH-1 in vivo, we created C terminal gene fusions between HoxF and Green Fluorescent Protein (GFP), and NdhM and Yellow Fluorescent Protein (YFP), with the HoxF protein being expressed at the native chromosomal *hoxF* locus and NdhM being expressed at the native *ndhM* locus. As previously shown [21], HoxF–GFP was still fully functional being incorporated into fully assembled complexes and H_2 production was still comparable to the wild type strain. NdhM–YFP was also fully functional as demonstrated [22]. Confocal laser scanning microscopy (CLSM) was utilized to determine the localization of both HoxF–GFP and NdhM–YFP relative to the thylakoid membrane, identified from chlorophyll fluorescence. The resolution of CLSM is not sufficient to localize complexes to the nm scale and to judge

if both are in direct vicinity. However, it is possible to acquire a measure of the cellular distribution of proteins/complexes and subsequently calculate the match of two different populations of complexes. To this end, multiple images were acquired per condition and the post bleach image was subtracted from the pre bleach image; images were analyzed in Image J utilizing the colocalization function to calculate the Manders coefficient. This value gives a measure of the congruence of the two distributions [38,39].

Under photoautotrophic growth at 30 μmol photon $m^{-2} s^{-1}$, Hox colocalized to the thylakoid membrane as identified from the chlorophyll fluorescence (Figure S9A–L), with a Manders coefficient (M) of 0.7107 (Figure 8). Hox was present in two populations; a dispersed population, and puncta, correlating with the distribution already seen [21] (Figure S9D). Following the addition of glucose, glucose oxidase and catalase to induce anoxia, Hox showed a higher degree of colocalization with the photosynthetic pigments (M = 0.9989) (Figure S9J), indicating a change in the distribution of Hox. The increase in Hox and photosynthetic pigment colocalization under anoxia compared to photoautotrophic growth was significantly enhanced ($p = 0.09$). NDH-1 was shown to be colocalized with the photosynthetic pigments under photoautotrophic growth (M = 0.7099) (Figure S9E). Following anoxia, the colocalization of NDH-1 and chlorophyll increased (M = 0.9979), which was similar to the redistribution seen in Hox (Figure S9K). NDH-1 colocalization with the photosynthetic pigments was significantly increase upon anoxia ($p = 0.018$). Both Hox and NDH-1 were colocalized at the thylakoid membrane (M = 0.7782) under photoautotrophic growth (Figure S9F). This association increased following anoxia (M = 0.9212), however both Hox and NDH-1 still maintained two distinct sub populations (Figure S9L). Overall, anoxia significantly increased the colocalization of Hox and NDH-1 ($p = 0.002$).
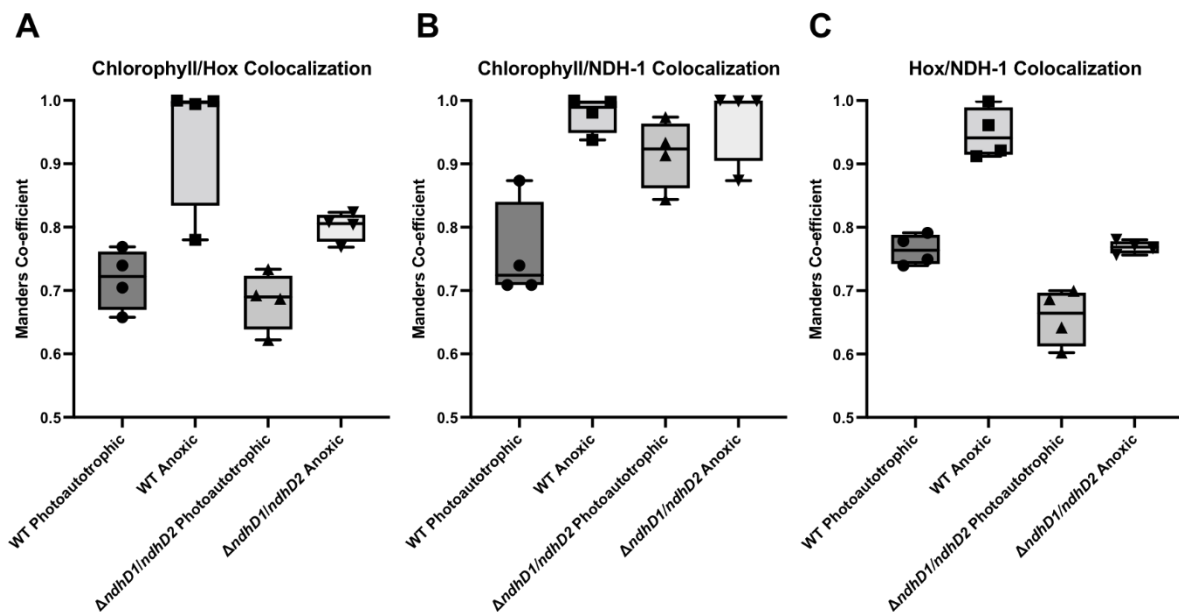
Figure 8. Spatial analysis of Hox, NDH-1 and chlorophyll distribution in cells grown photoautotrophically at 30 μmole photons $m^{-2} s^{-1}$ and following anoxia. The distribution of Hox relative to photosynthetic pigment (**A**), The distribution of NDH-1 relative to photosynthetic pigment (**B**), and the distribution of Hox relative to NDH-1 (**C**). Box plot displays the minimum and maximum Manders coefficient values with individual data points. Data represent four biological replicates: WT photoautotrophic $n = 898$, WT anoxic $n = 1010$, Δ*ndhD1*Δ*ndhD2* photoautotrophic $n = 912$, Δ*ndhD1*Δ*ndhD2* anoxic $n = 750$.

The significantly lower level of colocalization between Hox and NDH-1 upon anoxia in Δ*ndhD1*Δ*ndhD2* when compared to WT ($p = 0.0001$) suggests the loss of a functional interaction between NDH-1_1 and NDH-1_2 and the hydrogenase.

4. Discussion

Hydrogenases are extremely versatile redox enzymes mediating electron transfer between a large number of different reaction partners. The cyanobacterial bidirectional hydrogenase is an excellent example since it mediates electron transfer between hydrogen, ferredoxin/flavodoxin and $NAD(P)^+$ [11,14–16]. It has long been known that this enzyme is also responsible for H_2 uptake under anaerobic conditions when PSII is blocked/not working or at low light intensities [13,40,41]. In principle the electrons from H_2 oxidation could follow three different routes. The first could involve the hydrogenase reducing ferredoxin, which would be re-oxidized by the NDH-1 complex; in the second, it could reduce NAD^+ that in turn should be re-oxidized by the type 2 dehydrogenases since these are the only known to oxidize NADH in *Synechocystis* [35]; and in the third either electrons are directly delivered to the NDH-1 complex by the hydrogenase or they are passed on to the complex by another unknown mediator.

Our experiments indicate that H_2 oxidation is closely linked to functional NDH-1_1 and NDH-1_2 complexes (Figure 4). The absence of these two complexes results in hydrogen uptake with the same apparent K_M but a significantly lower V_{max}. Since the $\Delta ndhD1 \Delta ndhD2$ has a similar amount of hydrogenase (Figure S4) this indicates that it uses a different electron acceptor compared to wild type cells and all the other strains such as the $\Delta ndhD1$ and $\Delta ndhD2$ single deletion mutants, the $\Delta ndhD3 \Delta ndhD4$, and the strain without the type 2 dehydrogenase (Δndh-2). Calculations of the ATP/NADPH ratio produced by the different strains according to their electron flow (Figure 4 and Table 1) indicates that all of them except the $\Delta ndhD1 \Delta ndhD2$ double mutant must use their NDH-1 complex during H_2 uptake. If they do not use this complex, the ratio will not fit the requirements of the CBB cycle.

The fact that hydrogen concentration drops to zero during H_2 uptake excludes ferredoxin 1 as electron mediator since its redox potential under physiological conditions would not allow complete consumption of hydrogen (Figure S7). Since in the $\Delta ndhD1 \Delta ndhD2$ electron transfer does not involve the NDH-1 complex, this strain must use an alternative route. Most plausibly the hydrogenase reduces NAD^+ in this strain and the formed NADH in turn is oxidize by the type 2 dehydrogenases. The latter are known to have a low activity and were originally assigned a purely regulatory role [42]. However, recent investigations on a *Synechocystis* strain that expresses the soluble hydrogenase of a Knallgas bacterium, which is known to exclusively reduce NAD^+ [43] already suggested that the cells must be able to oxidize NADH during H_2 oxidation [44].

In summary, our H_2 oxidation experiments indicate that plastoquinone receives the electrons via the NDH-1 complex in all the strains tested but in the $\Delta ndhD1 \Delta ndhD2$, NAD^+ is used instead.

We found that the addition of glucose to wild type cells performing normal photosynthesis results in diminished oxygen evolution (Figure 5). Under these photomixotrophic conditions, glucose breakdown results in a higher number of electrons entering the intersystem chain. On the one hand this is seen by the slightly increased number of electrons passing PSI (Figure S8) and on the other by the excitation energy redistribution away from PSII (Figure 5). Whether the lowered oxygen evolution is only due to decreased PSII activity or results from a combination of reduced water oxidation and increased oxygen consumption remains to be seen. Since the mutant lacking all respiratory terminal oxidases (Figure 5) shows normal wild type behavior, respiratory oxygen uptake is not involved. Thus, an increased water–water cycle would be the only other alternative for diminished net oxygen evolution.

It is not surprising that the mutant without the two NDH-1 complexes ($\Delta ndhD1 \Delta ndhD2$) known to be part of respiratory electron transport [19] does not reduce oxygen evolution upon glucose addition (Figure 5). Since the entry of reducing equivalents into the electron transfer is impaired in this strain (also see the electron flow through PSI, Figure S8) the PQ pool would not become more reduced and would not force the strain to dimmish its PSII activity or to channel surplus electrons into oxygen reduction. However, it is surprising that

the two single mutants Δ*ndhD1* and Δ*ndhD2* already show the same effect while glucose addition on the state change in the Δ*ndhD1* is stronger compared to the Δ*ndhD2*. On the other hand, both single mutants do increase their electron flow through PSI after glucose addition when PSII is blocked (Figure 7). The Δ*ndhD3*Δ*ndhD4* increases oxygen consumption at low light intensities but not at higher light (Figure 5). It performs state change although it does not increase electron flow significantly after glucose addition in the presence of DCMU (Figure 7). Taken together, these results indicate that there is still a lot to learn about the functioning of the NDH-1 complexes and that there is some unknown cooperation between NDH-1$_1$ and NDH-1$_2$ causing a kind of interdependence between both and also between the complexes NDH-1$_3$ and NDH-1$_4$ and the two aforementioned complexes.

Most unexpectedly the *hox* mutants also do not decrease oxygen evolution after glucose addition (Figure 5), they do not increase electron flow (Figure S8), and they do not redistribute the excitation energy as strongly as wild type cells (Figure 6). This indicates that electron transfer into the intersystem chain is impaired in these strains. Glucose breakdown yields reducing equivalents predominately at the level of the pyridine nucleotides (NAD(P)H). Under photomixotrophic conditions, glucose is fed mainly via the PGI shunt into the CCB cycle. This shunt does not yield any reducing equivalents. However, some glucose is channeled into the OPP and yields two NADPH for each molecule [45]. These investigations have been conducted under steady state conditions. Although our measurements depict the reaction of the cells upon sudden glucose addition, suppression of oxygen evolution also indicates that more reducing equivalents become available. Obviously, the cells need to oxidize the surplus NAD(P)H formed and the strains without a functional bidirectional hydrogenase (Δ*hoxEFUYH* and Δ*hoxH*) are unable to. In this case it must be the diaphorase module that is responsible for this activity, and could deliver electrons from NADH oxidation into the electron transport chain, but also from NADPH, albeit with a lower affinity.

The fact that the Δ*hoxH* strain also behaves similarly to the mutant without any of the *hox* genes (Figures 5 and 6) during dark respiration (Figure 2) indicates that the stability or assembly of the complex might be impaired in the absence of the hydrogenase large subunit. Especially in this case the redox potential of NAD(P)H does not allow for the reduction of ferredoxin and thus delivery from the hydrogenase to the NDH-1 complex must either be direct or conducted by a different mediator with a more positive redox potential (Figure 9).
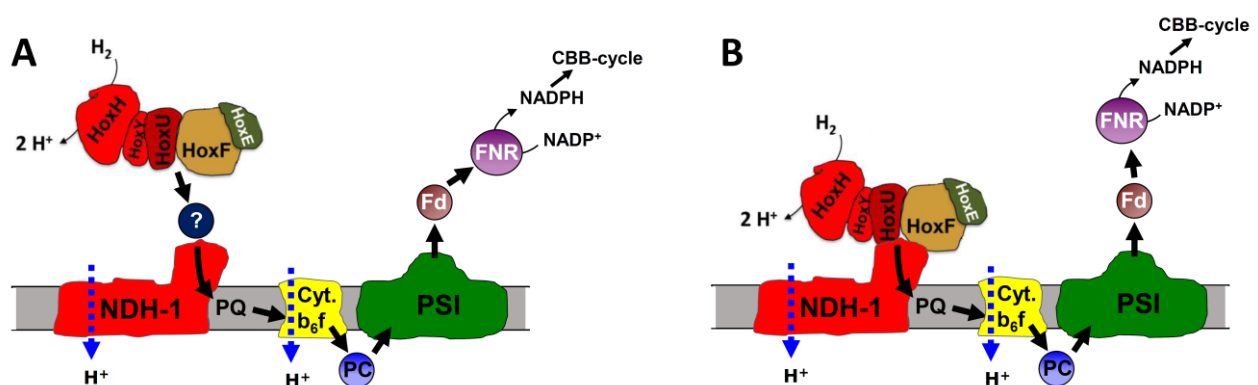
Figure 9. Scheme of the two possible scenarios suggested by our results. (**A**) The bidirectional hydrogenase might either reduce an unknown mediator that needs to have a lower redox potential as ferredoxin 1 or (**B**) it directly binds to the NDH-1 complex and H$_2$ oxidation is directly linked to PQ reduction.

We interpret a very recent study that found a higher NADPH level and higher contents of glycogen and polyhydroxybutyrate (PHB) in a Δ*hoxYH* deletion strain [46] in the same way as stated above. The interaction of the bidirectional hydrogenase including its diaphorase is probably impaired in this strain, resulting in an inhibition of the respiratory

breakdown of all these metabolites. Under the conditions used in this study we never found measurable H_2 levels thus excluding the idea that the hydrogenase is used to drain low potential electrons to H_2 as stated by the authors.

In a previous investigation, it was shown that a population of the bidirectional hydrogenase is localized to the membrane while others are soluble inside the cytoplasm [21]. Membrane binding was attributed to the HoxE subunit although it is a soluble protein. Our colocalization studies suggest that attachment to the membrane is not direct but due to the interaction of the hydrogenase with the NDH-1 complex (Figure 8). The loss of HoxE might perturb this interaction, as the loss of HoxH also seems to impair the functioning of the whole enzyme (result in Figures 2, 5 and 6).

It is highly attractive to hypothesize that attachment of the hydrogenase to the NDH-1 complex is controlled by the redox conditions inside the cell and thereby the reactions it catalyzes. The changes seen in the colocalization of NDH-1 and photosynthetic pigments (Figure 9B) are interesting in this respect as well, as they might indicate that the complex might migrate closer to the photosystems in one condition and in others move in membrane areas with less photosystems. In case of the *ndhD1ΔndhD2* one could hypothesize that the NDH-1_3 and NDH-1_4 complex are usually closer to the photosystems (Figure 9B) since they are involved in the CCM and needed during photosynthesis while the NDH-1_1 and NDH-1_2 complexes are also part of respiration and would not all be required close to the photosynthetic electron transport.

The scenario of NAD(P)H oxidation by the diaphorase of the hydrogenase as described above holds when PSII is active and also under dark conditions. In contrast to this, the inhibition of PSII in the light (Figure 7) or glucose addition in the dark (Figure S2) results in an increase in electron flow and increased respiration to wild type levels also in the *hox* mutants. We attribute this effect to an FNR working in the inverse direction since either the PQ pool is strongly oxidized in the light, the reductive power caused by rising NAD(P)H levels should increase, or both should occur. This effect could open an alternative pathway that would circumvent the diaphorase of the hydrogenase.

The photosynthetic complex I is part of a zoo of different transmembrane complexes that share a similar architecture in their membrane module (sometimes termed P-module). This module is used for proton translocation, which is driven by different redox reactions. It seems plausible that these complexes developed by increasing the redox potential difference these reactions span and thereby gradually increased the number of protons they can translocate [7,8]. One of the most primitive of these complexes might have oxidized reduced ferredoxin and produced H_2 [47]. A step further in evolution reduced ferredoxin might have been oxidized using polysulfide (e.g., in *Pyrococcus* species) or F_{420} was oxidized by methanophenazine (as in some methanogens). When oxygen became available quinones could be reduced by NADH and the largest potential difference could be used by oxidizing ferredoxin by plastoquinone as it is performed by the photosynthetic complex I [7].

Hydrogen is an attractive electron donor and has about the same redox potential as ferredoxin 1. It seems consequential to harness its reducing power to drive plastoquinone reduction through a complex consisting of the photosynthetic complex I and the bidirectional hydrogenase. In addition, this complex, in combination with the hydrogenase, might have played an important role during the evolution of photosynthesis as an intermediary step for cells with a type I reaction center that used H_2 as electron donor [48]. Today this is still performed by green sulfur bacteria, acidobacteria and heliobacteria [49]. Our results suggest that cyanobacteria are able to introduce H_2 into the photosynthetic electron transport via a NDH-1 hydrogenase complex.

References

1. Sazanov, L.A. A giant molecular proton pump: Structure and mechanism of respiratory complex I. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 375–388. [CrossRef] [PubMed]
2. Schuller, J.M.; Birrell, J.A.; Tanaka, H.; Konuma, T.; Wulflhorst, H.; Cox, N.; Schuller, S.K.; Thiemann, J.; Lubitz, W.; Setif, P.; et al. Structural adaptations of photosynthetic complex I enable ferredoxin-dependent electron transfer. *Science* **2019**, *363*, 257. [CrossRef] [PubMed]
3. Shikanai, T. Chloroplast NDH: A different enzyme with a structure similar to that of respiratory NADH dehydrogenase. *Biochim. Biophys. Acta-Bioenerg.* **2016**, *1857*, 1015–1022. [CrossRef] [PubMed]
4. Zhang, C.; Shuai, J.; Ran, Z.; Zhao, J.; Wu, Z.; Liao, R.; Wu, J.; Ma, W.; Lei, M. Structural insights into NDH-1 mediated cyclic electron transfer. *Nat. Commun.* **2020**, *11*, 888. [CrossRef]
5. Appel, J.; Schulz, R. Sequence analysis of an operon of a NAD(P)-reducing nickel hydrogenase from the cyanobacterium Synechocystis sp PCC 6803 gives additional evidence for direct coupling of the enzyme to NAD(P)H-dehydrogenase (complex I). *Biochim. Biophys. Acta-Protein Struct. Mol. Enzymol.* **1996**, *1298*, 141–147. [CrossRef]
6. Schmitz, O.; Bothe, H. The diaphorase subunit HoxU of the bidirectional hydrogenase as electron transferring protein in cyanobacterial respiration? *Naturwissenschaften* **1996**, *83*, 525–527. [CrossRef] [PubMed]
7. Schut, G.J.; Zadvornyy, O.; Wu, C.H.; Peters, J.W.; Boyd, E.S.; Adams, M.W. The role of geochemistry and energetics in the evolution of modern respiratory complexes from a proton-reducing ancestor. *Biochim. Biophys. Acta* **2016**, *1857*, 958–970. [CrossRef]

8. Yu, H.; Schut, G.J.; Haja, D.K.; Adams, M.W.W.; Li, H. Evolution of complex I–like respiratory complexes. *J. Biol. Chem.* **2021**, *296*, 100740. [CrossRef] [PubMed]

9. Greening, C.; Biswas, A.; Carere, C.R.; Jackson, C.J.; Taylor, M.C.; Stott, M.B.; Cook, G.M.; Morales, S.E. Genomic and metagenomic surveys of hydrogenase distribution indicate H2 is a widely utilised energy source for microbial growth and survival. *ISME J.* **2016**, *10*, 761–777. [CrossRef]

10. Greening, C.; Grinter, R. Microbial oxidation of atmospheric trace gases. *Nat. Rev. Microbiol.* **2022**. [CrossRef]

11. Cournac, L.; Guedeney, G.; Peltier, G.; Vignais, P.M. Sustained photoevolution of molecular hydrogen in a mutant of *Synechocystis* sp. strain PCC 6803 deficient in the type I NADPH-dehydrogenase complex. *J. Bacteriol.* **2004**, *186*, 1737–1746. [CrossRef]

12. Gutekunst, K.; Schulz, R. CHAPTER 4 The Physiology of the Bidirectional NiFe-hydrogenase in Cyanobacteria and the Role of Hydrogen Throughout the Evolution of Life. In *Microalgal Hydrogen Production: Achievements and Perspectives*; The Royal Society of Chemistry: London, UK, 2018; pp. 107–138.

13. Gutthann, F.; Egert, M.; Marques, A.; Appel, J. Inhibition of respiration and nitrate assimilation enhances photohydrogen evolution under low oxygen concentrations in Syneehocystis sp PCC 6803. *Biochim. Biophys. Acta-Bioenerg.* **2007**, *1767*, 161–169. [CrossRef]

14. Gutekunst, K.; Chen, X.; Schreiber, K.; Kaspar, U.; Makam, S.; Appel, J. The bidirectional NiFe-hydrogenase in *Synechocystis* sp. PCC 6803 is reduced by flavodoxin and ferredoxin and is essential under mixotrophic, nitrate-limiting conditions. *J. Biol. Chem.* **2014**, *289*, 1930–1937. [CrossRef]

15. Artz, J.H.; Tokmina-Lukaszewska, M.; Mulder, D.W.; Lubner, C.E.; Gutekunst, K.; Appel, J.; Bothner, B.; Boehm, M.; King, P.W. The structure and reactivity of the HoxEFU complex from the cyanobacterium *Synechocystis* sp. PCC 6803. *J. Biol. Chem.* **2020**, *295*, 9445–9454. [CrossRef]

16. Aubert-Jousset, E.; Cano, M.; Guedeney, G.; Richaud, P.; Cournac, L. Role of HoxE subunit in Synechocystis PCC6803 hydrogenase. *FEBS J.* **2011**, *278*, 4035–4043. [CrossRef]

17. Appel, J.; Phunpruch, S.; Steinmuller, K.; Schulz, R. The bidirectional hydrogenase of *Synechocystis* sp PCC 6803 works as an electron valve during photosynthesis. *Arch. Microbiol.* **2000**, *173*, 333–338. [CrossRef]

18. Burgstaller, H.; Wang, Y.; Caliebe, J.; Hueren, V.; Appel, J.; Boehm, M.; Leitzke, S.; Theune, M.; King, P.W.; Gutekunst, K. *Synechocystis* sp. PCC 6803 Requires the Bidirectional Hydrogenase to Metabolize Glucose and Arginine Under Oxic Conditions. *Front. Microbiol.* **2022**, *13*, 896190. [CrossRef]

19. Ohkawa, H.; Pakrasi, H.B.; Ogawa, T. Two types of functionally distinct NAD(P)H dehydrogenases in *Synechocystis* sp. strain PCC6803. *J. Biol. Chem.* **2000**, *275*, 31630–31634. [CrossRef]

20. Shibata, M.; Ohkawa, H.; Kaneko, T.; Fukuzawa, H.; Tabata, S.; Kaplan, A.; Ogawa, T. Distinct constitutive and low-CO$_2$-induced CO$_2$ uptake systems in cyanobacteria: Genes involved and their phylogenetic relationship with homologous genes in other organisms. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 11789–11794. [CrossRef]

21. Burroughs, N.J.; Boehm, M.; Eckert, C.; Mastroianni, G.; Spence, E.M.; Yu, J.; Nixon, P.J.; Appel, J.; Mullineaux, C.W.; Bryan, S.J. Solar powered biohydrogen production requires specific localization of the hydrogenase. *Energy Environ. Sci.* **2014**, *7*, 3791–3800. [CrossRef]

22. Liu, L.-N.; Bryan, S.J.; Huang, F.; Yu, J.; Nixon, P.J.; Rich, P.R.; Mullineaux, C.W. Control of electron transport routes through redox-regulated redistribution of respiratory complexes. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 11431–11436. [CrossRef] [PubMed]

23. Wang, Y.; Chen, X.; Spengler, K.; Terberger, K.; Boehm, M.; Appel, J.; Barske, T.; Timm, S.; Battchikova, N.; Hagemann, M.; et al. Pyruvate:ferredoxin oxidoreductase and low abundant ferredoxins support aerobic photomixotrophic growth in cyanobacteria. *eLife* **2022**, *11*, e71339. [CrossRef] [PubMed]

24. Makowka, A.; Nichelmann, L.; Schulze, D.; Spengler, K.; Wittmann, C.; Forchhammer, K.; Gutekunst, K. Glycolytic Shunts Replenish the Calvin-Benson-Bassham Cycle as Anaplerotic Reactions in Cyanobacteria. *Mol. Plant* **2020**, *13*, 471–482. [CrossRef]

25. Gibson, D.G.; Young, L.; Chuang, R.Y.; Venter, J.C.; Hutchison, C.A.; Smith, H.O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **2009**, *6*, 343-U341. [CrossRef]

26. Williams, J.G.K. [85] Construction of specific mutations in photosystem II photosynthetic reaction center by genetic engineering methods in Synechocystis 6803. In *Methods Enzymol*; Academic Press: Cambridge, MA, USA, 1988; Volume 167, pp. 766–778.

27. Fatihi, A.; Latimer, S.; Schmollinger, S.; Block, A.; Dussault, P.H.; Vermaas, W.F.; Merchant, S.S.; Basset, G.J. A Dedicated Type II NADPH Dehydrogenase Performs the Penultimate Step in the Biosynthesis of Vitamin K1 in Synechocystis and Arabidopsis. *Plant Cell* **2015**, *27*, 1730–1741. [CrossRef]

28. Theune, M.L.; Hildebrandt, S.; Steffen-Heins, A.; Bilger, W.; Gutekunst, K.; Appel, J. In-vivo quantification of electron flow through photosystem I—Cyclic electron transport makes up about 35% in a cyanobacterium. *Biochim. Biophys. Acta (BBA)-Bioenerg.* **2021**, *1862*, 148353. [CrossRef] [PubMed]

29. Spence, E.; Sarcina, M.; Ray, N.; Moller, S.G.; Mullineaux, C.W.; Robinson, C. Membrane-specific targeting of green fluorescent protein by the Tat pathway in the cyanobacterium Synechocystis PCC6803. *Mol. Microbiol.* **2003**, *48*, 1481–1489. [CrossRef] [PubMed]

30. Mcdonald, J.H.; Dunn, K.W. Statistical tests for measures of colocalization in biological microscopy. *J. Microsc.* **2013**, *252*, 295–302. [CrossRef]

31. Carrillo, N.; Ceccarelli, E.A. Open questions in ferredoxin-NADP+ reductase catalytic mechanism. *Eur. J. Biochem.* **2003**, *270*, 1900–1915. [CrossRef]

32. Kramer, D.M.; Evans, J.R. The Importance of Energy Balance in Improving Photosynthetic Productivity. *Plant Physiol.* **2011**, *155*, 70–78. [CrossRef]

33. Knoop, H.; Steuer, R. A Computational Analysis of Stoichiometric Constraints and Trade-Offs in Cyanobacterial Biofuel Production. *Front. Bioeng. Biotechnol.* **2015**, *3*, 47. [CrossRef] [PubMed]

34. Belkin, S.; Packer, L. Determination of pH gradients in intact cyanobacteria by electron spin resonance spectroscopy. *Methods Enzymol.* **1988**, *167*, 677–685. [CrossRef] [PubMed]

35. Cooley, J.W.; Vermaas, W.F.J. Succinate dehydrogenase and other respiratory pathways in thylakoid membranes of *Synechocystis* sp. strain PCC 6803: Capacity comparisons and physiological function. *J. Bacteriol.* **2001**, *183*, 4251–4258. [CrossRef]

36. Mullineaux, C.W.; Emlyn-Jones, D. State transitions: An example of acclimation to low-light stress. *J. Exp. Bot.* **2004**, *56*, 389–393. [CrossRef] [PubMed]

37. Nikkanen, L.; Santana Sánchez, A.; Ermakova, M.; Rögner, M.; Cournac, L.; Allahverdiyeva, Y. Functional redundancy between flavodiiron proteins and NDH-1 in *Synechocystis* sp. PCC 6803. *Plant J.* **2020**, *11*, 14812. [CrossRef] [PubMed]

38. Dunn, K.W.; Kamocka, M.M.; McDonald, J.H. A practical guide to evaluating colocalization in biological microscopy. *Am. J. Physiol.-Cell Physiol.* **2011**, *300*, C723–C742. [CrossRef]

39. Herce, H.D.; Casas-Delucchi, C.S.; Cardoso, M.C. New image colocalization coefficient for fluorescence microscopy to quantify (bio-)molecular interactions. *J. Microsc.* **2013**, *249*, 184–194. [CrossRef]

40. Frenkel, A.; Gaffron, H.; Battley, E.H. Photosynthesis and photoreduction by the blue green alga, synechococcus-elongatus, nag. *Biol. Bull.* **1950**, *99*, 157–162. [CrossRef]

41. Houchins, J.P. The physiology and biochemistry of hydrogen metabolism in cyanobacteria. *Biochim. Biophys. Acta* **1984**, *768*, 227–255. [CrossRef]

42. Howitt, C.A.; Udall, P.K.; Vermaas, W.F.J. Type 2 NADH dehydrogenases in the cyanobacterium Synechocystis sp strain PCC 6803 are involved in regulation rather than respiration. *J. Bacteriol.* **1999**, *181*, 3994–4003. [CrossRef]

43. Schneider, K.; Schlegel, H.G. Purification and properties of soluble hydrogenase from Alcaligenes eutrophus H 16. *Biochim. Biophys. Acta* **1976**, *452*, 66–80. [CrossRef]

44. Lupacchini, S.; Appel, J.; Stauder, R.; Bolay, P.; Klähn, S.; Lettau, E.; Adrian, L.; Lauterbach, L.; Bühler, B.; Schmid, A.; et al. Rewiring cyanobacterial photosynthesis by the implementation of an oxygen-tolerant hydrogenase. *Metab. Eng.* **2021**, *68*, 199–209. [CrossRef] [PubMed]

45. Schulze, D.; Kohlstedt, M.; Becker, J.; Cahoreau, E.; Peyriga, L.; Makowka, A.; Hildebrandt, S.; Gutekunst, K.; Portais, J.-C.; Wittmann, C. GC/MS-based 13C metabolic flux analysis resolves the parallel and cyclic photomixotrophic metabolism of *Synechocystis* sp. PCC 6803 and selected deletion mutants including the Entner-Doudoroff and phosphoketolase pathways. *Microb. Cell Factories* **2022**, *21*, 69. [CrossRef] [PubMed]

46. Sukkasam, N.; Incharoensakdi, A.; Monshupanee, T. Disruption of Hydrogen Gas Synthesis Enhances the Cellular Levels of NAD(P)H, Glycogen, Poly(3-hydroxybutyrate) and Photosynthetic Pigments Under Specific Nutrient Condition(s) in Cyanobacterium Synechocystis sp. PCC 6803. *Plant Cell Physiol.* **2021**, *63*, 135–147. [CrossRef]

47. Hedderich, R.; Forzi, L. Energy-converting NiFe hydrogenases: More than just H-2 activation. *J. Mol. Microbiol. Biotechnol.* **2005**, *10*, 92–104. [CrossRef]

48. Gutekunst, K. Hypothesis on the Synchronistic Evolution of Autotrophy and Heterotrophy. *Trends Biochem. Sci.* **2018**, *43*, 402–411. [CrossRef]

49. Hohmann-Marriott, M.F.; Blankenship, R.E. Evolution of Photosynthesis. *Annu. Rev. Plant Biol.* **2011**, *62*, 515–548. [CrossRef]

Zam Is a Redox-Regulated Member of the RNB-Family Required for Optimal Photosynthesis in Cyanobacteria

Patrick E. Thomas [1,2], Colin Gates [2,†] , William Campodonico-Burnett [1,2] and Jeffrey C. Cameron [1,2,3,*]

1   Department of Biochemistry, University of Colorado, Boulder, CO 80309, USA;
     patrick.thomas@colorado.edu (P.E.T.); will.campo@colorado.edu (W.C.-B.)
2   Renewable and Sustainable Energy Institute, University of Colorado, Boulder, CO 80309, USA;
     cgates4@luc.edu
3   National Renewable Energy Laboratory, Golden, CO 80401, USA
*   Correspondence: jeffrey.c.cameron@colorado.edu
†   Current Address: Department of Chemistry and Biochemistry, Loyola University Chicago,
     Chicago, IL 60660, USA.

Abstract: The *zam* gene mediating resistance to acetazolamide in cyanobacteria was discovered thirty years ago during a drug tolerance screen. We use phylogenetics to show that Zam proteins are distributed across cyanobacteria and that they form their own unique clade of the ribonuclease II/R (RNB) family. Despite being RNB family members, multiple sequence alignments reveal that Zam proteins lack conservation and exhibit extreme degeneracy in the canonical active site—raising questions about their cellular function(s). Several known phenotypes arise from the deletion of *zam*, including drug resistance, slower growth, and altered pigmentation. Using room-temperature and low-temperature fluorescence and absorption spectroscopy, we show that deletion of *zam* results in decreased phycocyanin synthesis rates, altered PSI:PSII ratios, and an increase in coupling between the phycobilisome and PSII. Conserved cysteines within Zam are identified and assayed for function using in vitro and in vivo methods. We show that these cysteines are essential for Zam function, with mutation of either residue to serine causing phenotypes identical to the deletion of Zam. Redox regulation of Zam activity based on the reversible oxidation-reduction of a disulfide bond involving these cysteine residues could provide a mechanism to integrate the 'central dogma' with photosynthesis in cyanobacteria.

Keywords: reduction-oxidation; photosynthesis; cyanobacteria; photosystem II; phycobilisome; RNase; light-harvesting

1. Introduction

Cyanobacteria are photosynthetic prokaryotes that play major roles in the biosphere—serving as the major fixers of nitrogen in the ocean and as primary producers [1]. As their ancestors gave rise to modern chloroplasts, cyanobacteria serve as model systems for both carbon fixation and oxygenic photosynthesis. Cyanobacteria are also considered a promising chassis for metabolic engineering due to their diverse metabolisms [2]. While decades of research have elucidated processes by which cyanobacteria regulate gene expression and metabolism, the mechanisms underlying many regulatory processes remain unknown. However, there is growing evidence that in photosynthetic organisms, redox regulation plays a crucial role in linking the cellular energy state with downstream regulatory networks involved in gene expression and RNA turnover.

The resistance to acetazolamide (*zam*) gene was first identified in the cyanobacteria *Synechocystis* sp. PCC 6803 (hereafter PCC 6803) [3]. Either a deletion or an insertion mutation in *zam* confers resistance to lethal concentrations of the carbonic anhydrase inhibitor acetazolamide to PCC 6803 [4]. Zam was also identified as a redox-responsive protein in both PCC 6803 and *Synechococcus* sp. PCC 7002 (hereafter PCC 7002) in two

separate redox-proteomics experiments [5,6]. It was initially presumed that *zam* encoded a transporter which affected the accessibility of acetazolamide to one or more of the cellular carbonic anhydrases; however, multiple sequence alignments show that Zam is actually a member of the RNase II/R family [4,7]. Redox-based control of RNA modulating enzymes is not unprecedented in cyanobacteria. For example, the expression of the cyanobacterial RNA helicase gene, *crhR*, is controlled by the cellular redox state and photosynthetic electron transport chain [8]. The CrhR protein, important for cold-tolerance, interacts with multiple RNA species involved in photosynthesis and the degradosome, suggesting a regulatory feedback loop [9–11].

RNase II/R proteins are, with few exceptions, 3′-5′ exonucleases with multiple roles in RNA degradation and processing [12,13]. These proteins often work in collaboration with other nucleases and have been found to be part of the degradosome, a multi-protein complex for degrading RNA, in diverse bacterial systems [14–16]. The catalytic activity of these proteins comes from the RNase II domain (RNB), the active site of which is essentially invariant across all three domains of life [12,17–19]. The best studied RNB proteins in prokaryotes are the eponymous *E. coli* RNase II (*E. coli* Rnb) and *E. coli* RNase R (*E. coli* Rnr). At the protein level *E. coli* Rnb and *E. coli* Rnr share 60% amino acid sequence identity and have identical domain architecture, with the only unique sequence level feature of the proteins being a long basic region at the C-termini of *E. coli* Rnr that is not found in *E. coli* Rnb [19]. Despite these similar structures, *E. coli* Rnr and Rnb exhibit unique catalytic activities; Rnr is capable of digesting through structured regions of RNA while Rnb is not [19–21]. While Zam proteins are clearly RNase II/R family proteins, whether they are RNase II or RNase R type enzymes is currently unknown.

Due to the similarities between RNase II type and RNase R type enzymes, identification of homologs based on amino acid sequence identity alone is difficult and additional biochemical tests with purified enzymes are often necessary [19,22]. Cyanobacteria encode two RNB proteins on their chromosome, one of which is typically annotated as Zam; some species also encode RNB homologs on endogenous plasmids [7]. Here we use RNB domain protein sequences from multiple cyanobacterial genomes to gain insight into the evolutionary history of Zam and its place in the larger RNase II/R family. Observations based on our sequence analysis led us to multiple hypotheses about the function of Zam, which we tested using various physiological and biophysical assays.

2. Materials and Methods

2.1. Reciprocal Best Hit, Multiple Alignment, and Phylogenetics

Protein sequences of the two RNB-containing proteins from PCC 7002_A0574 (Rnb) and PCC_A1543 (Zam) were downloaded from the National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/, accessed on 28 February 2019). The sequences were used as query sequences to search for homologs within a select set of diverse cyanobacterial genomes using NCBI BLAST [23]. Top hits from each genome were then used as BLAST query sequences against the PCC 7002 genome to identify reciprocal best hits (RBH) for Zam and Rnb for each species. Each RBH was considered a putative ortholog of either Zam and Rnb and are designated as such throughout this work.

Sequences of Zam or Rnb homologs were downloaded using batch ENTREZ on 28 February 2019. Multiple sequence alignment was performed using the Clustal-Omega algorithm through the EMBL server. Sequences of Zam and Rnb homologues were pooled and aligned with the sequences of *E. coli* Rnr, *E. coli* Rnb, and both chloroplastic and nuclear RNB sequences for *C. reinhardtii.* MEGA-X was used to generate a bootstrapped, maximum likelihood, and consensus tree based on 500 replicates.

2.2. Domain Identification

The sequence of each Zam ortholog was input into the EMBL-HMMER protein domain search tool using the default parameters and results were downloaded. Domain maps were created to scale using Adobe Illustrator.

2.3. Homology Mapping and Visualization

Multiple alignments were performed as described above for the Zam ortholog group or the Rnb ortholog group with the sequence of *E. coli* Rnr present in each group. For both alignments, in UCSF-Chimera an alignment was associated with the crystal structure of *E. coli* Rnr (PDB:5XGU) and the percent conservation at every residue was mapped onto the structure [21,24].

2.4. Computation Modeling

Threading was performed using the iterative threading assembly refinement (I-TASSER) server [25]. The sequence of Zam was uploaded using the default parameters which allowed the software to identify the best template crystal structure available in the protein database (PDB). *E. coli* Rnr was identified by the algorithm as the best match, and a structural model based on this was output accordingly.

2.5. Cell Growth and Maintenance

PCC 7002 strains were grown under constant illumination (~150 μmol photons m^{-2} s^{-1}) by cool white fluorescent lamps at 37 °C in AL-41 L4 Environmental Chambers (Percival Scientific, Perry, IA, USA). Cultures were grown either in liquid A+ medium in 125 mL baffled flasks closed with foam stoppers (Jaece Identi-Plug), or on A+ media solidified with Bacto Agar (1%; w/v) [26]. Antibiotics (kanamycin, 100 μg/mL; spectinomycin, 100 μg/mL) were used to grow strains with appropriate resistance cassettes.

2.6. Plasmid and Strain Construction

The plasmid used for *zam* deletion (pJCC254) was supplied by Jeffrey Cameron for use in this work [7]. All other plasmids for gene insertion and protein expression were cloned using Gibson Assembly [27]. Assembled plasmids were made using various backbones and inserts amplified by PCR using Phusion polymerase (Thermo Fisher Scientific, Waltham, MA, USA). After assembly, plasmids were transformed, isolated, and sequenced using standard molecular biology protocols.

Cyanobacterial strains were created by transformation through homologous recombination. Exponentially growing cultures were inoculated with 0.5–2 ng/mL of plasmid containing a desired insert flanked by 500 base pair homology arms. Cells and plasmid were incubated in 1.7 mL microfuge tubes overnight at constant illumination at 37 °C without shaking. After incubation, transformants were selected using appropriate antibiotic either on a pre-warmed A+ agar plate or in liquid A+ shake flasks. After initial selection, potentially positive transformants were streaked on plates, isolated after growth, and verified by colony PCR. Positive colonies were passaged twice on antibiotic containing media and then checked for segregation. Insert regions within segregated colonies were amplified by PCR and products were sent for sequencing to verify that the insert carried the correct sequence.

2.7. Spot Plates

PCC 7002 cells of each line were taken from freshly grown plates and used to inoculate 50 mL of liquid A+ supplemented with appropriate antibiotics. This primary preculture was grown overnight. After growth, precultures were diluted in new flasks to OD_{750nm} = 0.05 in A+ without antibiotics and allowed to grow overnight. Dilution and growth were repeated to ensure that there was no antibiotic present in the media. Final precultures were diluted to OD_{750nm} = 0.05. The diluted precultures were then used to make 10-fold serial dilutions. Spots (5 μL) of each dilution were placed on an A+ agar plates and allowed to dry for 30 min before being placed in 37 °C incubator.

2.8. Nitrogen Starvation and Repletion

Three flasks of each cell line used were precultured as described for spot plates. Final precultures were pelleted at 4300× *g* for 15 min. Cells were washed twice by resuspension

and pelleting with A+ -NO$_3$. After washing cells were resuspended in A+ -NO$_3$ and diluted to OD$_{750nm}$ = 0.2. Cells were then grown under standard conditions as described above. Triplicate samples were taken from each flask at multiple timepoints and absorption spectra recorded using a Tecan plate reader with a 5 nm step size from 400 nm to 750 nm. After 30 h, when the phycocyanin peak was gone, the cells in each flask were pelleted, resuspended in fresh A+ -NO$_3$, and diluted to OD$_{750nm}$ = 0.2 using A+ -NO$_3$. After cells were appropriately diluted, 500 mL of filter-sterilized 100 g/L NO$_3$ was added to each flask. Flask were placed back in the incubator and grown as during starvation. Samples were taken as during starvation.

2.9. 77 K Fluorescence Spectroscopy

Each cell line used was precultured as described for spot plates. The final preculture was inoculated at OD$_{750nm}$ = 0.05 and allowed to grow overnight. Cells were concentrated 5× in 1.7 mL microfuge tubes by low speed centrifugation with a clear top microfuge under constant illumination. A 0.5 mL quantity of each cell line was removed and placed in NMR tubes. NMR tubes with cells were placed back in the incubator under light conditions identical to growth for 30 min. The remaining 1 mL of sample was pelleted by centrifugation at 18,000× *g*.

Following this, 77 K fluorescence was performed on a Fluorolog-3 spectrofluorometer (Horiba, Kyoto, Japan) with a liquid nitrogen Dewar attachment to maintain sample temperature. Samples in NMR tubes were removed from light immediately before flash freezing in liquid nitrogen. Fluorescence emission spectra were collected from frozen samples following the excitation of chlorophyll (440 nm excitation; 600–800 nm emission) or phycocyanin (580 nm excitation; 600–800 nm emission). Data from chlorophyll excitation were normalized to the PSII emission maximum (692 nm) and data from phycocyanin excitation were normalized to PSI emission maximum (712 nm).

2.10. Quantitative Long-Term Time-Lapse Fluorescence Microscopy

Cells in late-exponential or linear growth were mixed together based on equal OD$_{750\,nm}$ and spotted on A+ medium solidified with 1.5% Bacto agar. Cells were equilibrated at 37 °C for 30 min in the dark prior to imaging using a custom Nikon TiE microscopic system and transilluminated growth light provided at 640 nm (50% power) as previously described [28]. Fluorescence of Chl-a and phycobilisomes were monitored following specific excitation using solid-state light sources at 640 nm or 555 nm (SpectraX, Lumencore, Beaverton, OR, USA). Emission of Chl-a and phycobilisome fluorescence was collected through standard Cy5 or RFP filter cubes (Nikon, Tokyo, Japan), respectively.

2.11. Protein Expression and Purification

E. coli (BL21-DE3) was transformed with a plasmid containing the coding sequence for MBP-*zam*-6xHis under a lactose inducible promoter. Transformants were selected on LB agar plates supplemented with kanamycin (50 μg/mL). Plates were kept at 37 °C overnight. A single colony was picked using a sterile toothpick, transferred into LB supplemented with kanamycin (50 μg/mL), and grown with shaking (200 rpm) for 6 h at 37 °C. The culture was then split, diluted 1:100 into fresh LB with kanamycin, and grown overnight to serve as starter cultures for expression.

Overnight cultures were used to inoculate (1:100) 5 L baffled flasks containing 600 mL pre-warmed LB supplemented with kanamycin (50 μg/mL) and glucose (0.2% *w/v*). Cells were grown ~3 h until OD$_{600nm}$ = 1.0, and then induced with 1 mM isopropyl β- d-1-thiogalactopyranoside (IPTG). Cells were harvested after 3 h by pelleting in a Beckman Coulter Allegra X-14R Centrifuge with a SX4750A rotor at 4300× *g*, 4 °C for 10 min. Cell pellets were flash frozen with liquid nitrogen and kept at −70 °C overnight. Cells were thawed on ice and resuspended in lysis buffer (20 mM Tris pH 7.9, 200 mM KCl, 1 mM MgCl$_2$, 10% glycerol) supplemented with 1× Halt protease inhibitor (ThermoFisher), 1 mM PMSF (phenylmethylsulfonyl fluoride), 1 mg/mL lysozyme, 0.01% TWEEN 20 (Sigma,

St. Louis, MO, USA), 6.25 U/mL Benzonase, and 10 mM 2-mercaptoethanol (βME). The resuspended pellet was incubated on ice for 10 min and lysed via sonication on ice with a Qsonica Q55 Sonicator Ultrasonic Homogenizer with Probe 55 W ($10\times$ pulses, 30 s on, 30 s off, 100% amplitude). Lysate was clarified by two rounds of centrifugation at $11,000\times g$ for 15 min. Clarified lysate was passed through a 0.45-micron filter to remove any additional clumps.

Lysate was mixed with Ni-NTA bead slurry (Thermo Fisher) which had been prepared for binding by preincubation with lysis buffer and incubated with agitation for 1 h at 4 °C. After incubation, the lysate/bead slurry were separated by flowing the mixture over a gravity column. Flow through was collected. Beads were then washed/eluted with lysis buffer containing progressively higher concentration of imidazole (10–500 mM). Elution samples were assayed for presence of protein via SDS-PAGE (Figure S1). Samples containing protein were pooled to be used for MBP-purification.

Pooled eluate from the nickel column was diluted 1:10 in lysis buffer and flowed over an amylose resin (New England Biolabs, Ipswich, MA, USA) gravity column prepared according to the resin manufacturer's specifications. Flow-through was collected. Amylose resin was washed with lysis buffer and then protein was eluted from column using lysis buffer supplemented with 10 mM maltose. Fractions were assayed by SDS-PAGE. Positive fractions were pooled and dialyzed overnight at 4 °C in $1\times$ binding buffer.

2.12. Analysis of CSD1 RNA Binging Potential

RNase II residues involved in binding were extracted from the literature and the RNase II-RNA bound crystal structure (PDB: 2IX1). Our Zam protein multiple alignment was uploaded to the UC-Berkeley WebLogo site and a sequence logo was generated to enhance visualization of conserved residues [29]. Conserved aromatic and positively charged residues were manually identified from the sequence logo and matched to their corresponding *E. coli* residues for evaluation. The positioning of secondary structure elements was performed using the threaded structure of Zam generated by I-Tasser.

3. Results

3.1. Domain Architecture and Conservation of Zam in Cyanobacteria

Comparative sequence and structural analysis was first performed to gain insight into the function of Zam. Pfam analysis shows that in addition to the RNB domain that defines the RNase II/R protein family, Zam also contains two cold-shock domains and an S1 domain (Figure 1A) and exhibits structural similarity with *E. coli* Rnr (Figure 1B). Both S1 and cold-shock domains are members of the larger OB-fold family and have been implicated in RNA binding, carbohydrate binding, and protein–protein interactions in multiple systems [30].

Given the similarity of the domain architecture within the Zam proteins, we wondered if there was extensive primary sequence conservation in any regions that might indicate shared function. To identify regions of interest, the percentage of amino acid conservation calculated from multiple sequence alignments of the cyanobacterial Rnb homologs (Figure 1C) or Zam homologs (Figure 1D) was mapped onto the *E. coli* Rnr crystal structure [21] using Chimera [24]. The Rnb homolog conservation was to serve as a control for general patterns of conservation seen across RNase II/R family proteins. However, patterns of conservation between the two protein groups were strikingly different. In the Zam proteins, conservation is clustered primarily around the three OB-fold containing domains (Figure 1D). In contrast, conservation is more limited in the OB-fold containing domains of Rnb proteins and is mostly found in the RNB nuclease domain (Figure 1C).

Comparing primary sequences within the RNB domain explains the stark conservation difference. The active sites of the Rnb proteins (Figure 1E) and the Zam proteins (Figure 1F) reveal that, while the canonical RNB domain is present in cyanobacteria, the Zam proteins have highly degenerate and poorly conserved 'active' sites. Most Zam proteins lack the four aspartic or glutamic acid residues required for coordination of the Mg^{2+} that is essential

for catalysis [31,32]. These acidic residues are essentially invariant in all known functional RNase II/R proteins, suggesting that all Zam proteins lack nuclease activity, but could retain RNA-binding activity.
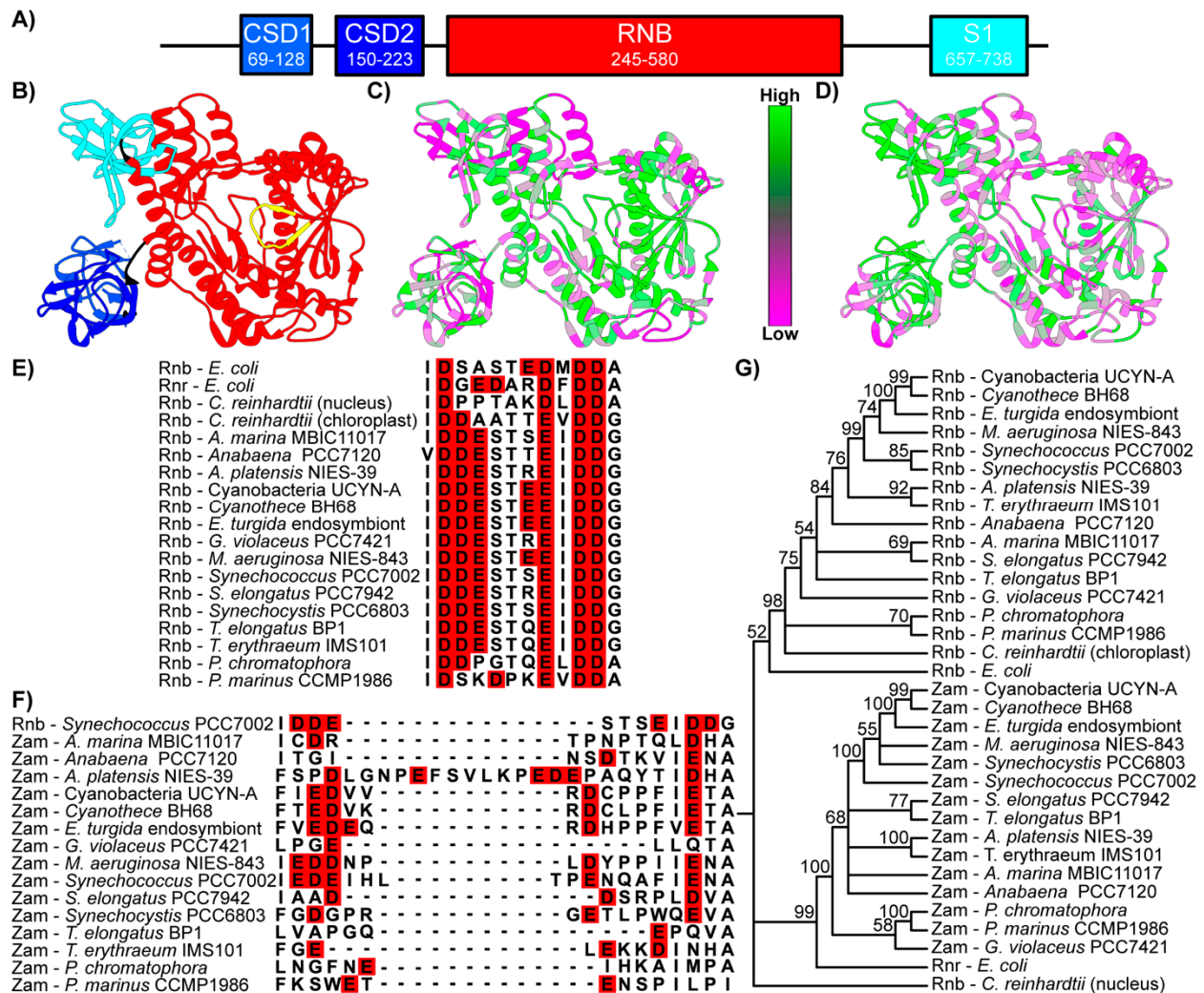
Figure 1. Domain architecture and conservation of Zam. (**A**) Map of PCC 7002 Zam domains as identified by Pfam. Each predicted domain of Zam corresponds to a domain found in *E. coli* Rnr. (**B**) *E. coli* Rnr crystal structure with domains color coded to match Zam domain map. The active site shown in alignments is indicated in yellow. (**C**) Cyanobacterial Rnb protein conservation mapped to *E. coli* Rnr crystal structure. Conservation is observed across the protein but is most concentrated around the RNB domain active site. (**D**) Cyanobacterial Zam conservation mapped onto *E. coli* Rnr crystal structure. Most conservation is observed in the RNA binding domains (CSD1, CSD2, S1). Low conservation is observed around the active site. (**E**) Multiple alignment of cyanobacterial Rnb protein active sites with outgroups shown as reference. Aspartic and glutamic acid residues color coded in red. (**F**) Multiple alignment of *zam* protein active sites. Aspartic and glutamic acid residues color coded in red. Active site regions were identified by alignment with active site region shown in Figure 1E. The sequence of Rnb from PCC 7002 is included to represent functional RNB domain active sites. (**G**) Bootstrap consensus tree of cyanobacterial Rnr, cyanobacterial Rnb, and outgroup homolog protein sequences. Cyanobacteria Rnb genes cluster as a distinct group along with chloroplastic Rnb genes and the *E. coli* Rnb gene. Zam proteins form their own cluster with *E. coli* Rnr. An Rnb domain containing protein from the nuclear genome of *C. reinhardtii* is provided as an outgroup.

3.2. Zam and Rnb Phylogeny

Using BLAST, we identified reciprocal best hits (RBH) for the annotated Zam and Rnb proteins from our model species *Synechococcus sp.* PCC 7002 in a diverse set of cyanobacterial genomes. We also searched for homologs to Zam and Rnb in the genomes of other photosynthetic organisms. Zam-like proteins were not found in any algae or higher plant genome queried. Interestingly, chloroplast genomes do contain an RBH for Rnb, but lack Zam. While many RNB domains containing proteins were identified in all species whose genomes were searched, an RBH for Zam was only found in *Paulinella chromatophora* and *Epithemia turgida*. Both *P. chromatophora* and *E. turgida* have experienced recent endosymbiotic events [33,34] and their Zam proteins are thus cyanobacterial in origin.

A phylogenetic tree for representative Zam and Rnb proteins was created (Figure 1G). As expected, the Rnb proteins cluster into their own clade, along with the chloroplastic nuclease from *C. reinhardtii*, and *E. coli* Rnb. This supports the idea that these proteins are genuine RNase II proteins. The Zam proteins also cluster into their own clade with *E. coli* Rnr. This clear divergence into distinct protein families, along with the presence of a Zam protein in every cyanobacterial genome searched, indicate that the Zam and the Rnb proteins diverged early in the cyanobacterial lineage, and that Zam is most likely a homolog of *E. coli* Rnr.

3.3. Zam Proteins Possess Conserved, Redox-Active Cysteines in CSD1

Given the extreme divergence of putative Zam active site regions, we chose to investigate the regions that were conserved. The first cold shock domain (CSD1) of Zam is highly conserved among homologs (Figure 1D). Of special interest in this domain are two invariant cysteines (C73 and C79 in PCC 7002) (Figure 2A) that were previously identified as reduction-oxidation (redox) active in global redox-proteomics studies of two different cyanobacterial species [5,6].

Due to the proximity of these cysteines to one another and the well-recognized role of disulfide bond formation in the redox regulation of proteins, we built a model of CSD1 to assess whether its conserved cysteines could potentially form a disulfide bridge [35]. With the *E. coli* Rnr crystal structure as a reference, threading was performed to create a model of PCC 7002 Zam. The modeled structure places these cysteines less than 4 Angstroms apart in an orientation favoring disulfide bond formation (Figure 2B).

While computational modeling and proteomics studies support the hypothesis that the conserved cysteines in Zam could undergo disulfide bond formation, we wanted to validate that PCC 7002 Zam was redox responsive in vitro. In *E. coli*, we expressed PCC 7002 Zam fused to an N-terminal maltose binding protein and a C-terminal 6x-histidine tag (Figure S1). The recombinant Zam was purified and analyzed by SDS-PAGE under both reducing and oxidizing conditions (Figure 2C). The mobility of recombinant Zam is influenced by the redox state of the protein, with the reduced form of the protein running at an apparent higher molecular weight when compared to the oxidized form. This result is consistent with a protein capable of undergoing reduction of cysteines and leads us to conclude that Zam likely forms a disulfide bond [36]. Further evidence for disulfide bond formation comes from recent structural predictions of Zam from PCC 6803 using AlphaFold [37,38]. The predicted 3D structure (https://alphafold.ebi.ac.uk/entry/Q46363, accessed on 13 May 2022) includes a disulfide bond between C73 and C79 and a per-residue confidence score (pLDDT) of greater than 90 at these residues, indicating a very high model confidence.
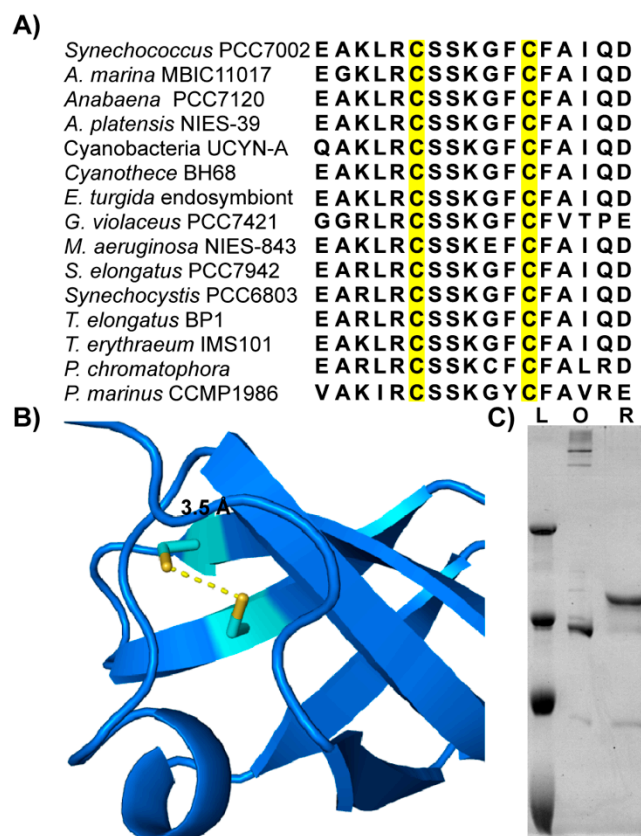
Figure 2. Identification of Redox-Active Cysteines in Zam: (**A**) Multiple sequence alignment of cyanobacterial Zam CSD1 domain focused on the CSD1 OB-fold. Conserved cysteines are highlighted in yellow. (**B**) Threaded structure prediction of PCC 7002 Zam CSD1 domain with conserved cysteines indicated in yellow. (**C**) SDS-PAGE of purified PCC 7002 Zam in oxidizing (O) and reducing (R) conditions. A quantity of 1.3 μM protein was incubated in 10 mM oxidized glutathione and then was diluted 1:20 in buffer without glutathione (oxidized) or in the presence of 710 mM βME (Reduced) and incubated at room temperature for 1 h. The samples were boiled for 5 min prior to loading. Ladder (L) shows apparent molecular weights (250 kDa, 130 kDa, 100 kDa, 70 kDa). Reduced protein runs at an apparent higher molecular weight.

3.4. Analysis of CSD1 RNA Binding Potential

OB-fold domains are known to bind to numerous different ligands including single-stranded DNA, RNA, and proteins [30]. In other members of the RNase II/R family, the CSD1 OB-fold is known to bind RNA in in vitro studies and to contact RNA in crystal structures [32,39,40]. In the *E. coli* sRNase II co-crystalized RNA–protein complex, the major contact point between CSD1 and RNA occurs at the loop between the first and second β-strand (L12) of CSD1 [32]. The contacts in L12 of the *E. coli* protein appear to be mediated by four residues—two lysine residues and two phenylalanine residues, which are commonly used in the recognition of ssRNA [32].

In the Zam protein these four residues are also conserved in the L12 loop, though the first lysine has been replaced by an arginine (Figure 3). It is also important to note that it is within the L12 loop of Zam that the conserved Zam cysteines are found. Given the importance of the L12 loop in nucleic acid binding by CSD1 in RNase II-type proteins, it seems likely that these cysteines are involved in the nucleic acid binding process. A possible mechanism of action would be the direct regulation of RNA binding through stabilization of the L12 loop via regulated disulfide formation. Alternatively, other post-translational modifications of cysteines besides disulfides can occur and could serve to enhance or abolish RNA binding at this site.
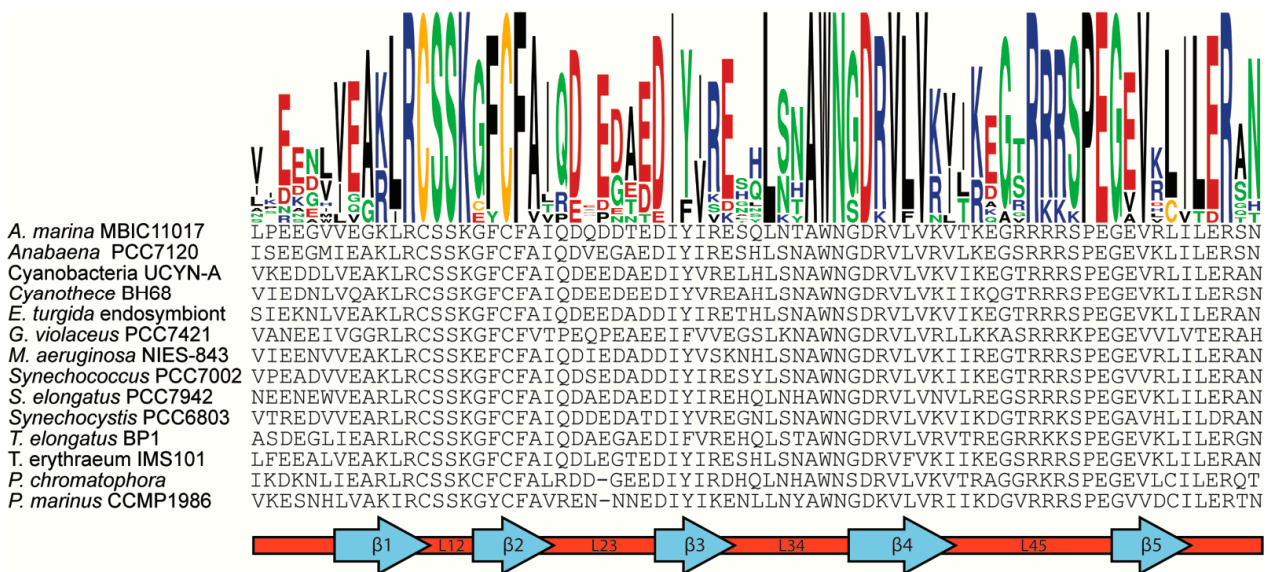
Figure 3. Annotated alignment of CSD1 OB-fold domain from Zam proteins. Conservation is illustrated with a sequence logo colorized by amino acid class (Black = Hydrophobic, Green = Polar, Red = Acidic, Blue = Basic, Yellow = Cysteine). The position of canonical OB-fold secondary structures within the sequence are annotated below. B-strands are labeled (β1–β4) with the intervening loops described by the strands they connect.

While the L12 loop is essential for RNA binding in the CSD1 of RNase II, this region is not responsible for RNA binding in all RNase II/R family members. Rrp44, a yeast RNase II/R family member, has been crystalized with RNA [40]. In the Rrp44 structure, RNA enters the protein active site via a side channel formed by the RNB domain and the two CSD domains instead of passing through the central RNA channel utilized by RNA in RNase II [40]. Thus, in Rrp44, a different loop of CSD1 is exposed to RNA to facilitate binding. While the CSD1 of Rrp44 is partially unresolved in the crystal structure, it is apparent that the loop between β-strand 2 and β-strand 3 (L23) contains at least some of the key residues required for binding. In Zam, L23 does not contain any well conserved positively charged or aromatic residues and thus likely does not use L23 as an RNA binding interface.

The mechanism of RNA binding in RNase II/R family members is fairly well understood, but not known for every member. In RNase R it has been suggested that RNA might use either the RNase II-like or the Rrp44-like binding mode [21]. Additionally, since relatively short RNA molecules have been co-crystalized with these proteins, how other regions within the CSD1 domain might interact with RNA remain unknown. Other potential RNA binding residues are conserved within the Zam CSD1, including a tryptophan in loop region L34 and an arginine repeat in L45 (Figure 3). These conserved residues appear to be surface exposed the structural model of Zam (Figure 1D) and could potentially make additional RNA contacts.

3.5. CSD1 Cysteine to Serine Mutants Phenocopy Zam Deletion

Proteomics, bioinformatic analysis, computational modeling, and in vitro experimentation suggest that the conserved CSD1 cysteines are involved in Zam function. Therefore, we created and characterized strains with C73 and C79 mutations in vivo. We first created a PCC 7002 Zam deletion line (ΔZ) and a complemented deletion line (ΔZ+) in which PCC 7002 Zam is expressed in a different location in the genome but under its native promoter, as previously described [7]. Additionally, we created three complemented lines (ΔZ+ C73S, ΔZ+ C79S, ΔZ+ DCS) containing mutations that convert one or both native cysteine residues to serine residues.

Spot plate analysis shows that the ΔZ and cysteine mutant lines all exhibit a mild growth defect compared to wild-type (WT) and ΔZ+ (Figure 4A). Additionally, colonies

of the ΔZ and cysteine mutants are chlorotic. This chlorotic phenotype is also observed in liquid culture and is caused by a statistically significant decrease in the ratio of phycocyanin relative to chlorophyll in the ΔZ and cysteine mutant lines (Figure 4B,C).
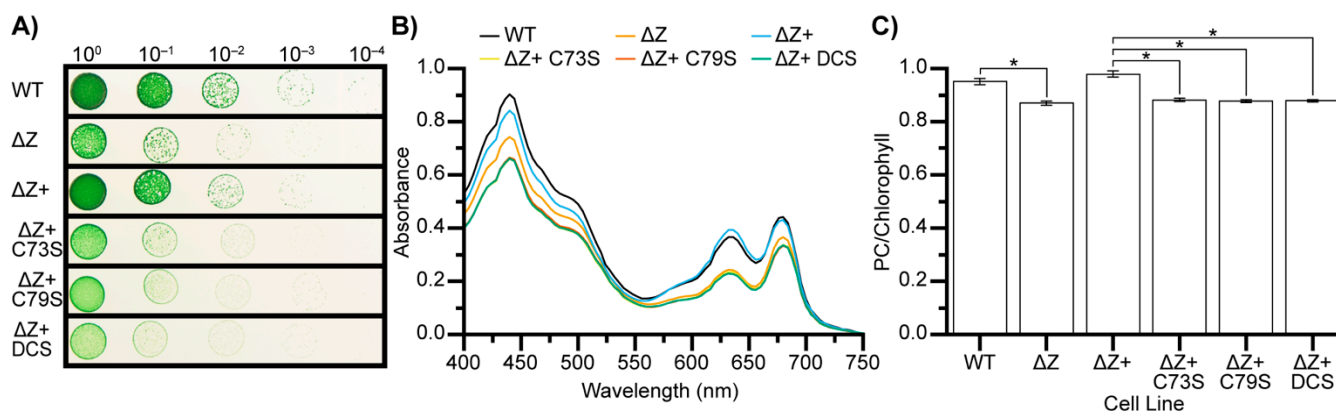
Figure 4. C73 and C79 are required for Zam function in vivo. (**A**) Spot plate showing relative growth of cell lines used in this study. As previously reported ΔZ exhibits a mild growth defect. The mutant complemented lines show the same defect. (**B**) Representative absorption spectrums of cell lines used in this study. ΔZ cells and the mutant complement lines show decreased phycocyanin relative to chlorophyll. (**C**) Average ratio of phycocyanin absorption to chlorophyll absorption for each cell line (*n* = 3). Statistical difference determined by Student's *t*-test at *p* = 0.001 indicated by *.

3.6. ΔZ and ΔZ+ DCS Exhibit Decreased Synthesis of Phycocyanin

Since the decreased steady state level of phycocyanin in ΔZ and the mutant complement lines could arise from increased phycocyanin degradation, decreased phycocyanin synthesis, or a combination of the two, we sought to isolate and observe both processes. WT, ΔZ, ΔZ+ and ΔZ+ DCS cells were grown in shaking flasks, starved of nitrate to induce the degradation of phycocyanin, and then resupplied with nitrate to reinitiate the synthesis of phycocyanin (Figure 5) [41]. Phycocyanin and chlorophyll levels were monitored spectroscopically. During nitrogen starvation, all four cell lines behave similarly, with a phycocyanin degradation pattern consistent with previous nitrogen starvation experiments [41,42]. However, upon the addition of nitrogen, WT and ΔZ+ exhibit faster initiation, increased rate, and increased yield of phycocyanin synthesis compared to the ΔZ and the cysteine double-mutant (Figure 5). This leads us to conclude that the lower steady state phycocyanin abundance in ΔZ and the cysteine mutants are a result of lower phycocyanin synthesis, and not due to increased phycocyanin turnover.

3.7. ΔZ and Cysteine Mutants Exhibit Altered Photosynthetic Energy Transfer

The ΔZ and cysteine mutant cell lines showed altered pigment abundance and slower growth rates compared to the controls. Therefore, we reasoned that they could exhibit altered light harvesting or energy transfer phenotypes. We conducted 77 K fluorescence spectroscopy with excitation at either 440 nm to preferentially excite chlorophyll or 580 nm to preferentially excite phycocyanin. Fluorescence emission spectra (600–800 nm) containing peaks originating from terminal emitters of phycocyanin (PC), photosystem II (PSII), and photosystem I (PSI) were then collected (Figure 6). Chlorophyll excitation shows an increased ratio of PSI:PSII in the ΔZ and cysteine mutants compared to WT and ΔZ+ strains (Figure 6A). Phycocyanin excitation shows increased energy transfer from the phycobilisome to PSII in the ΔZ and cysteine mutants compared to WT and ΔZ+. (Figure 6B). Taken together, these results suggest that altered coupling of phycobilisomes and PSII could balance energy flow to compensate for the altered ratio of PSI:PSII. Alternatively, changes to the antenna coupling could affect photosystem stoichiometry.
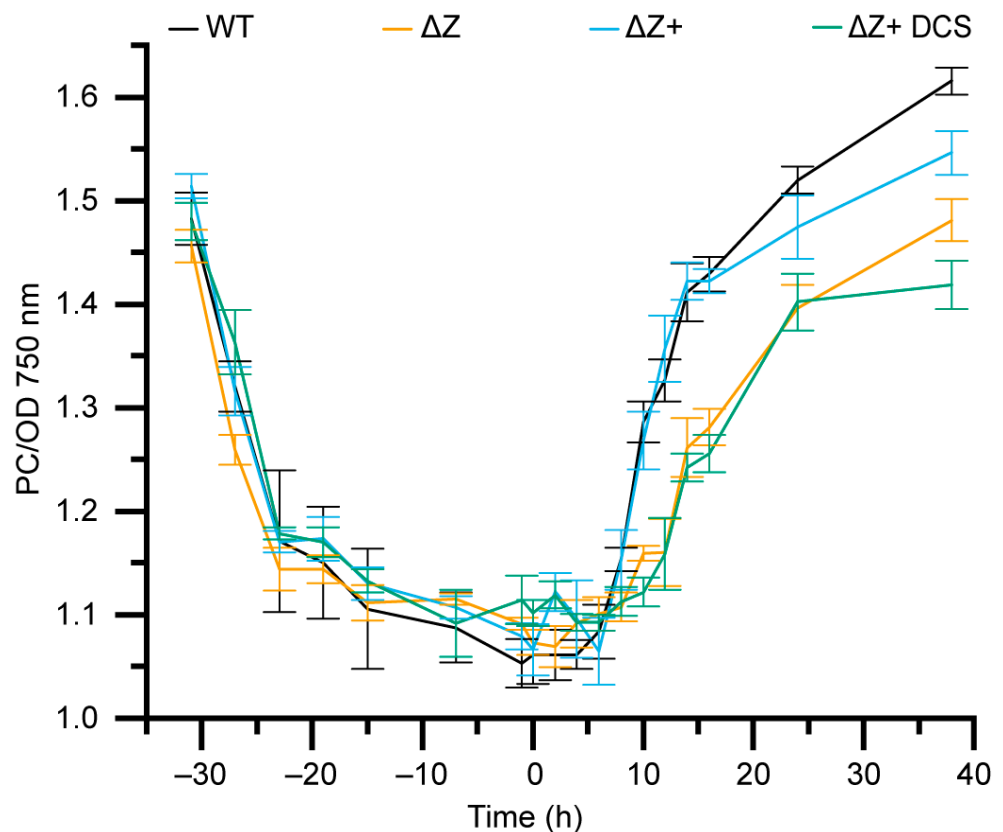
Figure 5. Time-course of nitrogen depletion and repletion. Ratio of phycocyanin to optical density 750 nm for cell lines growing under nitrogen starvation for 30 h followed by nitrogen repletion. Nitrate was added to samples at t = 0.

3.8. Comparison of WT and ΔZ Growth and Fluorescence at Single-Cell Resolution

Bulk fluorescence and growth experiments revealed altered pigment profiles and reduced growth capacity of ΔZ compared to the WT. The growth and fluorescence properties of ΔZ and a WT control strain expressing GFP (WT-GFP) were also compared at single(sub)-cellular resolution using long-term time-lapse fluorescence microscopy (Figure 6C and Supplementary Movies S1–S3) on a custom microscope we recently developed for growth and imaging of photosynthetic organisms [28,43]. By confining cells in a 2D layer between the coverslip and pad of growth media solidified with agar, spectrally defined light provided by a solid-state transillumination source (640 nm) can be uniformly applied to the cells (as sole energy source) while simultaneously monitoring the fluorescence dynamics of chlorophyll (Chl-a) and phycobilisomes. Spike-in WT-GFP served as an internal growth control as it has very reproducible growth in addition to producing a distinctive and expected colony morphology at the 16-cell stage [28]. Notably, the ΔZ strain exhibited lower fluorescence in both the Chl-a and phycobilisome channel at the initial timepoint compared to WT-GFP cells at the same cell-cycle stage (Figure 6C upper panel and Figure S2). After 18.5 h growth, the WT-GFP and ΔZ strain both produced 16 cells and typical colony morphologies, further confirming the subtle growth defects observed on solid medium and as shown in Figure 4A. After 18.5 h of growth, the fluorescence intensity of Chl-a and phycobilisomes in ΔZ cells recovered to levels similar to that of WT-GFP. However, there are subtle differences in the appearance and localization of fluorescence between WT-GFP and ΔZ. In particular, WT-GFP exhibits more fluorescence (both Chl-a and phycobilisome) at the periphery of the cell in association with the thylakoid membrane whereas it is more evenly distributed in ΔZ, which could indicate a change in the sub-cellular distribution of phycobilisomes or PSII.
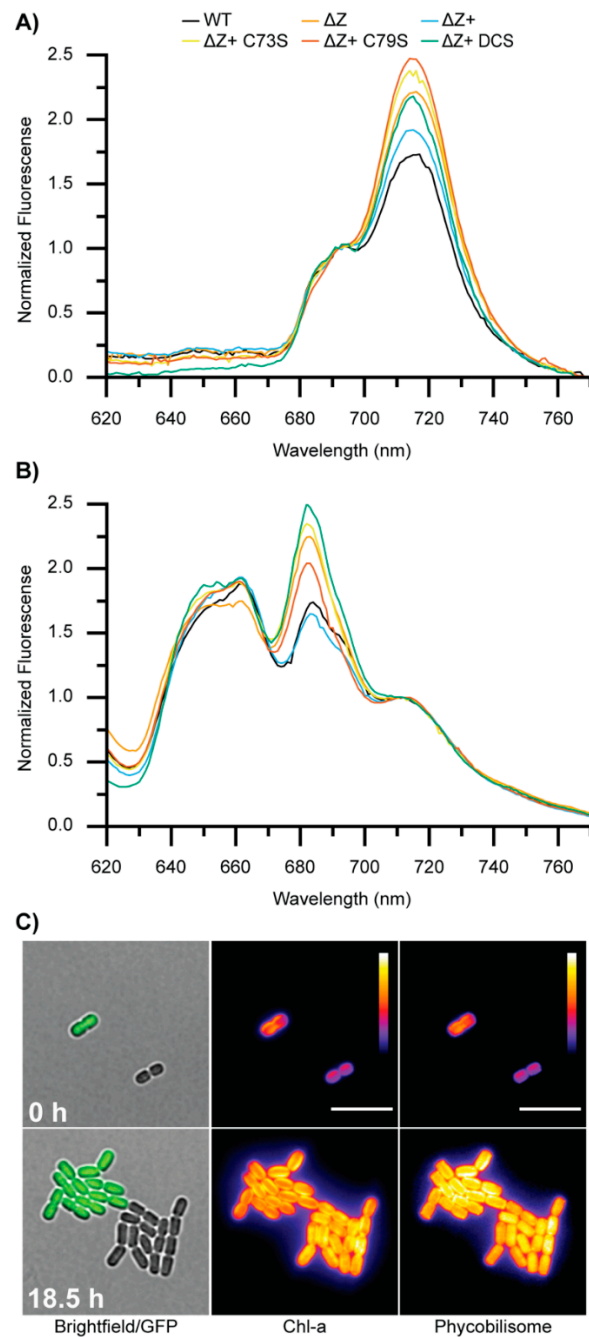
Figure 6. Bulk 77 K emission spectra and single-cell analysis of growth and fluorescence at room temperature. (**A**) 77 K fluorescence emission spectra for each cell line produced by chlorophyll excitation at 440 nm normalized to PSII max emission (692 nm). Increased PSI:PSII ratio is observed for ΔZ and mutant complement lines relative to ΔZ and ΔZ+. (**B**) 77 K fluorescence emission spectra produced by phycobilisome excitation at 580 nm normalized to PSI max emission (712 nm). Increased energy transfer to PSII is observed in ΔZ and cysteine mutants. (**C**) Snapshots from time-lapse imaging (Supplementary Movies S1–S3) showing growth and fluorescent properties of ΔZ and control WT-GFP cells under identical conditions for 18.5 h. Specific excitation/emission filters were used to monitor fluorescence of GFP (green), Chl-a (Intensity scale = 177–3000 A.U.; purple-white), and phycobilisomes (Intensity scale = 177–1096 A.U.; purple-white). Scale bar = 10 μm.

4. Discussion

Based on homology and domain architecture, one would presume that the Zam proteins were functional exonucleases, and in many genomes they have been annotated as such. However, closer examination of the protein sequences via multiple sequence alignment shows that the canonical RNB active site is not present in any Zam protein. Without this key region, Zam likely cannot function as a nuclease. The absence of nuclease activity by Zam is consistent with in vivo knockout data from PCC 7002 [7]. In *E. coli*, Rnr, Rnb, and polynucleotide phosphorylase (PNPase) work cooperatively and redundantly to provide the cell with 3′-5′ exonuclease activity. Deletion of any one of those proteins produces only mild phenotypes. However, deletion of PNPase in concert with either Rnr or Rnb is lethal [44–46]. In cyanobacteria, deletion of either the PNPase or Rnb in PCC 7002 has been shown to be lethal, while deletion of Zam produces the complex phenotypes outlined in this and other works [7]. Our data are consistent with the idea that Zam is not functional as a nuclease and cannot provide redundancy to the RNA degradation system. Indeed, purified PCC 7002 Zam (Figure S1) did not exhibit any nuclease or RNA binding activity when incubated with several purified Rnb substrates (data not shown). As our purified protein is prone to aggregation, any interpretation of these negative results must be done with an abundance of caution. However, they are consistent with the bioinformatic and in vivo findings.

If Zam proteins are not functioning as nucleases, they must be exerting their influence on the cell through another mechanism. The conservation of the cold shock and S1 domains in Zam proteins indicates they may represent the major functional domains of the protein. Our work indicates that the conserved cysteines in the CSD1 of Zam undergo redox modulation and are essential for in vivo function, with mutagenesis of the redox active cysteines in CSD1 being sufficient to produce phenotypes identical to that of gene knockout. We hypothesize that the conserved cysteines in CSD1 are subject to control by the cellular redox state and that locking them in the permanently reduced state renders Zam nonfunctional. CSD1 is canonically an RNA binding domain in RNase II/R proteins, and it is possible that cellular redox state is modulating the ability of Zam to bind RNA targets via the disulfide modulation of CSD1 [39]. However, a recent study shows that an RNase II/R protein (identified here as Zam) from *Anabaena* sp. PCC 7120 interacts with and enhances the activity of RNase E, the major degradosome enzyme, via its cold-shock and S1 domains [47]. This finding prompts us to propose a different model in which Zam binding to RNase E under oxidizing, but not reducing, conditions increases RNA turnover rates. Based on current data, it is unknown whether the interaction between Zam and RNase E is mediated fully or in part by RNA interactions, but as our analysis shows CSD1 maintains the canonical RNA binding residues, an RNA mediated binding modality is a distinct possibility.

A role for Zam in regulating RNA turnover via interaction with the degradosome would be consistent with both our analysis and the literature. The original phenotype of *zam* deletion in PCC 6803 was a tolerance of carbonic anhydrase inhibitors and we see this phenotype to a lesser degree in PCC 7002 (Figure S3) [3,4]. The authors of the papers characterizing the PCC 6803 mutant hypothesized that *zam* encoded a transporter that affected the access of the inhibitor to the carbonic anhydrase [4]. While this explanation is unlikely based on the evolutionary roots of Zam, it is simple to imagine a change in global RNA degradation rates affecting membrane protein abundance and changing transport. A similar phenomenon could explain the alteration of photosystem or phycobilisome complex abundance that we see in our *zam* deletion or mutant strains using bulk and single-cell analysis.

This work provides new insight into the RNase II/R family of cyanobacteria. We classify Zam proteins as non-functional nucleases and show that they have a role in controlling diverse phenotypes including growth rate, pigment biosynthesis, and the PSI:PSII ratio. The ability of Zam to function in vivo is dependent upon redox active cysteines located in CSD1, and this highly conserved region is almost certainly responsible

for controlling the function of Zam. While further experimentation is needed, a model where the redox state of Zam modulates both the interaction with and subsequent activation of RNase E seems likely based on our work and the literature.

References

1. Moisander, P.H.; Beinart, R.A.; Hewson, I.; White, A.E.; Johnson, K.S.; Carlson, C.A.; Montoya, J.P.; Zehr, J.P. Unicellular Cyanobacterial Distributions Broaden the Oceanic N$_2$ Fixation Domain. *Science* **2010**, *327*, 1512–1514. [CrossRef] [PubMed]
2. Hitchcock, A.; Hunter, C.N.; Canniffe, D.P. Progress and challenges in engineering cyanobacteria as chassis for light-driven biotechnology. *Microb. Biotechnol.* **2019**, *13*, 363–367. [CrossRef] [PubMed]
3. Bédu, S.; Peltier, G.; Sarrey, F.; Joset, F. Properties of a Mutant from *Synechocystis* PCC6803 Resistant to Acetazolamide, an Inhibitor of Carbonic Anhydrase. *Plant Physiol.* **1990**, *93*, 1312–1315. [CrossRef] [PubMed]
4. Beuf, L.; Bédu, S.; Cami, B.; Joset, F. A protein is involved in accessibility of the inhibitor acetazolamide to the carbonic anhydrase(s) in the cyanobacterium *Synechocystis* PCC 6803. *Plant Mol. Biol.* **1995**, *27*, 779–788. [CrossRef]
5. Ansong, C.; Sadler, N.C.; Hill, E.A.; Lewis, M.P.; Zink, E.M.; Smith, R.D.; Beliaev, A.S.; Konopka, A.E.; Wright, A.T. Characterization of protein redox dynamics induced during light-to-dark transitions and nutrient limitation in cyanobacteria. *Front. Microbiol.* **2014**, *5*, 325. [CrossRef] [PubMed]
6. Guo, J.; Nguyen, A.Y.; Dai, Z.; Su, D.; Gaffrey, M.J.; Moore, R.J.; Jacobs, J.M.; Monroe, M.E.; Smith, R.D.; Koppenaal, D.W.; et al. Proteome-wide Light/Dark Modulation of Thiol Oxidation in Cyanobacteria Revealed by Quantitative Site-specific Redox Proteomics. *Mol. Cell. Proteom.* **2014**, *13*, 3270–3285. [CrossRef]
7. Cameron, J.C.; Gordon, G.C.; Pfleger, B.F. Genetic and genomic analysis of RNases in model cyanobacteria. *Photosynth. Res.* **2015**, *126*, 171–183. [CrossRef]
8. Kujat, S.L.; Owttrim, G.W. Redox-Regulated RNA Helicase Expression. *Plant Physiol.* **2000**, *124*, 703–714. [CrossRef]
9. Migur, A.; Heyl, F.; Fuss, J.; Srikumar, A.; Huettel, B.; Steglich, C.; Prakash, J.S.S.; Reinhardt, R.; Backofen, R.; Owttrim, G.W.; et al. The temperature-regulated DEAD-box RNA helicase CrhR interactome: Autoregulation and photosynthesis-related transcripts. *J. Exp. Bot.* **2021**, *72*, 7564–7579. [CrossRef]
10. Rosana, A.R.R.; Ventakesh, M.; Chamot, D.; Patterson-Fortin, L.M.; Tarassova, O.; Espie, G.S.; Owttrim, G.W. Inactivation of a Low Temperature-Induced RNA Helicase in *Synechocystis* sp. PCC 6803: Physiological and Morphological Consequences. *Plant Cell Physiol.* **2012**, *53*, 646–658. [CrossRef]

11. Georg, J.; Rosana, A.R.R.; Chamot, D.; Migur, A.; Hess, W.R.; Owttrim, G.W. Inactivation of the RNA helicase CrhR impacts a specific subset of the transcriptome in the cyanobacterium *Synechocystis* sp. PCC 6803. *RNA Biol.* **2019**, *16*, 1205–1214. [CrossRef] [PubMed]

12. Arraiano, C.M.; Andrade, J.M.; Domingues, S.; Guinote, I.B.; Malecki, M.; Matos, R.G.; Moreira, R.N.; Pobre, V.; Reis, F.P.; Saramago, M.; et al. The critical role of RNA processing and degradation in the control of gene expression. *FEMS Microbiol. Rev.* **2010**, *34*, 883–923. [CrossRef] [PubMed]

13. Uesono, Y.; Toh-e, A.; Kikuchi, Y. Ssd1p of *Saccharomyces cerevisiae* Associates with RNA. *J. Biol. Chem.* **1997**, *272*, 16103–16109. [CrossRef]

14. Py, B.; Causton, H.; Mudd, E.A.; Higgins, C.F. A protein complex mediating mRNA degradation in *Escherichia coli*. *Mol. Microbiol.* **1994**, *14*, 717–729. [CrossRef] [PubMed]

15. Purusharth, R.I.; Klein, F.; Sulthana, S.; Jäger, S.; Jagannadham, M.V.; Evgunieva-Hackenberg, E.; Ray, M.K.; Klug, G. Exoribonuclease R Interacts with Endoribonuclease E and an RNA Helicase in the Psychrotrophic Bacterium *Pseudomonas syringae* Lz4W. *J. Biol. Chem.* **2005**, *280*, 14572–14578. [CrossRef]

16. Lu, F.; Taghbalout, A. The *Escherichia coli* major exoribonuclease RNase II is a component of the RNA degradosome. *Biosci. Rep.* **2014**, *34*, e00166. [CrossRef]

17. Mian, I.S. Comparative sequence analysis of ribonucleases HII, III, II PH and D. *Nucleic Acids Res.* **1997**, *25*, 3187–3195. [CrossRef]

18. Portnoy, V.; Schuster, G. RNA polyadenylation and degradation in different Archaea; roles of the exosome and RNase R. *Nucleic Acids Res.* **2006**, *34*, 5923–5931. [CrossRef]

19. Zuo, Y.; Deutscher, M.P. Exoribonuclease superfamilies: Structural analysis and phylogenetic distribution. *Nucleic Acids Res.* **2001**, *29*, 1017–1026. [CrossRef]

20. Cheng, Z.-F.; Deutscher, M.P. An Important Role for RNase R in mRNA Decay. *Mol. Cell* **2005**, *17*, 313–318. [CrossRef]

21. Chu, L.-Y.; Hsieh, T.-J.; Golzarroshan, B.; Chen, Y.-P.; Agrawal, S.; Yuan, H.S. Structural insights into RNA unwinding and degradation by RNase R. *Nucleic Acids Res.* **2017**, *45*, 12015–12024. [CrossRef] [PubMed]

22. Matos, R.G.; Fialho, A.M.; Giloh, M.; Schuster, G.; Arraiano, C.M. The rnb Gene of *Synechocystis* PCC6803 Encodes a RNA Hydrolase Displaying RNase II and Not RNase R Enzymatic Properties. *PLoS ONE* **2012**, *7*, e32690. [CrossRef] [PubMed]

23. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef]

24. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612. [CrossRef]

25. Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: A unified platform for automated protein structure and function prediction. *Nat. Protoc.* **2010**, *5*, 725–738. [CrossRef]

26. Stevens, S.E.; Patterson, C.O.P.; Myers, J. The Production of Hydrogen Peroxide by Blue-Green Algae: A Survey. *J. Phycol.* **1973**, *9*, 427–430. [CrossRef]

27. Gibson, D.G.; Young, L.; Chuang, R.-Y.; Venter, J.C.; Hutchison, C.A., III; Smith, H.O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **2009**, *6*, 343–345. [CrossRef]

28. Moore, K.A.; Altus, S.; Tay, J.W.; Meehl, J.B.; Johnson, E.B.; Bortz, D.M.; Cameron, J.C. Mechanical regulation of photosynthesis in cyanobacteria. *Nat. Microbiol.* **2020**, *5*, 757–767. [CrossRef]

29. Crooks, G.E. WebLogo: A Sequence Logo Generator. *Genome Res.* **2004**, *14*, 1188–1190. [CrossRef]

30. Theobald, D.L.; Mitton-Fry, R.M.; Wuttke, D.S. Nucleic Acid Recognition by OB-Fold Proteins. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 115–133. [CrossRef]

31. Amblar, M.; Arraiano, C.M. A single mutation in *Escherichia coli* ribonuclease II inactivates the enzyme without affecting RNA binding. *FEBS J.* **2004**, *272*, 363–374. [CrossRef] [PubMed]

32. Frazao, C.; McVey, C.; Amblar, M.; Barbas, A.; Vonrhein, C.; Arraiano, C.; Carrondo, M.A. Unravelling the dynamics of RNA degradation by ribonuclease II and its RNA-bound complex. *Nature* **2006**, *443*, 110–114. [CrossRef] [PubMed]

33. Nowack, E.C.; Melkonian, M.; Glöckner, G. Chromatophore Genome Sequence of *Paulinella* Sheds Light on Acquisition of Photosynthesis by Eukaryotes. *Curr. Biol.* **2008**, *18*, 410–418. [CrossRef] [PubMed]

34. Nakayama, T.; Kamikawa, R.; Tanifuji, G.; Kashiyama, Y.; Ohkouchi, N.; Archibald, J.M.; Inagaki, Y. Complete genome of a nonphotosynthetic cyanobacterium in a diatom reveals recent adaptations to an intracellular lifestyle. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 11407–11412. [CrossRef] [PubMed]

35. Buchanan, B.B.; Balmer, Y. Redox Regulation: A Broadening Horizon. *Annu. Rev. Plant Biol.* **2005**, *56*, 187–220. [CrossRef]

36. Lane, L.C. A simple method for stabilizing protein-sulfhydryl groups during SDS-gel electrophoresis. *Anal. Biochem.* **1978**, *86*, 655–664. [CrossRef]

37. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [CrossRef]

38. Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; et al. AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **2022**, *50*, D439–D444. [CrossRef]

39. Amblar, M.; Barbas, A.; Fialho, A.; Arraiano, C.M. Characterization of the Functional Domains of *Escherichia coli* RNase II. *J. Mol. Biol.* **2006**, *360*, 921–933. [CrossRef]

40. Lorentzen, E.; Basquin, J.; Tomecki, R.; Dziembowski, A.; Conti, E. Structure of the Active Subunit of the Yeast Exosome Core, Rrp44: Diverse Modes of Substrate Recruitment in the RNase II Nuclease Family. *Mol. Cell* **2008**, *29*, 717–728. [CrossRef]

41. Allen, M.M.; Smith, A.J. Nitrogen chlorosis in blue-green algae. *Arch. Mikrobiol.* **1969**, *69*, 114–120. [CrossRef] [PubMed]

42. Stevens, S.E.; Balkwill, D.L.; Paone, D.A.M. The Effects of Nitrogen Limitation on the Ultrastructure of the Cyanobacterium *Agmeneilum quadruplicatum*. *Arch. Microbiol.* **1981**, *130*, 204–212. [CrossRef]

43. Hill, N.C.; Tay, J.W.; Altus, S.; Bortz, D.M.; Cameron, J.C. Life cycle of a cyanobacterial carboxysome. *Sci. Adv.* **2020**, *6*, eaba1269. [CrossRef] [PubMed]

44. Donovan, W.P.; Kushner, S.R. Polynucleotide phosphorylase and ribonuclease II are required for cell viability and mRNA turnover in *Escherichia coli* K-12. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 120–124. [CrossRef] [PubMed]

45. Cheng, Z.-F.; Zuo, Y.; Li, Z.; Rudd, K.E.; Deutscher, M.P. The *vacB* Gene Required for Virulence in *Shigella flexneri* and *Escherichia coli* Encodes the Exoribonuclease RNase R. *J. Biol. Chem.* **1998**, *273*, 14077–14080. [CrossRef]

46. Cheng, Z.-F.; Deutscher, M.P. Quality control of ribosomal RNA mediated by polynucleotide phosphorylase and RNase R. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 6388–6393. [CrossRef]

47. Zhou, C.; Zhang, J.; Hu, X.; Li, C.; Wang, L.; Huang, Q.; Chen, W. RNase II binds to RNase E and modulates its endoribonucleolytic activity in the cyanobacterium *Anabaena* PCC 7120. *Nucleic Acids Res.* **2020**, *48*, 3922–3934. [CrossRef]

Article

Cyclic Electron Flow-Coupled Proton Pumping in *Synechocystis* sp. PCC6803 Is Dependent upon NADPH Oxidation by the Soluble Isoform of Ferredoxin:NADP-Oxidoreductase

Neil T. Miller [1], Ghada Ajlani [2] and Robert L. Burnap [1,*]

[1] Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, OK 74078, USA; ntmiller2@wisc.edu

[2] Institute for Integrative Biology of the Cell (I2BC), Université Paris-Saclay, CNRS, CEA, 91198 Gif-sur-Yvette, France; gajlani@cea.fr

* Correspondence: rob.burnap@okstate.edu; Tel.: +1-(405)-744-7445

Abstract: Ferredoxin:NADP-oxidoreductase (FNR) catalyzes the reversible exchange of electrons between ferredoxin (Fd) and NADP(H). Reduction of $NADP^+$ by Fd via FNR is essential in the terminal steps of photosynthetic electron transfer, as light-activated electron flow produces NADPH for CO_2 assimilation. FNR also catalyzes the reverse reaction in photosynthetic organisms, transferring electrons from NADPH to Fd, which is important in cyanobacteria for respiration and cyclic electron flow (CEF). The cyanobacterium *Synechocystis* sp. PCC6803 possesses two isoforms of FNR, a large form attached to the phycobilisome (FNR_L) and a small form that is soluble (FNR_S). While both isoforms are capable of NADPH oxidation or $NADP^+$ reduction, FNR_L is most abundant during typical growth conditions, whereas FNR_S accumulates under stressful conditions that require enhanced CEF. Because CEF-driven proton pumping in the light–dark transition is due to NDH-1 complex activity and they are powered by reduced Fd, CEF-driven proton pumping and the redox state of the PQ and NADP(H) pools were investigated in mutants possessing either FNR_L or FNR_S. We found that the FNR_S isoform facilitates proton pumping in the dark–light transition, contributing more to CEF than FNR_L. FNR_L is capable of providing reducing power for CEF-driven proton pumping, but only after an adaptation period to illumination. The results support that FNR_S is indeed associated with increased cyclic electron flow and proton pumping, which is consistent with the idea that stress conditions create a higher demand for ATP relative to NADPH.

Keywords: cyanobacteria; cyclic electron flow; ferredoxin-NADP reductase; NDH-1; photosynthesis; proton motive force; thylakoid

1. Introduction

Cyclic electron flow (CEF) is an important mechanism in photosynthetic organisms as a means of recycling excess reductant while simultaneously driving the synthesis of ATP by the generation of proton motive force (PMF) [1–3]. In plants, algae, and cyanobacteria the reducing power produced by the photosynthetic light reactions is stored in the form of NADPH and Fd_{red}, with the former being consumed by anabolic processes, mainly CO_2 fixation via the Calvin–Benson–Bassham (CBB) cycle. This process is driven by linear electron flow (LEF) since electrons derived from the oxidation of water by photosystem II (PSII) follow a linear sequence of transfers through the photosynthetic electron transport chain, on through photosystem I (PSI), and ultimately terminating in the reduction of inorganic substrates for biomass production. During CEF, however, accumulated NADPH and Fd_{red} follow an alternative path that returns the energized electrons back to the membrane where they are oxidized by protein redox complexes in the electron transport chain, where depending on the pathway, they may drive the formation of PMF and thereby contribute to ATP production without net consumption of NADPH and Fd_{red}. The fraction

of electrons recycled through CEF relative to LEF allows the cells to adjust the output ratio of ATP/NADPH to accommodate different metabolic demands and fluctuating environmental conditions. Recently, cellular ATP concentrations were shown to modify the rate of CEF through a competitive inhibition of Fd oxidation by the NDH and FQR/PGR5 pathways in plant chloroplasts, providing a relatively simple feedback mechanism for adjusting this output ratio of phosphorylating and reducing power [4].

Although alternative cyclic paths involving different membrane complexes have been discovered, in most cases electrons are transferred to the plastoquinone (PQ) pool, thereby re-entering the electron transport chain to be further utilized to produce NADPH and Fd_{red} via PSI and subsequently recycled via CEF. Moreover, the overall CEF pathway may have many potential electron sources and contributing components. In cyanobacteria, the major routes by which electrons re-enter the PQ pool via CEF are NDH-1 and PGR5 [1,2,5–7]. Previously, in the dark–light transition NDH-1 was shown to be the major contributor to ΔpH formation by CEF in *Synechocystis* sp. PCC 6803, with the mutants unable to pump protons as efficiently after the loss of either only the respiratory-like NDH-1 complexes or loss of all NDH-1 complexes present [8]. In addition to this, pseudocyclic electron flow via the flavodiiron proteins, Flv1/3, may act as a way to dissipate electrons from Fd_{red}/NADPH to reduce O_2-yielding water, and is an important photoprotective mechanism, especially under conditions of fluctuating light and inorganic carbon limitation [3,9,10]. Because flavodiiron protein reduction of O_2 occurs in the cytoplasm, the contributions to proton motive force, and hence ATP production, are likely restricted to enabling unfettered photosystem II water-oxidation activity, with the concomitant release of protons into the thylakoid lumen [3,10].

Cyclic electron flow reentering the PQ pool can be directly from the acceptor side of PSI via Fd, or indirectly from the oxidation of photosynthetically produced NADPH, or carbohydrates [11–14]. Both of these CEF inputs contribute, but with different apparent kinetics, for example, in the post-illumination rise in Chl fluorescence yield [12]. One way the redox state of the NADPH/Fd pool is maintained is by the redox activity of Ferredoxin:NADP-oxidoreductase (FNR). In many cyanobacteria, only one form of FNR is present (FNR_L), and it attaches to the phycobilisome (PBS), poising it to function in photosynthetic two-electron reduction of $NADP^+$ by Fd_{red} from the acceptor side of PSI [15–19]. FNR_L localization in proximity of the photosynthetic membrane via attachment to the PBS is thought to facilitate fast rates of LEF. Specifically, it would minimize macromolecular diffusion constraints because the interaction of PSI, Fd, and FNR allows for the rapid production of quickly diffusing NADPH, which is in high demand and undergoes fast recycling between the oxidized and reduced forms as it communicates reductant from photosynthetic electron transport to the CBB cycle and anabolic metabolism in the cytoplasm [20].

Some cyanobacteria, such as the glucose tolerant *Synechocystis* sp. PCC6803, possess two isoforms of FNR. The second isoform is soluble and smaller, and therefore is referred to as FNR_S. It was discovered to be produced by the same gene [16] and is produced by an alternative initiation of translation, via mRNA secondary structure formation [21]. The isoform primarily observed in cyanobacteria is FNR_L, which may have arisen due to a gene recombination event of FNR_S with a PBS-linker domain, eventually resulting in the genetic conversion FNR_S to FNR_L in most of these bacteria [16]. This is consistent with the observation that *Gloeobacter violaceus*, a primitive cyanobacterium, possesses only the FNR_S form [16]. Cyanobacteria capable of heterotrophy, on the other hand, have retained the ability to recapitulate the FNR_S isoform through regulated alternative translational initiation codon utilization, and it is expressed under conditions of heterotrophy, or stress conditions such as high light intensity or nutrient limitation [16], which necessitates CEF activity [1]. In deletion mutants of either FNR_L (FS1) or FNR_S (MI6), similar amounts of total FNR accumulate compared to the WT, and though a small amount of FNR is detected in the smaller band in MI6, this is either a proteolytic cleavage product or a functionally insignificant product beginning at codon 102 [16]. The fact that these organisms have kept the ability to express the FNR_S isoform suggests that it is an important part of their

bioenergetic apparatus. Data on the kinetics of these isoforms has led to the hypothesis that FNR_S operates as a soluble NADPH oxidase, while FNR_L acts more as an $NADP^+$ reductase associated with the PBS, though both isoforms are capable of performing either reaction [17]. This hypothesis is supported by the impairment of photoautotrophic growth in the mutant lacking FNR_L [16]. Because cyanobacterial NDH-1 complexes utilize Fd as their source of reductant [14,22] and their contribution to proton pumping can be observed by AO fluorescence [8], the coupling of electron transport and proton pumping data can be used to explore the dynamics of the relationship between these two isoforms, CEF, and proton pumping. Here, this relationship is explored by utilizing these techniques and strains lacking either FNR_S or FNR_L.

2. Materials and Methods

Strains of the glucose-tolerant *Synechocystis* sp. PCC 6803 (hereafter, *Synechocystis*) were maintained on pH 8 BG-11 [23] with 1.5% agar supplemented with 18mM sodium thiosulfate and containing the 25 µg/mL spectinomycin for the FS1 and MI6 mutants. They were grown under ~70 µE m^{-2} s^{-1} Cool White fluorescent lighting (GE) and air levels of CO_2. Experimental material was obtained from 100 mL cultures grown in 250 mL Erlenmeyer flasks with rotary shaking (200 rpm) under ~100 µE m^{-2} s^{-1} Cool White fluorescent lighting (GE). Cultures were harvested upon reaching an OD_{750} 0.7–1.0. The strains MI6 and FS1 were constructed as described previously [16].

Cells were harvested by centrifugation, washed with 50 mM Tricine pH 8 + 25 mM KCl (TCK) and resuspended to 5 µg/mL Chl, measured by UV-Vis spectrophotometer (Shimadzu, Kyoto, Japan) [24]. Measurements of Chl and NADPH fluorescence kinetics were acquired using the Dual PAM-100 (Walz, Effeltrich, Germany) with the 9-AA/NADPH module. For 5 min illumination-period measurements, a nominal actinic light intensity of 53 µE m^{-2} s^{-1} and wavelength of 635 nm was utilized with multiple turnover (MT, nominally 20,000 µE m^{-2} s^{-1}) flashes occurring before, during, and after illumination to visualize the approximate proportion of open PSII [25]. Chlorophyll fluorescence was monitored with pulse amplitude modulated (PAM) LED excitation at 620 nm with detection at >700 nm. NADPH fluorescence was followed with PAM LED excitation at 365 nm and detection of emission within the broad blue-green band (420–580 nm) [26]. The application of an MT flash in NADPH fluorescence measurements produces an artifactual feature indicative of a very fast oxidation and a gradual increase back to steady state after the flash, but it is unclear to what extent this feature is artifactual nor is its basis understood, therefore it was omitted from the graphical representation of the data in order to focus on the known features.

Acridine orange (AO) has been used as a fluorescent dye for measuring proton pumping across a membrane in both whole cyanobacterial cells and membrane vesicles [8,27–30] and the assay protocol outlined previously in [8] was utilized with minor modifications. Cells were grown as described above and washed with TCK buffer as above for Chl fluorescence measurements. Samples were diluted to 5.9 µg/mL Chl in TCK buffer and acridine orange added to 5 µM, and were incubated while shaking gently in the dark for 20 min to allow acridine orange penetration into the cell [29]. Samples of the cells were then added to a cuvette and the cells diluted to 5 µg/mL in TCK buffer and the inhibitors applied as described previously [8]. Samples were then stirred for 5 min in the dark, and the fluorescence measured with the JTS-100 (Biologic, Seyssinet-Pariset, France), described in greater detail below. A 534/20 nm bandpass filter (Edmund Optics, Barrington, NJ, USA) was placed between the sample and the detector.

Data points in the JTS-100 were collected once per second in the dark and once per 100 ms in the light, with 600 µE m^{-2} s^{-1} actinic light applied. Cells incubating in the dark were provided with an additional 5 min dark incubation with stirring after the addition of inhibitors. Samples were then illuminated with actinic light for 15 s, and then actinic illumination was ceased. For light-adapted samples, the cells were incubated with the inhibitors for 5 min stirring in the dark, followed by 4 min of illumination with 600 µE m^{-2} s^{-1} actinic

light, followed by a further 4 min stirring in the dark before measurement in the same manner described above. Samples were refreshed after each measurement as drift in the measurements was seen when using TCK buffer.

3. Results

3.1. FNR$_S$ Has a Large Contribution to NDH-1 Cyclic Electron Flow

Chlorophyll fluorescence kinetics have long been used to observe electron flow through the PQ pool in cyanobacteria [31,32], and the post-illumination kinetics are especially indicative of NDH-1 driven CEF [1,12–14,33,34]. In Figure 1, the chlorophyll fluorescence traces for the WT, MI6 (containing only FNR$_L$), and FS1 (containing only FNR$_S$) are shown. During the 5 min illumination in the TCK assay buffer, WT and FS1 had similar initial fluorescence kinetics immediately upon illumination, but the oxidation of PQ, indicated by fluorescence quenching, was outpaced by reduction of the PQ pool in FS1, which potentially obscured the small peak at ~20–30 s of illumination corresponding to activation of the CBB cycle in the light (Figure 1A,E) [12]. The WT then maintained a steady state of PQ redox after the CBB cycle activation, while FS1 showed a slow reduction in the PQ pool over illumination toward what appears to be close to a steady state with a comparatively more reduced PQ pool. The MI6 mutant, while showing similar broad features as the WT, exhibited a large initial fluorescence yield which was evident in the saturating flash in the dark period and by the large amplitude of the fluorescence yield upon continuous illumination. We attribute the high fluorescence yields in MI6 to the redox poising of cells in State 1, much like mutants of the NDH-1 complex, where a relatively oxidized PQ pool is observed in the dark [35]. Upon continuous illumination there is reduction in the PQ pool upon illumination followed by steady oxidation over the course of illumination toward values nearer the WT than FS1 (Figure 1A,C,E). This consistent with MI6 operating primarily via linear electron transport, although the influence of state transitions and other quenching processes are likely to also contribute to the progress quenching after the initial S-M rise [36,37]. In contrast to either the wild-type or MI6, there is a progressive increase in fluorescence yield in FS1, indicating that the FNR$_S$ isoform is highly efficient at Fd$_{ox}$ reduction and subsequent electron transfer to the PQ pool transfer via the NDH-1 complexes [14,22]. In this context, we note that pseudocyclic electron transfer to O$_2$ via the flavodiiron proteins occurs in the cytoplasm, thereby dissipating electrons from the entire system and tending oxidize the PQ pool as the electrons have no return path.

Illumination was ceased at 360 s, and the post-illumination increase in fluorescence yield is attributed to CEF. The WT and MI6 have very minimal fluorescence rises while FS1 has a large rise that appears to include both the fast and slow components. The fast components (<10 s post-illumination) are associated with direct CEF involving NADPH and reduced ferredoxin, whereas the broad slow component (~30 s post-illumination) reflects reoxidation of photosynthetically produced sugars (Figure 1A,C,E) [12]. This indicates that in the WT and MI6 under these conditions, electrons are not flooding back into the PQ pool via redox complexes, primarily NDH-1 [8,14], after illumination, or if they are, they are being removed from the PQ pool at a similar rate to that which they are entering. This is not the case in FS1, where electrons flood the PQ pool after illumination, and the input of electrons into the PQ pool outpaces the consumption of those electrons after illumination indicated by the pronounced post-illumination rise in fluorescence relative the wild type (Figure 1E). This reflects the hypothesized role of FNR$_S$ in its participation in CEF [16,18].

When KCN is added to the samples, preventing cytochrome oxidase (COX) activity, stark changes are seen in the fluorescence yield. Over the course of the 5 min illumination, WT starts out with a very reduced PQ pool, which becomes more reduced over the illumination so Chl fluorescence reaches to near the F$_m$ level (Figure 1B). MI6 shows a large initial reduction in PQ upon illumination, followed by the typical quenching associated with CBB cycle activation (Figure 1D), which is followed by a slight increase in fluorescence over the course of illumination, indicating a slight reduction in the PQ pool. FS1 shows the typical initial reduction followed by a brief oxidation and then consistent steady reduction of the

PQ pool over the course of illumination (Figure 1F). The post-illumination fluorescence rise is similar in the WT and MI6 in terms of magnitude. This rise is further investigated in Figure 2, compiling the post-illumination period of the measurements in Figure 1B,D,F. As can be seen, FS1 has the sharpest initial rise after illumination and reaches the highest fluorescence yield. This initial rise is also seen in MI6 and the WT, but not to the magnitude seen in FS1. After this initial rise, a second rise occurs over the course of ~12–30 s after the cessation of illumination. This slower rise takes place at similar rates for all the strains in these conditions, but the peak of the fluorescence yield and its dissipation is shifted in FS1, where even a slow dissipation does not start until ~30 s after cessation of illumination versus ~20 s in WT and MI6 (Figure 1B,D,F and Figure 2). This shows that FNR_L, the form present in MI6 and predominant in the WT, may participate in Fd_{ox} reduction for CEF, and that FNR_S appears to be more effective at this function than its counterpart.

To further investigate the post-illumination Chl fluorescence rise in the absence of inhibitors, 15 s illuminations on dark-adapted samples were performed. Illuminations of this length are long enough to generate a substantial pool of photosynthetic reductant as NADPH and reduced ferredoxin, yet too short for the full activation of the CBB cycle. The FS1 and WT had similar initial slopes of the post-illumination fluorescence rise; however, the FS1 had a greater magnitude, which is associated with greater CEF activity, especially via NDH-1 [13] (Figure 3). The MI6 mutant, on the other hand, had a small rise, with dissipation of the post-illumination fluorescence occurring well before either FS1 or the WT had reached their peak and begun their own dissipations (Figure 3). These results indicate that FS1 has enhanced CEF, MI6 has diminished CEF, and the WT is somewhere in between the two. Taken together, these results indicate that only having the FNR_L in MI6 tends to shunt the reductant to CO_2 reduction via the CBB with minimal CEF with the resultant more efficient oxidation of the PQ pool. With the ability to express only the short form of FNR_S in FS1, CEF is maximized, resulting in enhanced re-reduction of the PQ pool at the expense of LEF.
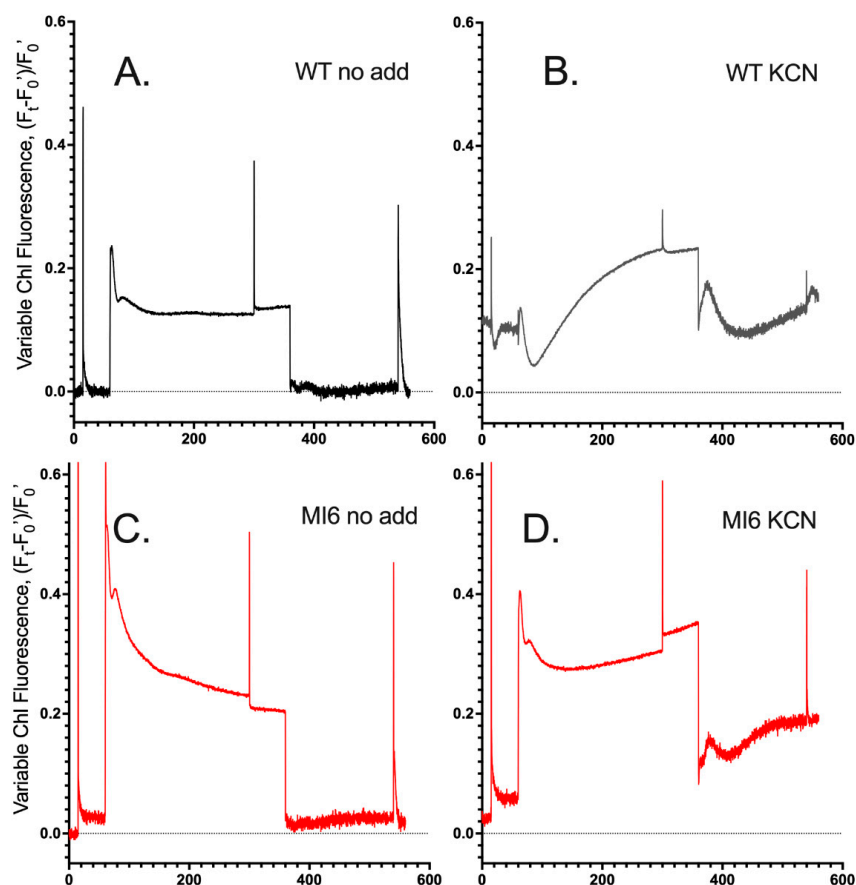
Figure 1. *Cont.*
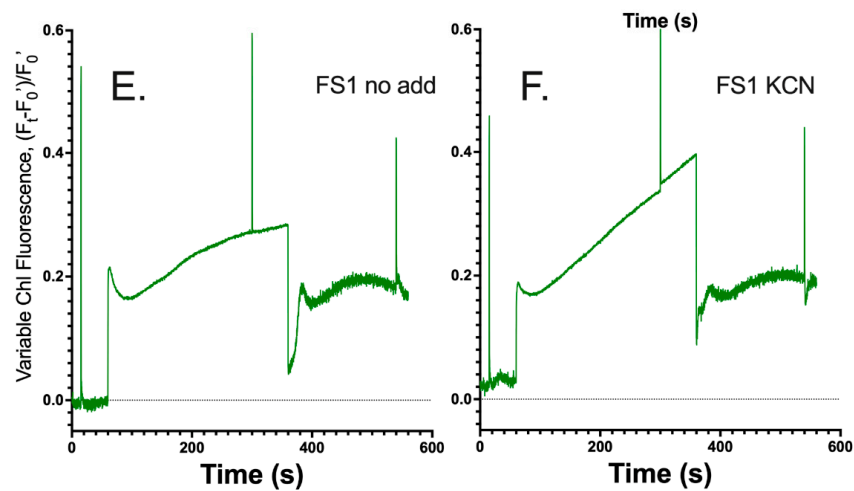
Figure 1. Chlorophyll fluorescence kinetics are distinct in strains lacking either FNR isoform. (**A**): Chlorophyll fluorescence in the DUAL PAM-100 of dark-adapted WT cells in TCK buffer upon illumination for 5 min with 53 μE m^{-2} s^{-1} actinic light with multiple turnover flashes before, during, and after illumination. (**B**): Chlorophyll fluorescence of the WT after treatment with KCN. The trace of the WT with and without KCN is repeated in panels (**C–F**) in grey for comparisons. (**C**): MI6 strain. (**D**): MI6 after treatment with KCN. (**E**): Chlorophyll fluorescence of the FS1 strain. (**F**): Chlorophyll fluorescence of FS1 after treatment with KCN.
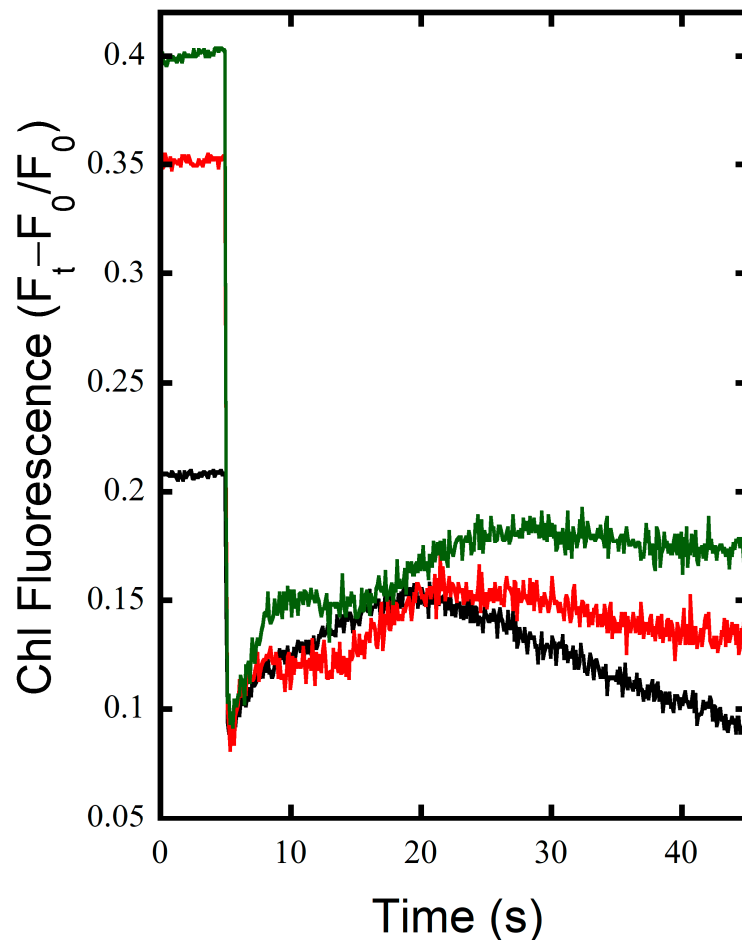
Figure 2. Post-illumination chlorophyll fluorescence rise is enhanced after addition of KCN. Chlorophyll fluorescence in the DUAL PAM-100 of WT (black), MI6 (red), and FS1 (green) cells in TCK buffer with 200 μM KCN upon termination of 5 min illumination with actinic illumination of 53 μE m^{-2} s^{-1}.
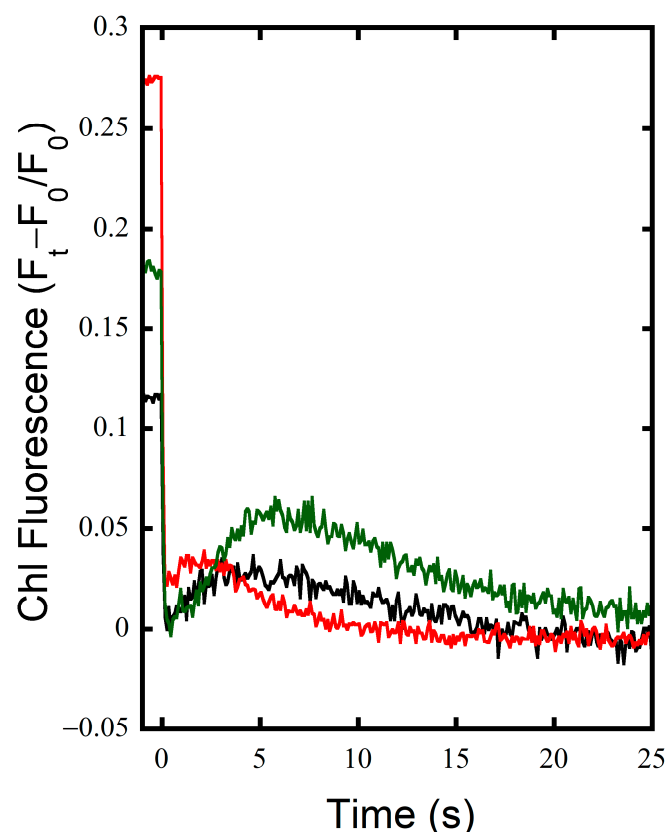
Figure 3. Post-illumination chlorophyll fluorescence rise after short illumination with no inhibitor additions. Chlorophyll fluorescence in the DUAL PAM-100 of dark-adapted WT (black), MI6 (red), and FS1 (green) cells in TCK buffer after actinic illumination with 53 μE m^{-2} s^{-1} for 15 s.

3.2. FNR$_S$ Enhances NADPH Oxidation during Illumination

NADPH fluorescence was measured concomitantly with chlorophyll fluorescence in the DualPAM-100 to observe the redox state of the soluble reductant pool during illumination. The MT flashes produce an artifact that was excluded in the graphics to emphasize the kinetics of the NADPH redox poise on in the seconds time frame. As seen in Figure 4A, the WT exhibits characteristics typical of NADPH fluorescence during an illumination of this length [12]. The rises in fluorescence indicate the reduction of NADP$^+$, and the quenching of fluorescence is indicative of NADPH oxidation. After the rapid reduction of the pool upon illumination, a small oxidation event is seen before a large transient reduction associated with the production of NADPH and representative of the State-2 to State-1 transition, where light harvesting energy is focused around PSII over PSI [38–41]. This is followed by a subsequent oxidation of NADPH associated with the activation of the CBB cycle and fixation of carbon after ~30 s of illumination [12]. After illumination ceases, there is a rise in fluorescence that starts ~10 s after illumination and continues for ~15–20 s, which is associated with the oxidation of sugars [12]. The delay before the rise correlates with the oxidation of trioses seen in the post-illumination rise of Chl fluorescence in KCN-treated cells (Figures 1 and 2), indicating that the delay in the rise is associated with Fd reduction before NADP$^+$ reduction by catabolism begins. The MI6 mutant has similar characteristics as the WT during and post-illumination; however, the overall NADPH pool was not as strongly reduced in the light, and its post-illumination rise is slightly less dramatic in magnitude (Figure 4C). The FS1 mutant on the other hand had distinct characteristics in its NADPH redox state over the course of illumination (Figure 4E). It had an initial reduction followed by a strong oxidation of NADPH to a steady state that was deeply oxidized compared to the WT or MI6, and even dropped below the baseline of fluorescence pre-illumination (Figure 4E). The transient reduction peak is shifted to ~15 s earlier than

it is in the WT or MI6, potentially indicating changes in the regulation of the redox pools. After illumination, a large and dramatic rise was seen in FS1 to approximately the level of the initial NADPH reduction peak in a biphasic manner, first occurring quickly and then rising slowly to a point before minor oxidation is observed in the dark. This occurred on approximately the same timescale as the WT or MI6, indicating the same mechanisms may be responsible for this dramatic rise, but perhaps with a greater flux of electrons through them than in either WT or MI6. This indicates that there is a strong oxidation of NADPH by FNR_S that is not observed when FNR_L is the dominant isoform.
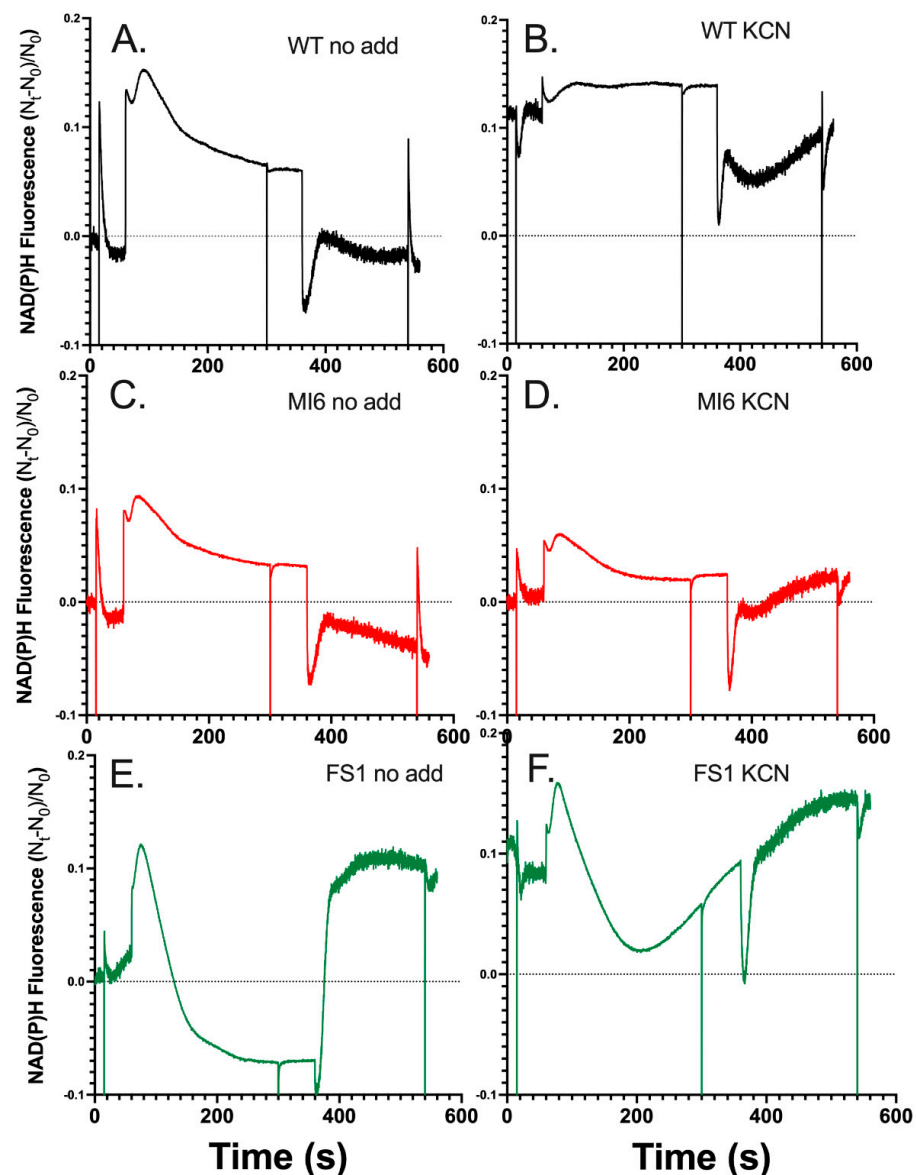
Figure 4. NADPH fluorescence in the WT and strains lacking either FNR isoform. (**A**) NADPH fluorescence in the DUAL PAM-100 of dark-adapted WT cells in TCK buffer upon illumination for 5 min with 53 μE m^{-2} s^{-1} and multiple turnover flashes before, during, and after illumination (cropped here for emphasis). (**B**) NADPH fluorescence of the WT after treatment with KCN. (**C**) NADPH fluorescence of the MI6 strain. (**D**) NADPH fluorescence of MI6 after treatment with KCN. (**E**) NADPH fluorescence of the FS1 strain. (**F**) NADPH fluorescence of FS1 after treatment with KCN.

Upon the addition of KCN, the WT almost entirely reduced its NADPH pool in the dark, as upon illumination there is little increase in fluorescence (Figure 4B). There is a transient oxidation, and then NADPH fluorescence rises slowly and reaches a steady

state for the rest of illumination. Given that the PQ pool is almost entirely reduced over the course of illumination (Figure 1B), NADPH might be as reduced as it can be as well, representing an acceptor side limitation. The MI6 has similar characteristics of fluorescence as before treatment with KCN; however, the initial rise in fluorescence upon illumination is diminished (Figure 4D). It also does not show strong reduction in the dark, or even during illumination, perhaps showing NADPH oxidase activity. A decrease in fluorescence is seen after the transient peak, indicating dissipation of reductant by CO_2 fixation. At the beginning of post-illumination NADPH fluorescence of MI6 rises to a greater magnitude than the WT but with a similar slope, and the pool continued to reduce over time. In the dark, FS1 had a similarly reduced NADPH pool to the WT, and interestingly, upon illumination and after its transient peak, had a strong oxidation of the NADPH pool, though not quite as quickly or deeply as without KCN, followed by a slow and steady rise in fluorescence which seems to continue at a similar rate after the oxidation and subsequent reduction of NADPH after termination of illumination (Figure 4F). The post-illumination rise was much smaller in magnitude than without KCN treatment, but still greater compared to the other strains (Figure 4E,F). Together with the chlorophyll fluorescence data (Figure 1), the NADPH fluorescence data indicate that FNR_S pulls electrons away from the NADPH pool and allows their input into the PQ pool to a greater degree than does FNR_L. While it is difficult to compare the physiological conditions exactly, this accords with the previous findings that the $NADP^+/NADPH$ ratios were about 2.6 and 2.0 for the wild type and MI6, respectively. In contrast, the reported ratio was about 4.3 in FS1, indicating a much more oxidized NADPH pool [17].

3.3. The Presence of FNR_s Enhances NDH-1 Powered Proton Pumping

To investigate the ability of these strains to power CEF-driven proton pumping and determine if Fd is indeed being reduced by the strong NADPH oxidation seen in FS1 to power CEF, ΔpH formation upon actinic illumination was measured by acridine orange (AO) fluorescence quenching. This assay was performed essentially as described previously [8]; however, the buffer used included 25 mM KCl (TCK buffer) rather than 25 mM NaCl (TCN). Upon switching the buffer, the rates of AO quenching dramatically reduced from those reported in TCN buffer (Supplementary Figure S1). The fast rate of quenching in the TCN buffer was shown to be Na^+ gradient-dependent, as the addition of 50 μM monensin dissipated the gradient, reducing the rates to approximately those seen in TCK buffer, indicating that the Na^+ gradient plays a part in light-induced PMF formation, perhaps by modulating the concentration of positive charges across the thylakoid. As discussed below, subtler changes in AO fluorescence quenching may be observed in TCK buffer, and it enhances the proportion of PMF formation driven by CEF. Because NDH-1 complexes have been shown to be the primary contributors to CEF-driven proton pumping in the dark–light transition in cells treated with DCMU [8], and their primary reductant source is reduced Fd [14], the action of FNR_S as an NADPH oxidizer should be most clearly seen when PSII is inhibited. While the action of an alternative CEF pathway such as PGR5 is unable to be excluded in its activity here, based on previous data it is likely that in these conditions the ΔpH formation is due to CEF via NDH-1 complexes [8]. The background inhibitors KCN, valinomycin, and DCCD were utilized as described previously to allow for the measurement of near-maximal rates of proton pumping [8]. In dark-adapted cells, FS1 has a significantly faster rate of quenching than the WT with only the background inhibitors, but the rates are not significantly different when DCMU is added to dark-adapted cells ($p = 0.88$) (Figure 5A,B and Table 1). The mutant MI6, lacking FNR_S, was not significantly different from the WT with only the background inhibitors in dark-adapted cells ($p = 0.35$); however, it was significantly slower than either the WT or FS1 when DCMU was added (Figure 5C and Table 1). The ability for CEF to compensate for the loss of LEF in MI6 is greatly reduced from that of the WT or FS1, operating at ~27% of the rate of the no-DCMU control, while WT and FS1 operate CEF-driven proton pumping at ~50% and 39% of the rate, respectively. These data indicate that in dark-adapted cells FNR_S is a better supplier

of Fd_{red} for NDH-1-driven proton pumping than FNR_L is, and that the WT can match the rate of FS1 due to its limited presence of FNR_S.

(A) (B) (C)
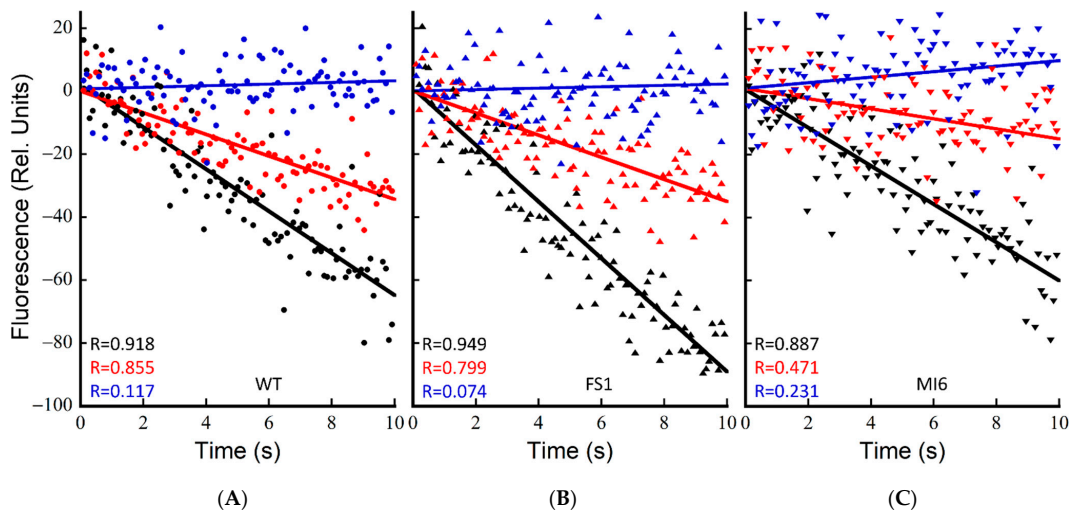
Figure 5. Rates of acidification, $j_H{}^+$, upon illumination of dark-adapted WT, FS1, and MI6. Acridine orange fluorescence of dark-adapted cells in the JTS-100. Cells were dark-adapted with acridine orange (5 µM) for 20 min in TCK buffer. Sample was prepared with Val (10 µM), DCCD (500 µM), and KCN (200 µM) and stirred in the dark for 5 min as the background inhibitors. Actinic illumination (630 nm, 600 µE) was applied for 15 s for measurement. These plots are typical of 3 technical replicates. WT (**A**), FS1 (**B**), and MI6 (**C**) after treatment with the background inhibitors (black symbols), 10 µM DCMU (red symbols), or DCMU and 250 µM CCCP (blue symbols). Data were normalized to the [Chl]/OD of the WT to account for differences in [Chl] per cell.

Table 1. Relative rates of acidification in dark- and light-adapted states. The rate of acidification, $j_H{}^+$, upon illumination in WT, FS1, and MI6 in the presence of the background inhibitors with either no additional inhibitors, with DCMU, or DCMU and CCCP added. Values are averages of three technical replicates per condition. Data are presented with the standard deviation. Data normalized to [Chl]/OD$_{750}$ of the WT to account for differences in the Chl content per cell in the different strains. Rates are divided into either dark-adapted conditions (top) or light-adapted conditions (bottom). Significance between measured rates are indicated by paired symbols. p values: † = 0.0003, * = 0.0003, ‡ = 1×10^{-8}, ** = 0.0003, ℂ = 0.02, ℓ = 0.004, ◇ = 0.02, • = 6×10^{-5}, ¤ = 0.005, + = 0.01, ¢ = 5×10^{-5}, ø = 0.05, § = 0.01, ℰ = 0.007, ⅏ = 0.001.

	Inhibitors Added	**Wild-Type**	**FS1**	**MI6**
Dark-adapted	Val + DCCD + KCN	6.67 ± 0.65 s^{-1} †+	9.00 ± 1.65 s^{-1} ‡+ ℰ	6.08 ± 1.27 s^{-1} ℂ• ℰ
	Val + DCCD + KCN + DCMU	3.44 ± 1.15 s^{-1} *†§	3.52 ± 0.48 s^{-1} **‡ℓ	1.63 ± 1.00 s^{-1} ◇•§ℓ
	Val + DCCD + KCN + DCMU + CCCP	-0.26 ± 0.46 s^{-1}	-0.022 ± 0.28 s^{-1}	-0.88 ± 0.55 s^{-1}
Light-adapted	Val + DCCD + KCN	8.95 ± 2.57 s^{-1}	10.60 ± 2.04 s^{-1}	9.36 ± 2.59 s^{-1} ℂ¤
	Val + DCCD + KCN + DCMU	6.81 ± 0.50 s^{-1} *¢ø	9.45 ± 0.72 s^{-1} **¢ ⅏	4.43 ± 2.22 s^{-1} ◇¤ø ⅏
	Val + DCCD + KCN + DCMU + CCCP	-0.08 ± 0.97 s^{-1}	0.21 ± 0.27 s^{-1}	-1.61 ± 1.67 s^{-1}

Upon repeated illumination of a sample in TCK buffer, increasing rates were seen in subsequent samples (not shown). This was hypothesized to be due to an increasing reductant pool feeding CEF proton pumping and allowing it to carry forward faster. To test this, cells were treated with the background inhibitors and DCMU/CCCP where appropriate, pre-illuminated for 2 min at 600 µE, let to sit in the dark for 4 min with stirring, and then subsequently measured for typical 15 s illuminations. As seen in Figure 6 and Table 1, pre-illumination of the cells resulted in increases in rates of proton pumping, especially in samples with DCMU. The proton-pumping rate upon DCMU treatment in light-adapted cells is nearly twice the rate of dark-adapted cells, while FS1 is ~2.7× as fast vs. dark-adapted cells. The MI6 increased in rates as well, essentially matching those of

WT and FS1 with only the background inhibitors (Figure 6 and Table 1). FS1 had the fastest proton-pumping rate with DCMU added, indicating that FNR_S is important for driving proton pumping via CEF (Figure 6C and Table 1). By treating the cells with illumination prior to measurement, the ability of the cells to compensate for the loss of LEF by DCMU treatment increased in all cases. The WT was able to achieve ~76% the light-adapted control rate when treated with DCMU, while it could only achieve ~50% the rate when dark-adapted (Figure 6A and Table 1). When light-treated, MI6 was able to achieve a rate of ~50% the control when treated with DCMU, while it could only achieve ~27% when dark-adapted (Figure 6B and Table 1). FS1 was able to achieve ~89% the rate of the control when light-adapted and treated with DCMU, indicating that in these conditions CEF-driven proton pumping may nearly entirely compensate for the loss of LEF in terms of PMF generation. Clearly, the abundance of FNRs in FS1 is able to drive high rates of CEF and the resultant NDH-1 mediated proton pumping.
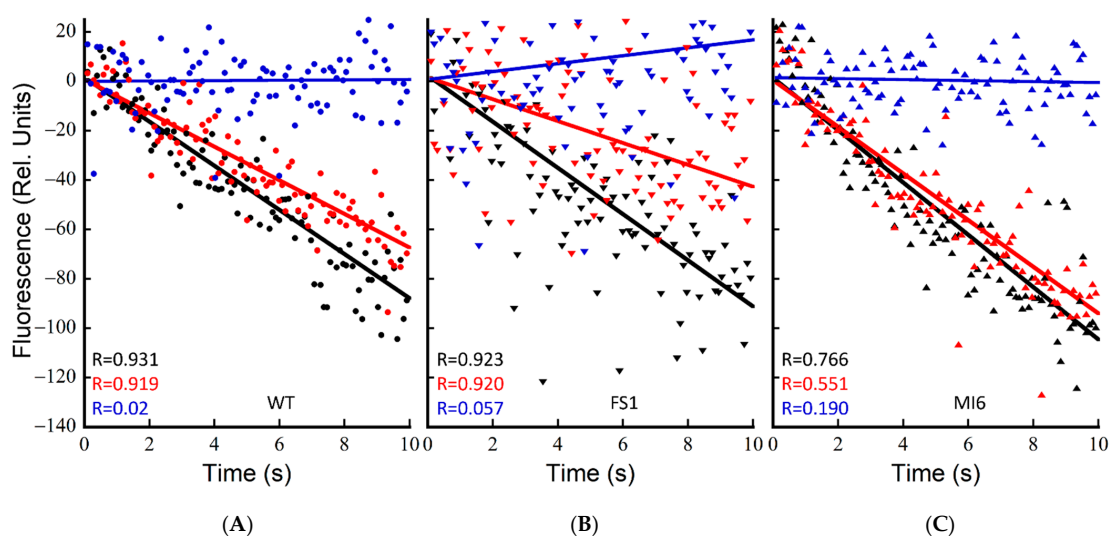

Figure 6. Rates of acidification, $j_H{}^+$, upon illumination of light-adapted WT, FS1, and MI6. Acridine orange fluorescence of dark-adapted cells in the JTS-100. Cells were dark-adapted with acridine orange (5 μM) for 20 min in TCK buffer. Sample was prepared with Val (10 μM), DCCD (500 μM), and KCN (200 μM) and stirred in the dark for 5 min as the background inhibitors. After the addition of inhibitors, cells were pre-illuminated with 600 μE actinic light for 2 min, and cells afforded 4 min in the dark before measurement with 15 s of illumination at the same light intensity. These plots are typical of 3 technical replicates. WT (**A**), FS1 (**B**), and MI6 (**C**) after treatment with the background inhibitors (black symbols), 10 μM DCMU (red symbols), or DCMU and 250 μM CCCP (blue symbols). Data were normalized to the [Chl]/OD of the WT to account for differences in [Chl] per cell.

4. Discussion

The FNR isoforms in cyanobacteria, despite having small differences in their kinetics in vitro [17], have very different roles in the regulation of the redox state of the NADPH/Fd pools. It has been shown before that the FS1 has a higher $NADP^+$/NADPH ratio and that the mutant does not exhibit a changed PSII/PSI ratio [17], indicating that the FNR_s isoform is involved in the reduction of Fd over the reduction of NADPH when it is the primary isoform. This is not the case in the WT, which under normal growth conditions has a larger proportion of FNR_L [16], the isoform possessed by the MI6 mutant. The FNR_S isoform is maximal during nutrient or light-stress conditions [8]. When the WT was exposed to conditions of high light or low CO_2, conditions that enhance CEF [42], the FNR_S isoform constitutes approximately $\frac{1}{4}$ of the total amount of FNR in species that can express both isoforms [21]. Presumably, the stress conditions shift the relative demand for the two primary products of the light reactions of photosynthesis: ATP and NADPH. Under stress conditions the demand of ATP increases relative to NADPH, resulting in the observed

regulatory modulation in the expression of FNR_S [16], which may work in parallel with more direct metabolic control mechanisms, such as the interaction of ATP with the NDH-1 complex [4].

Here, it was shown that CEF is enhanced in FS1, which contains only FNR_S (Figures 1E and 2). This was determined by observing the post-illumination rise of chlorophyll fluorescence which has two main components: a fast rise associated with CEF and NDH-1, and a slower rise associated with carbohydrate oxidation [12]. In FS1, a 5 min illumination produces reduction in the PQ pool over the course of illumination, rather than oxidation to a steady-state like in the WT as suggested by the Chl fluorescence transients shown in Figure 1A,E. This is indicative that there are electrons rushing into the PQ pool over the course of illumination in FS1 potentially as a result of recycling of electrons via CEF occurring during illumination such that the rise after the initial oxidation seen in WT and MI6 is either delayed heavily or never reached. However, this interpretation is complicated by the fact that Chl fluorescence yields are simultaneously being modulated by various other processes. These other processes include photoinhibition, orange carotenoid protein quenching, and other regulatory short-term processes, and also appear to be during the decay of fluorescence yields, especially after the M-T rise in *Synechocystis* sp. PCC 6803 [36,37,39–41]. After illumination, there is a strong rise in FS1, but not in the WT or MI6 (Figure 1A,C,E), indicative of enhanced electron flow to the PQ pool upon cessation of illumination. When treated with KCN, the PQ pool is generally reduced prior to illumination, yielding a high F_0, and the PQ pool reduces over the course of illumination, to a point near $F_{m'}$ in the WT, slightly in MI6, and slowly and constantly in FS1 (Figure 1B,D,F). After illumination, each of the strains experiences a strong post-illumination rise at similar rates; however, FS1 had the largest rise of the three strains (Figure 2). To observe the post-illumination rise without inhibition of COX, a short illumination with an MT flash was performed on dark-adapted cells, and again FS1 was shown to have the largest rise in terms of magnitude, and rises at approximately the same rate as the WT (Figure 3, black and green). The MI6 strain had a small shoulder that was quenched well before FS1 or the WT started their PQ pool oxidation (Figure 3, red). These results support the hypothesis established previously [16–18] that FNR_S contributes more strongly to cyclic electron flow than FNR_L does. Because the post-illumination rise is associated with NDH-1 activity [8,14], and because NDH-1 complexes depend on Fd_{red} to power their proton pumping activity [14,22,43], the oxidation state of the NADPH pool during illumination was investigated side-by-side with chlorophyll fluorescence.

The WT and MI6 had similar kinetics of NADPH fluorescence in the absence of KCN, with the WT having overall higher fluorescence during illumination, indicating a slightly more reduced NADPH pool during illumination (Figure 4A,C). Their post-illumination rises, associated with carbohydrate oxidation, are similar in both, being slightly larger in the WT. The FS1 mutant, on the other hand, had strikingly unique kinetics. It had a sharp rise in fluorescence after illumination, shifted ~15 s earlier than the peaks seen in WT and MI6 (Figure 4E). This might be FNR_S operating as an NADPH reductase early in illumination due to a large influx of photosynthetically reduced Fd produced by LEF. Because the CBB cycle is regulated by the redox state of the NADPH pool, activating upon heavy reduction [44], this could explain the shifting of the redox state of the NADPH pool by CBB activity to its earlier point in FS1 (Figure 4E). There is steady oxidation of NADPH after this initial peak that is sharper in its fluorescence decline than in the WT or MI6. This is likely due to CBB cycle activity drawing down NADPH, the activity of flavodiiron proteins and COX, as well as Fd reduction by FNR_S. This oxidation state of the NADPH pool then reaches a steady state where it is consumed as fast as it is produced with a higher $NADP^+/NADPH$ ratio than the WT or MI6 in the same conditions, matching earlier data [17,18]. Post-illumination fluorescence kinetics of FS1 are also interesting in these conditions, with a large post-illumination fluorescence rise (Figure 4E). It is unlikely that this peak can only be due to increased sugar-oxidation activity due to its magnitude; therefore, FNR_S may be acting to dissipate the reduced Fd now that it is no longer being

photosynthetically produced. Upon treatment with KCN, the NADPH pool of the WT is almost entirely reduced in the dark, with little increase in fluorescence seen when illuminated (Figure 4B). This is maintained throughout illumination and is followed by a small post-illumination rise. The MI6 has similar kinetics as it had before treatment with KCN; however, fluorescence does not increase as much upon illumination (Figure 4D). It reaches a steady state after the slow oxidation of NADPH by the CBB cycle, and post-illumination experiences a large increase in fluorescence. The FS1 had a fast reduction in NADPH followed by a rapid oxidation, although this was not as rapid as without KCN (Figure 4F). After ~2.5 min of illumination a trough is seen in the fluorescence, and it slowly and steadily rises for the rest of illumination. Due to the lack of COX as an electron sink in the dark, it is possible that the slower oxidation followed by steady reduction in NADPH could be due to a strong reduction in both NADPH and Fd pools in the dark, which would result in a slower exchange of electrons between the two as they become further reduced. The slow rise could then be the FNR_S operating to reduce the NADPH pool in conditions where Fd is over-reduced.

The establishment of proton motive force in cyanobacteria is dependent upon photosynthetic electron transport upon illumination. When PSII is inhibited, the major driver of proton pumping by CEF-driven electron transport involving NDH-1 [8]. Because NDH-1 is dependent upon reduced Fd to supply electrons [14,22,43], the ability for the proton gradient to be established in the absence of PSII activity may be indicative of the Fd reduction state. When dark-adapted, the WT has its fastest rate when PSII is active, with the addition of DCMU decreasing the rate by ~50% (Figure 5A and Table 1). FS1 was similar in rates to the WT when DCMU was added, but significantly faster ($p = 0.01$) with only the background inhibitors, including KCN, which yields a more reduced NADPH pool in the dark, potentially providing an increased pool of reductant for Fd reduction (Figures 4E, 6B and Table 1). The rates in MI6 were similar to the WT with only the background inhibitors, and significantly slower than the WT with DCMU added ($p = 0.01$) (Figure 5C and Table 1), achieving only ~47% the rate of the WT when PSII is inactive. When PSII is inhibited, the rate of MI6 is only 27% the rate of MI6 without DCMU, indicating a severe inhibition of CEF in these conditions, likely due to a lack of reduced Fd for consumption. These data indicate that FNR_S is an important component in the shifting from dark to light conditions, providing reductant for CEF and contributing to the formation of PMF, even when PSII is active.

Because repeated illumination of the same sample was seen to cause increases in the rate of AO quenching, a pre-illumination of 600 μE m^{-2} s^{-1} over a period of 2 min was applied to the cells, followed by darkness for 4 min, and then measurement of the typical 15 s illuminations at 600 μE m^{-2} s^{-1}. This caused rate increases in all the strains measured. The rates in the WT increased by ~1.3× without DCMU and by ~2× with DCMU when pre-illuminated, which when paired with the NADPH fluorescence data, indicates that a heavily reduced NADPH pool is beneficial to CEF-driven PMF generation (Figures 4B and 6A, Table 1). The rate increases in the MI6 mutant were quite dramatic, and without DCMU allowed it to nearly match the other strains. While pre-illumination still improved its rate when treated with DCMU, it remained slower than the other strains (Figure 5C and Table 1). This could be indicative that the FNR_L isoform is not able to perform NADPH oxidation in vivo as FNR_S. The FS1 mutant, had the most dramatic increase in rate with light adaptation, with a DCMU-treated rate 2.7× that in the dark-adapted state with DCMU (Figure 5B and Table 1). When compared to the rate with only the background inhibitors, FS1 was able to achieve ~89% the quenching rate, enough to be not significantly different in rate than the rate without DCMU ($p = 0.24$). These data indicate that the pre-illumination period in the presence of the background inhibitors results in the accumulation of reductant, which may then be consumed to generate PMF. Because the FS1 strain, containing only FNR_S, has such a dramatically increased rate compared to its counterparts, the FNR_S isoform must be providing reduced Fd for consumption by NDH-1 to produce PMF, as it also possesses a heavily oxidized NADPH pool (Figure 4E,F). Because the pre-illumination also increased the rate in MI6 to near WT levels and is the

primary FNR isoform present in cells in normal growth conditions, FNR_L may be able to provide reduced Fd for CEF and is the likely isoform doing so in the WT. It is, however, limited in its capacity to do so without a period of illumination to adapt to the new redox conditions. This indicates that these isoforms have different functions in maintaining the redox state of the NADP(H)/Fd pools. Because of the improved CEF-driven PMF formation in dark-adapted cells treated with DCMU that the WT possesses over MI6, while being not significantly different from FS1, it is possible that the WT utilizes the small amount of FNR_S it possesses in the dark–light transition.

The presence of alternately regulated FNR isoforms in cyanobacteria capable of heterotrophy suggests that they have different roles in their maintenance of the NADPH/Fd redox state (Figure 7). Because the NADPH pool is oxidized over the course of illumination in a mutant expressing only FNR_S (Figure 4E), it has enhanced CEF (Figures 2 and 3), and its presence allows for enhanced CEF-driven proton pumping both with and without adaptation to light (Figures 5 and 6, Table 1). This indicates that FNR_S is primarily involved in NADPH oxidation/Fd reduction, as reduced Fd is the source of reductant for NDH-1 complexes, the primary driver of proton pumping in the dark–light transition [8]. This also suggests that FNR_S is useful in the transition from dark to light, allowing faster adaptation to the light and establishing PMF more quickly to utilize the energy produced. When FNR_L is the only isoform present, the establishment of PMF is slower, but upon adaptation to illumination becomes faster in rate of AO quenching to values close to those observed in the WT. The data then suggest that FNR_S is useful for allowing for fast adaptation to illumination and likely provides reductant for proton pumping through carbohydrate oxidation in the dark, and FNR_L is useful for long-term adaptation to illumination and may be used to reduce Fd for CEF given some time in the light. The roles suggested here also raise a question of whether cyanobacteria that possess both FNR_L and FNR_S (those capable of heterotrophy) have a more robust response to fluctuating light conditions than their cousins that only possess FNR_L.
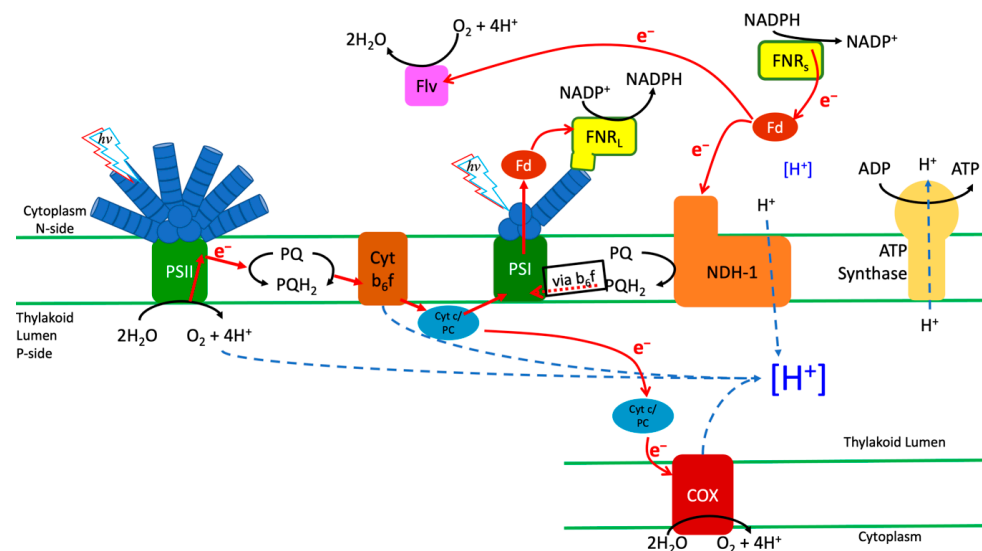
Figure 7. Schematic representation of the electron and proton transport pathways in *Synechocystis* sp. PCC6803 in relation to the large (FNR_L) and the small soluble (FNR_S) forms of the FNR protein (yellow). These alternative forms of FNR shift the balance of electron transfer (red arrows) toward either cyclic electron flow or linear electron flow. The mutant FS1, having exclusively FNR_S, tends to favor electron flow from NADPH to Fd_{ox}, which can donate to either the NDH-1 complex or flavodiioron proteins (e.g., Flv2/3). Either of these FNR_S-favoring pathways can lead to loss of electron from the system by reduction of O_2 either via cytochrome oxidase (COX) or flavodiioron proteins (Flv). The phycobilisome-attached FNR_L appears to be primarily involved in linear electron flow, favoring electron transfer from Fd_{red} produced by PSI to NADPH for CO_2 fixation.

5. Conclusions

The FNR isoforms present in cyanobacteria capable of heterotrophy present an interesting and poorly understood system for regulating the redox state of the NADPH/Fd pools. FNR$_S$ is utilized in vivo most generally as an NADPH oxidase in the light, allowing for rapid adaptation to the dark–light transition and quick production of a proton gradient. This may allow these bacteria to acclimate more rapidly to altered conditions of photosynthetic electron, flow maintaining redox balance and powering of NDH-1-mediated CEF proton pumping to augment ATP production to match increase demands due to stress. Under typical growth conditions, production of the FNR$_L$ isoform is favored. As the FNR$_L$ isoform is mostly involved in LEF and the reduction of NADP$^+$, it appears to be less effective in reducing Fd even when enough reductant as NADPH is available. The fact that only a subset of cyanobacteria is capable of expressing both isoforms suggests that there is a functional reason why these isoforms are kept in some cyanobacteria, and it perhaps allows these cyanobacteria capable of heterotrophy to adapt more readily to changes in the environment that affect the internal redox pools.

References

1. Battchikova, N.; Wei, L.; Du, L.; Bersanini, L.; Aro, E.M.; Ma, W. Identification of novel Ssl0352 protein (NdhS), essential for efficient operation of cyclic electron transport around photosystem I, in NADPH:plastoquinone oxidoreductase (NDH-1) complexes of *Synechocystis* sp. PCC 6803. *J. Biol. Chem.* **2011**, *286*, 36992–37001. [CrossRef] [PubMed]
2. Bernat, G.; Appel, J.; Ogawa, T.; Rogner, M. Distinct roles of multiple NDH-1 complexes in the cyanobacterial electron transport network as revealed by kinetic analysis of P700+ reduction in various Ndh-deficient mutants of *Synechocystis* sp. strain PCC6803. *J. Bacteriol.* **2011**, *193*, 292–295. [CrossRef] [PubMed]
3. Nikkanen, L.; Solymosi, D.; Jokel, M.; Allahverdiyeva, Y. Regulatory electron transport pathways of photosynthesis in cyanobacteria and microalgae: Recent advances and biotechnological prospects. *Physiol. Plant* **2021**, *173*, 514–525. [CrossRef] [PubMed]
4. Fisher, N.; Bricker, T.M.; Kramer, D.M. Regulation of photosynthetic cyclic electron flow pathways by adenylate status in higher plant chloroplasts. *Biochim. Biophys. Acta Bioenerg.* **2019**, *1860*, 148081. [CrossRef] [PubMed]
5. Dann, M.; Leister, D. Evidence that cyanobacterial Sll1217 functions analogously to PGRL1 in enhancing PGR5-dependent cyclic electron flow. *Nat. Commun.* **2019**, *10*, 5299. [CrossRef]
6. Yeremenko, N.; Jeanjean, R.; Prommeenate, P.; Krasikov, V.; Nixon, P.J.; Vermaas, W.F.; Havaux, M.; Matthijs, H.C. Open reading frame ssr2016 is required for antimycin A-sensitive photosystem I-driven cyclic electron flow in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Cell Physiol.* **2005**, *46*, 1433–1436. [CrossRef]
7. Wang, C.; Yamamoto, H.; Shikanai, T. Role of cyclic electron transport around photosystem I in regulating proton motive force. *Biochim. Biophys. Acta* **2015**, *1847*, 931–938. [CrossRef]

8. Miller, N.T.; Vaughn, M.D.; Burnap, R.L. Electron flow through NDH-1 complexes is the major driver of cyclic electron flow-dependent proton pumping in cyanobacteria. *Biochim. Biophys. Acta Bioenerg.* **2021**, *1862*, 148354. [CrossRef]

9. Allahverdiyeva, Y.; Suorsa, M.; Tikkanen, M.; Aro, E.M. Photoprotection of photosystems in fluctuating light intensities. *J. Exp. Bot.* **2015**, *66*, 2427–2436. [CrossRef]

10. Nikkanen, L.; Santana Sanchez, A.; Ermakova, M.; Rogner, M.; Cournac, L.; Allahverdiyeva, Y. Functional redundancy between flavodiiron proteins and NDH-1 in *Synechocystis* sp. PCC 6803. *Plant J.* **2020**, *103*, 1460–1476. [CrossRef]

11. Mi, H.; Endo, T.; Schreiber, U.; Ogawa, T.; Asada, K. Electron donation from cyclic and respiratory flows to the photosynthetic intersystem chain is mediated by pyridine nucleotide dehydrogenase in the cyanobacterium *Synechocystis* PCC 6803. *Plant Cell Physiol.* **1992**, *33*, 1233–1237.

12. Holland, S.C.; Kappell, A.D.; Burnap, R.L. Redox changes accompanying inorganic carbon limitation in *Synechocystis* sp. PCC 6803. *Biochim. Biophys. Acta* **2015**, *1847*, 355–363. [CrossRef] [PubMed]

13. Mi, H.; Endo, T.; Ogawa, T.; Asada, K. Thylakoid Membrane-Bound, NADPH-Specific Pyridine Nucleotide Dehydrogenase Complex Mediates Cyclic Electron Transport in the Cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Cell Physiol.* **1995**, *36*, 661–668. [CrossRef]

14. Zhang, C.; Shuai, J.; Ran, Z.; Zhao, J.; Wu, Z.; Liao, R.; Wu, J.; Ma, W.; Lei, M. Structural insights into NDH-1 mediated cyclic electron transfer. *Nat. Commun.* **2020**, *11*, 888. [CrossRef] [PubMed]

15. Schluchter, W.M.; Bryant, D.A. Molecular characterization of ferredoxin NADP+ oxidoreductase in cyanobacteria: Cloning and sequence of the petH gene of *Synechococcus* sp. PCC 7002 and studies on the gene product. *Biochemistry* **1992**, *31*, 3092–3102, Erratum in *Biochemistry* **1992**, *31*, 5952. [CrossRef] [PubMed]

16. Thomas, J.C.; Ughy, B.; Lagoutte, B.; Ajlani, G. A second isoform of the ferredoxin:NADP oxidoreductase generated by an in-frame initiation of translation. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 18368–18373. [CrossRef]

17. Korn, A.; Ajlani, G.; Lagoutte, B.; Gall, A.; Setif, P. Ferredoxin:NADP+ oxidoreductase association with phycocyanin modulates its properties. *J. Biol. Chem.* **2009**, *284*, 31789–31797. [CrossRef]

18. Korn, A. Respective Roles of the Ferredoxin: NADP-Oxidoreductase Isoforms in the Cyanobacterium *Synechocystis* sp. PCC 6803. Université Paris Sud-Paris XI. 2010. Available online: https://tel.archives-ouvertes.fr/tel-00460234 (accessed on 19 April 2022).

19. Liu, H.; Weisz, D.A.; Zhang, M.M.; Cheng, M.; Zhang, B.; Zhang, H.; Gerstenecker, G.S.; Pakrasi, H.B.; Gross, M.L.; Blankenship, R.E. Phycobilisomes harbor FNR(L) in cyanobacteria. *MBio* **2019**, *10*, e00669-19. [CrossRef]

20. Moal, G.; Lagoutte, B. Photo-induced electron transfer from photosystem I to NADP(+): Characterization and tentative simulation of the in vivo environment. *Biochim. Biophys. Acta* **2012**, *1817*, 1635–1645. [CrossRef]

21. Omairi-Nasser, A.; de Gracia, A.G.; Ajlani, G. A larger transcript is required for the synthesis of the smaller isoform of ferredoxin:NADP oxidoreductase. *Mol. Microbiol.* **2011**, *81*, 1178–1189. [CrossRef]

22. Schuller, J.M.; Birrell, J.A.; Tanaka, H.; Konuma, T.; Wulfhorst, H.; Cox, N.; Schuller, S.K.; Thiemann, J.; Lubitz, W.; Setif, P.; et al. Structural adaptations of photosynthetic compolex I enable ferredoxin-dependent electron transfer. *Science* **2019**, *363*, 257–260. [CrossRef] [PubMed]

23. Allen, M.M. Simple conditions for growth of unicellular blue-green algae on plates. *J. Phycol.* **1968**, *4*, 1–4. [CrossRef] [PubMed]

24. Ritchie, R.J. Consistent sets of spectrophotometric chlorophyll equations for acetone, methanol and ethanol solvents. *Photosynth. Res.* **2006**, *89*, 27–41. [CrossRef] [PubMed]

25. Campbell, D.; Hurry, V.; Clarke, A.K.; Gustafsson, P.; Oquist, G. Chlorophyll fluorescence analysis of cyanobacterial photosynthesis and acclimation. *Microbiol. Mol. Biol. R* **1998**, *62*, 667–683. [CrossRef] [PubMed]

26. Schreiber, U.; Klughammer, C. *New NADPH/9-AA Module for the DUAL-PAM-100: Description, Operation and Examples of Application*; Heinz Walz GmbH: Effeltrich, Germany, 2009; pp. 1–13.

27. Palmgren, M.G. Acridine orange as a probe for measuring pH gradients across membranes: Mechanism and limitations. *Anal. Biochem.* **1991**, *192*, 316–321. [CrossRef]

28. Nakamaru-Ogiso, E.; Kao, M.C.; Chen, H.; Sinha, S.C.; Yagi, T.; Ohnishi, T. The membrane subunit NuoL(ND5) is involved in the indirect proton pumping mechanism of Escherichia coli complex I. *J. Biol. Chem.* **2010**, *285*, 39070–39078. [CrossRef]

29. Teuber, M.; Roegner, M.; Berry, S. Fluorescent probes for non-invasive bioenergetic studies of whole cyanobacterial cells. *BBA* **2001**, *1506*, 31–46. [CrossRef]

30. Checchetto, V.; Segalla, A.; Allorent, G.; La Rocca, N.; Leanza, L.; Giacometti, G.M.; Uozumi, N.; Finazzi, G.; Bergantino, E.; Szabo, I. Thylakoid potassium channel is required for efficient photosynthesis in cyanobacteria. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 11043–11048. [CrossRef]

31. Schreiber, U.; Endo, T.; Mi, H.; Asada, K. Quenching Analysis of Chlorophyll Fluorescence by the Saturation Pulse Method: Particular Aspects Relating to the Study of Eukaryotic Algae and Cyanobacteria. *Plant Cell Physiol.* **1995**, *36*, 873–882. [CrossRef]

32. Renger, G.; Schreiber, U.; Amesz, J.; Fork, D. Practical applications of fluorometric methods to algae and higher plant research. In *Light Emission by Plants and Bacteria*; Govindjee, Amesz, J., Eds.; Academic Press: Orlando, FL, USA, 1986; pp. 589–616.

33. Artier, J.; Holland, S.C.; Miller, N.; Zhang, M.; Burnap, R.L. Synthetic DNA system for structure function studies of the high affinity CO_2 uptake (CupA) protein complex in cyanobacteria. *BBA-Bioenerg.* **2018**, *1859*, 1108–1118. [CrossRef]

34. Shikanai, T.; Endo, T.; Hashimoto, T.; Yamada, Y.; Asada, K.; Yokota, A. Directed disruption of the tobacco *ndhB* gene impairs cyclic electron flow around photosystem I. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 9705–9709. [CrossRef] [PubMed]

35. Ogawa, T.; Harada, T.; Ozaki, H.; Sonoike, K. Disruption of the ndhF1 gene affects Chl fluorescence through state transition in the cyanobacterium *Synechocystis* sp. PCC 6803, resulting in apparent high efficiency of photosynthesis. *Plant Cell Physiol.* **2013**, *54*, 1164–1171. [CrossRef] [PubMed]

36. Bernát, G.; Steinbach, G.; Kaňa, R.; Govindjee; Misra, A.N.; Prašil, O. On the origin of the slow M-T chlorophyll a fluorescence decline in cyanobacteria: Interplay of short-term light-responses. *Photosynth. Res.* **2018**, *136*, 183–198. [CrossRef] [PubMed]

37. Stirbet, A.; Lazár, D.; Papageorgiou, G.C.; Govindjee. Chapter 5-Chlorophyll a Fluorescence in Cyanobacteria: Relation to Photosynthesis ☆. In *Cyanobacteria*; Mishra, A.K., Tiwari, D.N., Rai, A.N., Eds.; Academic Press: Cambridge, MA, USA, 2019; pp. 79–130.

38. Mullineaux, C.W.; Allen, J.F. State 1-State 2 transitions in the cyanobacterium *Synechococcus* 6301 are controlled by the redox state of electron carriers between photosystems I and II. *Photosynth. Res.* **1990**, *23*, 297–311. [CrossRef] [PubMed]

39. Bhatti, A.F.; Choubeh, R.R.; Kirilovsky, D.; Wientjes, E.; van Amerongen, H. State transitions in cyanobacteria studied with picosecond fluorescence at room temperature. *Biochim. Biophys. Acta Bioenerg.* **2020**, *1861*, 148255. [CrossRef]

40. Ranjbar Choubeh, R.; Wientjes, E.; Struik, P.C.; Kirilovsky, D.; van Amerongen, H. State transitions in the cyanobacterium Synechococcus elongatus 7942 involve reversible quenching of the photosystem II core. *Biochim. Biophys. Acta Bioenerg.* **2018**, *1859*, 1059–1066. [CrossRef]

41. Kana, R.; Kotabova, E.; Komarek, O.; Sediva, B.; Papageorgiou, G.C.; Govindjee; Prasil, O. The slow S to M fluorescence rise in cyanobacteria is due to a state 2 to state 1 transition. *Biochim. Biophys. Acta* **2012**, *1817*, 1237–1247. [CrossRef]

42. Deng, Y.; Ye, J.; Mi, H. Effects of low CO_2 on NAD(P)H dehydrogenase, a mediator of cyclic electron transport around Photosystem I in the cyanobacterium *Synechocystis* PCC6803. *Plant Cell Physiol.* **2003**, *44*, 534–540. [CrossRef]

43. Laughlin, T.G.; Savage, D.F.; Davies, K.M. Recent advances on the structure and function of NDH-1: The complex I of oxygenic photosynthesis. *Biochim. Et Biophys. Acta (BBA)-Bioenerg.* **2020**, *11*, 148254. [CrossRef]

44. Tamoi, M.; Miyazaki, T.; Fukamizo, T.; Shigeoka, S. The Calvin cycle in cyanobacteria is regulated by CP12 via the NAD(H)/NADP(H) ratio under light/dark conditions. *Plant J.* **2005**, *42*, 504–513. [CrossRef]

Review

Advances in the Understanding of the Lifecycle of Photosystem II

Virginia M. Johnson and Himadri B. Pakrasi *

Department of Biology, Washington University, St. Louis, MO 63130, USA; virginia.johnson@wustl.edu
*   Correspondence: pakrasi@wustl.edu

Abstract: Photosystem II is a light-driven water-plastoquinone oxidoreductase present in cyanobacteria, algae and plants. It produces molecular oxygen and protons to drive ATP synthesis, fueling life on Earth. As a multi-subunit membrane-protein-pigment complex, Photosystem II undergoes a dynamic cycle of synthesis, damage, and repair known as the Photosystem II lifecycle, to maintain a high level of photosynthetic activity at the cellular level. Cyanobacteria, oxygenic photosynthetic bacteria, are frequently used as model organisms to study oxygenic photosynthetic processes due to their ease of growth and genetic manipulation. The cyanobacterial PSII structure and function have been well-characterized, but its lifecycle is under active investigation. In this review, advances in studying the lifecycle of Photosystem II in cyanobacteria will be discussed, with a particular emphasis on new structural findings enabled by cryo-electron microscopy. These structural findings complement a rich and growing body of biochemical and molecular biology research into Photosystem II assembly and repair.

Keywords: Photosystem II; oxygenic photosynthesis; photosynthetic reaction center; cyanobacteria; cryo-electron microscopy

1. Introduction: Photosystem II and Cyanobacteria

Photosystem II (PSII) is a multi-subunit membrane protein complex located in the thylakoid membrane of plants, cyanobacteria, and algae. It is the enzyme responsible for splitting water, providing the oxygen that fuels life on Earth as we recognize it. PSII performs this demanding photochemical reaction through sequential oxidation of a catalytic Mn_4CaO_5 cluster, which accumulates four oxidizing equivalents, cycling through five so-called S_n-states of oxidation [1] before oxidizing water in a concerted manner into molecular oxygen, producing four protons and four electrons in the process.

Cyanobacteria, oxygenic photosynthetic prokaryotes, have frequently been used as model organisms to study oxygenic photosynthetic processes due to their relatively fast growth and ease of genetic manipulation as compared with plants. For study of Photosystem II, three organisms have been widely utilized: *Synechocystis* sp. PCC 6803 (S6803), *Thermosynechococcus elongatus*, and *Thermosynechococcus vulcanus*. The mesophilic S6803 has been utilized for physiological studies and studies that require genetic manipulation of photosynthetic proteins, as it is a facultative heterotroph. The ability to grow S6803 on glucose allows for mutations and deletions to be made in genes necessary for photoautotrophic growth. PSII from either *T. elongatus* or *T. vulcanus*, thermophilic cyanobacterial strains, has been utilized for all high-resolution X-ray diffraction structures of PSII. Proteins from thermophilic organisms are often used for structural studies, as these proteins are more stable than their mesophilic homologs. Currently, due to these structural studies, our 'static' picture of active PSII is quite detailed. Atomic-resolution X-ray diffraction structures are available from *T. elongatus* and *T. vulcanus* [2–4] and X-ray free-electron laser (XFEL) technology and sample preparations have advanced to allow the structural determination of each of the Kok's S-states of the Mn_4CaO_5 cluster up to S_3 [5]. Recently, cryo-electron microscopy has enabled elucidation of the structure of active PSII from S6803 as well [6].

The biogenesis and repair (PSII lifecycle) of this membrane-protein-pigment complex are under active investigation. The assembly of PSII, a 20+ subunit membrane-protein-pigment complex, is an intricate process, involving cellular localization of protein subunits, pigment insertion, subunit assembly and processing, and cofactor assembly. Furthermore, under normal growth conditions, PSII is routinely damaged by reactive oxygen species that form within the protein and in its immediate vicinity [7,8]. To maintain a high level of photosynthetic activity, PSII must be repaired when damage occurs: damaged protein subunits must be replaced while the undamaged ones remain protected, and the biogenesis and repair processes must be coordinated within the cell.

In this review, we will discuss what is known about PSII assembly and repair in cyanobacteria, with an emphasis on new findings since the topic was last reviewed [9–17]. Recently, the field has been greatly advanced by high-resolution cryo-electron microcopy (cryo-EM). Cryo-EM has enabled several PSII assembly intermediate structures to be solved, as well as structures from the mesophilic S6803. These advances, in addition to other progress in the field, warrant a review and discussion.

2. PSII Structure

Active PSII has a modular architecture, consisting of what can be described as four 'modules' (Figure 1). These include the membrane-intrinsic reaction center and chlorophyll-binding antennas, and the membrane-extrinsic thylakoid lumenal proteins. For a thorough discussion of the PSII structure-function relationship, see Müh et al. [18].
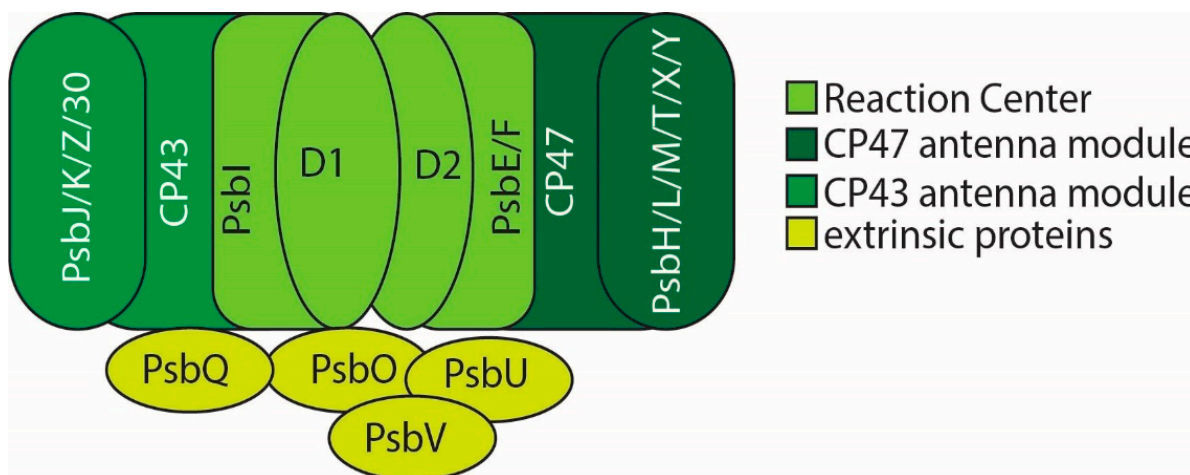
Figure 1. PSII monomer (PSII forms a dimer in vivo). Modules are colored as indicated.

The PSII reaction center (RC) is localized on the D1 and D2 protein subunits. These subunits bind the P680 reaction center chlorophyll special pair as well as the primary electron transfer chain, including plastoquinones Q_A and Q_B and the catalytic Mn_4CaO_5 water-oxidizing complex (WOC or OEC for oxygen evolving center). The PsbI subunit and cytochrome b559 (made up of the PsbE and PsbF subunits) complete the reaction center. This reaction center complex has been isolated independently and is the smallest PSII sub-complex capable of charge separation [19–21]. It is notable, however, that the CP43 subunit also provides a ligand to the manganese cluster, and so, while capable of charge separation, the reaction center proteins alone are not sufficient to bind the catalytic metal center or perform water-oxidizing chemistry.

The two antenna modules bind chlorophyll to absorb light energy that is funneled into the reaction center. These modules consist of the CP47 and CP43 proteins and their associated low molecular weight subunits (which are also referred to as low-molecular mass or small membrane-intrinsic subunits). PsbH, PsbL, PsbM, PsbT, PsbX, and PsbY bind to CP47. CP43 is associated with the small transmembrane subunits PsbJ, PsbK, and PsbZ and Psb30 (Ycf12). These low-molecular weight subunits can be individually deleted

without a loss of photosynthetic capacity, but stabilize and optimize PSII photochemistry, and some may play a regulatory role in the PSII lifecycle.

The last module consists of membrane-extrinsic proteins which bind PSII on the lumenal side of the thylakoid membrane and contribute to PSII stability and optimal performance. This module is the site of greatest variability between and among plants, algae, and cyanobacteria [22–25].

In addition to the active PSII subunits, there are assembly factors and accessory proteins involved in de novo biogenesis and repair following photodamage that are not part of the active PSII complex. Knowledge of these is almost certainly not complete, and they number as many if not more than the protein subunits of the active complex. These proteins, especially the Psb27 subunit, have been the focus of much recent research effort.

3. PSII Assembly

PSII assembly is thought to follow a highly ordered series of steps. The first step in PSII assembly is the formation of two pre-complexes, one containing pD1 and PsbI [26–28], and one containing D2 and cytochrome b559 [27–31] (Figure 2A, cytochrome b559 is composed of PsbE and PsbF). pD1 is a pre-processed form of D1. It has a 16 amino acid c-terminal extension that is cleaved by the protease **CtpA** (see Table 1 for details on bolded proteins) prior to assembly of the Mn_4CaO_5 cofactor [32,33]. The pre-D1 complex consists of the pD1 subunit and the PsbI subunit, as well as numerous translation and assembly chaperones that aid in ribosome binding to the *psbA* (gene coding for D1) transcript, insertion of the nascent peptide into the membrane, and chlorophyll binding to the newly inserted peptide. One such protein, **PratA**, is involved with the insertion of pD1 to the membrane and coordination of an initial Mn^{2+} ion to D1 [34–36]. It is also thought to be associated with specialized regions of the thylakoid membrane where PSII is synthesized and repaired [34,37]. The deletion of PratA was shown to be deficient in c-terminal processing of D1 [36], leading to the hypothesis that it facilitates access of the c-terminal protease, CtpA, to pD1. The rubredoxin **RubA** [21,31,38] also binds to D1 prior to its association with D2 and has been implicated in protection from photodamage to the reaction center during PSII assembly in the green alga *Chlamydomonas reinhardtii* [38].

Knoppova et al. found that Ycf39 and two HLIP family chlorophyll-binding proteins [39], **HliC** (ScpB) and **HliD** (ScpE), associate with D1 co-translationally [28]. The HLIP (high light-inducible) family of proteins is a family of chlorophyll-binding proteins related to the light-harvesting complex family of proteins in eukaryotes. In cyanobacteria, they are known to bind chlorophyll and stabilize the assembly of chlorophyll-binding proteins [39,40]. Ycf39, HliC and HliD form a complex which contains chlorophyll and β-carotene and is necessary for formation of the reaction center. This Ycf39-Hlip complex is also proposed to be involved in the delivery of recycled chlorophyll to newly synthesized D1 and protection from photodamage, as it interacts with the YidC/Alb3 membrane insertase complex as well as chlorophyll synthase [41,42]. **Pam68**, along with **Ycf48** and **Ycf39**, is implicated in co-translational insertion of chlorophyll to the reaction center proteins D1 and D2 [28,37,43,44]. Ycf48 is a soluble thylakoid membrane-associated protein associated with the D1 pre-complex and the reaction center complex [21,27,31,45–49], and it contains a conserved arginine patch that mediates its binding to integral membrane proteins [48]. These findings, in addition to the finding that apo-D1 does not accumulate [29], provide evidence that chlorophyll is co-translationally inserted into D1 as it is translocated into the membrane.
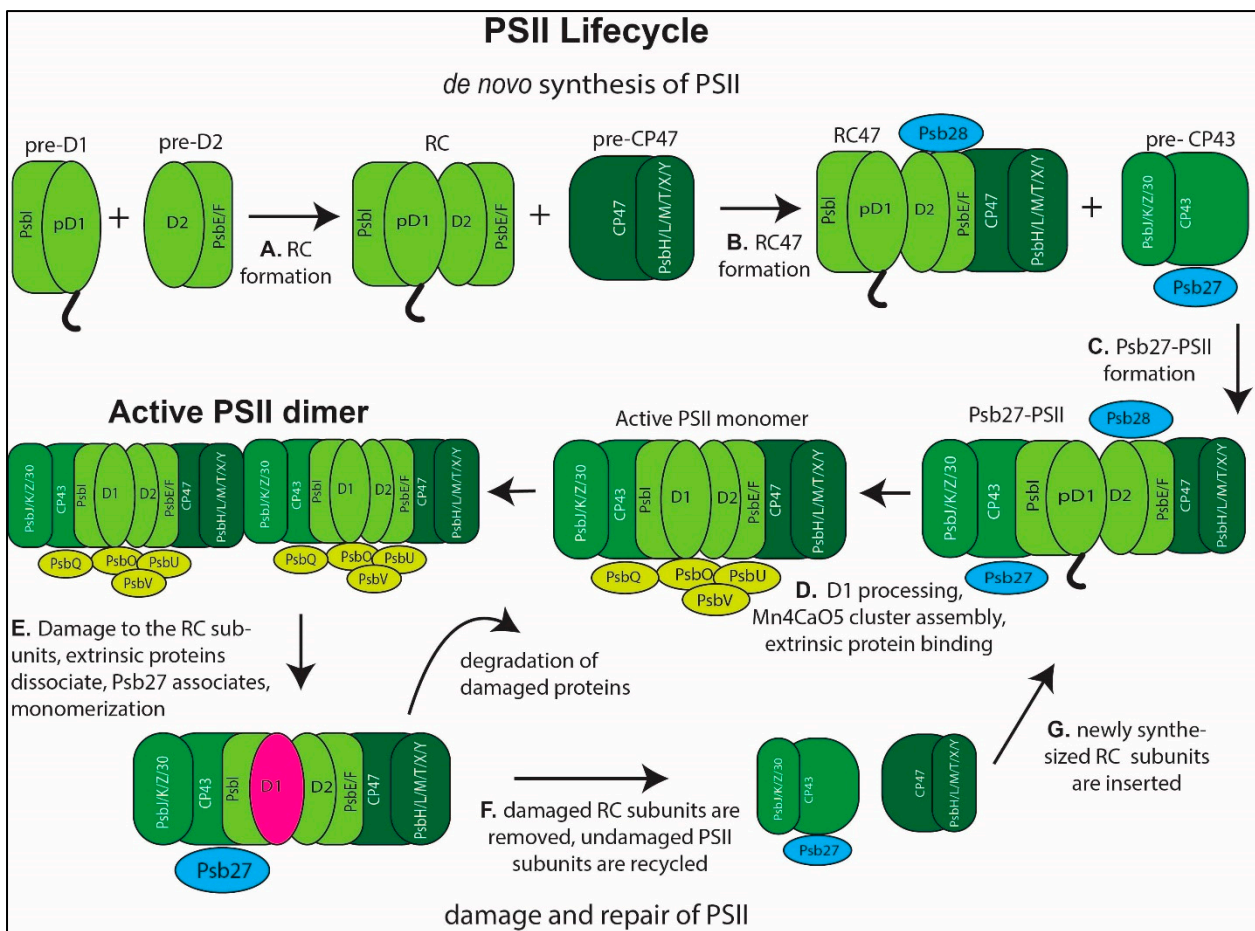
Figure 2. PSII Lifecycle, simplified. Not all accessory assembly chaperones are illustrated. (**A–D**) illustrate de novo synthesis, while (**D–G**) illustrate the repair cycle. (**A**) The pre-D1 and pre-D2 complexes come together to form the RC module. (**B**) RC and pre-CP47 modules come together to form the RC47 complex. Psb28 binds at this stage. (**C**) Psb27-PSII forms from pre-CP43 module and RC47. (**D**) The active PSII monomer is formed in a series of steps, which include processing of the D1 C-terminus, photo-assembly of the Mn_4CaO_5 cluster, and binding of extrinsic proteins PsbO, PsbV, PsbU, and PsbQ. The active PSII dimer forms following complete monomer assembly. (**E**) Following a photodamage event, the extrinsic proteins dissociate, Psb27 binds, and the dimer dissociates into monomers. (**F**) Damaged subunits are removed and proteolytically degraded. (**G**) Newly synthesized PSII subunits are inserted into recycled subunits to form a Psb27-PSII complex and the repair pathway re-joins the de novo synthesis pathway.

The reaction center forms (Figure 2A) [29,30,50,51] when the D1 and D2 pre-complexes bind together. **PsbP** (CyanoP), which binds to the D2 pre-complex [52], is thought to be involved in RC formation [22,53,54]. Recently, an RC complex was isolated from a strain of S6803 lacking CP47 that contains **PsbN** as well as the protein products of *slr0575* and *slr1470* [21]. This complex is on the de novo synthesis pathway, as active PSII does not form in this strain. However, a pre-D2 complex was also isolated from this strain which was associated with the protease FtsH2/FtsH3. This finding indicates that free-D2 levels are regulated by FtsH2/FtsH3.

The CP47 pre-complex contains PsbH, PsbL, PsbT, and possibly PsbM, PsbX, and PsbY [55,56]. The assembly factor **Psb35** also associates with the CP47 pre-complex and was shown to stabilize it and its association with **HliA** and **HliB** (ScpC and SpcD) [57]. Psb35 itself is a homolog of the cyanobacterial one helix domain/HLIP family of proteins, but it is unknown whether it binds chlorophyll. Pam68, in addition to its association with

the pre-D1 complex, has been shown to be involved with co-translational insertion of chlorophyll into CP47 prior to PsbH binding [58].

In the next step of PSII assembly, the RC joins together with the CP47 pre-complex to form the RC47 complex (Figure 2B). RC47 is bound by HliA/HliB and **Psb28** [56,59–64]. While the deletion mutant of Psb28 is photoautotrophic, it is deficient in recovery from photodamage at high temperature [63] and in light/dark conditions [64], indicating a role in stabilizing PSII intermediate assembly and repair complexes. Recent structural studies are consistent with the role of Psb28 as stabilizing assembly intermediates and protecting them from photodamage as well (discussed below).

Table 1. PSII assembly factors. Psb subunits are listed alphabetically, followed by additional assembly factors, listed alphabetically.

Name	Function	S6803 Locus Tag	Homolog in *A. thaliana* or *C. reinhardtii*	Phenotype of Inactivation	Citations
PsbN	RC formation	smr0009	PsbN	Not significant in cyanobacteria	[21,49,65]
PsbP (cyanoP)	RC formation	sll1418	PsbP	Reduced O$_2$ evolution, severe phenotype in low CaCl$_2$	[22,52,54]
Psb27	Binds to CP43 during assembly	slr1645	Psb27	Defective photoactivation, sensitive to high light	[45,66–74]
Psb28	Binds to CP47/cytochrome b559 in RC complex and Psb27-PSII complex. Alters electron transfer properties to increase photoprotection	sll1398	Psb28	Susceptible to photoinhibition in high light	[46,56,59,60,63,64,71,73]
Psb29	Accumulation of FtsH2/FtsH3	sll1414	Psb29/Thf1	Impaired growth in high light, lower PSII efficiency	[75,76]
Psb32	?	sll1390	TLP18.3	Sensitive to photoinhibition	[77–80]
Psb34	Binds to RC47 and PSII-I prior to activation	ssl1498	?	N/A	[61,62,81]
Psb35	Pre-CP47	ssl2148	?	Lower CP47 accumulation, faster bleaching in dark	[57]
CtpA	C-terminal processing of D1	slr0008	CtpA	not photosynthetic, no Mn$_4$CaO$_5$ cluster formation	[32,66]
HliA	CP47 formation	ssl2542	?	Inhibited growth in high light	[39]
HliB	CP47 formation	ssr2595	?	Inhibited growth in high light	[39]

Table 1. *Cont.*

Name	Function	S6803 Locus Tag	Homolog in *A. thaliana* or *C. reinhardtii*	Phenotype of Inactivation	Citations
HliC/ScpB	Chlorophyll insertion	ssl1633	CAB/HLIP/ELIP family (counterpart of OHP1/OHP2)	Inhibited growth in high light, depleted chlorophyll	[21,28,39,40,82]
HliD/ScpE	Chlorophyll insertion	ssr1789	CAB/HLIP/ELIP family (counterpart to OHP1/OHP2)	Inhibited growth in high light, depleted chlorophyll	[21,28,39–41]
Pam68	Translation of CP47 and insertion of chlorophyll	sll0933	PAM68	Sensitive to high light, low temperature, fluctuating light	[43,44,58]
PratA	Mn^{2+} loading to D1 and D1 processing; thylakoid and plasma membrane connection	slr2048	LPA1	abnormal membranes, reduced PSII accumulation	[34–37]
RubA	D1/D2 assembly	slr2033	RBD1, *At1g54500*	Reduced PSII level and activity	[31,83]
Slr0144-Slr0152	PSII assembly associated	slr0144-slr0152	?	Slower growth and lower PSII activity	[47,84,85]
Ycf39	Pre-D2 stabilization, chlorophyll insertion	slr0399	HCF244	Decrease in thermotolerance	[21,28,41,44]
Ycf48	Insertion of D1, chlorophyll into D1, replacement of damaged D1, RC formation	slr2034	HCF136	Decrease in D1, PSII, increase in susceptibility to photoinhibition	[21,27,31,44,45,54,86]

RC47 binds to the CP43 pre-complex to form Psb27-PSII (Figure 2C). The CP43 pre-complex is composed of CP43, the **Psb27** assembly protein, PsbK, Psb30, and possibly PsbZ [55,70,72,87]. At this stage, the Psb27-PSII complex consists of most of the intrinsic membrane subunits of the PSII monomer. Sll0606 is involved in RC47 binding to CP43, and its deletion leads to a loss of photoautotrophy [88]. **Psb34** (Ssl1498), an HLIP-family protein with no pigment binding site, is also suggested to be involved in RC47 binding to CP43 [61,62,81], as it is found in both RC47 and Psb27-PSII. This subunit is suggested to be involved in later stages of PSII assembly following HLIP dissociation, as it seems to compete for binding sites on PSII intermediates, and in its deletion strain HliA/HliB increase [81]. It should be noted that, in [89], the subunit name Psb34 is assigned to a PSII subunit from *C. gracilis*. However, the sequence of that subunit was unable to be determined. Therefore, we defer to more recent publications that have assigned Psb34 to a PSII subunit with a known sequence and gene, Tsl0063 in *T. elongatus* or Ssl1498 in S6803, which is in a different position in the PSII structure from the unknown subunit.

In Figure 2D, the Psb27-PSII complex becomes the active PSII monomer. This step is presumably shared between the assembly and repair stages of the PSII lifecycle and requires several events to occur, the order of which are still unclear. These include the processing of the D1 C-terminus by CtpA, assembly of the Mn_4CaO_5 cofactor (photoactivation [90] discussed further below), dissociation of Psb28, dissociation of Psb27 [66,68,71], and assembly of the extrinsic proteins PsbO, PsbU, PsbV, and PsbQ (also called cyanoQ) [23,24,91].

It is also unclear at what stage some subunits found in the active PSII complex bind. In particular, PsbJ has not been found in intermediate complexes. PsbJ binds near the Q_B site and possibly does not bind until after the formation of Psb27-PSII, when it triggers Psb28

dissociation. The ΔPsbJ mutant retains Psb28 in the active PSII monomer and accumulates more of a Psb27-and Psb28-bound PSII complex [73,74]. This mutant also has a much longer lifetime of reduced Q_A as compared to the wild type. Therefore, it was proposed that PsbJ plays a role in forward electron flow from reduced Q_A to the plastoquinone pool [92,93]. This observation led to the hypothesis that binding of PsbJ is a regulatory step in PSII formation, only occurring following the release of Psb28.

In addition to the assembly factors described above, several genes have been implicated in PSII assembly and/or maintenance whose specific interactions with PSII have not been fully elucidated. The operon containing *slr0144-slr0152* [47,84,85] is implicated in optimal PSII function. Slr0151, in particular, is involved in D1 association with CP43 [84]. Its deletion causes lower PSII levels, impaired D1 replacement, and the reordering of the thylakoid membranes. **Psb32** is a transmembrane protein which was shown to minimize photodamage in cyanobacteria [78], but how it does so is unknown. **Psb29** was shown to contribute to optimal PSII maintenance in both S6803 and *Arabidopsis thaliana* [75] by interacting with the FtsH2/FtsH3 protease, which removes damaged PSII subunits.

Following assembly of the active PSII dimer, higher-level organization of antenna complexes and photosystems into supercomplexes takes place. In cyanobacteria, this involves phycobilisome docking and PSI-PSII-Phycobilisome supercomplex formation [64,94]. These inter-complex molecular interactions are likely to be important for optimal cellular metabolism and the highest photosynthetic activity.

4. PSII Repair

The repair cycle of Photosystem II is less well studied than PSII synthesis due to the inherently difficult nature of experimentally separating the two spatially and temporally overlapping processes. Deletion mutants of key photosynthetic proteins have been a crucial part of the elucidation of de novo PSII synthesis but are halted at certain points in the assembly of active PSII, and so never form active PSII to be damaged and subsequently repaired. What is well-established is that the lifetimes of PSII subunits are not all the same [40,95,96]. Experiments done by Vermaas and coworkers used the stable isotope ^{15}N to determine the average lifetimes PSII subunits. They labeled cells with ^{15}N for varying lengths of time to determine the rate of incorporation into various PSII subunits using bottom-up mass spectrometry. Subunits with a high rate of turnover had the highest rate of ^{15}N incorporation. They found that the D1 subunit is turned over most frequently, with a half-life of less than an hour in cyanobacteria. This is followed by D2 with a half-life of about 3 h, CP43 at 6.5 h and CP47 around 11 h. Other proteins, such as Photosystem I subunits, are much more stable under normal conditions. Importantly, a similar labeling experiment was performed with ^{15}N incorporation into chlorophyll, and it was found that it has a lifetime of 300 h in WT S6803 [96]. The high turnover rate of D1, and the stark contrast in lifetime between the PSII subunits, is of considerable interest, as it implicates that D1 is removed selectively from PSII and replaced, while other PSII proteins remain stably bound together. Additionally, chlorophyll is apparently recycled repeatedly among PSII and PSI complexes. Complementary experiments by Rögner and coworkers found that Psb27 is associated with both PSII assembly and repair intermediates using a similar strategy [67]. How the cell signals that certain subunits are damaged and selectively replaced, or if certain subunits are constitutively replaced at different rates, is not known.

As for how the damaged protein subunits are selectively replaced, the assumption has been, due to the loose(r) association of the CP43 module and the existence of the RC47 complex, that the CP43 module dissociates, allowing for proteolytic degradation of D1 and re-synthesis and insertion at the RC47 level [51,97]. It is then assumed that, following re-insertion of D1 (and D2), that repair follows the same series of steps as de novo synthesis (Figure 2C,D). However, whether D1 insertion takes place at the dimer or monomer level has yet to be determined. A dimeric Psb27-PSII intermediate has been identified [69,74], and dimeric RC47 has been identified in a ΔCP43 mutant of S6803 [72], indicating that replacement of D1 may happen on the level of an RC47 dimer.

Recently, a complex that was a combination of the CP47 module and CP43 modules, called NRC (for No Reaction Center) [98], was isolated. Its abundance increased in the absence of protein synthesis and in high light, conditions that enhance PSII repair. This complex was hypothesized to be part of the repair cycle. Whether the CP47 and CP43 pre-complexes form a stable complex is disputed [99], but this is a topic that warrants further investigation.

5. Structural Advances in PSII Research

Cryo-electron microscopy (cryo-EM) has brought forth some of the most significant advances in our understanding of PSII assembly and function in recent years. An area of intense research focus and progress has been the elucidation of PSII structures from S6803 and also intermediate PSII complexes. Cryo-EM has provided an excellent means of studying such membrane protein complexes, as the technique does not require protein crystallization and allows for greater heterogeneity within samples.

Protein crystals of PSII from thermophilic cyanobacteria have been extensively characterized by X-ray crystallography. However, the structure of PSII from the mesophilic S6803, on which many physiological and biophysical studies have been performed, had been elusive. Gisriel et al. [6] solved the cryo-EM structure of a dimeric, active PSII from S6803 to 1.93 Å. This structure is an important step in our understanding of PSII function, as many studies that model PSII based on biophysical data from S6803 assume that the structure and function is conserved between cyanobacterial species. However, as the authors note, there are significant differences in sequence between PSII from S6803 and the thermophilic organisms, and membrane proteins generally have different intermolecular interactions in mesophilic and thermophilic organisms. Importantly, this structure definitively shows the binding site of PsbQ, which was known to associate with active PSII in S6803 [77,91], but had not been captured in any cyanobacterial structure. PsbQ is found associated with the most highly active isolated PSII complexes [91]. However, the authors found that PsbQ binding does not affect the conformation of PSII when bound, hypothesizing that PsbQ stabilizes interactions of other membrane-extrinsic proteins with the PSII core, contributing to high activity in isolated complexes. This structure will certainly be of great value as a basis for computational studies to interpret the mechanism of PSII water-splitting, and the impact of point mutants in S6803.

Many intermediate PSII complexes have been isolated and characterized, but until recently, structural details of these complexes and their implications for PSII function were unknown. To better understand the role of Psb27 in the PSII lifecycle, Huang et al. [74] isolated a dimeric Psb27-PSII complex from *T. vulcanus* and determined the cryo-EM structure to 3.78 Å. The complex was isolated from a deletion mutant of the PsbV subunit, as increased assembly and repair intermediates are present in that strain. This dimeric Psb27-PSII does not have extrinsic proteins bound and it does not have an assembled Mn_4CaO_5 cluster. Psb27 binds to the E-loop of CP43, as evidenced previously by mass-spectrometry cross-linking data [70,72,87]. The authors found that there are limited molecular interactions between Psb27 and CP43, indicating relatively weak binding of the assembly factor. They demonstrate that Psb27 prevents binding of PsbO and PsbV by sterically hindering their association sites, which is consistent with data that report that Psb27 binds to PSII in Psb27-PSII to allow photoassembly of the Mn_4CaO_5 cluster to occur by preventing binding of extrinsic subunits [68]. Additionally, the authors did not find electron density corresponding to PsbY or PsbJ. PsbY has frequently been absent in PSII structures, which has been explained by the fact that it may easily dissociate due to its location on the periphery of PSII. PsbJ has been found to be absent in assembly intermediates that include Psb27. The authors suggest that structural perturbations from active PSII amongst subunits in Psb27-PSII may cause dissociation or prevent the association of PsbJ.

The dimeric Psb27-PSII in [74] may be the same species previously described by Grasse et al. [69] that was isolated from cyanobacteria grown at low temperature and high light. Because that dimeric Psb27-PSII was proposed to be part of the PSII repair cycle [69],

and because the C-terminus of the D1 protein was unable to be modeled beyond Arg334 (the final residue in the mature peptide is Ala344) in the cryo-EM structure, the authors hypothesized that D1 in their sample may be partially proteolytically degraded as part of the first step of PSII repair. To test this hypothesis, they analyzed D1 c-terminal peptides from their sample through mass spectrometry and found that about 1% are proteolytically cleaved prior to the c-terminal Alanine 344 of the mature peptide. This finding confirms that the invisibility of the D1 C-terminus in their structure is due to its flexibility in the absence of a bound manganese cluster and extrinsic proteins, and not because it is degraded. However, the small amount of proteolytically cleaved D1 subunits does not preclude the possibility that this complex is part of the repair cycle, consisting of damaged PSII that has been partially disassembled. Interestingly, Zabret et al. [61] found that the D1 C-terminus of their monomeric Psb27-PSII structure (detailed below) can be modeled and is bound more closely to CP43 than in mature PSII. This difference may be due to a higher overall resolution of the structure, or to different stages of the complexes in the PSII lifecycle.

Zabret et al. [61] purified a PSII assembly intermediate containing Psb27, Psb28, and Psb34 from the ΔPsbJ mutant of *T. elongatus*, based on the observation from [73] that this mutant contains an increased amount of this intermediate sub-complex. They solved the cryo-EM structure of this complex, which they term PSII-I, to 2.9Å. In the structure, they found that the C-terminus of mature D1 is bound more closely to CP43 than in the active structure. They propose that this is to hold the D1 C-terminus in a conformation that favors photoassembly of the OEC. One cation was modeled at the OEC site, which could either be a Mn^{2+} or a Ca^{2+} cation. In addition to the difference in the D1 C-terminus described above, there is a slight discrepancy with the finding in the Huang Psb27-PSII structure in that they do not find a direct steric conflict with PsbO binding. PsbO has been found to be associated with Psb27-PSII in [70], so perhaps this also a difference between the dimeric, repair intermediate and the monomeric, Psb28-bound complex which may be the state of PSII just prior to photoactivation on the assembly pathway. On the acceptor side of PSII, they find that the D1 DE loop interaction with CP47 in active PSII is disturbed by Psb28 binding and that the CP47 C-terminus forms a β-sheet with Psb28. In addition, the Q_B binding site and the non-heme iron hydrogen bond network are altered from the active PSII structure, resulting in the absence of the bicarbonate ion and Q_B. They found that these changes lead to changes in the PSII electron transfer pathway that reduce singlet O_2 formation (discussed below). These findings suggest that Psb28 binds to PSII during its assembly to protect PSII from photodamage. Interestingly, Eaton-Rye found that PsbT interacts with the D1 DE loop in active PSII, and PsbT deletion causes increased photodamage [100–102], indicating a regulatory role of this loop in the repair cycle.

Xiao et al. also solved the structure of a Psb28 bound-PSII, as well as a Psb28-bound RC47 [62], to 3.14 Å resolution. They used a His-tagged Psb28 to isolate PSII intermediate complexes from both a ΔPsbV mutant and WT strain of *T. vulcanus*. Both the 28-PSII and the RC47 structure have Psb28, Psb34, and an unknown subunit bound that they were not able to model. This additional subunit was not found in the similar PSII-I structure in [61], but may be a second Psb34 copy. The authors found that Psb28 and Psb34 associate similarly to in the PSII-I structure, and consistent with the Zabret paper, there are changes in the quinone binding sites and H-bond network around the non-heme iron, with absence of the bicarbonate. They also observed a similar shift of the D1 C-terminus and increased flexibility. CP43, and the CP43 module as a whole is also shifted relative to its position in active PSII, consistent with [61,103], which the authors describe as a looser attachment. Additionally, alpha helices in D1, D2, and CP47 are changed to loops near the Psb28 binding on the cytoplasmic side. Overall, the PSII-I and Psb28-PSII structures are consistent with the binding position of Psb28 determined by Weisz et al. [59], although the precise attachment is slightly altered (see below). Intriguingly, the WT Psb28-PSII structure has Psb27 bound, but ΔPsbV does not. This finding led the authors to propose that Psb27 associates following Psb28 dissociation. However, it is not consistent with multiple findings that CP43 and

Psb27 associate with one another prior to joining with RC47. PsbJ and PsbY were not found in either structure.

Xiao [62] and Zabret [61] both observed a distorted Q_B binding pocket and bicarbonate and heme iron binding site in the Psb28-bound PSII. These structures are consistent with the position of Psb28 that was identified by [59], and they find that, as proposed by Weisz et al., the donor side of PSII is perturbed by the presence of Psb28. Discrepancies between the crosslinking data and the structure could also have resulted from the fact that the location in the crosslinking paper was modeled off the active structure, while it was found that Psb28 binding causes significant structural changes in the acceptor (cytoplasmic) side of PSII. These structural alterations may slow electron transfer to the plastoquinone pool and change the midpoint redox potential of $Q_A/Q_A^{-\bullet}$, as proposed by Brinkert et al. [104], who showed that bicarbonate loss alters the *Em* of $Q_A/Q_A^{-\bullet}$ to a more positive value. This positive shift is predicted to increase the potential between $P_{680}^{+\bullet}/Q_A^{-\bullet}$ and $P_{680}^{+\bullet}/Pheo^{-\bullet}$, which may favor direct recombination between $P_{680}^{+\bullet}/Q_A^{-\bullet}$ decrease back-reactions through $P_{680}^{+\bullet}/Pheo^{-\bullet}$ recombination, reducing the probability of the formation of an RC triplet state [7,105–107]. Overall, this may lead to a decrease in the production of singlet O_2, consistent with the findings in Zabret et al. Additionally, photoactivation is also thought to decrease the potential of $Q_A/Q_A^{-\bullet}$ [108], so intermediate complexes such as the Psb28-bound PSII structures lacking OEC would also be protected by a similar mechanism.

Higher levels of Psb27-PSII are also associated with higher NPQ [109]. This finding would be consistent with an overall model where Psb27 and Psb28 binding to PSII intermediates induces structural perturbations that protect the not-fully-assembled PSII from photodamage. An unanswered question in the role of Psb27 is that of Psb27 N-terminal lipidation and how it relates to the interaction between PSII and Psb27. Nowaczyk and coworkers initially suggested that Psb27 had an n-terminal lipid modification based on the presence of a lipobox motif next to the N-terminal cysteine of the mature peptide and confirmed a lipid attachment through mass spectrometry [67]. Recently, more focused mass spectrometry has provided a more detailed picture of the lipid attachment [110]. However, this lipid attachment was not resolved in the recent structures of Psb27-PSII, so how it mediates attachment of Psb27 to the thylakoid membrane and/or to PSII remains unclear.

6. Advances in Photoactivation

Assembly of the Mn_4CaO_5 cluster takes place through sequential binding and oxidation of Mn^{2+} ions, as well as the binding of Ca^{2+} and the deprotonation of water, using the same photochemical machinery for charge separation as active PSII. Photoactivation must take place during both de novo PSII assembly and repair, when the D1 subunit is replaced. This phenomenon has long been known and studied [12,90,108,111–114], but its exact mechanism is unclear. Recent PSII structures and studies have begun to describe this process in greater detail.

In one such structural study, Zhang et al. [115] solved the structure of apo-PSII (PSII without an intact Mn_4CaO_5 cluster) from *T. elongatus* using X-ray crystallography and found that it aligned closely to the active PSII structure, but without electron density corresponding to the OEC. The authors in this study intentionally removed the Mn_4CaO_5 cluster by treating crystals with hydroxylamine and EDTA, but the extrinsic PSII subunits were maintained. They propose that the lack of structural alteration of the OEC ligands in its absence means that the protein scaffold is already 'pre-organized' prior to Mn_4CaO_5 cluster assembly, and significant rearrangement does not need to take place for it to form. However, in another study, Gisriel et al. solved a structure of apo-PSII from S6803 [116] using cryo-EM and did find structural differences in the soluble domain of CP43 and the D1 C-terminus between their structure and that of active PSII. It is likely that this structure from S6803 lost the Mn_4CaO_5 cluster during sample preparation, unlike the Psb28-and Psb27-PSII structures described above, which are most likely assembly complexes. However, the study does provide insight into photoactivation, as the S6803 apo-PSII lacks the OEC and extrinsic PSII subunits and CP43 is shifted, as in [61,62,74]. These features suggest that the

donor side of PSII does have some structural flexibility prior to photoactivation, and that rearrangement needs to take place during the assembly of the OEC. The S6803 apo-PSII structure is also lacking PsbY, PsbJ, and PsbZ, although whether they were lost in sample preparation is unknown. It is possible that the flexibility observed in the S6803 apo-PSII is not observed in the apo-PSII crystal structure due to the constraints of the crystals, as the OEC was removed after crystallization.

Computational modeling can complement structural studies and provide insight into structural perturbations and dynamics. Narzi et al. used a molecular dynamics simulation to find that structural rearrangements do need to occur after the first Mn^{2+} oxidation [117] for the manganese binding sites to form in apo-PSII. Their results are in agreement with the flexible nature of the D1 C-terminus found in [61,74,116], and they propose that alternate protonation of OEC-liganding residues contributes to flexibility in the D1 C-terminus as well. These results, together with Gisriel et al., also suggest that the structure solved by Zhang et al. was constrained by crystal contacts and does not represent the native structure of PSII prior to photoactivation. This study also agrees with a wealth of biophysical data that suggests a structural rearrangement of OEC ligands takes place during photoactivation [90,108,111–114,118].

The conditions necessary for optimal photoassembly, or photoactivation, of the manganese cluster have been the focus of recent work [118–120]. Photoactivation requires Ca^{2+} and is known to have a much lower quantum efficiency than photochemistry in active PSII leading to O_2 evolution. Low quantum efficiency of photoactivation has been explained by the need for rearrangement of the ligand shell after the first Mn^{2+} oxidation, but also the need for acceptor side rearrangement to downshift $Q_A/Q_A^{-\bullet}$ potential [108]. Cl^- is necessary for photoactivation as well. Vinyard and coworkers found, by tracking photoactivation using EPR, that it tunes the pK_a of residues around the OEC site, allowing for the necessary deprotonation [119]. In another recent study, Avramov et al. examined an over-expression strain of the Psb27 protein (OE27), as well as a deletion mutant of the PsbO protein for photoactivation. They found that, while both the ΔPsbO and OE27 strains had an increased optimal ratio of Ca^{2+}/Mn^{2+}, the quantum efficiency of photoactivation also increased in OE27, implicating a role for Psb27 beyond just blocking binding of extrinsic proteins as previously proposed [68]. The authors suggest Psb27 has a role in maintaining an optimal ligand configuration for photoassembly, in agreement with the altered CP43/D1 c-terminal conformation in [61,62]. How these results relate to the findings that Psb28- and Psb27- bound PSII have altered photochemistry, and when these subunits dissociate during the process, is yet to be determined.

7. Use of CRISPR and CRISPR Interference for Study of Photosynthetic Protein Complexes

CRISPR and its associated techniques have emerged as powerful tools to manipulate genes and gene expression in cyanobacteria [121–125]. Specifically in cyanobacteria, the nucleases Cas9 and Cas12a (Cpf1), and their DNase-dead counterparts have been used to great effect to efficiently make modifications. While these techniques are of great interest for metabolic engineering of cyanobacteria, they also have the potential to expedite and enable basic photosynthesis research. One advantage of CRISPR editing over more traditional genetic manipulation is that it is markerless, enabling successive mutations to be made without the need for multiple antibiotic resistance cassettes to be stacked or cells to be maintained on multiple antibiotics. Additionally, CRISPR editing allows for an expedited segregation process, which is an important consideration in an organism that has multiple genome copies.

Of recent interest in terms of Photosystem II research is CRISPR-interference, whereby a DNase-dead CRISPR-associated protein is targeted to a specific genomic locus (or loci) to block transcription. CRISPRi has emerged as a strategy to selectively inhibit photosynthetic genes. CRISPRi enables a particularly intriguing strategy for manipulation of photosynthetic proteins due to its reversibility and its potential to multiplex targets through

incorporation of multiple guide RNAs [126,127]. In one example of this type of strategy, DNase-dead Cpf1 was introduced into *S. elongatus* UTEX 2973 as part of a CRISPRi system to target Photosystem I for inhibition [128]. Additionally, a plasmid-based CRISPRi system was introduced into S6803 [129] to inhibit the expression of the D1 protein of photosystem II, and Liu et al. [130] introduced a reversible CRISPRi system into S6803 by using a Rhamnose-theophylline inducible promoter system. This CRISPRi system was able to almost completely knock out PSII by targeting the D2 (*psbD* gene) protein, and it was fully reversible. These inhibition strategies introduce a new paradigm for the study of membrane protein complexes, in particular PSII, whereby its assembly can be reversibly halted at varying stages, depending on the gene targeted for inhibition.

8. Conclusions and Future Perspectives

A common point that has been further emphasized in recent studies is that subtle perturbations in PSII structure by either assembly factors or low-molecular-weight peripheral subunits have a significant impact on photochemistry. These perturbations can lead to overall efficiencies on a cellular level by protecting assembly intermediates from photodamage when necessary and optimizing photochemistry in active PSII. Along that theme, the impact of detergent solubilization on our current understanding of PSII function is not well described. A few alternate solubilization techniques that may preserve a more native membrane lipid arrangement, including lipid/protein nanodiscs [131–133] and SMALPs [134,135] have been explored, but it is an area that warrants more attention.

A more complete structural and functional picture of the lifecycle of PSII is coming into focus, but there are most likely PSII assembly and repair factors that have not yet been identified. Open questions remain about several aspects of the lifecycle. These aspects include the mechanism of cofactor insertion into nascent PSII subunits and what signals that PSII is damaged and should be repaired [136,137]. Additionally, some PSII assembly factors have been identified whose function has yet to be fully determined. The study of these questions will lend insight into PSII functions and membrane protein complex assembly more broadly.

While knowledge about structures of isolated PSII intermediates has been expanded in recent studies, an open area of investigation is the cellular localization of PSII biogenesis and repair in cyanobacteria. The topic of localization of PSII biogenesis and repair is related to the question of the structure of the cyanobacterial membrane system. Liberton et al. [138] found through serial sectioning and electron tomography that the thylakoids form a completely independent membrane system from the plasma membrane. However, the thylakoid sheets do converge at certain cellular locations near the plasma membrane [139–142]. Whether these convergence zones are connected to the plasma membrane is debated, but they have been proposed to be the site of PSII biogenesis and/or repair [34]. An earlier study determined that reaction center proteins were localized to the plasma membrane of cyanobacteria, indicating that certain PSII proteins are inserted to the plasma membrane and then trafficked to the thylakoid to regulate PSII assembly, but that has since been disputed [143,144]. In plants, damaged PSII migrates to the thylakoid stroma lamellae from the grana stacks to be repaired [13], and the 'thylakoid convergence zones' are thought to be the location of an analogous process. Supporting this hypothesis is the fact that ribosomes are found at these sites, while phycobilisomes are not [142]. PratA, which binds to pD1, is found to be associated with an intermediate density membrane fraction between the plasma membrane and the thylakoid membrane [34–37]. The membrane fraction which contains PratA was proposed to be the site of the 'thylakoid convergence zones' (also called variously biogenesis centers, thylakoid centers, or thylapses [142]), as it also contained other PSII and chlorophyll biogenesis proteins. These studies were performed by separating membrane fractions by density, but an alternative approach was recently taken by Dahlgren et al. [145], which is a promising avenue for this type of investigation. The authors of that study utilized a proximity-based proteomics approach to identify components of different intracellular compartments. In this case, they used the APEX2 protein, a modified

ascorbate peroxidase that catalyzes a reaction between biotin-phenol (BP) and hydrogen peroxide (H_2O_2) to create a BP radical that covalently attaches to proteins, fused to a known thylakoid lumen protein, PsbU, to selectively biotinylate proteins localized in the thylakoid lumen of *Synechococcus* sp. PCC 7002. Approaches such as this have great potential for answering questions of cellular localization and reinforcing findings by other methods.

Atomic force microscopy and hyperspectral confocal [146] studies of thylakoid fragments from *Thermosynechococcus elongatus*, *Synechococcus* sp. PCC 7002, and *Synechocystis* sp. PCC 6803 suggest that there are defined regions of the thylakoids that are differentiated by the arrangement of photosynthetic and respiratory complexes in cyanobacteria. This finding is corroborated by Casella et al. [147] in *Synechococcus elongatus* PCC 7942. The purpose of these differentiated membrane regions is yet to be determined, but the arrangement certainly has an impact on the diffusion of photosynthetic reactants and intermediates, such as quinones, as well as the spatial organization of PSII biogenesis and repair. However, it should also be noted that thylakoid arrangement varies drastically among cyanobacterial species [148], and so if and how these findings apply generally to other cyanobacteria is also unclear. Understanding the relative mobility of protein complexes and supercomplexes and assembly intermediates and their coordination will be key to understanding the purpose of these processes. Further advances in imaging, specifically electron microscopy correlated with light microscopy, and cryo-electron tomography or cryo-TEM on sections of cells, will shed light on this issue. One example of the potential of this kind of study is the in situ structure of a phycobilisome-PSII supercomplex from red algae [149]. New technologies, and new applications of technologies, will shed insight into our knowledge of PSII assembly, the process of membrane protein complex assembly more generally, and photosynthetic organisms on a systems level.

References

1. Kok, B.; Forbush, B.; McGloin, M. Cooperation of charges in photosynthetic O_2 evolution-I. A linear four step mechanism. *Photochem. Photobiol.* **1970**, *11*, 457–475. [CrossRef] [PubMed]
2. Umena, Y.; Kawakami, K.; Shen, J.-R.; Kamiya, N. Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 Å. *Nature* **2011**, *473*, 55–60. [CrossRef]
3. Suga, M.; Akita, F.; Hirata, K.; Ueno, G.; Murakami, H.; Nakajima, Y.; Shimizu, T.; Yamashita, K.; Yamamoto, M.; Ago, H.; et al. Native structure of photosystem II at 1.95 A resolution viewed by femtosecond X-ray pulses. *Nature* **2015**, *517*, 99–103. [CrossRef]
4. Young, I.D.; Ibrahim, M.; Chatterjee, R.; Gul, S.; Fuller, F.; Koroidov, S.; Brewster, A.S.; Tran, R.; Alonso-Mori, R.; Kroll, T.; et al. Structure of photosystem II and substrate binding at room temperature. *Nature* **2016**, *540*, 453–457. [CrossRef]
5. Kern, J.; Chatterjee, R.; Young, I.D.; Fuller, F.D.; Lassalle, L.; Ibrahim, M.; Gul, S.; Fransson, T.; Brewster, A.S.; Alonso-Mori, R.; et al. Structures of the intermediates of Kok's photosynthetic water oxidation clock. *Nature* **2018**, *563*, 421–425. [CrossRef] [PubMed]
6. Gisriel, C.J.; Wang, J.; Liu, J.; Flesher, D.A.; Reiss, K.M.; Huang, H.L.; Yang, K.R.; Armstrong, W.H.; Gunner, M.R.; Batista, V.S.; et al. High-resolution cryo-electron microscopy structure of photosystem II from the mesophilic cyanobacterium, *Synechocystis* sp. PCC 6803. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2116765118. [CrossRef]
7. Vass, I. Molecular mechanisms of photodamage in the Photosystem II complex. *Biochim. Biophys. Acta* **2012**, *1817*, 209–217. [CrossRef]

8. Zavafer, A. A theoretical framework of the hybrid mechanism of photosystem II photodamage. *Photosynth. Res.* **2021**, *149*, 107–120. [CrossRef]

9. Nixon, P.J.; Michoux, F.; Yu, J.; Boehm, M.; Komenda, J. Recent advances in understanding the assembly and repair of photosystem II. *Ann. Bot.* **2010**, *106*, 1–16. [CrossRef] [PubMed]

10. Komenda, J.; Sobotka, R.; Nixon, P.J. Assembling and maintaining the Photosystem II complex in chloroplasts and cyanobacteria. *Curr. Opin. Plant. Biol.* **2012**, *15*, 245–251. [CrossRef] [PubMed]

11. Nickelsen, J.; Rengstl, B. Photosystem II assembly: From cyanobacteria to plants. *Annu. Rev. Plant Biol.* **2013**, *64*, 609–635. [CrossRef]

12. Vinyard, D.J.; Ananyev, G.M.; Dismukes, G.C. Photosystem II: The reaction center of oxygenic photosynthesis. *Annu. Rev. Biochem.* **2013**, *82*, 577–606. [CrossRef]

13. Jarvi, S.; Suorsa, M.; Aro, E.M. Photosystem II repair in plant chloroplasts–Regulation, assisting proteins and shared components with photosystem II biogenesis. *Biochim. Biophys. Acta* **2015**, *1847*, 900–909. [CrossRef]

14. Weisz, D.A.; Gross, M.L.; Pakrasi, H.B. The Use of Advanced Mass Spectrometry to Dissect the Life-Cycle of Photosystem II. *Front. Plant Sci.* **2016**, *7*, 617. [CrossRef] [PubMed]

15. Heinz, S.; Liauw, P.; Nickelsen, J.; Nowaczyk, M. Analysis of photosystem II biogenesis in cyanobacteria. *Biochim. Biophys. Acta* **2016**, *1857*, 274–287. [CrossRef]

16. Theis, J.; Schroda, M. Revisiting the photosystem II repair cycle. *Plant. Signal. Behav.* **2016**, *11*, e1218587. [CrossRef]

17. Barber, J. Photosystem II: The water splitting enzyme of photosynthesis and the origin of oxygen in our atmosphere. *Q. Rev. Biophys.* **2016**, *49*, e14. [CrossRef]

18. Muh, F.; Zouni, A. Structural basis of light-harvesting in the photosystem II core complex. *Protein Sci.* **2020**, *29*, 1090–1119. [CrossRef]

19. Nanba, O.; Satoh, K. Isolation of a photosystem II reaction center consisting of D-1 and D-2 polypeptides and cytochrome b-559. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 109–112. [CrossRef] [PubMed]

20. Ikeuchi, M.; Inoue, Y. A new photosystem II reaction center component (4.8 kDa protein) encoded by chloroplast genome. *FEBS Lett.* **1988**, *241*, 99–104. [CrossRef]

21. Knoppova, J.; Sobotka, R.; Yu, J.; Beckova, M.; Pilny, J.; Trinugroho, J.P.; Csefalvay, L.; Bina, D.; Nixon, P.J.; Komenda, J. Assembly of D1/D2 complexes of photosystem II: Binding of pigments and a network of auxiliary proteins. *Plant Physiol.* **2022**, kiac045. [CrossRef]

22. Thornton, L.E.; Ohkawa, H.; Roose, J.L.; Kashino, Y.; Keren, N.; Pakrasi, H.B. Homologs of plant PsbP and PsbQ proteins are necessary for regulation of photosystem ii activity in the cyanobacterium *Synechocystis* 6803. *Plant. Cell* **2004**, *16*, 2164–2175. [CrossRef]

23. Roose, J.L.; Wegener, K.M.; Pakrasi, H.B. The extrinsic proteins of Photosystem II. *Photosynth. Res.* **2007**, *92*, 369–387. [CrossRef]

24. Bricker, T.M.; Roose, J.L.; Fagerlund, R.D.; Frankel, L.K.; Eaton-Rye, J.J. The extrinsic proteins of Photosystem II. *Biochim. Biophys. Acta* **2012**, *1817*, 121–142. [CrossRef]

25. Gisriel, C.J.; Brudvig, G.W. Comparison of PsbQ and Psb27 in photosystem II provides insight into their roles. *Photosynth. Res.* **2022**. [CrossRef]

26. Dobakova, M.; Tichy, M.; Komenda, J. Role of the PsbI protein in photosystem II assembly and repair in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Physiol.* **2007**, *145*, 1681–1691. [CrossRef]

27. Komenda, J.; Nickelsen, J.; Tichy, M.; Prasil, O.; Eichacker, L.A.; Nixon, P.J. The cyanobacterial homologue of HCF136/YCF48 is a component of an early photosystem II assembly complex and is important for both the efficient assembly and repair of photosystem II in *Synechocystis* sp. PCC 6803. *J. Biol. Chem.* **2008**, *283*, 22390–22399. [CrossRef] [PubMed]

28. Knoppova, J.; Sobotka, R.; Tichy, M.; Yu, J.; Konik, P.; Halada, P.; Nixon, P.J.; Komenda, J. Discovery of a chlorophyll binding protein complex involved in the early steps of photosystem II assembly in *Synechocystis*. *Plant Cell* **2014**, *26*, 1200–1212. [CrossRef]

29. Muller, B.; Eichacker, L.A. Assembly of the D1 precursor in monomeric photosystem II reaction center precomplexes precedes chlorophyll a-triggered accumulation of reaction center II in barley etioplasts. *Plant Cell* **1999**, *11*, 2365–2377.

30. Komenda, J.; Reisinger, V.; Muller, B.C.; Dobakova, M.; Granvogl, B.; Eichacker, L.A. Accumulation of the D2 protein is a key regulatory step for assembly of the photosystem II reaction center complex in *Synechocystis* PCC 6803. *J. Biol. Chem.* **2004**, *279*, 48620–48629. [CrossRef]

31. Kiss, E.; Knoppova, J.; Aznar, G.P.; Pilny, J.; Yu, J.; Halada, P.; Nixon, P.J.; Sobotka, R.; Komenda, J. A Photosynthesis-Specific Rubredoxin-Like Protein Is Required for Efficient Association of the D1 and D2 Proteins during the Initial Steps of Photosystem II Assembly. *Plant Cell* **2019**, *31*, 2241–2258. [CrossRef]

32. Anbudurai, P.R.; Mor, T.S.; Ohad, I.; Shestakov, S.V.; Pakrasi, H.B. The ctpA gene encodes the C-terminal processing protease for the D1 protein of the photosystem II reaction center complex. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 8082–8086. [CrossRef]

33. Komenda, J.; Kuvikova, S.; Granvogl, B.; Eichacker, L.A.; Diner, B.A.; Nixon, P.J. Cleavage after residue Ala352 in the C-terminal extension is an early step in the maturation of the D1 subunit of Photosystem II in *Synechocystis* PCC 6803. *Biochim. Biophys. Acta* **2007**, *1767*, 829–837. [CrossRef] [PubMed]

34. Stengel, A.; Gugel, I.L.; Hilger, D.; Rengstl, B.; Jung, H.; Nickelsen, J. Initial steps of photosystem II de novo assembly and preloading with manganese take place in biogenesis centers in *Synechocystis*. *Plant Cell* **2012**, *24*, 660–675. [CrossRef]

35. Schottkowski, M.; Gkalympoudis, S.; Tzekova, N.; Stelljes, C.; Schunemann, D.; Ankele, E.; Nickelsen, J. Interaction of the periplasmic PratA factor and the PsbA (D1) protein during biogenesis of photosystem II in *Synechocystis* sp. PCC 6803. *J. Biol. Chem.* **2009**, *284*, 1813–1819. [CrossRef]

36. Klinkert, B.; Ossenbuhl, F.; Sikorski, M.; Berry, S.; Eichacker, L.; Nickelsen, J. PratA, a periplasmic tetratricopeptide repeat protein involved in biogenesis of photosystem II in *Synechocystis* sp. PCC 6803. *J. Biol. Chem.* **2004**, *279*, 44639–44644. [CrossRef]

37. Rengstl, B.; Oster, U.; Stengel, A.; Nickelsen, J. An intermediate membrane subfraction in cyanobacteria is involved in an assembly network for Photosystem II biogenesis. *J. Biol. Chem.* **2011**, *286*, 21944–21951. [CrossRef]

38. Garcia-Cerdan, J.G.; Furst, A.L.; McDonald, K.L.; Schunemann, D.; Francis, M.B.; Niyogi, K.K. A thylakoid membrane-bound and redox-active rubredoxin (RBD1) functions in de novo assembly and repair of photosystem II. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 16631–16640. [CrossRef]

39. Komenda, J.; Sobotka, R. Cyanobacterial high-light-inducible proteins–Protectors of chlorophyll-protein synthesis and assembly. *Biochim. Biophys. Acta* **2016**, *1857*, 288–295. [CrossRef]

40. Yao, D.C.; Brune, D.C.; Vavilin, D.; Vermaas, W.F. Photosystem II component lifetimes in the cyanobacterium *Synechocystis* sp. strain PCC 6803: Small Cab-like proteins stabilize biosynthesis intermediates and affect early steps in chlorophyll synthesis. *J. Biol. Chem.* **2012**, *287*, 682–692. [CrossRef]

41. Chidgey, J.W.; Linhartova, M.; Komenda, J.; Jackson, P.J.; Dickman, M.J.; Canniffe, D.P.; Konik, P.; Pilny, J.; Hunter, C.N.; Sobotka, R. A cyanobacterial chlorophyll synthase-HliD complex associates with the Ycf39 protein and the YidC/Alb3 insertase. *Plant Cell* **2014**, *26*, 1267–1279. [CrossRef]

42. Ossenbuhl, F.; Inaba-Sulpice, M.; Meurer, J.; Soll, J.; Eichacker, L.A. The *Synechocystis* sp. PCC 6803 oxa1 homolog is essential for membrane integration of reaction center precursor protein pD1. *Plant Cell* **2006**, *18*, 2236–2246. [CrossRef] [PubMed]

43. Armbruster, U.; Zühlke, J.; Rengstl, B.; Kreller, R.; Makarenko, E.; Rühle, T.; Schünemann, D.; Jahns, P.; Weisshaar, B.; Nickelsen, J.; et al. The *Arabidopsis* thylakoid protein PAM68 is required for efficient D1 biogenesis and photosystem II assembly. *Plant Cell* **2010**, *22*, 3439–3460. [CrossRef] [PubMed]

44. Knoppova, J.; Komenda, J. Sequential deletions of photosystem II assembly factors Ycf48, Ycf39 and Pam68 result in progressive loss of autotrophy in the cyanobacterium *Synechocystis* PCC 6803. *Folia Microbiol.* **2019**, *64*, 683–689. [CrossRef] [PubMed]

45. Jackson, S.A.; Hervey, J.R.; Dale, A.J.; Eaton-Rye, J.J. Removal of both Ycf48 and Psb27 in *Synechocystis* sp. PCC 6803 disrupts Photosystem II assembly and alters Q(A)(-) oxidation in the mature complex. *FEBS Lett.* **2014**, *588*, 3751–3760. [CrossRef]

46. Mabbitt, P.D.; Wilbanks, S.M.; Eaton-Rye, J.J. Structure and function of the hydrophilic Photosystem II assembly proteins: Psb27, Psb28 and Ycf48. *Plant Physiol. Biochem.* **2014**, *81*, 96–107. [CrossRef]

47. Rast, A.; Rengstl, B.; Heinz, S.; Klingl, A.; Nickelsen, J. The Role of Slr0151, a Tetratricopeptide Repeat Protein from *Synechocystis* sp. PCC 6803, during Photosystem II Assembly and Repair. *Front. Plant Sci.* **2016**, *7*, 605. [CrossRef]

48. Yu, J.; Knoppova, J.; Michoux, F.; Bialek, W.; Cota, E.; Shukla, M.K.; Straskova, A.; Pascual Aznar, G.; Sobotka, R.; Komenda, J.; et al. Ycf48 involved in the biogenesis of the oxygen-evolving photosystem II complex is a seven-bladed beta-propeller protein. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E7824–E7833. [CrossRef]

49. Plochinger, M.; Schwenkert, S.; von Sydow, L.; Schroder, W.P.; Meurer, J. Functional Update of the Auxiliary Proteins PsbW, PsbY, HCF136, PsbN, TerC and ALB3 in Maintenance and Assembly of PSII. *Front. Plant Sci.* **2016**, *7*, 423. [CrossRef]

50. Vermaas, W.F.; Ikeuchi, M.; Inoue, Y. Protein composition of the photosystem II core complex in genetically engineered mutants of the cyanobacterium *Synechocystis* sp. PCC 6803. *Photosynth. Res.* **1988**, *17*, 97–113. [CrossRef]

51. van Wijk, K.J.; Roobol-Boza, M.; Kettunen, R.; Andersson, B.; Aro, E.M. Synthesis and assembly of the D1 protein into photosystem II: Processing of the C-terminus and identification of the initial assembly partners and complexes during photosystem II repair. *Biochemistry* **1997**, *36*, 6178–6186. [CrossRef]

52. Knoppova, J.; Yu, J.; Konik, P.; Nixon, P.J.; Komenda, J. CyanoP is involved in the Early Steps of Photosystem II Assembly in the Cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Cell Physiol.* **2016**, *57*, 1921–1931. [CrossRef]

53. Suorsa, M.; Regel, R.E.; Paakkarinen, V.; Battchikova, N.; Herrmann, R.G.; Aro, E.M. Protein assembly of photosystem II and accumulation of subcomplexes in the absence of low molecular mass subunits PsbL and PsbJ. *Eur. J. Biochem.* **2004**, *271*, 96–107. [CrossRef] [PubMed]

54. Jackson, S.A.; Eaton-Rye, J.J. Characterization of a *Synechocystis* sp. PCC 6803 double mutant lacking the CyanoP and Ycf48 proteins of Photosystem II. *Photosynth. Res.* **2015**, *124*, 217–229. [CrossRef] [PubMed]

55. Boehm, M.; Romero, E.; Reisinger, V.; Yu, J.; Komenda, J.; Eichacker, L.A.; Dekker, J.P.; Nixon, P.J. Investigating the early stages of photosystem II assembly in *Synechocystis* sp. PCC 6803: Isolation of CP47 and CP43 complexes. *J. Biol. Chem.* **2011**, *286*, 14812–14819. [CrossRef]

56. Boehm, M.; Yu, J.; Reisinger, V.; Beckova, M.; Eichacker, L.A.; Schlodder, E.; Komenda, J.; Nixon, P.J. Subunit composition of CP43-less photosystem II complexes of *Synechocystis* sp. PCC 6803: Implications for the assembly and repair of photosystem II. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2012**, *367*, 3444–3454. [CrossRef]

57. Pascual-Aznar, G.; Konert, G.; Beckov, M.; Kotabov, E.; Gardian, Z.; Knoppov, J.; Bucinsk, L.; Kana, R.; Sobotka, R.; Komenda, J. Psb35 Protein Stabilizes the CP47 Assembly Module and Associated High-Light Inducible Proteins during the Biogenesis of Photosystem II in the Cyanobacterium *Synechocystis* sp. PCC6803. *Plant. Cell Physiol.* **2021**, *62*, 178–190. [CrossRef]

58. Bucinska, L.; Kiss, E.; Konik, P.; Knoppova, J.; Komenda, J.; Sobotka, R. The Ribosome-Bound Protein Pam68 Promotes Insertion of Chlorophyll into the CP47 Subunit of Photosystem II. *Plant Physiol.* **2018**, *176*, 2931–2942. [CrossRef]

59. Weisz, D.A.; Liu, H.; Zhang, H.; Thangapandian, S.; Tajkhorshid, E.; Gross, M.L.; Pakrasi, H.B. Mass spectrometry-based cross-linking study shows that the Psb28 protein binds to cytochrome b559 in Photosystem II. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 2224–2229. [CrossRef] [PubMed]

60. Dobakova, M.; Sobotka, R.; Tichy, M.; Komenda, J. Psb28 protein is involved in the biogenesis of the photosystem II inner antenna CP47 (PsbB) in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Physiol.* **2009**, *149*, 1076–1086. [CrossRef]

61. Zabret, J.; Bohn, S.; Schuller, S.K.; Arnolds, O.; Moller, M.; Meier-Credo, J.; Liauw, P.; Chan, A.; Tajkhorshid, E.; Langer, J.D.; et al. Structural insights into photosystem II assembly. *Nat. Plants* **2021**, *7*, 524–538. [CrossRef] [PubMed]

62. Xiao, Y.; Huang, G.; You, X.; Zhu, Q.; Wang, W.; Kuang, T.; Han, G.; Sui, S.F.; Shen, J.R. Structural insights into cyanobacterial photosystem II intermediates associated with Psb28 and Tsl0063. *Nat. Plants* **2021**, *7*, 1132–1142. [CrossRef] [PubMed]

63. Sakata, S.; Mizusawa, N.; Kubota-Kawai, H.; Sakurai, I.; Wada, H. Psb28 is involved in recovery of photosystem II at high temperature in *Synechocystis* sp. PCC 6803. *Biochim. Biophys. Acta* **2013**, *1827*, 50–59. [CrossRef]

64. Beckova, M.; Gardian, Z.; Yu, J.; Konik, P.; Nixon, P.J.; Komenda, J. Association of Psb28 and Psb27 Proteins with PSII-PSI Supercomplexes upon Exposure of *Synechocystis* sp. PCC 6803 to High Light. *Mol. Plant* **2017**, *10*, 62–72. [CrossRef] [PubMed]

65. Kashino, Y.; Koike, H.; Yoshio, M.; Egashira, H.; Ikeuchi, M.; Pakrasi, H.B.; Satoh, K. Low-molecular-mass polypeptide components of a photosystem II preparation from the thermophilic cyanobacterium *Thermosynechococcus vulcanus*. *Plant Cell Physiol.* **2002**, *43*, 1366–1373. [CrossRef] [PubMed]

66. Roose, J.L.; Pakrasi, H.B. Evidence that D1 processing is required for manganese binding and extrinsic protein assembly into photosystem II. *J. Biol. Chem.* **2004**, *279*, 45417–45422. [CrossRef]

67. Nowaczyk, M.M.; Hebeler, R.; Schlodder, E.; Meyer, H.E.; Warscheid, B.; Rogner, M. Psb27, a cyanobacterial lipoprotein, is involved in the repair cycle of photosystem II. *Plant Cell* **2006**, *18*, 3121–3131. [CrossRef]

68. Roose, J.L.; Pakrasi, H.B. The Psb27 protein facilitates manganese cluster assembly in photosystem II. *J. Biol. Chem.* **2008**, *283*, 4044–4050. [CrossRef] [PubMed]

69. Grasse, N.; Mamedov, F.; Becker, K.; Styring, S.; Rogner, M.; Nowaczyk, M.M. Role of novel dimeric Photosystem II (PSII)-Psb27 protein complex in PSII repair. *J. Biol. Chem.* **2011**, *286*, 29548–29555. [CrossRef] [PubMed]

70. Liu, H.; Huang, R.Y.; Chen, J.; Gross, M.L.; Pakrasi, H.B. Psb27, a transiently associated protein, binds to the chlorophyll binding protein CP43 in photosystem II assembly intermediates. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 18536–18541. [CrossRef]

71. Liu, H.; Roose, J.L.; Cameron, J.C.; Pakrasi, H.B. A genetically tagged Psb27 protein allows purification of two consecutive photosystem II (PSII) assembly intermediates in *Synechocystis* 6803, a cyanobacterium. *J. Biol. Chem.* **2011**, *286*, 24865–24871. [CrossRef]

72. Komenda, J.; Knoppova, J.; Kopecna, J.; Sobotka, R.; Halada, P.; Yu, J.; Nickelsen, J.; Boehm, M.; Nixon, P.J. The Psb27 assembly factor binds to the CP43 complex of photosystem II in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Physiol.* **2012**, *158*, 476–486. [CrossRef]

73. Nowaczyk, M.M.; Krause, K.; Mieseler, M.; Sczibilanski, A.; Ikeuchi, M.; Rogner, M. Deletion of psbJ leads to accumulation of Psb27-Psb28 photosystem II complexes in *Thermosynechococcus elongatus*. *Biochim. Biophys. Acta* **2012**, *1817*, 1339–1345. [CrossRef] [PubMed]

74. Huang, G.; Xiao, Y.; Pi, X.; Zhao, L.; Zhu, Q.; Wang, W.; Kuang, T.; Han, G.; Sui, S.F.; Shen, J.R. Structural insights into a dimeric Psb27-photosystem II complex from a cyanobacterium *Thermosynechococcus vulcanus*. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2018053118. [CrossRef]

75. Keren, N.; Ohkawa, H.; Welsh, E.A.; Liberton, M.; Pakrasi, H.B. Psb29, a conserved 22-kD protein, functions in the biogenesis of Photosystem II complexes in *Synechocystis* and *Arabidopsis*. *Plant Cell* **2005**, *17*, 2768–2781. [CrossRef] [PubMed]

76. Bec Kova, M.; Yu, J.; Krynicka, V.; Kozlo, A.; Shao, S.; Konik, P.; Komenda, J.; Murray, J.W.; Nixon, P.J. Structure of Psb29/Thf1 and its association with the FtsH protease complex involved in photosystem II repair in cyanobacteria. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2017**, *372*, 20160394. [CrossRef]

77. Kashino, Y.; Lauber, W.M.; Carroll, J.A.; Wang, Q.; Whitmarsh, J.; Satoh, K.; Pakrasi, H.B. Proteomic analysis of a highly active photosystem II preparation from the cyanobacterium *Synechocystis* sp. PCC 6803 reveals the presence of novel polypeptides. *Biochemistry* **2002**, *41*, 8004–8012. [CrossRef]

78. Wegener, K.M.; Bennewitz, S.; Oelmuller, R.; Pakrasi, H.B. The Psb32 protein aids in repairing photodamaged photosystem II in the cyanobacterium *Synechocystis* 6803. *Mol. Plant* **2011**, *4*, 1052–1061. [CrossRef] [PubMed]

79. Sirpio, S.; Allahverdiyeva, Y.; Suorsa, M.; Paakkarinen, V.; Vainonen, J.; Battchikova, N.; Aro, E.M. TLP18.3, a novel thylakoid lumen protein regulating photosystem II repair cycle. *Biochem. J.* **2007**, *406*, 415–425. [CrossRef]

80. Jarvi, S.; Isojarvi, J.; Kangasjarvi, S.; Salojarvi, J.; Mamedov, F.; Suorsa, M.; Aro, E.M. Photosystem II Repair and Plant Immunity: Lessons Learned from *Arabidopsis* Mutant Lacking the THYLAKOID LUMEN PROTEIN 18.3. *Front. Plant Sci.* **2016**, *7*, 405. [CrossRef]

81. Rahimzadeh-Karvansara, P.; Pascual-Aznar, G.; Beckova, M.; Komenda, J. Psb34 protein modulates binding of high-light-inducible proteins to CP47-containing photosystem II assembly intermediates in the cyanobacterium *Synechocystis* sp. PCC 6803. *Photosynth. Res.* **2022**, 1–14. [CrossRef] [PubMed]

82. Kufryk, G.; Hernandez-Prieto, M.A.; Kieselbach, T.; Miranda, H.; Vermaas, W.; Funk, C. Association of small CAB-like proteins (SCPs) of *Synechocystis* sp. PCC 6803 with Photosystem II. *Photosynth. Res.* **2008**, *95*, 135–145. [CrossRef]

83. Calderon, R.H.; Garcia-Cerdan, J.G.; Malnoe, A.; Cook, R.; Russell, J.J.; Gaw, C.; Dent, R.M.; de Vitry, C.; Niyogi, K.K. A conserved rubredoxin is necessary for photosystem II accumulation in diverse oxygenic photoautotrophs. *J. Biol. Chem.* **2013**, *288*, 26688–26696. [CrossRef] [PubMed]

84. Yang, H.; Liao, L.; Bo, T.; Zhao, L.; Sun, X.; Lu, X.; Norling, B.; Huang, F. Slr0151 in *Synechocystis* sp. PCC 6803 is required for efficient repair of photosystem II under high-light condition. *J. Integr. Plant Biol.* **2014**, *56*, 1136–1150. [CrossRef] [PubMed]

85. Wegener, K.M.; Welsh, E.A.; Thornton, L.E.; Keren, N.; Jacobs, J.M.; Hixson, K.K.; Monroe, M.E.; Camp, D.G.; Smith, R.D.; Pakrasi, H.P.; et al. High sensitivity proteomics assisted discovery of a novel operon involved in the assembly of photosystem II, a membrane protein complex. *J. Biol. Chem.* **2008**, *283*, 27829–27837. [CrossRef]

86. Knoppova, J.; Yu, J.; Janouškovec, J.; Halada, P.; Nixon, P.J.; Whitelegge, J.P.; Komenda, J. The Photosystem II Assembly Factor Ycf48 from the Cyanobacterium *Synechocystis* sp. PCC 6803 Is Lipidated Using an Atypical Lipobox Sequence. *Int. J. Mol. Sci.* **2021**, *22*, 3733. [CrossRef]

87. Liu, H.; Chen, J.; Huang RY, C.; Weisz, D.; Gross, M.L.; Pakrasi, H.B. Mass spectrometry-based footprinting reveals structural dynamics of loop E of the chlorophyll-binding protein CP43 during photosystem II assembly in the cyanobacterium *Synechocystis* 6803. *J. Biol. Chem.* **2013**, *288*, 14212–14220. [CrossRef]

88. Zhang, S.; Frankel, L.K.; Bricker, T.M. The Sll0606 protein is required for photosystem II assembly/stability in the cyanobacterium *Synechocystis* sp. PCC 6803. *J. Biol. Chem.* **2010**, *285*, 32047–32054. [CrossRef] [PubMed]

89. Pi, X.; Zhao, S.; Wang, W.; Liu, D.; Xu, C.; Han, G.; Kuang, T.; Sui, S.-F.; Shen, J.-R. The pigment-protein network of a diatom photosystem II-light-harvesting antenna supercomplex. *Science* **2019**, *365*, eaax4406. [CrossRef]

90. Bao, H.; Burnap, R.L. Photoactivation: The Light-Driven Assembly of the Water Oxidation Complex of Photosystem II. *Front. Plant Sci.* **2016**, *7*, 578. [CrossRef]

91. Roose, J.L.; Kashino, Y.; Pakrasi, H.B. The PsbQ protein defines cyanobacterial Photosystem II complexes with highest activity and stability. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 2548–2553. [CrossRef]

92. Shi, L.X.; Hall, M.; Funk, C.; Schroder, W.P. Photosystem II, a growing complex: Updates on newly discovered components and low molecular mass proteins. *Biochim. Biophys. Acta* **2012**, *1817*, 13–25. [CrossRef]

93. Ohad, I.; Dal Bosco, C.; Herrmann, R.G.; Meurer, J. Photosystem II proteins PsbL and PsbJ regulate electron flow to the plastoquinone pool. *Biochemistry* **2004**, *43*, 2297–2308. [CrossRef]

94. Liu, H.; Zhang, H.; Niedzwiedzki, D.M.; Prado, M.; He, G.; Gross, M.L.; Blankenship, R.E. Phycobilisomes supply excitations to both photosystems in a megacomplex in cyanobacteria. *Science* **2013**, *342*, 1104–1107. [CrossRef] [PubMed]

95. Yao, D.C.; Brune, D.C.; Vermaas, W.F. Lifetimes of photosystem I and II proteins in the cyanobacterium *Synechocystis* sp. PCC 6803. *FEBS Lett.* **2012**, *586*, 169–173. [CrossRef]

96. Vavilin, D.; Brune, D.C.; Vermaas, W. 15N-labeling to determine chlorophyll synthesis and degradation in *Synechocystis* sp. PCC 6803 strains lacking one or both photosystems. *Biochim. Biophys. Acta* **2005**, *1708*, 91–101. [CrossRef]

97. Krynicka, V.; Shao, S.; Nixon, P.J.; Komenda, J. Accessibility controls selective degradation of photosystem II subunits by FtsH protease. *Nat. Plants* **2015**, *1*, 15168. [CrossRef] [PubMed]

98. Weisz, D.A.; Johnson, V.M.; Niedzwiedzki, D.M.; Shinn, M.K.; Liu, H.; Klitzke, C.F.; Gross, M.L.; Blankenship, R.E.; Lohman, T.M.; Pakrasi, H.B. A novel chlorophyll protein complex in the repair cycle of photosystem II. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 21907–21913. [CrossRef] [PubMed]

99. Beckova, M.; Sobotka, R.; Komenda, J. Photosystem II antenna modules CP43 and CP47 do not form a stable 'no reaction centre complex' in the cyanobacterium *Synechocystis* sp. PCC 6803. *Photosynth. Res.* **2022**. [CrossRef] [PubMed]

100. Bentley, F.K.; Luo, H.; Dilbeck, P.; Burnap, R.L.; Eaton-Rye, J.J. Effects of inactivating psbM and psbT on photodamage and assembly of photosystem II in *Synechocystis* sp. PCC 6803. *Biochemistry* **2008**, *47*, 11637–11646. [CrossRef] [PubMed]

101. Fagerlund, R.D.; Forsman, J.A.; Biswas, S.; Vass, I.; Davies, F.K.; Summerfield, T.C.; Eaton-Rye, J.J. Stabilization of Photosystem II by the PsbT protein impacts photodamage, repair and biogenesis. *Biochim. Biophys. Acta Bioenerg.* **2020**, *1861*, 148234. [CrossRef]

102. Forsman, J.A.; Eaton-Rye, J.J. The Interaction between PsbT and the DE Loop of D1 in Photosystem II Stabilizes the Quinone-Iron Electron Acceptor Complex. *Biochemistry* **2021**, *60*, 53–63. [CrossRef] [PubMed]

103. Tokano, T.; Kato, Y.; Sugiyama, S.; Uchihashi, T.; Noguchi, T. Structural Dynamics of a Protein Domain Relevant to the Water-Oxidizing Complex in Photosystem II as Visualized by High-Speed Atomic Force Microscopy. *J. Phys. Chem. B* **2020**, *124*, 5847–5857. [CrossRef] [PubMed]

104. Brinkert, K.; De Causmaecker, S.; Krieger-Liszkay, A.; Fantuzzi, A.; Rutherford, A.W. Bicarbonate-induced redox tuning in Photosystem II for regulation and protection. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 12144–12149. [CrossRef] [PubMed]

105. Vass, I.; Styring, S.; Hundal, T.; Koivuniemi, A.; Aro, E.; Andersson, B. Reversible and irreversible intermediates during photoinhibition of photosystem II: Stable reduced QA species promote chlorophyll triplet formation. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 1408–1412. [CrossRef]

106. Vass, I.; Styring, S. Spectroscopic characterization of triplet forming states in photosystem II. *Biochemistry* **1992**, *31*, 5957–5963. [CrossRef] [PubMed]

107. Vass, I.; Styring, S. Characterization of chlorophyll triplet promoting states in photosystem II sequentially induced during photoinhibition. *Biochemistry* **1993**, *32*, 3334–3341. [CrossRef] [PubMed]

108. Johnson, G.N.; Rutherford, A.W.; Krieger, A. A change in the midpoint potential of the quinone QA in Photosystem II associated with photoactivation of oxygen evolution. *Biochim. Biophys. Acta BBA Bioenerg.* **1995**, *1229*, 202–207. [CrossRef]

109. Johnson, V.M.; Biswas, S.; Roose, J.L.; Pakrasi, H.B.; Liu, H. Psb27, a photosystem II assembly protein, enables quenching of excess light energy during its participation in the PSII lifecycle. *Photosynth.* **2022**. [CrossRef]

110. Lambertz, J.; Liauw, P.; Whitelegge, J.P.; Nowaczyk, M.M. Mass spectrometry analysis of the photosystem II assembly factor Psb27 revealed variations in its lipid modification. *Photosynth. Res.* **2021**. [CrossRef]

111. Dasgupta, J.; Ananyev, G.M.; Dismukes, G.C. Photoassembly of the Water-Oxidizing Complex in Photosystem II. *Coord. Chem. Rev.* **2008**, *252*, 347–360. [CrossRef]

112. Hwang, H.J.; Nagarajan, A.; McLain, A.; Burnap, R.L. Assembly and disassembly of the photosystem II manganese cluster reversibly alters the coupling of the reaction center with the light-harvesting phycobilisome. *Biochemistry* **2008**, *47*, 9747–9755. [CrossRef] [PubMed]

113. Radmer, R.; Cheniae, G.M. Photoactivation of the manganese catalyst of 0 2 evolution II. A two-guantum mechanism. *Biochim. Biophys. Acta* **1971**, *253*, 182–186. [CrossRef]

114. Cheniae, G.M.; Martin, I.F. Photoactivation of the manganese catalyst of O 2 evolution. I. Biochemical and kinetic aspects. *Biochim. Biophys. Acta* **1971**, *253*, 167–181. [CrossRef]

115. Zhang, M.; Bommer, M.; Chatterjee, R.; Hussein, R.; Yano, J.; Dau, H.; Kern, J.; Dobbek, H.; Zouni, A. Structural insights into the light-driven auto-assembly process of the water-oxidizing Mn4CaO5-cluster in photosystem II. *Elife* **2017**, *6*, e26933. [CrossRef] [PubMed]

116. Gisriel, C.J.; Zhou, K.F.; Huang, H.L.; Debus, R.J.; Xiong, Y.; Brudvig, G.W. Cryo-EM Structure of Monomeric Photosystem II from *Synechocystis* sp. PCC 6803 Lacking the Water-Oxidation Complex. *Joule* **2020**, *4*, 2131–2148. [CrossRef]

117. Narzi, D.; Guidoni, L. Structural and dynamic insights into Mn4Ca cluster-depleted Photosystem II. *Phys. Chem. Chem. Phys.* **2021**, *23*, 27428–27436. [CrossRef] [PubMed]

118. Avramov, A.P.; Hwang, H.J.; Burnap, R.L. The role of Ca^{2+} and protein scaffolding in the formation of nature's water oxidizing complex. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 28036–28045. [CrossRef]

119. Russell, B.P.; Vinyard, D.J. Chloride facilitates Mn(III) formation during photoassembly of the Photosystem II oxygen-evolving complex. *Photosynth. Res.* **2021**. [CrossRef] [PubMed]

120. Vinyard, D.J.; Badshah, S.L.; Riggio, M.R.; Kaur, D.; Fanguy, A.R.; Gunner, M.R. Photosystem II oxygen-evolving complex photoassembly displays an inverse H/D solvent isotope effect under chloride-limiting conditions. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 18917–18922. [CrossRef] [PubMed]

121. Ungerer, J.; Pakrasi, H.B. Cpf1 is a versatile tool for CRISPR genome editing across diverse species of Cyanobacteria. *Sci. Rep.* **2016**, *6*, 39681. [CrossRef] [PubMed]

122. Yao, L.; Cengic, I.; Anfelt, J.; Hudson, E.P. Multiple Gene Repression in Cyanobacteria Using CRISPRi. *ACS Synth. Biol.* **2016**, *5*, 207–212. [CrossRef] [PubMed]

123. Behler, J.; Vijay, D.; Hess, W.R.; Akhtar, M.K. CRISPR-Based Technologies for Metabolic Engineering in Cyanobacteria. *Trends Biotechnol.* **2018**, *36*, 996–1010. [CrossRef] [PubMed]

124. Santos-Merino, M.; Singh, A.K.; Ducat, D.C. New Applications of Synthetic Biology Tools for Cyanobacterial Metabolic Engineering. *Front. Bioeng. Biotechnol.* **2019**, *7*, 33. [CrossRef]

125. Naduthodi, M.I.S.; Barbosa, M.J.; van der Oost, J. Progress of CRISPR-Cas Based Genome Editing in Photosynthetic Microbes. *Biotechnol. J.* **2018**, *13*, e1700591. [CrossRef]

126. Yao, L.; Shabestary, K.; Bjork, S.M.; Asplund-Samuelsson, J.; Joensson, H.N.; Jahn, M.; Hudson, E.P. Pooled CRISPRi screening of the cyanobacterium *Synechocystis* sp. PCC 6803 for enhanced industrial phenotypes. *Nat. Commun.* **2020**, *11*, 1666. [CrossRef]

127. Santos, M.; Pacheco, C.C.; Yao, L.; Hudson, E.P.; Tamagnini, P. CRISPRi as a Tool to Repress Multiple Copies of Extracellular Polymeric Substances (EPS)-Related Genes in the Cyanobacterium *Synechocystis* sp. PCC 6803. *Life* **2021**, *11*, 1198. [CrossRef]

128. Knoot, C.J.; Biswas, S.; Pakrasi, H.B. Tunable repression of key photosynthetic processes using Cas12a CRISPR Interference in the fast-growing cyanobacterium *Synechococcus* sp. UTEX 2973. *ACS Synth. Biol.* **2020**, *9*, 132–143. [CrossRef] [PubMed]

129. Kirtania, P.; Hodi, B.; Mallick, I.; Vass, I.Z.; Feher, T.; Vass, I.; Kos, P.B. A single plasmid based CRISPR interference in *Synechocystis* 6803—A proof of concept. *PLoS ONE* **2019**, *14*, e0225375. [CrossRef]

130. Liu, D.; Johnson, V.M.; Pakrasi, H.B. A Reversibly Induced CRISPRi System Targeting Photosystem II in the Cyanobacterium *Synechocystis* sp. PCC 6803. *ACS Synth. Biol.* **2020**, *9*, 1441–1449. [CrossRef]

131. Lu, Y.; Zhang, H.; Niedzwiedzki, D.M.; Jiang, J.; Blankenship, R.E.; Gross, M.L. Fast Photochemical Oxidation of Proteins Maps the Topology of Intrinsic Membrane Proteins: Light-Harvesting Complex 2 in a Nanodisc. *Anal. Chem.* **2016**, *88*, 8827–8834. [CrossRef] [PubMed]

132. Denisov, I.G.; Sligar, S.G. Nanodiscs for structural and functional studies of membrane proteins. *Nat. Struct. Mol. Biol.* **2016**, *23*, 481–486. [CrossRef] [PubMed]

133. Marty, M.T. Nanodiscs and Mass Spectrometry: Making Membranes Fly. *Int. J. Mass Spectrom.* **2020**, *458*, 116436. [CrossRef] [PubMed]

134. Korotych, O.I.; Nguyen, T.T.; Reagan, B.C.; Burch-Smith, T.M.; Bruce, B.D. Poly(styrene-co-maleic acid)-mediated isolation of supramolecular membrane protein complexes from plant thylakoids. *Biochim. Biophys. Acta Bioenerg.* **2021**, *1862*, 148347. [CrossRef]

135. Brady, N.G.; Workman, C.E.; Cawthon, B.; Bruce, B.D.; Long, B.K. Protein Extraction Efficiency and Selectivity of Esterified Styrene-Maleic Acid Copolymers in Thylakoid Membranes. *Biomacromolecules* **2021**, *22*, 2544–2553. [CrossRef]

136. He, Q.; Vermaas, W. Chlorophyll a availability affects psbA translation and D1 precursor processing in vivo in *Synechocystis* sp. PCC 6803. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 5830–5835. [CrossRef]

137. Chotewutmontri, P.; Williams-Carrier, R.; Barkan, A. Exploring the Link between Photosystem II Assembly and Translation of the Chloroplast psbA mRNA. *Plants* **2020**, *9*, 152. [CrossRef] [PubMed]

138. Liberton, M.; Howard Berg, R.; Heuser, J.; Roth, R.; Pakrasi, H.B. Ultrastructure of the membrane systems in the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6803. *Protoplasma* **2006**, *227*, 129–138. [CrossRef]

139. van de Meene, A.M.; Hohmann-Marriott, M.F.; Vermaas, W.F.; Roberson, R.W. The three-dimensional structure of the cyanobacterium *Synechocystis* sp. PCC 6803. *Arch. Microbiol.* **2006**, *184*, 259–270. [CrossRef]

140. Nevo, R.; Charuvi, D.; Shimoni, E.; Schwarz, R.; Kaplan, A.; Ohad, I.; Reich, Z. Thylakoid membrane perforations and connectivity enable intracellular traffic in cyanobacteria. *EMBO J.* **2007**, *26*, 1467–1473. [CrossRef]

141. Huokko, T.; Ni, T.; Dykes, G.F.; Simpson, D.M.; Brownridge, P.; Conradi, F.D.; Beynon, R.J.; Nixon, P.J.; Mullineaux, C.W.; Zhang, P.; et al. Probing the biogenesis pathway and dynamics of thylakoid membranes. *Nat. Commun.* **2021**, *12*, 3475. [CrossRef]

142. Rast, A.; Schaffer, M.; Albert, S.; Wan, W.; Pfeffer, S.; Beck, F.; Plitzko, J.M.; Nickelsen, J.; Engel, B.D. Biogenic regions of cyanobacterial thylakoids form contact sites with the plasma membrane. *Nat. Plants* **2019**, *5*, 436–446. [CrossRef] [PubMed]

143. Selao, T.T.; Zhang, L.; Knoppovaa, J.; Komenda, J.; Norling, B. Photosystem II Assembly Steps Take Place in the Thylakoid Membrane of the Cyanobacterium *Synechocystis* sp. PCC6803. *Plant Cell Physiol.* **2016**, *57*, 878. [CrossRef]

144. Zak, E.; Norling, B.; Andersson, B.; Pakrasi, H.B. Subcellular localization of the BtpA protein in the cyanobacterium *Synechocystis* sp. PCC 6803. *Eur. J. Biochem.* **1999**, *261*, 311–316. [CrossRef]

145. Dahlgren, K.K.; Gates, C.; Lee, T.; Cameron, J.C. Proximity-based proteomics reveals the thylakoid lumen proteome in the cyanobacterium *Synechococcus* sp. PCC 7002. *Photosynth. Res.* **2021**, *147*, 177–195. [CrossRef]

146. MacGregor-Chatwin, C.; Sener, M.; Barnett, S.F.H.; Hitchcock, A.; Barnhart-Dailey, M.C.; Maghlaoui, K.; Barber, J.; Timlin, J.A.; Schulten, K.; Hunter, C.N. Lateral Segregation of Photosystem I in Cyanobacterial Thylakoids. *Plant Cell* **2017**, *29*, 1119–1136. [CrossRef]

147. Casella, S.; Huang, F.; Mason, D.; Zhao, G.Y.; Johnson, G.N.; Mullineaux, C.W.; Liu, L.N. Dissecting the Native Architecture and Dynamics of Cyanobacterial Photosynthetic Machinery. *Mol. Plant* **2017**, *10*, 1434–1448. [CrossRef]

148. Mares, J.; Strunecky, O.; Bucinska, L.; Wiedermannova, J. Evolutionary Patterns of Thylakoid Architecture in Cyanobacteria. *Front Microbiol.* **2019**, *10*, 277. [CrossRef] [PubMed]

149. Li, M.; Ma, J.; Li, X.; Sui, S.F. In situ cryo-ET structure of phycobilisome-photosystem II supercomplex from red alga. *Elife* **2021**, *10*, e69635. [CrossRef]

Article

Translational Frameshifting in the *chlD* Gene Gives a Clue to the Coevolution of the Chlorophyll and Cobalamin Biosyntheses

Stepan Kuznetsov [1], Alexander Milenkin [1] and Ivan Antonov [2,3,*]

1   Moscow Institute of Physics and Technology, 141701 Dolgoprudny, Russia;
     stepan.v.kuznetsov@phystech.edu (S.K.); milenkin.aa@phystech.edu (A.M.)
2   Institute of Bioengineering, Research Center of Biotechnology, Russian Academy of Science,
     117312 Moscow, Russia
3   Laboratory of Bioinformatics, Faculty of Computer Science, National Research University Higher School of
     Economics, 101000 Moscow, Russia
*   Correspondence: ivan.antonov@gatech.edu

Abstract: Today, hundreds of prokaryotic species are able to synthesize chlorophyll and cobalamin (vitamin B12). An important step in the biosynthesis of these coenzymes is the insertion of a metal ion into a porphyrin ring. Namely, Mg-chelatase ChlIDH and aerobic Co-chelatase CobNST are utilized in the chlorophyll and vitamin B12 pathways, respectively. The corresponding subunits of these enzymes have common evolutionary origin. Recently, we have identified a highly conserved frameshifting signal in the *chlD* gene. This unusual regulatory mechanism allowed production of both the small and the medium chelatase subunits from the same gene. Moreover, the *chlD* gene appeared early in the evolution and could be at the starting point in the development of the chlorophyll and B12 pathways. Here, we studied the possible coevolution of these two pathways through the analysis of the chelatase genes. To do that, we developed a specialized Web database with comprehensive information about more than 1200 prokaryotic genomes. Further analysis allowed us to split the coevolution of the chlorophyll and B12 pathway into eight distinct stages.

Keywords: chlorophyll; cobalamin; vitamin B12; frameshifting; evolution; chelatase; chlIDH; cobNST

1. Introduction

The early evolution of photosythesis has been an intriguing but complex research subject for many years [1,2]. Clearly, the appearance of photosynthesis required the ability of the ancient prokaryotes to synthesize chlorophyll (CHL). Chlorophyll is a relatively complex organic compound that includes Mg^{2+} ion in its tetrapyrrole ring [3]. Modern organisms utilize a number of different enzymes to produce chlorophyll from a metal-free tetrapyrrole intermediate Uroporphyrinogen III (Uro) [4]. On the other hand, other species utilize several alternative pathways that can turn Uro into cobalamin (vitamin B12), heme, or coenzyme F430 (Figure 1). These coenzymes contain different metal ions in their tetrapyrrole rings [5]. Namely, cobalamine (vitamin B12), haem, and coenzyme F430 include Co^{2+}, Fe^{2+} and Ni^{2+}, respectively. The enzymes that insert those ions into the corresponding organic rings are called chelatases.

The biosynthesis of Mg^{2+}-containing chlorophylls and bacteriochlorophylls requires the magnesium chelatase that performs the insertion of the magnesium ion (Mg^{2+}) into protoporphyrin IX [6]. This enzyme consists of small (I), medium (D), and large (H) subunits [7,8]. The corresponding genes are called *chlI*, *chlD*, and *chlH* in the genomes of chlorophyll-producing organisms and *bchI*, *bchD*, and *bchH* in bacteriochlorophyll producers. To reduce the number of different gene names in this work, we will use the *chlI*, *chlD*, and *chlH* terminology for both groups of genes.
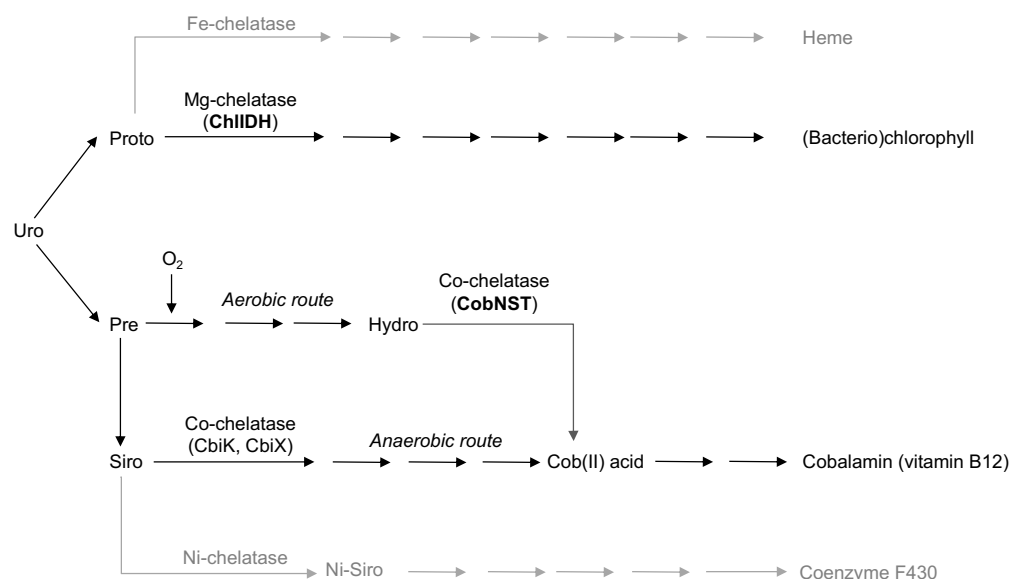
Figure 1. Overview of the existing biochemical pathways that produce (Bacterio)chlorophyll, Heme, Cobalamin, and Coenzyme F430 from Uroporphyrinogen III (Uro). The magnesium-chelatase ChlIDH inserts Mg^{2+} ion in the Protoporphyrin IX (Proto). There are two alternative pathways to synthesize Cobalamin. The CobNST chelatase inserts Co^{2+} in the aerobic path. It has been suggested that ChlIDH and CobNST have common evolutionary origin.

Interestingly, the cobalamin biosynthesis pathway has two alternative routes—the anaerobic and aerobic ones (Figure 1) [9]. Strict anaerobes only posses the anaerobic B12 pathway. On the other hand, cobalamin-producing facultative anaerobes or aerobic organisms usually possess both routes and switch between them depending on the current living conditions [10]. From the evolutionary point of view, it is logical to assume that the aerobic route could only evolve when free oxygen was available in the Earth's atmosphere, e.g., only after the CHL pathway and oxygenic photosynthesis have already existed. Moreover, the interplay between the CHL pathway and the aerobic B12 route is not limited to the oxygen itself. It has been well established that the cobalt chelatase CobNST from the aerobic route resembles the magnesium chelatase ChlIDH as it also consists of the small, medium and large subunits encoded by *cobS*, *cobT*, and *cobN* genes, respectively [11]. Additionally, the *chlI* and *chlD* from the chlIDH chelatase are likely to be homologous to the *cobS* and *cobT* genes of the cobNST chelatase [12]. The similarity between the large subunit genes (i.e., *chlH* and *cobN*) has also been reported [13].

Later studies have revealed additional existing strategies to encode the cobNST enzyme. First, Rodionov et al. [14] have suggested that in some species, the products of the *chlI* and *chlD* can replace the *cobS* and *cobT* genes and function as the small and medium subunits of the aerobic Co-chelatase. Moreover, a functional programmed ribosomal frameshifting (PRF) found in some *chlD* genes may allow the production of both the small and the medium subunits of the cobNST chelatase from the same gene [15,16]—see Supplementary Figure S1. Altogether, this indicates a close connection between the CHL pathway and the aerobic B12 route suggesting their possible coevolution [17].

Here, we attempted to organize all the different strategies to encode cobNST and chlIDH enzymes and suggest a logical model of the CHL and aerobic B12 pathways coevolution. Our strategy was to pinpoint possible prokaryotic genomes that may still be using the ancient strategies to encode Mg- and/or Co-chelatases. Such "living fossil" species can be used to reconstruct the early stages of the CHL and B12 pathway evolution.

2. Materials and Methods

The Chelatase DB was developed using the Django framework implemented in Python. To populate the database the tBLASTn [18] tool was used to identify the genomes containing *chlD* gene(s). The translations of the three *chlD* genes with the validated frameshifting signals were used as queries [16]. To allow the identification of the other frameshifted *chlD* genes, two separate tBASTn searches (E-value threshold = 10^{-6}) were performed using the N-terminal (i.e., before the frameshift) or the C-terminal (after the frameshift) parts of each query protein. The genomic regions with adjacent hits (in the correct order) with up to one frameshift were classified as *chlD* genes.

In order to identify the small and the large chelatase subunit genes as well as the other genes from the chlorophyll and cobalamin biosynthesis pathways, tBLASTn search was performed for a set of reference query proteins. Given the similarities between different chelatase subunits, we imposed additional constraints on their lengths. Namely, we required that the small chelatase subunits were between 250 aa and 500 aa, medium were between 500 aa and 800 aa, and large were between 1000 aa and 2000 aa (these values corresponded to the lengths of the annotated chelatase subunits). The identified genes were automatically annotated (i.e., assigned a gene name such as *chlD*, *bchD*, or *cobT*) by the best reciprocal hit approach.

We used the identified representative sets of the genes from the CHL and B12 pathways to predict whether a given prokaryotic genome possessed the corresponding biochemical pathway. Additionally, the automatically assigned genotype of the species mentioned in the text was double-checked with the information from the KEGG database (the "Porphyrin metabolism" section) [19]. The phylogenetic tree was constructed using RAxML version 8.2.12 [20]. Subsequent phylogenetic analyses and visualizations were performed using the "ape" R package [21]. The RNA secondary structures were predicted using the RNAfold web server [22] available at http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi (accessed on 1 April 2022).

All the data files and scripts used to generate figures and tables from the article are available at https://github.com/vanya-antonov/article-chelatase-db (accessed on 1 April 2022).

3. Results

Our recent analysis [16] as well as earlier studies [8,14,23] have suggested that there were several different strategies to encode aerobic cobalt chelatase cobNST. Interestingly, in a number of species, the products of the Mg-chelatase genes (*chlD* and/or *chlI*) functioned as corresponding subunits of the cobNST complex (Supplementary Figure S1). To explore the diversity of the strategies utilized by different prokaryotes to encode chelatases, we developed a specialized Web database called Chelatase DB (http://ivanya.com/chelatase/ (accessed on 1 April 2022)). For each prokaryotic genome, the Chelatase DB provided information about the number of the small, medium, and large subunit genes, the existence of the chlorophyll and/or B12 biosynthesis pathway and reference to the KEGG database for detailed analysis. Additionally, the database contained a collection of the frameshifting signals predicted in some *chlD* genes.

To identify the possible starting point in the evolution of the CHL pathway, the *chlD* gene encoding medium subunit of the Mg-chelatase was analyzed. We chose this gene because it was likely to appear early in the evolution, since it may have been present in the genome of the last universal common ancestor (LUCA) [24]. Thus, it was possible that the modern genes encoding medium chelatase subunits were descendants of the putative *chlD* gene from LUCA. Our previous analysis [16] of the frameshifted *chlD* (fs-*chlD*) genes has demonstrated that such genes were present in all major prokaryotic phyla (including archaea, proteobacteria, and actinobacteria)—Supplementary Figure S2. Indeed, on the phylogenetic tree of 286 medium subunit genes all the fs-*chlD* genes were located near the root of the tree (Supplementary Figure S3). Thus, the modern *chlD*, *bchD*, and *cobT* genes may have originated from the ancient fs-*chlD* gene.
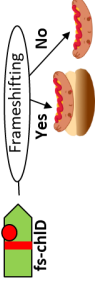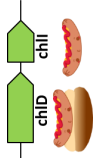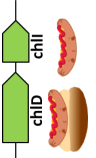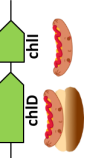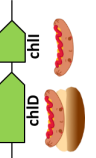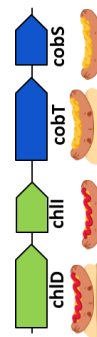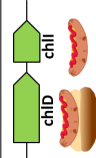
Our next goal was to use the functionality of the Chelatase DB to reveal a possible evolutionary trajectory of the CHL and B12 pathways with respect to the corresponding chelatase genes. Namely, we inspected different chlIDH/cobNST genotypes and matched them with the corresponding predicted phenotypes (i.e., the ability to synthesize CHL and/or B12). This allowed us to define eight stages in the evolution of CHL and B12 pathways as well as *chlD* gene—see Table 1. Since fs-*chlD* gene had an ancient origin, it was logical to assume that it appeared in the evolution before the complete CHL and B12 pathways existed. Therefore, we search the Chelatase DB for the genomes with fs-*chlD* genes that did not have the genes from the CHL and B12 pathways. Indeed, we identified several such species including *Spirochaeta thermophyla* and *Brevefilum fermentans* (Table 1, stage 1). All of these bacteria that corresponded to the conditions of the prechlorophyll era were anaerobic [25]. Moreover, according to the global prokaryotic phylogentic tree, both Spirochaetes and Chloroflexi were among the most ancient prokaryotic phyla that are present today [26]. The presence of the frameshifting signals in the fs-*chlD* genes from *S. thermophyla* and *B. fermentans* suggested that they produced two proteins from the same mRNA (Supplementary Figure S4). In the *chlD* genes from various taxa −1 frameshifts have been observed more frequently than +1 frameshifts [16]. The recoding signals in these genes usually consisted of AT-rich "slippery sites" together with a stimulatory RNA secondary structures located downstream. More info about the frameshifting mechanism can be found elsewhere [27]. Due to the stochastic nature of the programmed frameshifting, ribosomes reading the same mRNA can produce two different proteins that allows cotranslational assembly of protein complexes. Consequently, it was likely that the two proteins produced from fs-*chlD* mRNA interacted with each other—similarly to the CobS and CobT subunits [28]. However, the function of this protein complex was unclear since the genomes of the corresponding bacteria completely lacked the CHL and B12 pathways as well as the large chelatase subunit gene (either *chlH* or *cobN*). Together, these observations indicated that the *S. thermophyla* and *B. fermentans* could be living prokaryotic fossils and their genotypes can be considered as the starting point in the evolution of the CHL and B12 pathways.

It has previously been reported that the *chlI* gene encoding the small chelatase subunit was homologous to the N-terminal part of the *chlD* gene [29]. Therefore, the next putative evolutionary event was duplication of the N-terminal part of the fs-*chlD* gene and formation of a separate *chlI* gene. In this case, the frameshifting signal in the fs-*chlD* was likely to disappear since there was no need to produce both proteins from a single gene. Accordingly, we found another prokaryotic group that had *chlD* and *chlI* genes but still lacked the CHL and B12 pathways (Table 1, stage 2). Interestingly, the ability of the cotranslational interaction between the products of these proto-*chlI* and proto-*chlD* genes was possible since they were located next to each other on the genome and can be present in the same polycistronic mRNA. Although the function of the proteins produced from these two genes was still unclear, the corresponding genotypes can be considered as the next stage in the early evolution of CHL biosynthesis. Moreover, we identified several Chloroflexi species (from the Anaerolineaceae group) with a potential intermediate genotype between Stage 1 and 2. Namely, *Anaerolinea thermophila UNI-1*, *Ornatilinea apprima*, *Longilinea arvoryzae*, and *Levilinea saccharolytica* possessed the fs-chlD genes with characteristic frameshifting signals but also had a separate *chlI* gene. Interestingly, these species were also strict anaerobes and thermophiles suggesting their ancient origin [30]. Thus, there were a number of living prokaryotes that provided insights into the early evolution of the modern Mg-chelatase enzyme.

A good representative case of the stage 3 was photosynthetic proteobacteria *Polynucleobacter duraquae* that had the *chlH* gene encoding the large subunit of the Mg-chelatase (Table 1, stage 3). Similar to the genomes from stage 2, all three chelatase subunit genes were located at the same genomic locus, possibly allowing cotraslational assembly of the Mg-chelatase from a polycystronic mRNA. According to the KEGG database, *P. duraquae* possessed all the major genes required for CHL biosynthesis. It was likely that this bacteria was only capable of anoxygenic photosynthesis [31] supporting its relatively early position in our proposed evolutionary system.

Table 1. The possible stages of the evolution of the genes encoding Mg-/Co-chelatases as well as CHL and B12 pathways.

Stage	Example Species (Phila)	Synthesis of the Medium and Small Proteins	Function of the Medium and Small Proteins	B12 Pathway (Anaerobic)	Large Subunit (chlH/cobN)	CHL Pathway	B12 Pathway (Aerobic)
1	*Spirochaeta thermophila* (Spirochaetes) *Brevefilum fermentans* (Chloroflexi)	fs-chlD (Frameshifting Yes/No)	Unknown	-	-	-	-
2	*Thermodesulfobacterium commune* (Thermodesulfobacteria) *Thermodesulfovibrio yellowstonii* (Nitrospirae)	chlD, chlI	Unknown	-	-	-	-
3	*Polynucleobacter duraquae* (Proteobacteria)	chlD, chlI	Mg-chelatase (small and medium subunits)	-	chlH	Yes	-
4	*Halorhodospira halophila* (Proteobacteria) *Thiocystis violascens* (Proteobacteria)	chlD, chlI	Mg-(Co-?)chelatase (small and medium subunits)	Yes	chlH	Yes	Yes?
5	*Synechococcus elongatus* (Cyanobacteria) *Prochlorococcus marinus* (Cyanobacteria)	chlD, chlI	Mg- and Co- chelatases (small and medium subunits)	Yes	chlH + cobN	Yes	Yes
6	*Rhodospirillum rubrum* (Proteobacteria) *Dinoroseobacter shibae* (Proteobacteria) *Rhodovulum sulfidophilum* (Proteobacteria)	chlD, chlI, cobT, cobS	Mg- and Co- chelatases (small and medium subunits)	Yes	chlH + cobN	Yes	Yes
7	*Mycobacterium tuberculosis* (Actinobacteria) *Prauserella marina* (Actinobacteria)	chlD, chlI	Small and medium subunits of Co- chelatases	Yes	cobN	-	Yes
8	*Pseudomonas aeruginosa* (Proteobacteria) *Methanocaldococcus fervens* (Archaea)	fs-chlD (Frameshifting Yes/No)	Small and medium subunits of Co- chelatases	Yes	cobN	-	Yes

At stage 4, we placed anoxygenic proteobacteria [32] that in addition to chlorophyll biosynthesis also had the B12 pathway (Table 1, stage 4). These bacteria possessed anaerobic B12 pathway as well as many genes from the aerobic route. However, it was unclear if they were able to synthesize cobalamin using oxygen since these species were anaerobic. Additionally, their genome did not encode the large subunit of the aerobic Co-chelatase (the *cobN* gene). Thus, if the B12 biosynthesis was possible through the aerobic route, then the Mg-chelatase from these species might be able to function as a Co-chelatase as well.
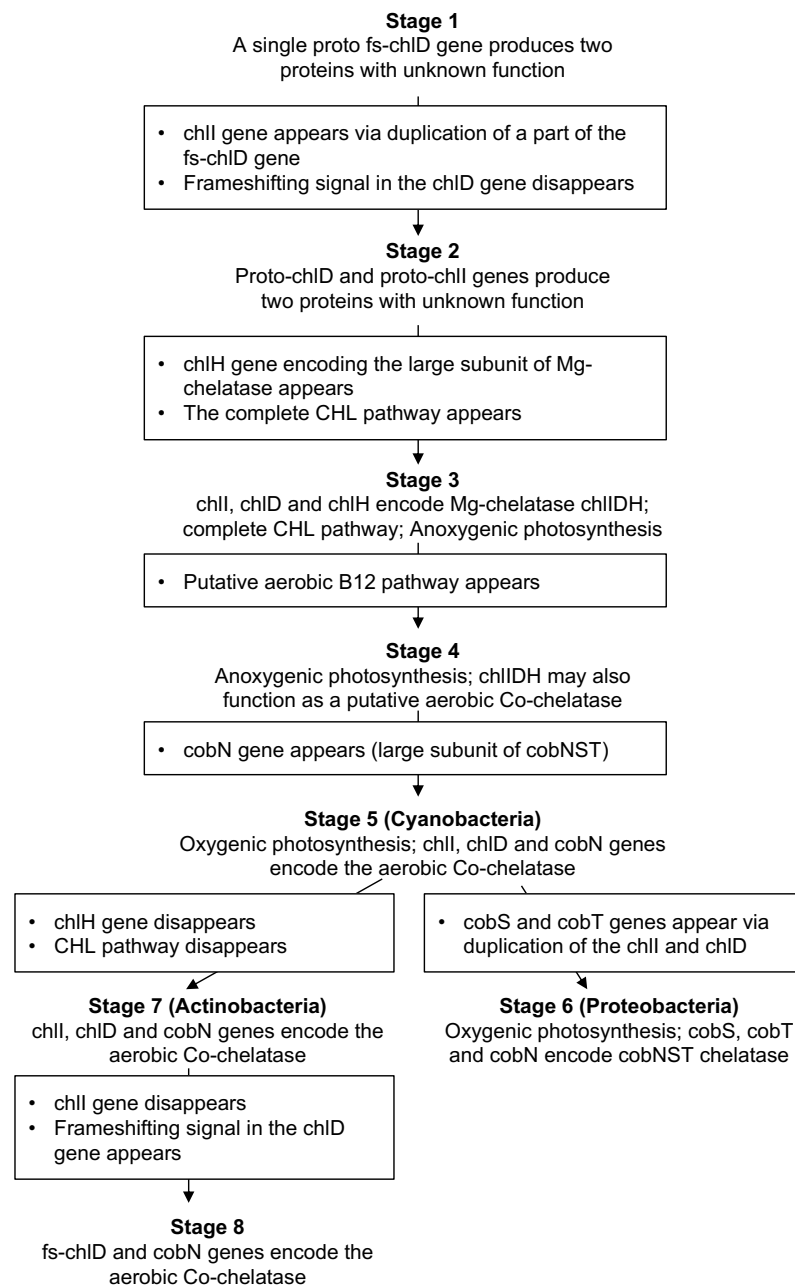
Modern Cyanobacteria (such as *Synechococcus elongatus*) were at the next stage in our model because they were able to perform oxygenic photosynthesis, encoded large subunits for both Mg- and Co-chelatases, and could produce cobalamin via anaerobic and aerobic routes (Table 1, stage 5). Interestingly, Cyanobacteria did not have the *cobS* and *cobT* genes that encode the small and the medium subunits of the cobNST chelatase. It has been suggested that the corresponding subunits from the Mg-chelatase (*chlI* and *chlD*) could substitute them [14].

A number of oxygenic phototrophic Proteobacteria had the most advanced genotype with respect to the Mg- and Co-chelatases (Table 1, stage 6). In comparison to Cyanobacteria, their genomes contained *cobS* and *cobT* genes that were likely duplicated from the *chlI* and *chlD* and subsequently specialized to function as subunits of the cobNST chelatase. Since many of these bacteria (e.g., *Rhodospirillum rubrum*) were able to live under both aerobic and aerobic conditions, their genomes encoded both routes of B12 biosynthesis [33].

Starting from stage 1 to stage 6, we observed gradual development of the gene sets encoding chlIDH and cobNST chelatases. On the other hand, a number of modern heterotrophic Actinobacteria (such as *Mycobacterium tuberculosis*) did not have the CHL pathway but encoded both pathways of B12 biosynthesis (Table 1, stage 7). However, they utilized the *chlI* and *chlD* genes, rather than *cobN* and *cobS*, to encode the small and the medium subunits of their aerobic Co-chelatase. This suggested that their chelatase genotypes originated from the Cyanobacteria where *chlI* and *chlD* genes performed similar function. Thus, we hypothesized that Actinobacteria from stage 7 adopted some genes from the Cyanobacterial CHL pathway (stage 5) and lost the rest of the CHL biosynthesis genes. Finally, stage 8 represented further simplification of the chelatase genotype (Table 1, stage 8) where a frameshifting signal in the fs-*chlD* gene allowed the production of both the small and the medium subunits of the Co-chelatase for the aerobic route of B12 biosythesis [16].

4. Discussion

In this paper, we presented a model for the possible coevolution of the CHL and B12 (aerobic route) pathways (Figure 2). To develop this model, we assumed that among the thousands of the sequenced prokaryotic genomes, there were examples of so-called "living fossils", i.e., the direct descendants of the species from the early days of cellular life on Earth. Our analysis was based on the *chlD* gene due to its likely early appearance in the evolution [24]. We hypothesized that the frameshifted version of this gene was the starting point in the early evolution of CHL pathway. Namely, we identified this gene in several bacteria that could be considered as living fossils (stage 1). We also identified several other bacteria where a part of *chlD* gene was probably duplicated producing the proto-*chlI* gene (stage 2). Importantly, the stage 1 and 2 phlyla (Spirochaetes, Chloroflexi, Thermodesulfobacteria, and Nitrospirae) that appeared in the evolution before the Proteobacteria from the stage 3 and 4 [26] supporting our suggested evolutionary order. Interestingly, species from stage 1 and 2 did not have B12 or CHL pathways as well as either large subunit genes. On the other hand, the fs-*chlD*, *chlI* and *chlD* genes from these species can produce two proteins that may interact with each other and form a separate complex [28]. To our knowledge, the function of the *chlD* + *chlI* complex (without the a subunit protein) has not been studied and its function is unknown. Thus, further studies of these genomes may reveal the new function for the corresponding proteins.

Stage 1
A single proto fs-chlD gene produces two
proteins with unknown function

- chlI gene appears via duplication of a part of the
  fs-chlD gene
- Frameshifting signal in the chlD gene disappears

Stage 2
Proto-chlD and proto-chlI genes produce
two proteins with unknown function

- chlH gene encoding the large subunit of Mg-
  chelatase appears
- The complete CHL pathway appears

Stage 3
chlI, chlD and chlH encode Mg-chelatase chlIDH;
complete CHL pathway; Anoxygenic photosynthesis

- Putative aerobic B12 pathway appears

Stage 4
Anoxygenic photosynthesis; chlIDH may also
function as a putative aerobic Co-chelatase

- cobN gene appears (large subunit of cobNST)

Stage 5 (Cyanobacteria)
Oxygenic photosynthesis; chlI, chlD and cobN genes
encode the aerobic Co-chelatase

- chlH gene disappears
- CHL pathway disappears

- cobS and cobT genes appear via
  duplication of the chlI and chlD

Stage 7 (Actinobacteria)
chlI, chlD and cobN genes encode the
aerobic Co-chelatase

Stage 6 (Proteobacteria)
Oxygenic photosynthesis; cobS, cobT
and cobN encode cobNST chelatase

- chlI gene disappears
- Frameshifting signal in the chlD
  gene appears

Stage 8
fs-chlD and cobN genes encode the
aerobic Co-chelatase

Figure 2. A putative chain of events that could happen between different stages during the evolution of the CHL and aerobic B12 pathways.

The appearance of the CHL and the aerobic B12 pathways as well as oxygenic photosynthesis was observed in Cyanobacteria at stage 5 of our system. Clearly, the aerobic B12 pathway could only appear after the oxygenic photosynthesis; however, we were not able to find any oxygenic phototroph without the aerobic B12 pathway. Although we believe that such putative bacteria existed at some evolutionary point, they probably adopted the more efficient aerobic B12 pathway after its appearance and therefore could not be found today. Proteobacteria at stage 6 represented the most advanced chelatase genotypes where all the subunits of the chlIDH and cobNST enzymes were encoded by specialized genes. On the other hand, a possible reduction of the Cyanobacterial genomes could gave rise to the modern Actinobacteria that utilized *chlI* and *chlD* genes to encode small and medium cobNST subunits. Indeed, according to the global phylogenetic tree, Actinobacteria had a common ancestor with Cyanobacteria which appeared in the evolution first [26]. Finally, we observed that some prokaryotes utilized the ancient strategy to encode both the small

and medium subunits of the aerobic Co-chelatase using frameshifting signal in the fs-*chlD* gene (stage 8). The *chlD* genes with frameshifting signals in these genomes likely originated via horizontal gene transfer rather than the signals evolving once again in the conventional *chlD*/*bchD* genes. This is supported by the fact the fs-*chlD* genes do not group with the other genes without frameshifts (Supplementary Figure S3).

Another interesting question was related to the anaerobic cobalamin biosynthesis pathway—e.g., whether it appeared before or after the CHL biosynthesis pathway. Interestingly, the Chelatase DB included several archaeal species (such as *Methanobacterium paludis* and *Thermoplasma acidophilum*) that contained the *chlD* gene (without a frameshift) as well as the anaerobic B12 pathway. Again, the function of the *chlD* gene in these organisms was unclear, but the presence of the anaerobic B12 route may suggest that this pathway existed before the development of the CHL pathway.

To conclude, our analysis suggested a logical model for the coevolution of CHL and B12 pathways. We pinpointed specific prokaryotic species that can be considered as living fossils and their further study may shed a new light on the early evolution of these pathways. To our knowledge, the function of the *chlI* and *chlD* proteins without the large subunit (*chlH* or *cobN*) remained unexplored, and here, we suggested some model organisms where this could be studied.

Abbreviations

The following abbreviations are used in this manuscript:

B12	vitamin B12 or cobalamine (contains Co^{2+})
BCHL	bacteriochlorophyll (contains Mg^{2+})
CHL	chlorophyll (contains Mg^{2+})
F430	coenzyme F430 (contains Ni^{2+})
fs-*chlD* gene	frameshifted *chlD* gene
Proto	Protoporphyrin IX
Uro	Uroporphyrinogen III

References

1. Blankenship, R.E.; Hartman, H. The origin and evolution of oxygenic photosynthesis. *Trends Biochem. Sci.* **1998**, *23*, 94–97. [CrossRef]
2. Blankenship, R.E. Early evolution of photosynthesis. *Plant Physiol.* **2010**, *154*, 434–438. [CrossRef]
3. Willows, R.; Hansson, M. Mechanism, structure, and regulation of magnesium chelatase. *Porphyr. Handb. II* **2003**, *113*, 1–48.
4. Thweatt, J.L.; Canniffe, D.P.; Bryant, D.A. Biosynthesis of chlorophylls and bacteriochlorophylls in green bacteria. *Adv. Bot. Res.* **2019**, *90*, 35–89.
5. Guilard, R.; Kadish, K.M.; Smith, K.M.; Guilard, R. *The Porphyrin Handbook*; Academic Press: New York, NY, USA, 2003; Volume 18.
6. Walker, C.J.; Willows, R.D. Mechanism and regulation of Mg-chelatase. *Biochem. J.* **1997**, *327 Pt 2*, 321–333. [CrossRef]
7. Bollivar, D.W.; Suzuki, J.Y.; Beatty, J.T.; Dobrowolski, J.M.; Bauer, C.E. Directed mutational analysis of bacteriochlorophyll a biosynthesis in *Rhodobacter capsulatus*. *J. Mol. Biol.* **1994**, *237*, 622–640. [CrossRef]
8. Gibson, L.; Willows, R.D.; Kannangara, C.G.; von Wettstein, D.; Hunter, C.N. Magnesium-protoporphyrin chelatase of *Rhodobacter sphaeroides*: Reconstitution of activity by combining the products of the bchH,-I, and-D genes expressed in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 1941–1944. [CrossRef]
9. Fang, H.; Kang, J.; Zhang, D. Microbial production of vitamin B12: A review and future perspectives. *Microb. Cell Factories* **2017**, *16*, 1–14. [CrossRef]
10. Crespo, A.; Blanco-Cabra, N.; Torrents, E. Aerobic vitamin B12 biosynthesis is essential for *Pseudomonas aeruginosa* class II ribonucleotide reductase activity during planktonic and biofilm growth. *Front. Microbiol.* **2018**, *9*, 986. [CrossRef]
11. Fodje, M.; Hansson, A.; Hansson, M.; Olsen, J.; Gough, S.; Willows, R.; Al-Karadaghi, S. Interplay between an AAA module and an integrin I domain may regulate the function of magnesium chelatase. *J. Mol. Biol.* **2001**, *311*, 111–122. [CrossRef]
12. Escalante-Semerena, J.C.; Warren, M.J. Biosynthesis and use of cobalamin (B12). *EcoSal Plus* **2008**, *3*. [CrossRef]
13. Lundqvist, J.; Elmlund, D.; Heldt, D.; Deery, E.; Söderberg, C.A.; Hansson, M.; Warren, M.; Al-Karadaghi, S. The AAA+ motor complex of subunits CobS and CobT of cobaltochelatase visualized by single particle electron microscopy. *J. Struct. Biol.* **2009**, *167*, 227–234. [CrossRef]
14. Rodionov, D.A.; Vitreschak, A.G.; Mironov, A.A.; Gelfand, M.S. Comparative genomics of the vitamin B12 metabolism and regulation in prokaryotes. *J. Biol. Chem.* **2003**, *278*, 41148–41159. [CrossRef]
15. Antonov, I.; Coakley, A.; Atkins, J.F.; Baranov, P.V.; Borodovsky, M. Identification of the nature of reading frame transitions observed in prokaryotic genomes. *Nucleic. Acids Res.* **2013**, *41*, 6514–6530. [CrossRef]
16. Antonov, I.V. Two Cobalt Chelatase Subunits Can Be Generated from a Single chlD Gene via Programed Frameshifting. *Mol. Biol. Evol.* **2020**, *37*, 2268–2278. [CrossRef]
17. Dailey, H.A. Illuminating the black box of B12 biosynthesis. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 14823–14824. [CrossRef]
18. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 421. [CrossRef]
19. Kanehisa, M.; Araki, M.; Goto, S.; Hattori, M.; Hirakawa, M.; Itoh, M.; Katayama, T.; Kawashima, S.; Okuda, S.; Tokimatsu, T.; et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **2007**, *36*, D480–D484. [CrossRef]
20. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [CrossRef]
21. Paradis, E.; Schliep, K. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **2018**, *35*, 526–528. [CrossRef]
22. Lorenz, R.; Bernhart, S.H.; Zu Siederdissen, C.H.; Tafer, H.; Flamm, C.; Stadler, P.F.; Hofacker, I.L. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **2011**, *6*, 26. [CrossRef]
23. Debussche, L.; Couder, M.; Thibaut, D.; Cameron, B.; Crouzet, J.; Blanche, F. Assay, purification, and characterization of cobaltochelatase, a unique complex enzyme catalyzing cobalt insertion in hydrogenobyrinic acid a, c-diamide during coenzyme B12 biosynthesis in *Pseudomonas denitrificans*. *J. Bacteriol.* **1992**, *174*, 7445–7451. [CrossRef]
24. Weiss, M.C.; Sousa, F.L.; Mrnjavac, N.; Neukirchen, S.; Roettger, M.; Nelson-Sathi, S.; Martin, W.F. The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* **2016**, *1*, 16116. [CrossRef]
25. Angelov, A.; Liebl, S.; Ballschmiter, M.; Bomeke, M.; Lehmann, R.; Liesegang, H.; Daniel, R.; Liebl, W. Genome sequence of the polysaccharide-degrading, thermophilic anaerobe *Spirochaeta thermophila* DSM 6192. *J. Bacteriol.* **2010**, *192*, 6492–6493. [CrossRef]
26. Zhu, Q.; Mai, U.; Pfeiffer, W.; Janssen, S.; Asnicar, F.; Sanders, J.G.; Belda-Ferre, P.; Al-Ghalith, G.A.; Kopylova, E.; McDonald, D.; et al. Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea. *Nat. Commun.* **2019**, *10*, 1–14. [CrossRef]
27. Atkins, J.F.; Loughran, G.; Bhatt, P.R.; Firth, A.E.; Baranov, P.V. Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use. *Nucleic Acids Res.* **2016**, *44*, 7007–7078. [CrossRef]
28. Cameron, B.; Guilhot, C.; Blanche, F.; Cauchois, L.; Rouyez, M.; Rigault, S.; Levy-Schil, S.; Crouzet, J. Genetic and sequence analyses of a *Pseudomonas denitrificans* DNA fragment containing two cob genes. *J. Bacteriol.* **1991**, *173*, 6058–6065. [CrossRef]
29. Sousa, F.L.; Shavit-Grievink, L.; Allen, J.F.; Martin, W.F. Chlorophyll biosynthesis gene evolution indicates photosystem gene duplication, not photosystem merger, at the origin of oxygenic photosynthesis. *Genome Biol. Evol.* **2013**, *5*, 200–216. [CrossRef]

30. Sekiguchi, Y.; Yamada, T.; Hanada, S.; Ohashi, A.; Harada, H.; Kamagata, Y. *Anaerolinea thermophila* gen. nov., sp. nov. and *Caldilinea aerophila* gen. nov., sp. nov., novel filamentous thermophiles that represent a previously uncultured lineage of the domain Bacteria at the subphylum level. *Int. J. Syst. Evol. Microbiol.* **2003**, *53*, 1843–1851. [CrossRef]

31. Hahn, M.W.; Schmidt, J.; Pitt, A.; Taipale, S.J.; Lang, E. Reclassification of four Polynucleobacter necessarius strains as representatives of Polynucleobacter asymbioticus comb. nov., Polynucleobacter duraquae sp. nov., *Polynucleobacter yangtzensis* sp. nov. and *Polynucleobacter sinensis* sp. nov., and emended description of Polynucleobacter necessarius. *Int. J. Syst. Evol. Microbiol.* **2016**, *66*, 2883.

32. Challacombe, J.F.; Majid, S.; Deole, R.; Brettin, T.S.; Bruce, D.; Delano, S.F.; Detter, J.C.; Gleasner, C.D.; Han, C.S.; Misra, M.; et al. Complete genome sequence of Halorhodospira halophila SL1. *Stand. Genom. Sci.* **2013**, *8*, 206–214. [CrossRef]

33. Schultz, J.; Weaver, P. Fermentation and anaerobic respiration by *Rhodospirillum rubrum* and *Rhodopseudomonas capsulata*. *J. Bacteriol.* **1982**, *149*, 181–190. [CrossRef]

Article

Zeta-Carotene Isomerase (Z-ISO) Is Required for Light-Independent Carotenoid Biosynthesis in the Cyanobacterium *Synechocystis* sp. PCC 6803

Matthew S. Proctor [1,*], Felix S. Morey-Burrows [1], Daniel P. Canniffe [2], Elizabeth C. Martin [1], David J. K. Swainsbury [1,3], Matthew P. Johnson [1], C. Neil Hunter [1], George A. Sutherland [1,*] and Andrew Hitchcock [1,*]

1 School of Biosciences, University of Sheffield, Sheffield S10 2TN, UK
2 Institute of Systems, Molecular & Integrative Biology, University of Liverpool, Liverpool L69 7BE, UK
3 School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK
* Correspondence: matt.proctor@sheffield.ac.uk (M.S.P.); g.sutherland@sheffield.ac.uk (G.A.S.); a.hitchcock@sheffield.ac.uk (A.H.)

Abstract: Carotenoids are crucial photosynthetic pigments utilized for light harvesting, energy transfer, and photoprotection. Although most of the enzymes involved in carotenoid biosynthesis in chlorophototrophs are known, some are yet to be identified or fully characterized in certain organisms. A recently characterized enzyme in oxygenic phototrophs is 15-*cis*-zeta(ζ)-carotene isomerase (Z-ISO), which catalyzes the *cis*-to-*trans* isomerization of the central 15–15′ *cis* double bond in 9,15,9′-tri-*cis*-ζ-carotene to produce 9,9′-di-*cis*-ζ-carotene during the four-step conversion of phytoene to lycopene. Z-ISO is a heme B-containing enzyme best studied in angiosperms. Homologs of Z-ISO are present in organisms that use the multi-enzyme poly-*cis* phytoene desaturation pathway, including algae and cyanobacteria, but appear to be absent in green bacteria. Here we confirm the identity of Z-ISO in the model unicellular cyanobacterium *Synechocystis* sp. PCC 6803 by showing that the protein encoded by the slr1599 open reading frame has ζ-carotene isomerase activity when produced in *Escherichia coli*. A *Synechocystis* Δslr1599 mutant synthesizes a normal quota of carotenoids when grown under illumination, where the photolabile 15–15′ *cis* double bond of 9,15,9′-tri-*cis*-ζ-carotene is isomerized by light, but accumulates this intermediate and fails to produce 'mature' carotenoid species during light-activated heterotrophic growth, demonstrating the requirement of Z-ISO for carotenoid biosynthesis during periods of darkness. In the absence of a structure of Z-ISO, we analyze AlphaFold models of the *Synechocystis*, *Zea mays* (maize), and *Arabidopsis thaliana* enzymes, identifying putative protein ligands for the heme B cofactor and the substrate-binding site.

Keywords: carotenoid; cyanobacteria; *Synechocystis*; zeta-carotene isomerase (Z-ISO); photosynthesis

1. Introduction

Carotenoids are pigments with important structural, light-harvesting, and photoprotective roles in photosynthetic pigment-protein complexes and membranes [1,2]. Carotenoid biosynthesis in phototrophs (for reviews see [3,4]) begins with the condensation of the products of the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway: isopentyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). This generates geranyl pyrophosphate (GPP), which is converted to geranylgeranyl pyrophosphate (GGPP) (via farnesyl-pyrophosphate) by condensation with two further IPP molecules, prior to the production of 15-*cis*-phytoene (phytoene) by the condensation of two molecules of GGPP. How phytoene is converted to carotenoids with longer conjugation length (*N*) polyene chains differs between phototrophic organisms. In phototrophic bacteria, phytoene desaturase (CrtI) catalyzes sequential desaturations of phytoene (colorless with three conjugated carbon–carbon double bonds; *N* = 3) to produce neurosporene (yellow, three desaturations of

phytoene; *N* = 9) or lycopene (orange/red; four desaturations of phytoene; *N* = 11) [3,4]. Neurosporene and/or lycopene are subsequently modified by additional enzymes to produce the carotenoids that are utilized for light harvesting and photoprotection. Conversely, oxygenic phototrophs (cyanobacteria, algae, and plants), green sulfur bacteria (GSB), and Chloroacidobacteria use a multi-enzyme pathway to synthesize lycopene from phytoene (Figure 1); the four enzymes are phytoene desaturase (PDS/CrtP; catalyzes two desaturations of lycopene), zeta(ζ)-carotene isomerase (Z-ISO), ζ-carotene desaturase (ZDS/CrtQ; catalyzes two desaturations of ζ-carotene), and the prolycopene isomerase (CRT-ISO/CrtH) [3,4].
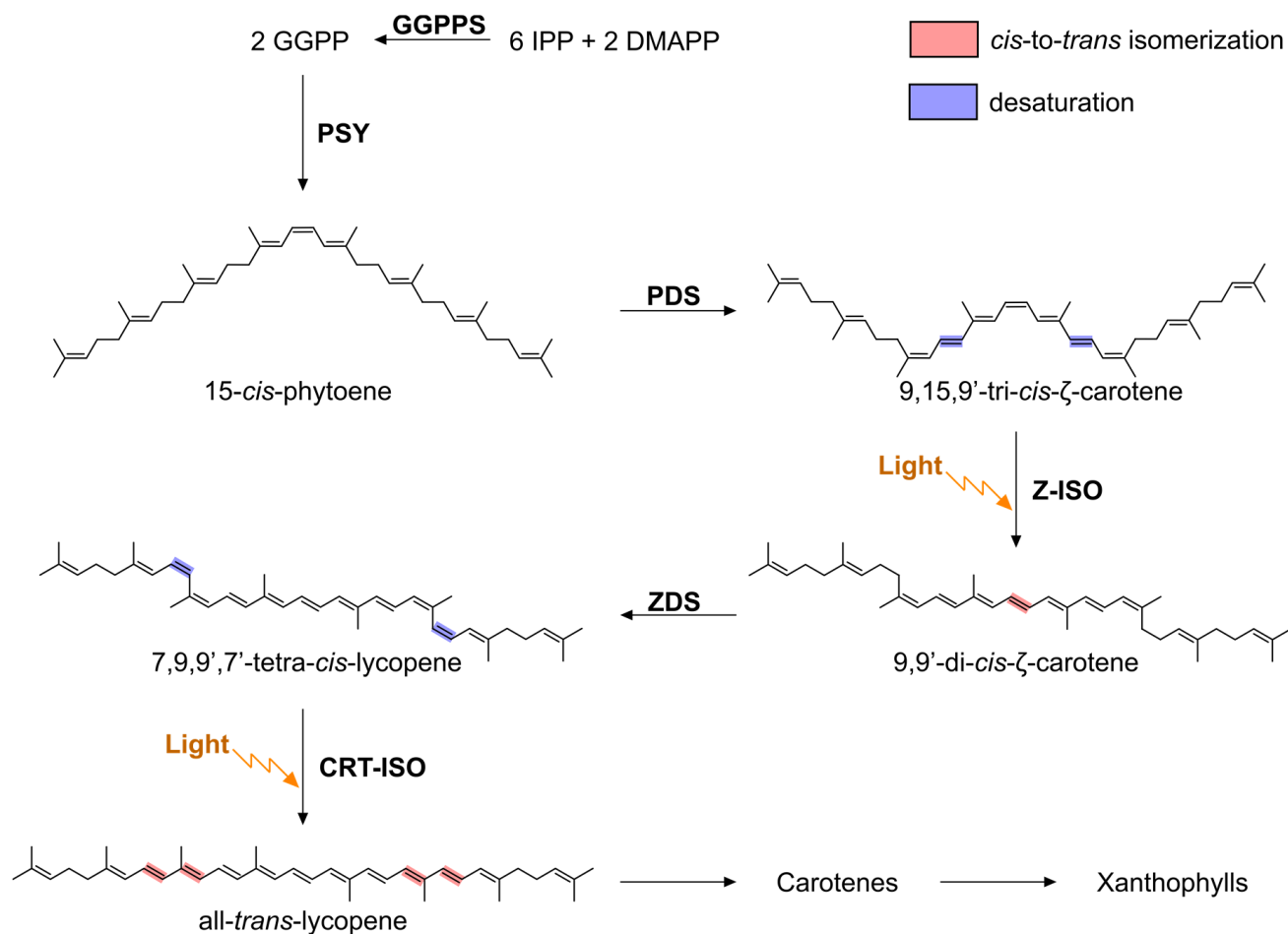
Figure 1. Overview of carotenoid biosynthesis in oxygenic phototrophs focusing on the steps in the conversion of phytoene to lycopene. The products of the MEP pathway, IPP and DMAPP, are condensed to form GGPP by the enzyme GGPP synthase (GGPPS). Two molecules of GGPP are subsequently condensed by the enzyme phytoene synthase (PSY), producing 15-*cis*-phytoene. Phytoene desaturase (PDS) and ζ-carotene desaturase (ZDS) introduce four double bounds by desaturation of the carotenoid backbone (blue highlight) to produce all-*trans*-lycopene, with *cis*-to-*trans* isomerization of bonds in the intermediate species (red highlight) carried out by Z-ISO and CRT-ISO; these *cis*-bonds are photo-labile and can be isomerized non-enzymatically by light (indicated by orange arrows). Subsequent modification of all-*trans*-lycopene (enzymatic steps not shown) yields the carotene and xanthophyll species used for light harvesting, photoprotection, and structural stabilization of proteins and lipid membranes.

Z-ISO is the most recently identified early carotenoid biosynthesis enzyme; the *Zea mays* (maize) enzyme has been demonstrated to be a heme B-containing integral membrane protein that catalyzes the redox-regulated *cis*-to-*trans* conversion of the 15-*cis* double bond in 9,15,9′-tri-*cis*-ζ-carotene to form 9,9′-di-*cis*-ζ-carotene [5,6]. Z-ISO homologs are present in plants, algae, diatoms, and cyanobacteria; the enzymes from maize and *Arabidopsis thaliana* (hereafter *Arabidopsis*) are the best studied [5,7,8] but Z-ISO has also been experimentally identified in *Oryza sativa* (rice) [9,10], the filamentous cyanobacterium *Arthospira* [11], the eukaryotic microalga Euglena [7,11], and the fruits *Citrus sinensis* [12] and tomato [13]. Interestingly, GSB and species of the *Chloracidobacterium* genus appear to lack Z-ISO [8,11]. Maize and *Arabidopsis* mutants accumulate 9,15,9′-tri-*cis*-ζ-carotene and lack carotenoids in the dark, as well as having delayed greening and a lower carotenoid content when exposed to light [8]. In rice, Z-ISO links carotenoid and plant hormone biosynthesis to regulate/modulate photosynthesis and tillering [9,10]. The silencing of *Z-ISO* mRNA in a tomato plant results in a reduction in lycopene and an increase in phytoene, phytofluene, and ζ-carotene [14] and a non-functional splice variant of Z-ISO is associated with the yellow pigmentation of the "Pinalate" sweet orange mutant, which accumulates unusual proportions of 9,15,9′-tri-*cis*- and 9,9′-di-*cis*-ζ-carotene, a phenotype partially restored by exposure to high light [12].

Carotenoid biosynthesis has been extensively characterized in model cyanobacteria such as *Synechocystis* sp. PCC 6803 (hereafter *Synechocystis*), where carotenoids play important functional and structural roles in photosystems (PS) I and II [15,16], and are also found in the cytochrome b_6f complex [17,18] and the NADH dehydrogenase-like photosynthetic complex I [19,20]. *Synechocystis* mutants lacking *pds* (*crtP*, slr1254) or *zds* (*crtQ*, slr0940) accumulate phytoene or ζ-carotene, respectively, display very slow glucose-dependent growth, are acutely light sensitive, and lack PSII [21]. Conversely, the deletion of sll0033 (*crtH*), which encodes the CRT-ISO enzyme that converts 7,9,7′,9′-tetra-*cis*-lycopene (prolycopene) to all-*trans*-lycopene, results in a strain that can produce mature carotenoids in the light, albeit in different ratios to the WT organism, due to the photo-isomerization of prolycopene [22–24]. However, the ΔcrtH strain predominantly accumulates *cis*-isomers of carotenes and does not make β-carotene under light-activated heterotrophic growth (LAHG) conditions (darkness with a short pulse of light every 24 h), and therefore it cannot assemble PSII [25].

CRT-ISO isomerizes the adjacent 7,9 and 7′9′ *cis*-pairs of double bonds in 7,9,7′,9′-tetra-*cis*-lycopene but does not act on the 15,15′ *cis* double bond of 9,15,9′-tri-*cis*-ζ-carotene [26]. This 15-*cis* double bond is also photo-labile and can be photo-isomerized, but this appears to be less efficient than the enzyme-catalyzed reaction [7,8]. As Z-ISO is essential for carotenogenesis in dark plant tissues such as the roots, etiolated leaves, and endosperm [7], we predicted that the same would be true for carotenoid biosynthesis in cyanobacteria during periods of darkness. Here, we verify the ζ-carotene isomerase activity of the *Synechocystis* Z-ISO homolog (encoded by the slr1599 locus) by heterologous production of Slr1599 in a 9,15,9′-tri-*cis*-ζ-carotene-accumulating strain of *Escherichia coli*. We also show that Z-ISO is needed for carotenoid biosynthesis in the absence of light in a *Synechocystis* Δslr1599 mutant, which to our knowledge is the first characterization of a cyanobacterial *Z-ISO* deletion strain. Despite repeated attempts, we were unable to purify the *Synechocystis* Z-ISO for spectral analysis and structure determination, but a comparison of AlphaFold structural models of the *Synechocystis* and plant enzymes shows conservation of putative heme-ligating residues and allows us to predict the 9,15,9′-tri-*cis*-ζ-carotene-binding site.

2. Materials and Methods

2.1. Generation of Constructs to Test Synechocystis Z-ISO Activity in Escherichia coli

To test the isomerase activity of Slr1599 in *E. coli*, the slr1599 gene was PCR-amplified from the *Synechocystis* genome and cloned into the NdeI and XhoI sites of MCS2 of pCDFDuet^Tm-1 (Novagen). The resulting sequence-verified construct (pAH592) was introduced into *E. coli* BL21(DE3) along with the pAC-ZETAipi plasmid [27] (purchased from

Addgene, plasmid #53284) with selection on LB agar containing streptomycin (50 μg mL^{-1}) and chloramphenicol (34 μg mL^{-1}). pCDFDuet^{Tm}-1 constructs containing the maize *Z-ISO* gene (synthesized as a gBLOCK by Integrated DNA Technologies) at MCS2 (NdeI/XhoI sites; pAH595), *Synechocystis zds* (*crtQ*; slr0940) at MCS1 (NcoI/SalI sites; pAH585), *zds* at MCS1 and slr1599 at MCS2 (pAH587), and *zds* at MCS1 and the maize *Z-ISO* gene at MSC2 (pAH590) were also generated. Further details of plasmids and primers used in this study are provided in Supplementary Table S1 and Table S2, respectively. PCRs were performed using Q5® Hot Start High-Fidelity 2x Master Mix (New England Biolabs (UK) Ltd.) and restriction digests using Thermo Scientific™ FastDigest enzymes. Ligation reactions were conducted using T4 DNA ligase (New England Biolabs (UK) Ltd.) and all cloning steps used *E. coli* JM109 competent cells (Promega). DNA was purified using the FastGene Gel/PCR Extraction Kit or Plasmid Mini Kit (both from Nippon Genetics), as appropriate, and automated Sanger sequencing was performed by Eurofins.

2.2. Carotenoid Analysis by Reverse-Phase High-Performance Liquid Chromatography (RP-HPLC)

For production of carotenoids in *E. coli*, 300 μL of a 5 mL overnight culture was used to inoculate 15 mL auto-induction Terrific Broth (TB) base without trace elements (Formedium) with relevant antibiotics in 50 mL opaque falcon tubes. These cultures were incubated in darkness at 37 °C for 24 h in a rack set at a 45° angle on a shaking platform rotating at 220 rpm. All subsequent steps were performed in the dark. Cells were harvested by centrifugation at 4600× *g* for 10 min at 4 °C, washed with ice-cold 30 mM HEPES pH 7.4, and transferred to 2 mL screw-capped tubes. Pigments were extracted from cell pellets by adding 700 μL of 7:2 acetone:methanol (*v/v*), vortex-mixing for 30 s, and incubating on ice for 20 min in the dark. The extracts were clarified by centrifugation (15,000× *g* for 5 min at 4 °C), supernatants were transferred to fresh tubes to which 30 μL 5 M NaCl and 700 μL hexane were added, mixed, and allowed to partition. The upper hexane phase was transferred to a 2 mL glass vial and evaporated to dryness in a vacuum centrifuge at ambient temperature. Dried pigments were dissolved in 150 μL acetonitrile, and immediately analyzed on an Agilent 1200 HPLC system. ζ-carotene species were separated on a YMC30 C30 RP column (3 μm particle size; 250 mm × 4.6 mm). An isocratic buffer composed of 40:30:30 acetonitrile:methanol:tetrahydrofuran (*v/v/v*) was used to separate carotenoids at 1 mL min^{-1} at 40 °C, and elution of pigments was monitored by absorbance at 425 nm. Lycopene species were separated on a Discovery HS C18 column (5 μm particle size; 250 mm × 4.6 mm) according to the slightly modified method described in our previous work [28].

Synechocystis carotenoids were extracted from cell pellets washed with ice-cold 30 mM HEPES pH 7.4 by vortex mixing in 100% methanol for 30 s, the extracts were clarified by centrifugation (as above), and the supernatants were analyzed following the same protocol used for analysis of lycopene species in *E. coli*.

2.3. Generation and Growth of Synechocystis Strains

Synechocystis strains were grown photoautotrophically under ~50 μmol photons m^{-2} s^{-1} constant illumination at 30 °C in sterile BG11 medium [29] buffered with 10 mM TES (N-Tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid)-KOH pH 8.2 (BG11-TES). For LAHG, cultures were grown in BG11-TES supplemented with 5 mM glucose in complete darkness apart from a 5-min pulse of white light (~40 μmol photons m^{-2} s^{-1}) once every 24 h [30]. For growth on solid media, 1.5% (*w/v*) bactoagar and 0.3% (*w/v*) sodium thiosulphate were added to BG11-TES. Antibiotics were included where appropriate, as detailed below. For attempted purification of FLAG-tagged Slr1599, cultures were grown either photoautotrophically (~50–100 μmol photons m^{-2} s^{-1}) or under LAHG conditions (as above) in 8 L vessels mixed by bubbling with sterile air and maintained at 30 °C by a temperature coil connected to a thermostat-controlled circulating water bath.

To make a Δslr1599 deletion mutant, a linear mutagenesis construct was generated from three individual PCR product templates by overlap-extension (OLE)-PCR. The first fragment consisted of 427 bp upstream of the slr1599 start codon (see Section 3.1), the first 80 bp of the slr1599 gene, and 26 bp of homology to the start of the erythromycin-resistance cassette from pERY [31] and was amplified with primers AH912 and AH913. The second fragment consisted of the erythromycin resistance gene, which was amplified from pERY using primers AH595 and AH596. The third fragment consisted of 26 bp of homology to the end of the erythromycin resistance cassette, the final 80 bp of slr1599, and 391 bp downstream of the slr1599 stop codon and was amplified with primers AH914 and AH915. The purified fragments were mixed in equimolar amounts and joined by OLE-PCR using primers AH912 and AH915, and the resulting construct was sequenced to confirm correct assembly (Eurofins). Following introduction of the OLE-PCR product into wild-type (WT) *Synechocystis* by natural transformation, transformants were selected on BG11 agar containing 5 μg mL^{-1} erythromycin and single colonies were picked and streaked onto BG11 agar containing a progressively doubled concentration of erythromycin up to 40 μg mL^{-1}. Full segregation of the Δslr1599 mutant was confirmed by PCR amplification with primers AH912 and AH915 and Sanger sequencing of the resultant purified PCR product (Eurofins).

Complementation of Δslr1599 was performed by integrating slr1599 or the *Arabidopsis Z-ISO* gene at the *psbAII* locus using the pPD-*N*FLAG plasmid [32,33]. slr1599 was amplified from *Synechocystis* WT genomic DNA using the primers AH955 and AH956 and cloned into the NotI and BglII sites of pPD-*N*FLAG downstream of the sequence encoding an N-terminal 3×FLAG tag. Attempts to insert slr1599 into the pPD-CFLAG plasmid [34] —such that when integrated in place of *psbAII* in the *Synechocystis* genome the construct would encode Z-ISO with a C-terminal 3×FLAG tag—were unsuccessful, presumably due to toxicity of this fusion protein in the *E. coli* JM109 cells used for cloning. Therefore, we PCR-amplified a linear DNA fragment consisting of slr1599 followed by the sequence adding the C-terminal FLAG-tag and the kanamycin resistance gene and flanked by the *psbAII* homology arms from the ligation reaction using primers AH198 and AH199, which was verified by sequencing the PCR product. The *Arabidopsis Z-ISO* gene (At1g10830) was synthesized (Integrated DNA Technologies) lacking bases 4–174 (the sequence encoding the N-terminal chloroplast transit peptide) and with the sequence encoding a Ser-Ala linker followed by the Strep-tag® II added to the 3′ end of the gene, and then cloned into the NdeI and BglII sites of pPD-*N*FLAG. The sequence-verified plasmids (pAH445 and pAH436; Table S1) and linear PCR fragment were individually transformed into the Δslr1599 mutant and, following selection on BG11 agar containing 7.5 μg mL^{-1} kanamycin, colonies were picked and segregated by streaking onto increasing concentrations of kanamycin up to 30 μg mL^{-1}. Full segregation of the *psbAII* locus was determined by PCR with primers AH47 and AH48, and automated Sanger sequencing (Eurofins) confirmed the correct insertion of slr1599 or At1g10830. It was not possible to isolate kanamycin-resistant colonies of Δslr1599 for the C-terminally FLAG-tagged *Synechocystis* enzyme.

Synechocystis slr1599 point mutants were generated with the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies), according to the manufacturer's instructions, using pAH445 as template and the primer pairs detailed in Table S2. The resulting plasmids were sequence-verified prior to introduction to the *Synechocystis* Δslr1599 mutant and after PCR amplification of the segregated *psbAII* locus from *Synechocystis* strains (as above).

2.4. Purification of Recombinant Maize Z-ISO

Maize Z-ISO fused to maltose-binding protein (MBP) was purified following a protocol adapted from those described by Beltrán et al [5,35,36]. The sequence of the codon-optimized construct that was cloned into pET21a(+), yielding plasmid pAH545, is provided in Table S3. Twelve 500 mL cultures of *E. coli* BL21(DE3) harboring pAH545 were grown in 2xYT medium (10 g L^{-1} yeast extract, 16 g L^{-1} tryptone, and 5 g L^{-1} NaCl) containing

100 µg mL^{-1} ampicillin at 37 °C with shaking (200 rpm) to an optical density at 600 nm of 0.6, at which point 1 mM isopropyl β-d-1-thiogalactopyranoside was added to each culture and the temperature was lowered to 28 °C for overnight (~18 h) incubation. Cells were collected by centrifugation at 8671× *g* for 10 min at 4 °C and resuspended in 50 mM Tris-HCl pH 7.6, 300 mM NaCl, 5% (*w/v*) glycerol, 0.5 mM dithiothreitol (buffer A) supplemented with a Roche protease inhibitor tablet, and a small spatula each of DNase I and lysozyme. The cell suspension was cooled on ice and lysed by two passes through a chilled French press at 18,000 psi with 5 min of chilling on ice between cycles. Lysed cells were centrifuged at 7025× *g* for 20 min at 4 °C to remove cell debris and unbroken cells, and the supernatant was centrifuged again at 48,400× *g* for 30 min at 4 °C to pellet *E. coli* membranes. The pellet (membrane fraction) was resuspended in 80 mL buffer A and n-dodecyl-β-D-maltoside (β-DDM; Anatrace) was added to a final concentration of 1.5% (*w/v*) before a 1-h period of incubation in the dark at 4 °C with gentle agitation. The solubilized membrane solution was centrifuged at 48,400× *g* for 30 min at 4 °C and the supernatant was retained, diluted 2x in buffer A, and incubated with 3 mL of pre-washed amylose resin (New England Biolabs (UK) Ltd.) overnight at 4 °C with gentle agitation. The next day, the resin was collected by filtration through a 0.22 µm spin column, washed with 20 resin volumes of buffer A containing 0.04% (*w/v*) β-DDM, and bound protein was eluted in the same buffer supplemented with 10 mM maltose. Eluted protein was directly analyzed by absorbance spectroscopy in a UV–Vis cuvette (Sarstedt Inc.) using a Cary 60 UV-Vis spectrophotometer (Agilent Technologies). For redox difference spectra, the protein was oxidized by incubation with a few grains of potassium ferricyanide and then reduced by addition of a small quantity of sodium dithionite. Spectra were measured following 1 min of incubation with each reagent.

Point mutants in the maize *Z-ISO* gene were generated with the Quikchange II Site-Directed Mutagenesis Kit (Agilent Technologies) using pAH545 as template and the primer pairs detailed in Table S2. The variant proteins were purified in the same way as WT maize Z-ISO.

2.5. Attempted Purification of Recombinant and Native Slr1599

We attempted to purify Slr1599 as an MBP-fusion protein using the same growth conditions and purification procedures as for the maize enzyme (as described in Section 2.4). We initially generated a construct (pAH546) in which the maize *Z-ISO* gene in pAH545 was replaced with a codon-optimized (for *E. coli*) slr1599 gene (see Table S1 for plasmid details and Table S3 for the full annotated gene sequence). A second construct (pAH565, Table S1) in which the sequence encoding the single transmembrane helix (TMH) protein YnhF was fused to the 5′ end of slr1599 was also generated (see Table S3 for sequence). Constructs in which slr1599 was inserted into the commercial pMAL-c5X (pAH424) and pMAL-p5X (pAH707) plasmids (both New England Biolabs (UK) Ltd.) were also generated (see Tables S1–S3 for details).

Attempted immunoprecipitation of FLAG-tagged Slr1599 from β-DDM-solubilized membranes isolated from the Δslr1599 FLAG-Slr1599 strain of *Synechocystis* was performed as reported in our previous work [28,34].

2.6. Computational Modelling and Ligand Docking

AlphaFold models for the maize and *Arabidopsis* Z-ISO enzymes were downloaded from UniProt (https://www.uniprot.org/; peptide sequence identities B4FHU1 (AF-B4FHU1-F1) and Q9SAC0 (AF-Q9SAC0-F1), respectively (accessed on 1 June 2022). The UniProt entry for Slr1599 is N-terminally truncated (see Section 3.1) so the model for *Synechocystis* Z-ISO was generated using AlphaFold v2.1.2 [37] (git commit e93a9eb) with full database searching and the monomer model parameter, which was run on a Dell PowerEdge C4130, housed within the University of Sheffield SHARC high-performance computing cluster, and equipped with an Intel Xenon E5-2630 v3 CPU, 64 GB DDR4 ECC-RAM, and eight NVIDIA Tesla K80 GPUs. Mutated Z-ISO structures were generated in ChimeraX [38]

using the AlphaFold structures and scored using the Rosetta membrane protein-scoring suite [39]. Mutants were chosen based on previous biochemical studies [5,6] and rational interrogation of the AlphaFold structures.

The AlphaFold structures of the *Arabidopsis* and maize Z-ISO enzymes were used to dock heme B, and these heme-bound structures were subsequently used for docking of 9,15,9′-tri-*cis*-ζ-carotene. The 3D ligand structure of heme B (HEM-model spatial data file (SDF)) was downloaded from RCSB PDB (https://www.rcsb.org/). We could not find the 3D structure of 9,15,9′-tri-*cis*-ζ-carotene in any database and so generated one from the 2D SDF using ChimeraX, predicting bond lengths by bond order. AutoDock structure files containing structure, partial charges, and rotational freedoms (.pdbqt files) were generated with AutoDock Tools [40]. Rotational freedom was removed from any bonds involved in the 9,15,9′-tri-*cis*-ζ-carotene π-conjugation system. Heme B was docked into both structures by searching a 125,000 Å^3 volume, investigating 10,000 different binding modes. Using the *Arabidopsis* model, two binding sites were identified with estimated ~9 kcal mol^{-1} binding energy; however, only one was near prototypical bis-His ligands predicted by the AlphaFold structures. This informed a more targeted docking of heme B into the maize Z-ISO structure, where 1000 binding modes in a 2744 Å^3 volume around the putative bis-His heme-binding site were investigated. The lowest-energy heme B-binding mode was used to generate a new maize Z-ISO-heme B PDB structure, which was used to dock 9,15,9′-tri-*cis*-ζ-carotene, searching 10,000 binding modes in a 125,000 Å^3 volume.

3. Results and Discussion

3.1. Identification of a Z-ISO Homolog in Synechocystis

To our knowledge, at the time of writing, the experimental verification of a cyanobacterial Z-ISO has only been reported for the enzyme from the filamentous, non-N$_2$-fixing cyanobacterium *Arthrospira platensis* NIES-39, where activity was demonstrated by heterologous production in an *E. coli* strain that accumulated 9,15,9′-tri-*cis*-ζ-carotene in darkness [11]. The same authors also identified Z-ISO homologs in other *Arthrospira* species, the model cyanobacteria *Synechocystis* and *Synechococcus* sp. PCC 7002 (hereafter *Synechococcus* 7002), and *Prochlorococcus marinus* str. SS35 [11]. Zhu et al [41] had earlier identified the same 238 amino acid protein (SYNPCC7002_A1197) as the likely ζ-carotene isomerase in *Synechococcus* 7002, and the locus predicted to encode Z-ISO in the original genome-sequenced *Synechocystis* Kazusa substrain [42] is slr1599 [43].

In the UniProt (https://www.uniprot.org/uniprot/P72984), KEGG (https://www.genome.jp/dbget-bin/www_bget?syn:slr1599), and CYORF (http://cyano.genome.jp/cgi-bin/cyorf_view.cgi?ORG=syn&ACCESSION=slr1599) databases (accessed on 1 June 2022), slr1599 is predicted to encode a 188 amino acid protein, ~50 residues smaller than other cyanobacterial Z-ISO homologs [11,41]. This is due to an apparent misidentification of the start codon; the translation of the DNA immediately upstream of slr1599 in combination with analysis of other *Synechocystis* substrains (where the protein is predicted to range from 229–247 amino acids) and an alignment with known Z-ISO enzymes from other organisms indicates that the gene likely starts 162 bp upstream of the annotated GTG start codon, encoding a 242 amino acid protein (Figure 2A). Like other Z-ISO enzymes, Slr1599 is annotated as an NnrU domain-containing protein [8], and similar to other cyanobacterial Z-ISO homologs it is predicted to contain six TMHs (Figure 2B), compared to seven for the plant enzymes [5,12].

A

```
6803    ----------------------------------------------------------------
7002    ----------------------------------------------------------------
Ap      ----------------------------------------------------------------
Eg      EPNPQEAVPWEALVAPLLAVAMVLGLLRLPRWGMAATSGETDPFRTPLGEDAAVFDLQSQRLADWVKF    165
At      IPAHPRLGNSTSLLSSSSPVIRKILVRSTLREDQPIASDSESSPTLLIGEDSAAFELGKQKLVSWVYF     99
Zm      NPIRAPLPPLSRVLSHARPA-RAVGGGIEPKEGVVAEGDESGGGPVLVGEDSAAFELKDQSVASWAYF     98


6803    -------MFDISSPW-----------------PWLTPSHGIIAAYLLLFAIVHSGLAALRAWGESKIG     44
7002    -----------MTDL-----------------SWFSTSHWIMAGLLLGFAIAHSGLAALRPWGEQKIG     40
Ap      MELTLLSVGEVKMSD-----------------QWLSLSHFVMVGLLLCFAIAHSGLAALRPVVEQVIG     51
Eg      GAVFIAVMGGLFVTWVYPPTGLADEFVRSLEALMGGNSEATMLLILLIFGVAHSGLAGLRTQGEELIG    233
At      GVVLGVVLFILNVVWIDNSTGFGKSFIDAVSNI-SGSPEVAMLMLILIFAIVHSGLASLRDIGEKLIG    166
Zm      AGILGAVLVALNVLWIDPSTGVGTKFLDAVASV-SDSHEVVMLLLTIIFAVVHSGMASLRESGEKIVG    165
                                         .  :      : *.:.***:*.**    *. :*


6803    ARGYRVIFALVSIPLATGLIIYFFNHRYDGLQLWQVQGVTGVKPLVWILSALSFFFLFPATFNLLEIA    112
7002    SRLYRVLFALVSIPFATCLIIYFFNHRYDGLVLWQVQGDAWVKPTVWWLSAISFFFLYPSTFNLLEIA    108
Ap      PRLYRIFFALVSLSLAVVLIIYFFNHRYDGLQLWQLQGVPALKPIVWVLSAISFVFLYPSTFNLLEIA    119
Eg      ARAYRVLFGLISLPLALVAIVYFMNHRYDGVQLWSFRGVPGVHSFVWITTFISFFFLYPSTFNLLEVA    301
At      ERAFRVLFAGISLPLAMSTIVYFINHRYDGSQLWQLQGVPGVHEAIWVANFVSFFFLYPSTFNLLEVA    234
Zm      ERVYRVLFAGISLPLAVTTIVYFINHRYDGTQLWQVQGITGIHELLWFSSFISFFFLYPSTFNLLEVA    233
          *  :*::*. :*: :*    *:**:******   **..:*     ::   :*   .  :**.**:*:*******:*


6803    AIQKPEIHLYETGIIRICRHPQMVGQVIWCIAHTLWLGTTFTLVTSLGLIAHHCFAVWHGDRRWQHKY    180
7002    AVKQPQVHLYEKGIIRITRHPQMVGQLIWCVAHTLWIGSSFTLLTSIGLMAHHIFAVWHGDRRLRSRY    176
Ap      AIAKPQVHLHQTGIIRITRHPQMVGQVIWCIAHTLWIGTTFFTLVTSIGLVLHHLFGVWHGDRRLQARF    187
Eg      AVDQPKVHLWETGIMRITRHPQMVGQLLWCMAHTLWIGNTFAIVTSAGLMAHHLFGCWHGDRRLAAKY    369
At      AVDKPKMHLWETGIMRITRHPQMVGQIVWCLAHTLWIGNTVAASASLGLIAHHLFGAWNGDRRLAKRY    302
Zm      AVDKPKLHMWETGIMRITRHPQMVGQVIWCLAHTLWIGNSVAVAASVGLISHHLFGAWNGDRRLLSRY    301
        *:  :*::*:  :.**:** ********::**:*****:*.:.:   :* **: ** *. *:****   ::


6803    GEAFLAVKQRTSIIPFQAIWEGRQILV---WQEFIKPAYVGVMGFIALLWWGHPWLMVMTSRVNW     242
7002    GQTFLAVKERTSIIPFLAILQGRQQLK---LEEFLKPAYVGVTAFIALLWWGHPWLMVMTAKVPW     238
Ap      GQSFETLKSQTSVIPFLAIIQGRQTLV---IQEFLRPAYLGVGIFIIILWQAHPLLIRVTSNVGW     249
Eg      GPAFEEVKARTSVWPFVAILEGRQTLPPDYWREFARAPYLAITIGTVAAYLAHPLMQSSSYSLPW     434
At      GEDFESIKKRTSVIPFAAIFEGRQVLPEDYYKEFVRLPYLAITALTVGAYFAHPLMQGASFRLHW     367
Zm      GEAFEVLKKRTSVMPFAAIIDGRQKLPKDYHKEFFRLPYVAITMLTLGAYFAHPLMQASSYQLPW     366
        *   *   :* :**: ** ** :*** *       .** :   *:..:        : .** :    :   : *
```

B

Figure 2. Sequence alignment of Slr1599 with known Z-ISO enzymes from other oxygenic phototrophs. (**A**) Sequence alignment of Slr1599 (6803) with Z-ISO homologs from *Synechococcus* 7002 (7002), *A. platensis* (Ap), *E. gracilis* (Eg), *A. thaliana* (At), and *Z. mays* (Zm). The N-terminal sequence in

Slr1599 that is 'missing' in some databases is indicated in red font. Note that the eukaryotic enzymes (Eg, At and Zm) are N-terminally truncated and, where appropriate, the sequence corresponding to their chloroplast transit peptides is shown in blue. The position of predicted TMHs in Slr1599 (six; note that helices 4 and 5 are consecutive) and the *Z. mays* (seven) enzyme are highlighted in green and yellow, respectively. Residue numbers are shown on the left and the degree of conservation is shown below the sequence alignment, where '*' indicates the residue is identical, ':' = a conservative substitution, and '.' = a semi-conservative substitution. (**B**) Schematic showing the predicted number of TMHs for Slr1599 in comparison to the *Z. mays* Z-ISO. One fewer TMH in the *Synechocystis* enzyme means that the N-terminus is predicted to be localized to the thylakoid lumen, in contrast to the plant Z-ISO, where the extra TMH leads to a predicted location for the N-terminus on the stromal side of the membrane. Four residues previously implicated as having potential structural/functional roles in the maize Z-ISO enzyme [5,6], and investigated here in Slr1599 (Section 3.4), are indicated on the schematics.

3.2. Slr1599 Displays Z-ISO Activity in Escherichia coli

To test the function of Slr1599, we expressed slr1599 in an *E. coli* strain containing pAC-ZETAipi, which contains the *Pantoea agglomerans* (previous known as *Erwinia herbicola*) *crtE* (encodes GGPPS), *crtB* (encodes PSY), and *idi* (encodes IPP isomerase, which inter-converts IPP and DMAPP) genes and the *Synechococcus* sp. PCC 7942 *crtP* gene (encodes PDS) to drive the production of the substrate of Z-ISO, 9,15,9′-tri-*cis*-ζ-carotene [27]. In agreement with the previously reported data, *E. coli* cells containing pAC-ZETAipi and empty pCDFDuet^Tm-1 accumulated primarily 9,15,9′-tri-*cis*-ζ-carotene (Figure 3A, peak 1) when grown in darkness, with small additional peaks that we assigned as 9,9′-di-*cis*-ζ-carotene (peak 3) and all-*trans*-ζ-carotene (peak 4) based on previous analyses [11,44]. The same was true for pAC-ZETAipi plus pCDFDuet^Tm-1 with the *Synechocystis zds* (*crtQ*) gene, consistent with the fact that ZDS cannot act on 9,15,9′-tri-*cis*-ζ-carotene [26].

Upon the co-transformation of *E. coli* with pAC-ZETAipi and pCDFDuet^Tm-1 carrying either maize *Z-ISO* or slr1599, the size of the 9,15,9′-tri-*cis*-ζ-carotene peak (no. 1) was significantly decreased and that of the 9,9′-di-*cis*-ζ-carotene peak (no. 3) increased correspondingly (Figure 3A), showing that both enzymes can isomerize the 15–15′ *cis* double bond of 9,15,9′-tri-*cis*-ζ-carotene. Furthermore, when *zds* and slr1599 or *zds* and the maize *Z-ISO* gene were present in the pCDFDuet^Tm-1, ζ-carotene species were almost completely absent and a variety of new carotenoid species (Figure 3B, peaks 5–7), likely *cis*-isomers of lycopene [23], were produced; these *cis*-lycopene species were absent in the corresponding samples lacking Z-ISO, which accumulated 9,15,9′-tri-*cis*-ζ-carotene (peak 8, Figure 3B), as in the absence of *zds*. Taken together, these data demonstrate that slr1599 encodes an enzyme with 9,15,9′-tri-*cis*-ζ-carotene isomerase activity akin to previously characterized Z-ISO enzymes [5,7,8,11].
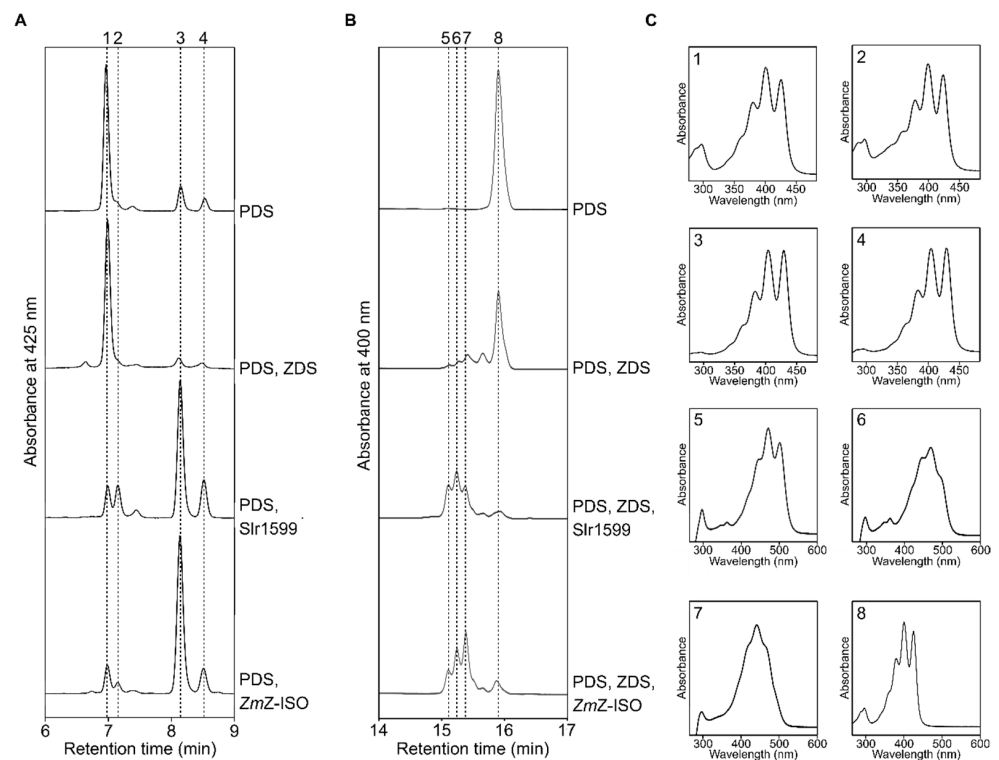
Figure 3. Slr1599 isomerizes 9,15,9′-tri-*cis*-ζ-carotene when produced in *E. coli*. (**A**) RP-HPLC analysis of carotenoids extracted from dark-grown *E. coli* cells containing pAC-ZETAipi and either empty pCDFDuet^Tm-1 (labeled PDS), pCDFDuet^Tm-1 containing *Synechocystis crtQ* (PDS, ZDS), pCDFDuet^Tm-1 containing *Synechocystis* slr1599 (PDS, Slr1599), or pCDFDuet^Tm-1 containing the maize *Z-ISO* gene (PDS, *Zm*Z-ISO). Eluted pigments were detected by monitoring absorbance at 425 nm. (**B**) RP-HPLC profile of carotenoids extracted from dark-grown *E. coli* cells harboring empty pCDFDuet^Tm-1 (PDS), pCDFDuet^Tm-1 containing *Synechocystis crtQ* (PDS, ZDS), pAC-ZETAipi and pCDFDuet^Tm-1 containing both *crtQ* and slr1599 (PDS, ZDS, Slr1599) or pAC-ZETAipi and pCDFDuet^Tm-1 containing both *crtQ* and the maize Z-ISO gene (PDS, ZDS, *Zm*Z-ISO). Eluted pigments were detected by monitoring absorbance at 400 nm. (**C**) Absorbance spectra of the numbered carotenoid species in panels (A) and (B). Peaks 1 and 8 = 9,15,9′-tri-*cis*-ζ-carotene; peak 2 = alternative central *cis*-isomer of ζ-carotene [44]; peak 3 = 9,9′-di-*cis*-ζ-carotene; peak 4 = all-*trans* ζ-carotene [44]; peaks 5–7 = *cis*-lycopene isomers [23].

3.3. The Synechocystis Z-ISO Is Required for Carotenoid Biosynthesis in Darkness In Vivo

We next examined the role of Slr1599 in *Synechocystis* carotenoid biosynthesis in vivo. A mutant lacking slr1599 was generated by the partial replacement of the open reading frame with an erythromycin resistant cassette (Figure 4A). Under photoautotrophic conditions (constant illumination), the Δslr1599 mutant grew similarly to the WT, and the absorbance spectra of both strains were almost identical (Figure 4B). The production of β-carotene was used to indicate Z-ISO activity; β-carotene levels were similar in the WT and Δslr1599 strains grown under constant illumination, indicating that the non-enzymatic photo-isomerization of the 15,15′ *cis* double bond of 9,15,9′-tri-*cis*-ζ-carotene can occur in the absence of Slr1599 (Figure 4C). This is analogous to the case with the *Synechocystis* Δ*crtH* mutant, where the loss of CRT-ISO does not result in a significant carotenoid defect in the light due to the photolability of the *cis* double bonds of 7,9,7′,9′-tetra-*cis*-lycopene, such that all-*trans*-lycopene is formed non-enzymatically [23].
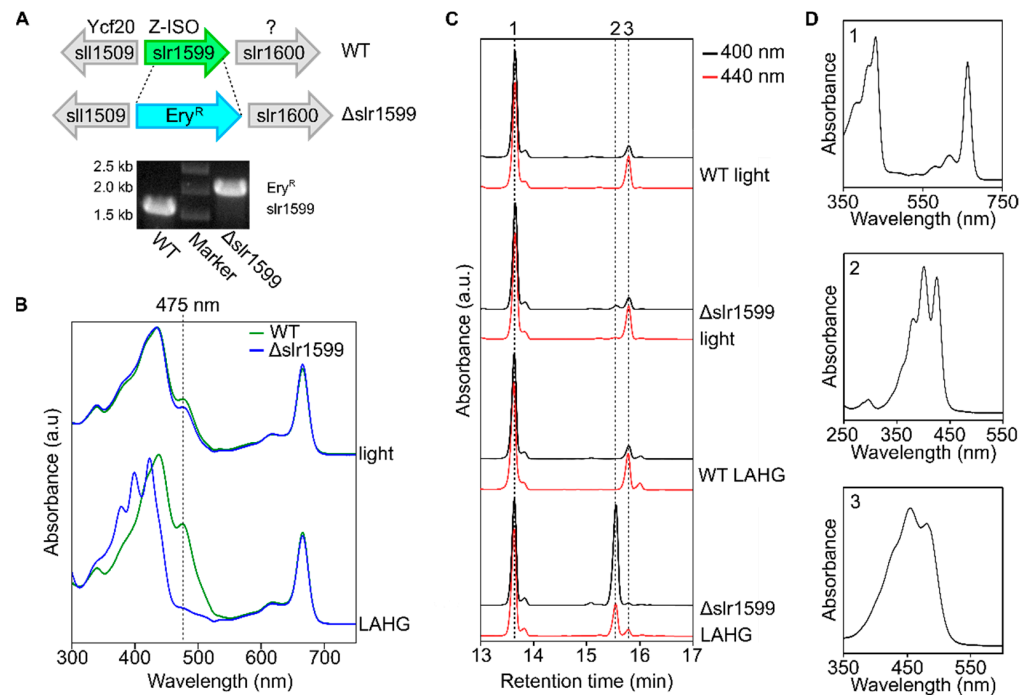
Figure 4. Carotenoid composition of WT and Δslr1599 strains of *Synechocystis* grown under photoautotrophic and LAHG conditions. (**A**) Schematic showing generation of the Δslr1599 *Synechocystis* mutant by partial replacement of slr1599 (green) with the erythromycin resistance gene (Ery[R]) by homologous recombination (indicated by the dotted lines). Segregation of the mutant was confirmed by PCR amplification of the slr1599 locus in comparison to the isogenic WT parent strain. (**B**) Absorbance spectra of methanolic extracts of the WT (green trace) and Δslr1599 (blue trace) strains grown under constant illumination (light) or LAHG conditions. The black dotted line highlights the contribution of 'mature' carotenoids to the WT spectra at ~475 nm, which is largely depleted in the mutant under LAHG conditions. (**C**) RP-HPLC analysis of pigments extracted from WT and Δslr1599 cells grown under photoautotrophic (light) or LAHG conditions. Chlorophyll (peak 1), 9,15,9′-tri-*cis*-ζ-carotene (peak 2) and β-carotene (peak 3) were detected by monitoring absorbance at 400 nm (black traces) and 440 nm (red traces); traces are normalized to the height of the Chl peak. (**D**) Absorbance spectra of the numbered pigment species in panel (**C**).

Synechocystis cannot grow in complete darkness even when supplemented with glucose due to its requirement for at least a short (~5 min) daily pulse of light [30]. The pulse of light that permits LAHG is not sufficient to photo-isomerize the *cis*-bonds in 7,9,7′,9′-tetra-*cis*-lycopene and so a Δ*crtH* mutant cannot produce carotenoids under these conditions [23]. We tested the Δslr1599 mutant under the same LAHG conditions and observed an almost complete loss of β-carotene accompanied by a significant accumulation of ζ-carotene (Figure 4C), showing that unlike the WT, the mutant is unable to convert 9,15,9′-tri-*cis*-ζ-carotene to 9,9′-di-*cis*-ζ-carotene in the absence of prolonged illumination. The defect in carotenoid biosynthesis in the Δslr1599 strain under LAHG is reflected by the loss of carotenoid absorbance at ~475 nm and the concomitant appearance of absorbance features in the 350–430 nm region of the absorbance spectrum due to the accumulation of 9,15,9′-tri-*cis*-ζ-carotene (Figure 4B).

We ruled out a polar effect of the disruption of slr1599 by the complementation of the Δslr1599 mutant; the *in trans* expression of slr1599 or the *Arabidopsis Z-ISO* gene (At1g10830 lacking its predicted 58 aa N-terminal chloroplast transit peptide) under the control of the native *psbAII* promoter restored β-carotene biosynthesis under LAHG (Figure S1).

3.4. Complementation of the Δslr1599 Mutant Allows Identification of Key Z-ISO Residues

Residues H150, C263, H266, and D294 have previously been implicated as being functionally important in the maize Z-ISO enzyme [5,6]. These residues appear to be highly conserved in Z-ISO homologs (see Figure 2) and we generated point mutants in the equivalent positions in the *Synechocystis* enzyme (H29, C142, H145 and D173) to see if they could complement the Δslr1599 mutant phenotype, as was the case with the *in trans* expression of the WT slr1599 gene discussed above.

As with previous observations of the maize enzyme [5], the *Synechocystis* H29A and H145A substitutions appear to have decreased Z-ISO activity, observed as a large reduction in β-carotene accumulation and the presence of 9,15,9′-tri-*cis*-ζ-carotene under LAHG conditions (Figure 5). Conversely, the C142S variant produced β-carotene but also accumulated some 9,15,9′-tri-*cis*-ζ-carotene (Figure 5), again consistent with the equivalent C263A variant of the maize enzyme [5,6]. Beltrán Zambano previously reported that D294 was essential for the activity of maize Z-ISO [6]; the equivalent D173A mutant of the *Synechocystis* enzyme produced β-carotene but also accumulated a significant amount of a *cis*-ζ-carotene species, suggesting some perturbation of function.

These data demonstrate the utility of the Δslr1599 mutant as a background strain for studying variants of Z-ISO, such as point mutants of the native or heterologous enzymes, or for testing the activity of candidate Z-ISOs from other organisms, in vivo. The functional implications of the results obtained with the H29A, C142S, H145A, and D173A variants of Slr1599 are discussed in relation to the modelled structures of Z-ISO in Section 3.6 below.
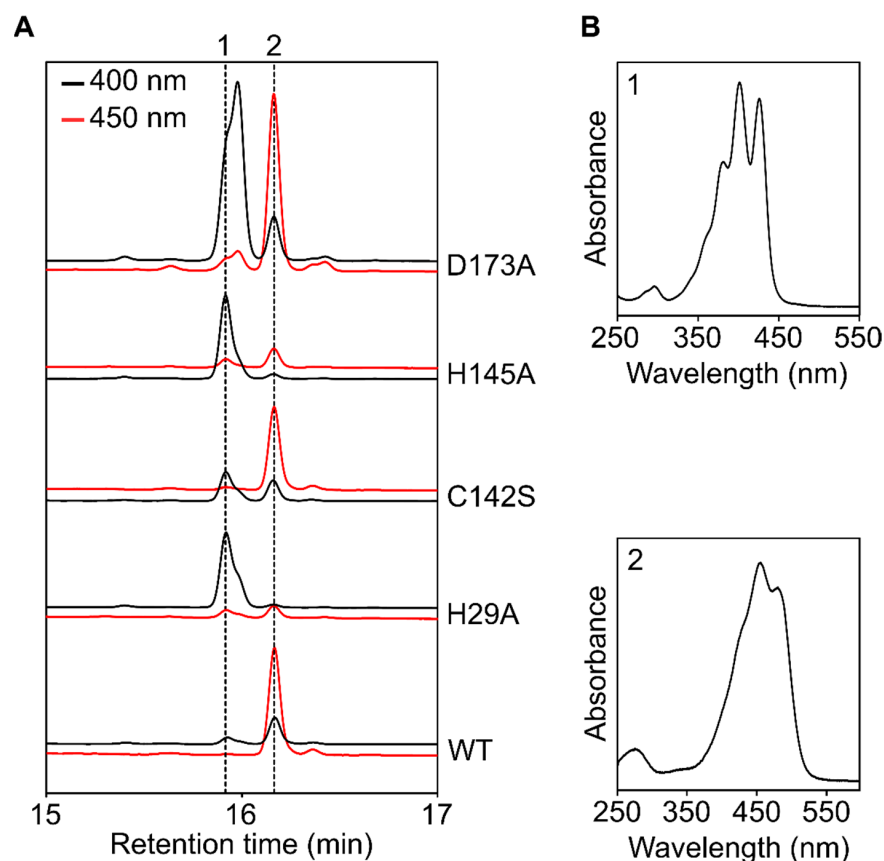
Figure 5. Complementation of the Δslr1599 mutant enables the identification of functionally important residues. (**A**) Analysis of the activity of the indicated Slr1599 variants produced in the Δslr1599 background. Strains were grown under LAHG conditions and carotenoids were extracted and separated by RP-HPLC, and detected by absorbance at 400- and 450 nm. Dotted lines and numbers correspond to the spectra in panel (**B**). (**B**) Absorbance spectra of the carotenoid peaks in panel (**A**). Peak 1 = 9,15,9′-tri-*cis*-ζ-carotene and peak 2 = β-carotene.

3.5. Attempted Purification of Recombinant and Native Synechocystis Z-ISO

The maize Z-ISO has previously been purified as a recombinant His-tagged MBP-fusion from detergent-solubilized *E. coli* membranes and shown to contain a heme B cofactor [5,35]. We generated a construct similar to that described by Beltrán Zambano [6] to produce N-terminally His-MBP-tagged maize Z-ISO, and a second construct in which the maize *Z-ISO* gene was replaced with slr1599 (fusion gene sequences are provided in Table S3). Following the protocol described by Wurtzel and Beltrán [35] (see Section 2.4), we were able to isolate the maize protein and demonstrate similar spectroscopic characteristics to those reported previously (Figure S2) [5]. However, we were unable to purify Slr1599 using the same approach; unlike with the maize enzyme, the *E. coli* pellets and membranes from cells carrying the slr1599 construct were not pink/red, indicating that significant quantities of a heme-containing protein were not produced, and the protein was not detected by SDS-PAGE or immunoblot analysis. Attempts to produce the *Synechocystis* protein using a second construct with the gene cloned into the commercial pMAL-c5X plasmid (Table S1) were also unsuccessful.

The maize Z-ISO and other plant enzymes are predicted to have seven TMHs with an N-terminus in/C-terminus out orientation [5,12], whereas the *Synechocystis* enzyme is predicted to contain one less TMH with both the N- and C-termini in the periplasm/thylakoid lumen (Figure 2). While cytoplasmic MBP is compatible with the N-terminal orientation of the maize enzyme, we hypothesized that the presence of MBP may prevent the localization of the N-terminus of the *Synechocystis* enzyme to the periplasm, preventing the proper folding, orientation, and insertion of Z-ISO into the membrane. To retain MBP and the N-terminus of Z-ISO in the cytoplasm and periplasm, respectively, we fused a single TMH protein, YnhF [45], between MBP and Z-ISO. Despite repeated attempts to produce and purify the fusion protein, we were unable to do so. We also inserted slr1599 into the commercial pMAL-p5x vector to N-terminally fuse *Synechocystis* Z-ISO to an MBP variant that is directed to the periplasm so that the N-terminus of Z-ISO is in the correct 'out' orientation (Table S1), but this also failed to yield detectable quantities of protein.

Due to the inability to isolate recombinant Slr1599 from *E. coli*, we changed tack and tried to purify the enzyme from the native cyanobacterial host. To attempt to immuno-precipitate Slr1599 from *Synechocystis*, we expressed the gene encoding the enzyme with a 3xFLAG tag from the *psbAII* locus in the Δslr1599 background, generating the strain Δslr1599 FLAG-Slr1599. It was only possible to perform this with the FLAG tag at the N-terminus as we were unable to isolate *Synechocystis* colonies for the C-terminally-tagged equivalent, which is consistent with previous findings that the C-terminus is important for the function of plant Z-ISO [6]. FLAG-tagged Slr1599 was functional as the Δslr1599 FLAG-Slr1599 strain produced β-carotene during LAHG (Figure S1); however, we were unable to purify the tagged enzyme from detergent-solubilized membranes from the cells grown under these conditions or under constant illumination. We speculate that the enzyme is quickly degraded in the light, either during growth under illumination, where it is not required and may be inactivated by the proposed redox-dependent ligand-switching mechanism [5], or during the subsequent cell breakage and purification process, even though these steps were performed under dim light. Consistent with this hypothesis, Slr1599 was not identified in our recent quantitative proteomics analysis of *Synechocystis* cells grown under constant illumination, conditions where PDS, ZDS, and CRT-ISO were all detected [46]. Our inability to isolate Slr1599 following LAHG suggests that the enzyme is present in low amounts even under conditions where we have shown that it is required for carotenoid biosynthesis (Figure 4).

3.6. Modelling of Z-ISO for In Silico Structure–Function Analysis

As a structure of Z-ISO remains elusive, we used a combination of AlphaFold [37], Autodock [40], and Rosetta [47,48] to probe the folding homology and structure–function relationships in Slr1599 and the two best-studied plant Z-ISO homologs in silico. While these approaches provide useful information, we stress that the AlphaFold structures,

free-energy estimates, and cofactor docking are all computational models and should not be considered replacements for experimentally determined structures, and we note the recent publication of a genomics-based approach to identify Z-ISO orthologs suitable for structural studies [49].

Model structures derived from AlphaFold simulations are available for the maize and *Arabidopsis* Z-ISO enzymes from UniProt (identifiers B4FHU1 and Q9SAC0, respectively) and the Slr1599 model was generated using AlphaFold v2.1.2 (see Section 2.6 for details); all the models have high pLDDT confidence scores and the structural coordinate files and per-residue model confidence values for Slr1599 are available for download in the Supplementary Information. The superimposition of the model structures with the transit signal peptides removed for the plant enzymes shows a very high homology (RMSD 0.4–0.9 Å; Figure 6A). In agreement with the previous predictions, the plant enzymes have seven TMHs and Slr1599 has six, with N- and C-termini on opposing sides of the membrane for the plant Z-ISOs and the same (lumen) side of the membrane for the *Synechocystis* enzyme.
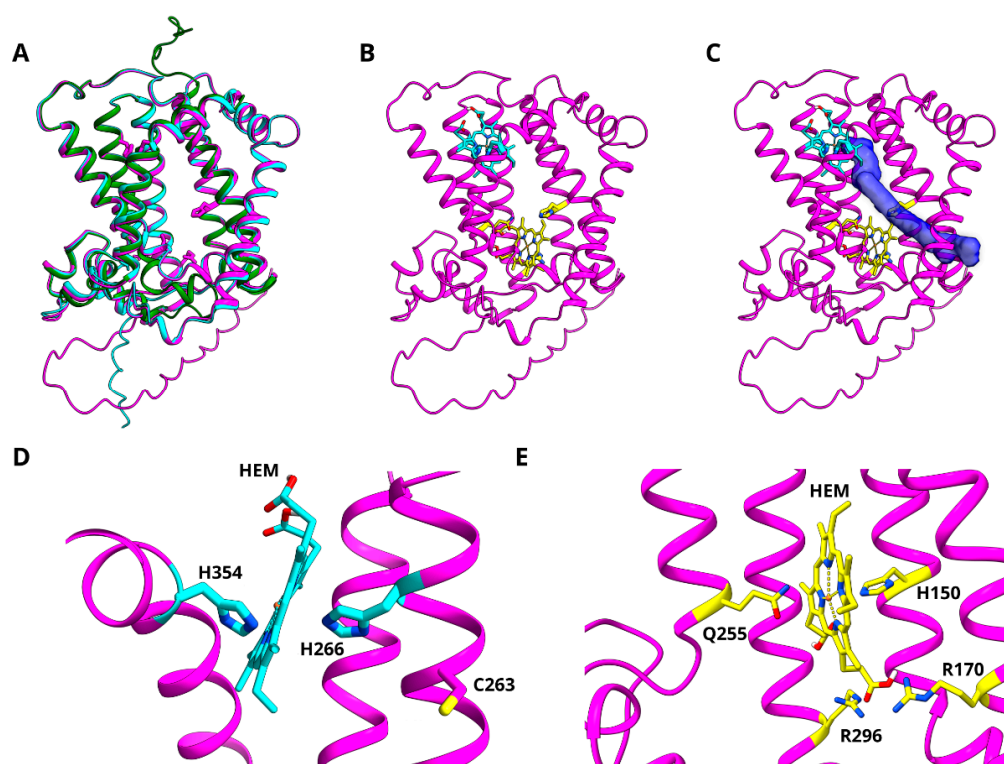
Figure 6. AlphaFold models of Z-ISO. (**A**) Backbone structures of the *Synechocystis* (green), *Arabidopsis* (cyan), and maize (magenta) Z-ISO enzymes. (**B**) Positions of the two putative heme (HEM) binding sites in the maize enzyme, with H266 (cyan) and H150 (yellow) acting as heme ligands. (**C**) Density for the lowest-energy mode identified for 9,15,9′-tri-*cis*-ζ-carotene (purple) binding to maize Z-ISO. (**D**,**E**) Expanded views of the putative H266 (panel (**D**)) and H150 (panel (**E**)) binding sites.

Two histidine residues, H150 and H266, and a cysteine, C263, have previously been implicated in heme binding/Z-ISO activity in work with the maize enzyme and these residues are conserved in Z-ISO homologs (Figure 2 and Figure S3). Mutagenesis studies showed that H150 and H266 are the only histidine residues required for function in maize Z-ISO [5] and we found the equivalent H29 and H145 residues are important for activity in Slr1599 (Figure 5). Based on homology modelling with maize Z-ISO, Beltràn et al proposed that H150 acts as the proximal ligand to the heme cofactor and H266 and C263 act as alternate distal heme ligands in a redox-regulated ligand-switching mechanism that activates/inactivates the enzyme [5]. However, the AlphaFold models predict that H150 and H266 are too far apart in the structures for bis-His coordination of the heme. We

therefore used Autodock Vina to dock heme B into the maize and *Arabidopsis* Z-ISO models to investigate possible alternatives for heme ligation.

First, a very broad search area (125,000 Å^3) was used for both structures (see Section 2.6 for details). In *Arabidopsis*, four potential heme binding sites were predicted, two of which were of particular interest (Figure S4); one site is oriented near H150 and another in close proximity to H266 (maize enzyme numbering). The H266 site contains a second possible axial ligand to coordinate the heme Fe-center, H354, and is proximal to the suggested alternative heme-coordinating residue, C263 [5], however we note that H354 was not required for Z-ISO function in previous studies [5,6]. Conversely, in the other site there is no residue to donate the requisite sixth coordinate bond to the Fe center, suggesting that H150 is involved in inter-domain H-bonding and has a structural role rather than facilitating cofactor binding (Figure S8B).

In the broad search regime, with the same sampling method used, Autodock Vina did not find the equivalent H266/H354 bis-His binding site in the maize enzyme model. This is possibly due to the rotamer position of H354 in the maize structure being more directly aligned with H266 (Figure S5), leading to a narrower binding cavity that is difficult to identify in the simulation. Given that heme binding would likely occur through axial ligation, a narrower sampling area (2744 Å^3) was used for the maize enzyme focusing on the H266 site. Under these conditions, the heme B group was reproducibly positioned with the central Fe positioned approximately between H266 and H354, albeit with a higher binding energy for all modes than the alternative H150 site (Figure S6). The lowest energy binding modes from each of the two putative heme binding sites are displayed for the maize structural model in Figure 6B,D,E.

The model with the highest-affinity binding mode for the putative H266/H354 site was subsequently used to dock 9,15,9′-tri-*cis*-ζ-carotene to investigate whether the putative cofactor binding site was close to the carotenoid substrate, thereby identifying a possible active site. The twenty binding modes identified in the simulation fell broadly into three groups of similar orientation, all with similar binding energies (Figure S7). One of these groups positions the substrate in proximity to both the predicted position of the heme B cofactor and the H150 residue and is displayed in Figure 6C. The proximity of H150 to the 15-*cis* bond of 9,15,9′-tri-*cis*-ζ-carotene and its known involvement in Z-ISO activity suggests that this residue may have a catalytic role in the *cis*–to-*trans* isomerization. A conserved serine (S177 in maize) is located ~3 Å away from H150, suggesting these two residues may form a catalytic dyad (Figure S8B).

Finally, D294 has also been implicated in Z-ISO activity [6]. The AlphaFold structures identified possible H-bond networks from this residue to residues on other secondary motifs, suggesting a role for this aspartate in protein folding (Figure S7A). Alanine substitutions of D294 and its predicted H-bond partners R252 and H253 gave increased free energy predictions relative to the WT protein (Figure S8A).

3.7. In Silico Mutagenesis and Biochemical Characterization of Putative Heme Binding Sites

We used the Rosetta score function to estimate free energy changes following the in silico mutagenesis of the two putative heme-binding sites (H150 and H266) in the maize Z-ISO AlphaFold model. Unlike AlphaFold, which scores protein-folding accuracy based on an internal confidence probability derived from the machine learning procedure, Rosetta uses a scoring system related to thermodynamic free-energy changes (known as Rosetta energy units, REU) based on calculations of physiologically relevant parameters [39]. The H150A substitution resulted in a small increase in relative free energy compared to WT (−1104 and −1109 REU, respectively), possibly indicating that this substitution alters protein folding, which may explain the decreased and altered heme binding by this mutant observed previously [5]. No inter-domain interactions were identified at the H266 site and the H266A substitution resulted in only a marginal free energy decrease relative to WT (−1110 and 1109 REU, respectively), likely due to the preference for hydrophobic residue packing in the absence of a cofactor.

We further investigated the prospective heme binding sites by performing in silico mutagenesis around the H150 and H266 residues and estimating the free-energy changes using Rosetta. We sought to change the chemical nature of residues (e.g., His to Ala) or introduce a steric clash (e.g., His to Trp) to block the binding cavity. We also aimed to maintain the protein structure using mutations that would have a minimal effect on folding (i.e., minor free-energy changes). For the H266 site, H354 (approximately axial to H266; Figure 6D) was mutated to Ala or Trp, with relatively minor free-energy changes (+1 and +5 REU; Figure S8C). For the H150 site, we analyzed mutations to three proximal residues (Figure 6E), R170, Q255, and R296, changing each to Ala, Phe, or Trp. The Ala and Trp substitution in R170 yielded the most minimal free-energy changes of the three residues (Figure S8C). Thus, the H354A, H354W, R170A, and R170W variants were selected for production in *E. coli*. The membrane extracts and crude protein preparations were visible colored for all four of these protein variants, suggesting that they all bound at least some heme. A comparison of the oxidized and reduced spectrum of the R170A and R170W crude protein extracts showed a typical shift upon the addition of sodium dithionite, suggesting that neither of these substitutions prevent heme binding to Z-ISO (Figure 7A,B). Although pigmentation was observed in the H354A and H354W mutants, no significant change was observed between the oxidized and reduced spectra (Figure 7C,D) [5]. This suggests that heme may be weakly bound by the remaining histidine ligand (H266), but without the requisite sixth coordination bond, the addition of sodium dithionite is insufficient to reduce the ferric center, and hence no spectral changes are observed. In combination with the above modeling, these preliminary results suggest that H266 and H354 are likely to coordinate heme B, while H150 may instead have a structural or catalytic role.
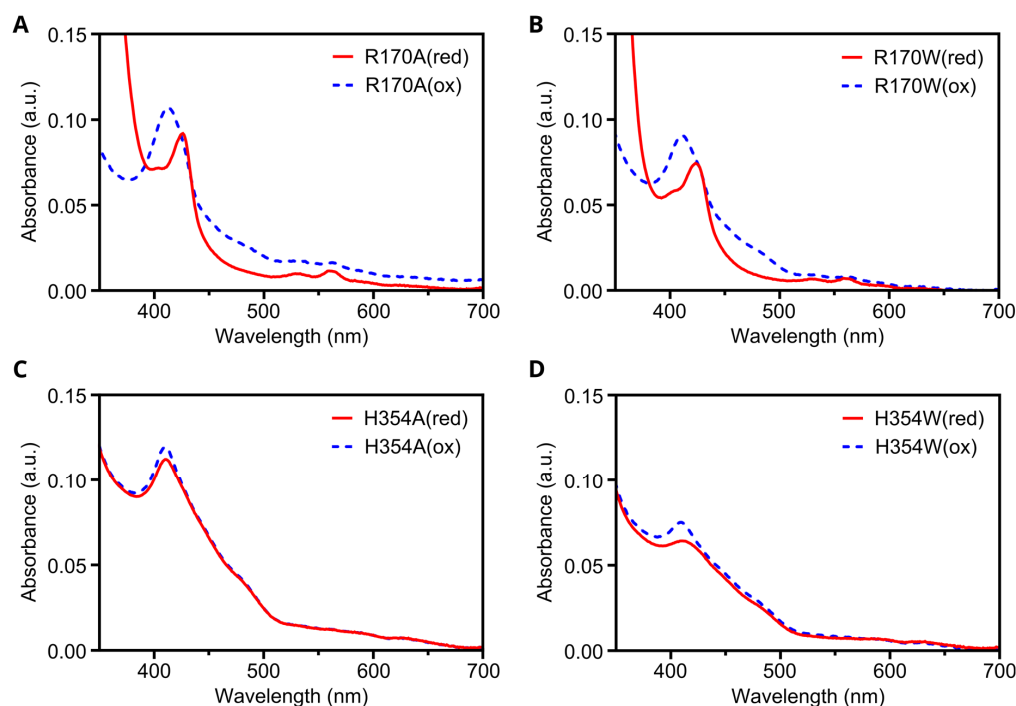
Figure 7. Oxidized and reduced spectra of maize Z-ISO mutants. MBP-*Zm*Z-ISO Arg170Ala (**A**), Arg170Trp (**B**), His354Ala (**C**), and His354Trp (**D**) proteins produced in *E. coli* were solubilized in β-DDM and partially purified by amylose affinity chromatography. The oxidized (ox—dotted blue trace) and reduced (red—solid red trace) absorbance spectra of the eluates were measured after addition of potassium ferricyanide or sodium dithionite, respectively.

4. Conclusions

In this study, we have shown that Slr1599 is the *Synechocystis* Z-ISO, and that it is required for the *cis*-to-*trans* isomerization of the central 15–15′ *cis* double bond of 9,15,9′-tri-

cis-ζ-carotene for the synthesis of carotenoids when the organism is grown in the absence of light. The experimental verification of Z-ISO completes the biosynthetic pathway from phytoene to lycopene, the common precursor of β-carotene and all the xanthophylls utilized for light absorption, energy transfer, and photoprotection, in this model cyanobacterium. In addition to revealing the phenotype of a cyanobacterial Δ*Z-ISO* mutant, we have demonstrated the utility of the Δslr1599 strain for in vivo testing of key residues or the confirmation of Z-ISO homologs from other organisms.

Despite adopting numerous approaches, we were unable to purify Slr1599 for spectroscopic or structural analysis, either as a recombinant MBP-fusion protein from *E. coli* or with a FLAG-tag from the native host. However, sequence alignments and site-directed mutagenesis of conserved residues suggest that Slr1599 contains a heme cofactor like the well-characterized maize enzyme [5]. We suggest that the heme has a role in redox sensing, where the ligand-switch activation model proposed by Beltrán and colleagues [5] would allow for Z-ISO to be inactivated when *Synechocystis* is grown under illumination. Under these conditions, Z-ISO activity can be replaced by photo-isomerization, and we predict that Z-ISO is quickly degraded as we were not able to detect the protein in the proteome of photoautotrophically grown cells [46]. Our failure to isolate the protein during LAHG further suggests that the protein is present at very low levels, even under conditions where we have shown that its activity is required for carotenogenesis.

In the absence of a structure of Z-ISO, we used AlphaFold models and in silico docking to investigate the heme- and substrate-binding sites of Slr1599 and the Z-ISO homologs from maize and *Arabidopsis*. The models suggest that the H150 (maize numbering) may not act as a heme ligand as previously suggested [5], but its position relative to the predicted substrate-binding site offers an explanation for why its substitution with alanine results in a loss of activity in both the maize and *Synechocystis* enzymes. The minimally disruptive mutagenesis strategy and preliminary biochemical analysis described here suggests that it is instead H354 that acts alongside H266 to ligate the heme B cofactor. However, we emphasize that our in silico analysis is not a substitute for an experimentally determined structure of Z-ISO and greater understanding of this complex enzyme requires further experimental work.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Where not included in the article or supplementary material, the data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cogdell, R.J.; Frank, H.A. How Carotenoids Function in Photosynthetic Bacteria. *Biochim. Biophys. Acta* **1987**, *895*, 63–79. [CrossRef]
2. Hashimoto, H.; Uragami, C.; Cogdell, R.J. Carotenoids and Photosynthesis. *Subcell. Biochem.* **2016**, *79*, 111–139. [PubMed]
3. Rosas-Saavedra, C.; Stange, C. Biosynthesis of Carotenoids in Plants: Enzymes and Color. *Subcell. Biochem.* **2016**, *79*, 35–69. [PubMed]
4. Canniffe, D.P.; Hitchcock, A. Carotenoids in photosynthesis—Structure and biosynthesis. In *Encyclopedia of Biological Chemistry III*, 3rd ed.; Jez, J., Ed.; Elsevier: Oxford, UK, 2021; pp. 163–185.
5. Beltrán, J.; Kloss, B.; Hosler, J.P.; Geng, J.; Liu, A.; Modi, A.; Dawson, J.H.; Sono, M.; Shumskaya, M.; Ampomah-Dwamena, C.; et al. Control of carotenoid biosynthesis through a heme-based *cis-trans* isomerase. *Nat. Chem. Biol.* **2015**, *11*, 598–605. [CrossRef] [PubMed]
6. Beltràn Zambrano, J.A. Functional Characterization of The Plant 15-Cis-Zeta-Carotene Isomerase Z-Iso. 2015. Available online: https://academicworks.cuny.edu/gc_etds/521 (accessed on 30 July 2022).
7. Li, F.; Murillo, C.; Wurtzel, E.T. Maize *Y9* encodes a product essential for 15-*cis*-zeta-carotene isomerization. *Plant Physiol.* **2007**, *144*, 1181–1189. [CrossRef]
8. Chen, Y.; Li, F.; Wurtzel, E.T. Isolation and characterization of the *Z-ISO* gene encoding a missing component of carotenoid biosynthesis in plants. *Plant Physiol.* **2010**, *153*, 66–79. [CrossRef]
9. Liu, X.; Hu, Q.; Yan, J.; Sun, K.; Liang, Y.; Jia, M.; Meng, X.; Fang, S.; Wang, Y.; Jing, Y.; et al. ζ-Carotene Isomerase Suppresses Tillering in Rice through the Coordinated Biosynthesis of Strigolactone and Abscisic Acid. *Mol. Plant* **2020**, *13*, 1784–1801. [CrossRef]
10. Zhou, H.; Yang, M.; Zhao, L.; Zhu, Z.; Liu, F.; Sun, H.; Sun, C.; Tan, L. HIGH-TILLERING AND DWARF 12 modulates photosynthesis and plant architecture by affecting carotenoid biosynthesis in rice. *J. Exp. Bot.* **2021**, *72*, 1212–1224. [CrossRef]
11. Sugiyama, K.; Takahashi, K.; Nakazawa, K.; Yamada, M.; Kato, S.; Shinomura, T.; Nagashima, Y.; Suzuki, H.; Ara, T.; Harada, J.; et al. Oxygenic Phototrophs Need ζ-Carotene Isomerase (Z-ISO) for Carotene Synthesis: Functional Analysis in *Arthrospira* and *Euglena*. *Plant Cell Physiol.* **2020**, *61*, 276–282. [CrossRef]
12. Rodrigo, M.J.; Lado, J.; Alós, E.; Alquézar, B.; Dery, O.; Hirschberg, J.; Zacarías, L. A mutant allele of ζ-carotene isomerase (*Z-ISO*) is associated with the yellow pigmentation of the "Pinalate" sweet orange mutant and reveals new insights into its role in fruit carotenogenesis. *BMC Plant Biol.* **2019**, *19*, 465. [CrossRef]
13. Efremov, G.I.; Shchennikova, A.V.; Kochieva, E.Z. Characterization of 15-*cis*-ζ-Carotene Isomerase Z-ISO in Cultivated and Wild Tomato Species Differing in Ripe Fruit Pigmentation. *Plants* **2021**, *10*, 2365. [CrossRef] [PubMed]
14. Fantini, E.; Falcone, G.; Frusciante, S.; Giliberto, L.; Giuliano, G. Dissection of tomato lycopene biosynthesis through virus-induced gene silencing. *Plant Physiol.* **2013**, *163*, 986–998. [CrossRef]
15. Gisriel, C.J.; Wang, J.; Liu, J.; Flesher, D.A.; Reiss, K.M.; Huang, H.L.; Yang, K.R.; Armstrong, W.H.; Gunner, M.R.; Batista, V.S.; et al. High-resolution cryo-electron microscopy structure of photosystem II from the mesophilic cyanobacterium, *Synechocystis* sp. PCC 6803. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2116765118. [CrossRef]
16. Malavath, T.; Caspy, I.; Netzer-El, S.Y.; Klaiman, D.; Nelson, N. Structure and function of wild-type and subunit-depleted photosystem I in *Synechocystis*. *Biochim. Biophys. Acta. Bioenerg.* **2018**, *1859*, 645–654. [CrossRef] [PubMed]
17. Kurisu, G.; Zhang, H.; Smith, J.L.; Cramer, W.A. Structure of the cytochrome b_6f complex of oxygenic photosynthesis: Tuning the cavity. *Science* **2003**, *302*, 1009–1014. [CrossRef] [PubMed]
18. Proctor, M.S.; Malone, L.A.; Farmer, D.A.; Swainsbury, D.J.K.; Hawkings, F.R.; Pastorelli, F.; Emrich-Mills, T.Z.; Siebert, C.A.; Hunter, C.N.; Johnson, M.P.; et al. Cryo-EM structures of the *Synechocystis* sp. PCC 6803 cytochrome b_6f complex with and without the regulatory PetP subunit. *Biochem. J.* **2022**, *479*, 1487–1503. [CrossRef] [PubMed]
19. Schuller, J.M.; Birrell, J.A.; Tanaka, H.; Konuma, T.; Wulfhorst, H.; Cox, N.; Schuller, S.K.; Thiemann, J.; Lubitz, W.; Sétif, P.; et al. Structural adaptations of photosynthetic complex I enable ferredoxin-dependent electron transfer. *Science* **2019**, *363*, 257–260. [CrossRef]
20. Schuller, J.M.; Saura, P.; Thiemann, J.; Schuller, S.K.; Gamiz-Hernandez, A.P.; Kurisu, G.; Nowaczyk, M.M.; Kaila, V.R. Redox-coupled proton pumping drives carbon concentration in the photosynthetic complex I. *Nat. Commun.* **2020**, *11*, 494. [CrossRef]
21. Bautista, J.A.; Rappaport, F.; Guergova-Kuras, M.; Cohen, R.O.; Golbeck, J.H.; Wang, J.Y.; Béal, D.; Diner, B.A. Biochemical and biophysical characterization of photosystem I from phytoene desaturase and zeta-carotene desaturase deletion mutants of *Synechocystis* Sp. PCC 6803: Evidence for PsaA- and PsaB-side electron transport in cyanobacteria. *J. Biol. Chem.* **2005**, *280*, 20030–20041. [CrossRef]

22. Breitenbach, J.; Fernández-González, B.; Vioque, A.; Sandmann, G. A higher-plant type zeta-carotene desaturase in the cyanobacterium *Synechocystis* PCC6803. *Plant Mol. Biol.* **1998**, *36*, 725–732. [CrossRef]

23. Masamoto, K.; Wada, H.; Kaneko, T.; Takaichi, S. Identification of a gene required for *cis*-to-*trans* carotene isomerization in carotenogenesis of the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Cell Physiol.* **2001**, *42*, 1398–1402. [CrossRef] [PubMed]

24. Tóth, T.N.; Chukhutsina, V.; Domonkos, I.; Knoppová, J.; Komenda, J.; Kis, M.; Lénárt, Z.; Garab, G.; Kovács, L.; Gombos, Z.; et al. Carotenoids are essential for the assembly of cyanobacterial photosynthetic complexes. *Biochim. Biophys. Acta* **2015**, *1847*, 1153–1165. [CrossRef] [PubMed]

25. Masamoto, K.; Hisatomi, S.; Sakurai, I.; Gombos, Z.; Wada, H. Requirement of carotene isomerization for the assembly of photosystem II in *Synechocystis* sp. PCC 6803. *Plant Cell Physiol.* **2004**, *45*, 1325–1329. [CrossRef]

26. Isaacson, T.; Ohad, I.; Beyer, P.; Hirschberg, J. Analysis *in vitro* of the enzyme CRTISO establishes a poly-cis-carotenoid biosynthesis pathway in plants. *Plant Physiol.* **2004**, *136*, 4246–4255. [CrossRef] [PubMed]

27. Cunningham, F.X., Jr.; Gantt, E. A portfolio of plasmids for identification and analysis of carotenoid pathway enzymes: *Adonis aestivalis* as a case study. *Photosynth. Res.* **2007**, *92*, 245–259. [CrossRef]

28. Proctor, M.S.; Pazderník, M.; Jackson, P.J.; Pilný, J.; Martin, E.C.; Dickman, M.J.; Canniffe, D.P.; Johnson, M.P.; Hunter, C.N.; Sobotka, R.; et al. Xanthophyll carotenoids stabilise the association of cyanobacterial chlorophyll synthase with the LHC-like protein HliD. *Biochem. J.* **2020**, *477*, 4021–4036. [CrossRef]

29. Rippka, R.; Derueles, J.; Waterbury, J.B.; Herdman, M.; Stainer, R.Y. Generic Assignments, Strain Histories and Properties of Pure Cultures of Cyanobacteria. *J. Gen. Microbiol.* **1979**, *111*, 1–61. [CrossRef]

30. Anderson, S.L.; McIntosh, L. Light-activated heterotrophic growth of the cyanobacterium *Synechocystis* sp. strain PCC 6803: A blue-light-requiring process. *J. Bacteriol.* **1991**, *173*, 2761–2767. [CrossRef]

31. Hitchcock, A.; Jackson, P.J.; Chidgey, J.W.; Dickman, M.J.; Hunter, C.N.; Canniffe, D.P. Biosynthesis of Chlorophyll *a* in a Purple Bacterial Phototroph and Assembly into a Plant Chlorophyll–Protein Complex. *ACS Synth. Biol.* **2016**, *5*, 948–954. [CrossRef]

32. Hollingshead, S.; Kopečná, J.; Jackson, P.J.; Canniffe, D.P.; Davison, P.A.; Dickman, M.J.; Sobotka, R.; Hunter, C.N. Conserved chloroplast open-reading frame *ycf54* is required for activity of the magnesium protoporphyrin monomethylester oxidative cyclase in *Synechocystis* PCC 6803. *J. Biol. Chem.* **2012**, *287*, 27823–27833. [CrossRef]

33. Chen, G.E.; Hitchcock, A.; Mareš, J.; Gong, Y.; Tichý, M.; Pilný, J.; Kovářová, L.; Zdvihalová, B.; Xu, J.; Hunter, C.N.; et al. Evolution of Ycf54-independent chlorophyll biosynthesis in cyanobacteria. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2024633118. [CrossRef] [PubMed]

34. Chidgey, J.W.; Linhartová, M.; Komenda, J.; Jackson, P.J.; Dickman, M.J.; Canniffe, D.P.; Koník, P.; Pilný, J.; Hunter, C.N.; Sobotka, R. A cyanobacterial chlorophyll synthase-HliD complex associates with the Ycf39 protein and the YidC/Alb3 insertase. *Plant Cell* **2014**, *26*, 1267–1279. [CrossRef] [PubMed]

35. Wurtzel, E.T.; Beltrán, J. Improved Expression and Purification of the Carotenoid Biosynthetic Enzyme Z-ISO. *Methods Mol. Biol.* **2020**, *2083*, 53–61. [PubMed]

36. Beltrán, J.; Wurtzel, E.T. Enzymatic isomerization of ζ-carotene mediated by the heme-containing isomerase Z-ISO. *Methods Enzymol.* **2022**, *671*, 153–170.

37. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [CrossRef]

38. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Meng, E.C.; Couch, G.S.; Croll, T.I.; Morris, J.H.; Ferrin, T.E. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **2021**, *30*, 70–82. [CrossRef]

39. Alford, R.F.; Leaver-Fay, A.; Jeliazkov, J.R.; O'Meara, M.J.; DiMaio, F.P.; Park, H.; Shapovalov, M.V.; Renfrew, P.D.; Mulligan, V.K.; Kappel, K.; et al. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **2017**, *13*, 3031–3048. [CrossRef]

40. Trott, O.; Olson, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461. [CrossRef]

41. Zhu, Y.; Graham, J.E.; Ludwig, M.; Xiong, W.; Alvey, R.M.; Shen, G.; Bryant, D.A. Roles of xanthophyll carotenoids in protection against photoinhibition and oxidative stress in the cyanobacterium *Synechococcus* sp. strain PCC 7002. *Arch. Biochem. Biophys.* **2010**, *504*, 86–99. [CrossRef]

42. Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirosawa, M.; Sugiura, M.; Sasamoto, S.; et al. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **1996**, *3*, 109–136. [CrossRef]

43. Trautner, C. Synechocystis mutants lacking genes potentially involved in carotenoid metabolism. 2011. Available online: https://keep.lib.asu.edu/items/149541 (accessed on 30 July 2022).

44. Niedzwiedzki, D.M.; Swainsbury, D.J.K.; Canniffe, D.P.; Hunter, C.N.; Hitchcock, A. A photosynthetic antenna complexes foregoes unity carotenoid to bacteriochlorophyll energy transfer efficiency to ensure photoprotection. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 6502–6508. [CrossRef] [PubMed]

45. Fontaine, F.; Fuchs, R.T.; Storz, G. Membrane localization of small proteins in *Escherichia coli*. *J. Biol. Chem.* **2011**, *286*, 32464–32474. [CrossRef]

46. Jackson, P.J.; Hitchcock, A.; Brindley, A.A.; Dickman, M.J.; Hunter, C.N. Absolute quantification of cellular levels of photosynthesis-related proteins in *Synechocystis* sp. PCC 6803. *Photosynth. Res.* **2022**, *under review*.

47. Rohl, C.A.; Strauss, C.E.M.; Misura, K.M.S.; Baker, D. Protein Structure Prediction Using Rosetta. *Methods Enzymol.* **2004**, *383*, 66–93. [PubMed]

48. Das, R.; Baker, D. Macromolecular modeling with rosetta. *Annu. Rev. Biochem.* **2008**, *77*, 363–382. [CrossRef] [PubMed]

49. Kloss, B. Genomics-based strategies toward the identification of a Z-ISO carotenoid biosynthetic enzyme suitable for structural studies. *Methods Enzymol.* **2022**, *671*, 171–205. [PubMed]

Communication

Expression of Scytonemin Biosynthesis Genes under Alternative Stress Conditions in the Cyanobacterium *Nostoc punctiforme*

Janine Bennett and Tanya Soule *

Department of Biological Sciences, Purdue University Fort Wayne, Fort Wayne, IN 46805, USA;
Janine.Bennett@osumc.edu
* Correspondence: soulet@pfw.edu; Tel.: +1-260-481-0229

Abstract: The indole-alkaloid scytonemin is a sunscreen pigment that is widely produced among cyanobacteria as an ultraviolet radiation (UVR) survival strategy. Scytonemin biosynthesis is encoded by two gene clusters that are known to be induced by long-wavelength radiation (UVA). Previous studies have characterized the transcriptome of cyanobacteria in response to a wide range of conditions, but the effect on the expression of scytonemin biosynthesis genes has not been specifically targeted. Therefore, the aim of this study is to determine the variable response of scytonemin biosynthesis genes to a variety of environmental conditions. Cells were acclimated to white light before supplementation with UVA, UVB, high light, or osmotic stress for 48 h. The presence of scytonemin was determined by absorbance spectroscopy and gene expression of representative scytonemin biosynthesis genes was measured using quantitative PCR. Scytonemin genes were up-regulated in UVA, UVB, and high light, although the scytonemin pigment was not detected under high light. There was no scytonemin or upregulation of these genes under osmotic stress. The lack of pigment production under high light, despite increased gene expression, suggests a time-dependent delay for pigment production or additional mechanisms or genes that may be involved in scytonemin production beyond those currently known.

Keywords: scytonemin; ultraviolet radiation; high light; cyanobacteria

1. Introduction

As phototrophic organisms, cyanobacteria must live in environments that are regularly exposed to solar ultraviolet radiation (UVR) [1]. Ultraviolet A radiation (UVA; 320–400 nm) is a major contributor to cell damage from solar energy [2] since more than 95% of solar UVR that reaches the Earth's surface is within the UVA range. The remainder of the UVR that reaches Earth comes from ultraviolet B radiation (UVB; 280–320 nm) [3,4]. UVA indirectly damages DNA by interaction with cellular chromophores, which lead to the production of DNA-damaging reactive oxygen species. These reactive oxygen species also go on to damage other biomolecules, resulting in UVA as a major contributor to cell damage [2,5–8].

Cyanobacteria have evolved several defense strategies against UVA radiation [1,9–11]. One strategy employed is the production of the indole-alkaloid sunscreen pigment, scytonemin. When scytonemin accumulates extracellularly following UVR exposure, it intercepts harmful photons before they damage cellular machinery. Scytonemin remains stable in the presence of abiotic stressors and is reported to block up to 90% of UVR from entering the cell [12,13]. In addition to its protective ability against UVR, scytonemin shows antioxidant capabilities and nontoxicity [14–17].

Scytonemin biosynthesis is encoded by two conserved gene clusters in the cyanobacterium *Nostoc punctiforme*. The main biosynthetic cluster includes 18 genes (Npun_R1276 to Npun_R1259) regulated by an upstream two-component regulatory system (TCRS; Npun_F1278 and Npun_F1277) (Figure 1) [18–21]. Interestingly, the sensor kinase, Npun_

F1277, contains PAS/PAC domains [22]. In other studies, these domains have been shown to bind to small molecules in response to stimuli, such as light, oxygen, pH, or salinity [23]. The other cluster of five genes (NpunF_5232 to NpunF_5236), known as the *ebo* cluster, is involved in precursor transport across the periplasm [18,24]. The downstream genes in the main gene cluster are involved in synthesizing amino acids, which are then utilized by the upstream genes *scyA* to *scyC* (Npun_R1276 to Npun_R1274) to synthesize specific scytonemin precursors [25–27]. Both gene clusters and the TCRS are highly conserved among closely related species of cyanobacteria, with the *ebo* cluster also occurring among diverse bacteria and algae [18,24].
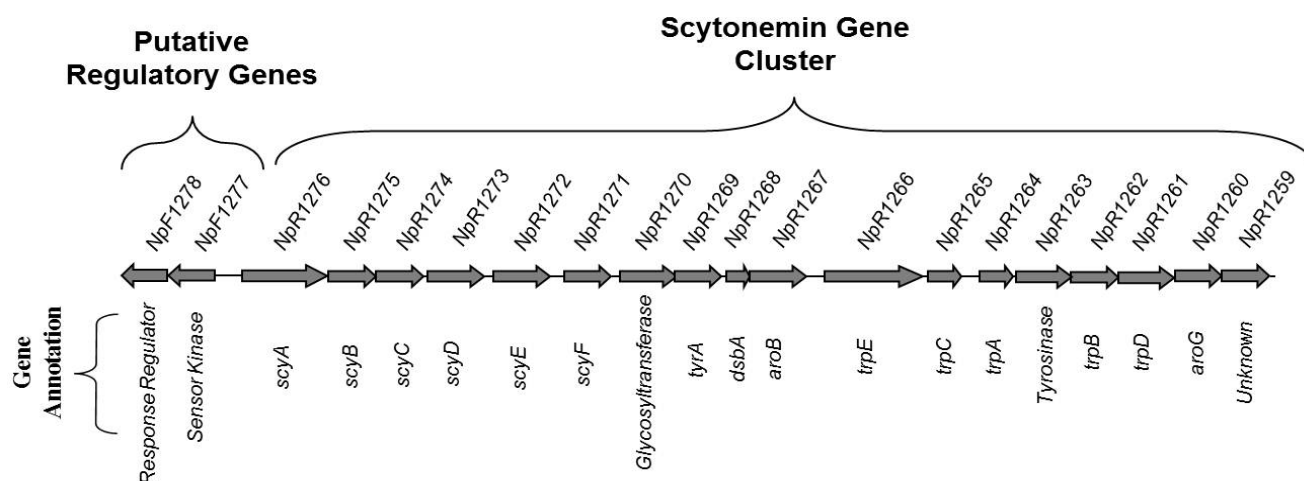
Figure 1. Main scytonemin biosynthetic gene cluster and regulatory genes in the *N. punctiforme* genome.

Although UVA alone can elicit the biosynthetic response, other studies have demonstrated that scytonemin induction can be enhanced by UVA supplemented with UVB, high light, and osmotic stress [13,28,29]. While these studies identified alternative stresses that can induce scytonemin biosynthesis, some were conducted in conjunction with UVA, while none of them were conducted at the gene expression level or in *Nostoc punctiforme*.

The objectives of this study are to evaluate the presence of scytonemin in *Nostoc punctiforme* and measure the transcriptional response of scytonemin biosynthetic genes in response to UVA, UVB, high light, and osmotic stress. It is widely accepted that UVA induces the production of scytonemin and a previous study revealed a synergistic or antagonistic effect, when cells were stressed with other environmental factors in conjunction with UVA [12,28]. Furthermore, expression of the main scytonemin biosynthetic gene cluster, the TCRS, and the *ebo* cluster have all been shown to increase in response to UVA stress [30,31]. Additionally, under short-term UVB and high light stress (20 to 60 min), the TCRS responded with an increase in gene expression [31]. Given these previous studies, it is expected that the light-associated environmental conditions (UVA, UVB, and high light) will induce the expression of genes from both gene clusters, while a non-light-associated condition, such as osmotic stress will downregulate these genes. To test this hypothesis, the expression of representative genes from each gene cluster were measured using quantitative-PCR following stress exposure. Genes *scyA* and *trpB* were measured to represent the main biosynthetic cluster and *eboE* was measured for the *ebo* gene cluster. A previous study showed that the main scytonemin gene cluster is transcribed as two main transcripts, the first starting with *scyA* and the downstream transcript starting with *trpE* [30]. Therefore, *scyA* and *trpB* were chosen to record expression levels from each transcript. The *eboE* gene (Npun_F5233) is the fourth gene in the cluster. It was selected since it showed the strongest expression in response to UVA out of any of the *ebo* genes in a previous study [30].

This study adds value to the previous body of research since it directly connects various elements of what we know regarding scytonemin biosynthesis and how it relates

to environmental conditions. For instance, there are studies that have focused on the induction of scytonemin under different conditions [28], others that have looked at the whole transcriptomic response to these conditions [32,33], and still others that connected scytonemin gene expression to only UVA stress [30]. To date, no single study has targeted the intersection of scytonemin biosynthesis, the associated gene expression response, and compared them under various conditions. Furthermore, any effort to mass produce scytonemin in a native host will need a thorough understanding of the regulatory mechanisms governing its synthesis to maximize efficiency. Overall, by identifying alternative ways to induce scytonemin at the gene expression level, this study will help us better understand the various ways that scytonemin may play a role in the cyanobacterial response to diverse environmental conditions.

2. Materials and Methods

2.1. Strain, Culture Conditions, and Pigment Extraction

Cultures of *Nostoc punctiforme* ATCC 29133 (PCC 73102; a gift from Ferran Garcia-Pichel) were grown under white light in sterile flasks containing the Allen and Arnon (AA/4) growth medium [34]. For high light and UVR exposure, *N. punctiforme* cells were filtered onto 90 mm polycarbonate membrane filters and the filters were placed floating on a nitrogen-free AA/4 medium in sterile glass dishes to allow for an even illumination and growth. All of the cultures remained under white light to acclimate and allow for initial growth. The conditions used for scytonemin induction studies were adapted from Dillon, Tatsumi, Tandingan, and Castenholz [28] and Soule, Garcia-Pichel, and Stout [30]: White light (~40 μmols photons m^{-2} s^{-1}) was supplemented with UVA (5 W m^{-2}; continuous), UVB (0.5 W m^{-2}; 3 h per day), or osmotic stress (5 g NaCl L^{-1}; continuous) and high light was provided at a final intensity of 100 μmols photons m^{-2} s^{-1} (continuous). For UVB stress, transparent glass lids were replaced with a plastic film for optimal UVB penetration. Each stress condition was provided for 48 h since previous data showed that expression of the scytonemin biosynthetic genes peak after 48 h of continuous UVA exposure [30]. Control cultures for each experimental variable were grown under standard (white light) conditions in the absence of the stress. After the stress period, cells were vortexed off the filter into a sterile AA/4 medium and then concentrated by centrifugation, frozen in liquid nitrogen, and stored at −80 °C until RNA processing.

To assess scytonemin pigment production following exposure, cells were opened by grinding with a glass mortar and pestle and scytonemin was extracted in 100% acetone in the dark overnight. The presence of scytonemin was distinguished from other pigments by its ability to absorb light at 384 nm [12].

2.2. Expression of Scytonemin-Associated Genes

Total RNA was isolated from the cells using bead-beating and LiCl precipitation [35]. DNA was degraded in each sample with the Ambion TURBO DNA-free™ Kit (Life Technologies), according to the manufacturer's protocol. DNase-treated RNA (2 μg) was converted to cDNA using the Sensi-Fast cDNA Synthesis Kit (Bioline). Transcript levels of the scytonemin-associated genes, *scyA* (Npun_R1276), *trpB* (Npun_R1262), and *eboE* (Npun_F5233) were measured in triplicate with quantitative-PCR on a CFX Connect Real-Time PCR System (Bio-Rad Laboratories) using the iScript™ Reverse Transcription Supermix for RT-PCR (Bio-Rad Laboratories) and primers from Soule, Garcia-Pichel, and Stout [30]. Samples were processed at 95 °C for 3 min, followed by 39 cycles at 95 °C for 10 s and 55 °C for 30 s. The reference gene was Npun_R0035, which encodes for DNA gyrase, a housekeeping protein that has been used in similar studies [30,31,36]. Expression levels were normalized to those of the reference gene and the fold change between stressed and unstressed cells for all of the treatments was determined using the ΔΔCq formula for calculating normalized expression [37] in the CFX Manager software (Bio-Rad Laboratories). Statistical comparisons were conducted with triplicate biological replicate data

using a student's *t*-test between the treated and untreated samples, with a *p*-value of <0.05 regarded as significant.

3. Results

3.1. Scytonemin Production

Scytonemin was extracted from the stressed cells to determine the relationship between scytonemin production and the expression of scytonemin biosynthetic genes. In response to 48 h of UVA and UVB exposure, scytonemin was produced by the cells (Figure 2). However, scytonemin was not produced in response to high light and osmotic stress using the same duration of exposure.
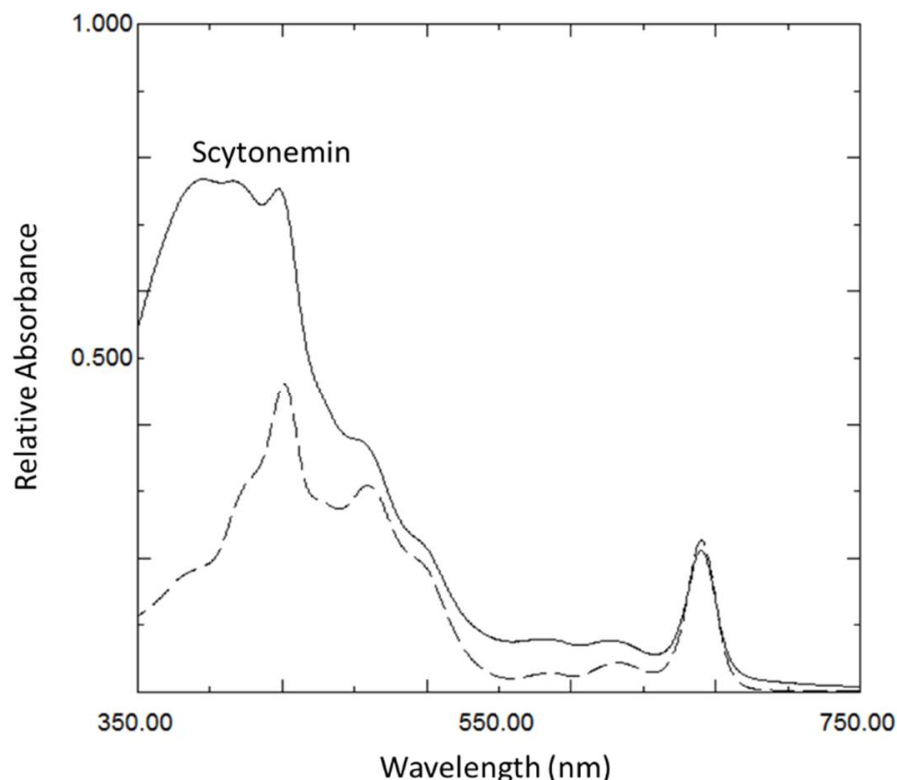
Figure 2. Representative absorption spectra of lipid-soluble pigments extracted from the cyanobacterium *Nostoc punctiforme* after exposure to UVA for 48 h with scytonemin (solid line) compared to high light for 48 h without scytonemin (dashed line). The strong absorption at 384 nm is indicative of scytonemin. Similar results were obtained for UVB (similar to the solid line) as well as high light and osmotic stress (similar to the dashed line).

3.2. Gene Expression

Experiments were conducted separately for each condition, with dedicated control experiments conducted in parallel for each condition. The gene expression analysis revealed the transcript abundance of scytonemin genes *scyA*, *trpB*, and *eboE* in response to each condition. Both *scyA* and *trpB* were upregulated in all of the conditions except for osmotic stress compared to the unstressed cells (Table 1). Interestingly, *trpB* had a stronger response for UVA, while *scyA* was stronger for UVB and high light stress. Osmotic-stressed samples showed a downregulation of *scyA* and *trpB* in comparison to the unstressed samples. The *eboE* gene expression response was more variable. There were no significant changes in gene regulation for *eboE* in response to UVA compared to the untreated samples. However, in cells treated with UVB and osmotic stress, *eboE* was downregulated, and for those exposed to high light, it was upregulated.

Table 1. Pigment production and transcript abundance of *scyA, trpB*, and *eboE* following 48 h of stress.

Stress	Scytonemin	Gene	Fold Change	*p*-Value
UVA	Yes	*scyA*	+26.05	0.049 *
		trpB	+77.53	0.039 *
		eboE	−1.63	0.278
UVB	Yes	*scyA*	+22.7	0.002 *
		trpB	+2.23	0.001 *
		eboE	−3.01	0.001 *
High Light	No	*scyA*	+28.59	<0.001 *
		trpB	+4.63	<0.001 *
		eboE	+3.59	<0.001 *
Osmotic	No	*scyA*	−4.90	0.002 *
		trpB	−3.12	0.001 *
		eboE	−1.53	0.010 *

* $p < 0.05$ indicates a significant difference in gene expression between triplicate treated and untreated cultures.

4. Discussion

Scytonemin is a highly effective molecule that plays an essential role in defending cyanobacteria against the harmful consequences of UVA exposure. It was hypothesized that the genes associated with sunscreen biosynthesis would be expressed in response to other environmental conditions. This was in part based on the broad ecological distribution of scytonemin-producing strains, as well as the previous literature [12,28]. The results mostly support the hypothesis by the upregulation of *scyA* and *trpB* genes in response to UVA, UVB, and high light. In response to osmotic stress, there was a downregulation in *scyA* and *trpB*. In a previous study, osmotic stress in conjunction with UVA showed an inhibitory effect on scytonemin production [28]. Although the previous study did not include a transcriptional abundance parameter, it is reasonable to deduce that scytonemin genes were not upregulated.

The *ebo* gene cluster was studied since it is related to scytonemin biosynthesis, specifically to the export of scytonemin precursors across the periplasm [24]. The *eboE* gene showed a moderate upregulation when exposed to high light and a moderate downregulation under UVB, but overall showed a greater decrease in fold change reaction as compared to *scyA* and *trpB*. This result is consistent with a previous study that showed that the expression of all the *ebo* genes followed the peak in the expression of the biosynthetic genes [30]. This temporal distinction between the biosynthetic genes and the *ebo* genes suggests that they play a post-biosynthetic role in the production of scytonemin, such as secretion or post-translation modification. Therefore, the role of the *ebo* cluster in export could explain the delayed gene expression response.

Although there was upregulation of the scytonemin genes in cultures treated with high light, scytonemin was not produced, in contrast to UVA and UVB stress. The differences in scytonemin production when comparing UVA, UVB, and high light suggest that other factors play a role in pigment production. One possibility is that the proteins involved in scytonemin biosynthesis are subject to post-translational modification and that high light does not regulate this process. While the post-translational regulatory mechanisms of these proteins are unknown, ScyF (Npun_R1271) contains a NHL domain subgroup (IPR013017) that could be involved in protein–protein interactions [27]. It is possible that this protein–protein interaction or others rely on post-translational modification. Alternatively, additional genes in the *N. punctiforme* genome could be involved in scytonemin biosynthesis, which are not expressed under high light. Still another possibility for the lack of pigment production under high light could be a time-dependent response for scytonemin production. The specific time point that transcripts and end products are measured is critical to the interpretation of the data, since transcript levels change rapidly and product synthesis can lag for days. In fact, scytonemin production has been shown to occur in response to high light over 99 µmols photons m^{-2} s^{-1} after 4 days of exposure, especially

with the addition of blue light [12]. Therefore, the failure to detect scytonemin could be due to the fact that the stress period of 2 days was not sufficient for translation and enzyme activity to occur. Time was also shown to be a factor in a temporal study of expression that showed expression of the scytonemin biosynthetic genes after 48 h of UVA stress, while the *ebo* genes were not upregulated until 96 h [30].

5. Conclusions

The light-associated conditions effectively induced the expression of scytonemin genes, while osmotic stress did not. While this was expected, it was also surprising since scytonemin was not produced under high light stress. This can possibly be explained by unknown post-translational modifications or additional genes necessary for pigment production, which are not induced by high light. Alternatively, it could simply be due to differences in the temporal response between expression and biosynthesis. Regardless of the specific reason that high light did not result in pigment production, several possibilities exist that could be explored in future studies. These results emphasize the significance of studies targeting the intersection of the physiological response and specific gene expression response, rather than inferring this relationship from other studies. Overall, our results support previous studies, and based on the expression of genes under high light that did not result in pigment production, the data also suggest that additional mechanisms may be involved in the biosynthesis and regulation of this pigment.

Author Contributions: Conceptualization and methodology, T.S.; investigation, J.B.; resources, T.S.; writing—original draft preparation, J.B.; writing—review and editing, T.S.; funding acquisition, T.S. and J.B. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study is available at https://users.pfw.edu/soulet/pubs.html (accessed on 23 December 2021).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Castenholz, R.W.; Garcia-Pichel, F. Cyanobacterial responses to UV radiation. In *Ecology of Cyanobacteria II*; Whitton, B.A., Ed.; Springer: Dordrecht, The Netherlands, 2012; pp. 481–502.
2. Jagger, J. *Solar-UV Actions on Living Cells*; Praeger: New York, NY, USA, 1985.
3. Kasting, J.F. *The Proterozoic Biosphere: A Multidisciplinary Study*; Cambridge University Press: Cambridge, UK, 1992.
4. Garcia-Pichel, F. Solar ultraviolet and the evolutionary history of cyanobacteria. *Orig. Life Evol. Biosph.* **1998**, *28*, 321–347. [CrossRef]
5. Jiang, Y.; Rabbi, M.; Kim, M.; Ke, C.; Lee, W.; Clark, R.L.; Mieczkowski, P.A.; Marszalek, P.E. UVA generates pyrimidine dimers in DNA directly. *Biophys. J.* **2009**, *96*, 1151–1158. [CrossRef] [PubMed]
6. Van Baalen, C. The effects of ultraviolet radiation on a coccoid blue-green alga: Survival, photosynthesis, and photoreactivation. *Plant Physiol.* **1968**, *43*, 1689–1695. [CrossRef] [PubMed]
7. Donkor, V.A.; Häder, D.P. Effects of ultraviolet irradiation on photosynthetic pigments in some filamentous cyanobacteria. *Aquat. Microbial. Ecol.* **1996**, *11*, 143–149. [CrossRef]
8. Tyrell, R.M. UVA (320–380 nm) radiation as an oxidative stress. In *Oxidative Stress: Oxidants and Antioxidants*; Sies, H., Ed.; Academic Press: Berkeley, CA, USA, 1991; pp. 57–83.
9. Ehling-Schulz, M.; Scherer, S. UV protection in cyanobacteria. *Eur. J. Phycol.* **1999**, *34*, 329–338. [CrossRef]
10. Bebout, B.M.; Garcia-Pichel, F. UV-B-induced vertical migrations of cyanobacteria in a microbial mat. *Appl. Environ. Microbiol.* **1995**, *61*, 4215–4222. [CrossRef] [PubMed]
11. Canfield, L.M.; Forage, J.W.; Valenzuela, J.G. Carotenoids as cellular antioxidants. *Proc. Soc. Exp. Biol. Med.* **1992**, *200*, 260–265. [CrossRef]
12. Garcia-Pichel, F.; Castenholz, R.W. Characterization and biological implications of scytonemin, a cyanobacterial sheath pigment. *J. Phycol.* **1991**, *27*, 395–409. [CrossRef]

13. Garcia-Pichel, F.; Sherry, N.D.; Castenholz, R.W. Evidence for an ultraviolet sunscreen role of the extracellular pigment scytonemin in the terrestrial cyanobacterium *Chlorogloeopsis* sp. *Photochem. Photobiol.* **1992**, *56*, 17–23. [CrossRef]

14. Matsui, K.; Nazifi, E.; Hirai, Y.; Wada, N.; Matsugo, S.; Sakamoto, T. The cyanobacterial UV-absorbing pigment scytonemin displays radical-scavenging activity. *J. Gen. Appl. Microbiol.* **2012**, *58*, 137–144. [CrossRef] [PubMed]

15. Stevenson, C.S.; Capper, E.A.; Roshak, A.K.; Marquez, B.; Eichman, C.; Jackson, J.R.; Mattern, M.; Gerwick, W.H.; Jacobs, R.S.; Marshall, L.A. The identification and characterization of the marine natural product scytonemin as a novel antiproliferative pharmacophore. *J. Pharmacol. Exp. Ther.* **2002**, *303*, 858–866. [CrossRef]

16. Duan, Z.F.; Ji, D.N.; Weinstein, E.J.; Liu, X.Z.; Susa, M.; Choy, E.; Yang, C.; Mankin, H.; Hornicek, F.J. Lentiviral shRNA screen of human kinases identifies PLK1 as a potential therapeutic target for osteosarcoma. *Cancer Lett.* **2010**, *293*, 220–229. [CrossRef]

17. Evans, J.; Jones, A.C.; Blumenthal, E.; Soule, T. Anti-proliferation of melanoma cells and immune stimulation by the cyanobacterial indole-alkaloid scytonemin. *Fine Focus* **2021**, *7*, 54–63. [CrossRef]

18. Soule, T.; Palmer, K.; Gao, Q.; Potrafka, R.; Stout, V.; Garcia-Pichel, F. A comparative genomics approach to understanding the biosynthesis of the sunscreen scytonemin in cyanobacteria. *BMC Gen.* **2009**, *10*, 336–346. [CrossRef] [PubMed]

19. Soule, T.; Stout, V.; Swingley, W.D.; Meeks, J.C.; Garcia-Pichel, F. Molecular genetics and genomic analysis of scytonemin biosynthesis in *Nostoc punctiforme* ATCC 29133. *J. Bacteriol.* **2007**, *189*, 4465–4472. [CrossRef]

20. Naurin, S.; Bennett, J.; Videau, P.; Philmus, B.; Soule, T. The response regulator Npun_F1278 is essential for scytonemin biosynthesis in the cyanobacterium *Nostoc punctiforme* ATCC 29133. *J. Phycol.* **2016**, *52*, 564. [CrossRef]

21. Garcia-Pichel, F.; Lombard, J.; Soule, T.; Dunaj, S.; Wu, S.H.; Wojciechowski, M.F.; Giovannoni, S.J. Timing the evolutionary advent of cyanobacteria and the later great oxidation event using gene phylogenies of a sunscreen. *mBio* **2019**, *10*, e00561-19. [CrossRef] [PubMed]

22. Ashby, M.; Houmard, J. Cyanobacterial two-component proteins: Structure, diversity, distribution, and evolution. *Microbiol. Mol. Biol. Rev.* **2006**, *70*, 472–509. [CrossRef]

23. Ponting, C.P.; Aravind, L. PAS: A multifunctional domain family comes to light. *Curr. Biol.* **1997**, *7*, R674–R677. [CrossRef]

24. Klicki, K.; Ferreira, D.; Hamill, D.; Dirks, B.; Mitchell, N.; Garcia-Pichel, F.; Greenberg, E.P.; Burnap, R.; Preston, G.; Vermaas, W. The widely conserved *ebo* cluster is involved in precursor transport to the periplasm during scytonemin synthesis in *Nostoc punctiforme*. *mBio* **2018**, *9*, e02266-18. [CrossRef]

25. Balskus, E.P.; Walsh, C.T. Investigating the initial steps in the biosynthesis of cyanobacterial sunscreen scytonemin. *J. Am. Chem. Soc.* **2008**, *130*, 15260–15261. [CrossRef]

26. Balskus, E.P.; Walsh, C.T. An enzymatic cyclopentyl[b]indole formation involved in scytonemin biosynthesis. *J. Am. Chem. Soc.* **2009**, *131*, 14648–14649. [CrossRef]

27. Ferreira, D.; Garcia-Pichel, F. Mutational Studies of Putative Biosynthetic Genes for the Cyanobacterial Sunscreen Scytonemin in *Nostoc punctiforme* ATCC 29133. *Front. Microbiol.* **2016**, *7*, 735. [CrossRef]

28. Dillon, J.G.; Tatsumi, C.M.; Tandingan, P.G.; Castenholz, R.W. Effect of environmental factors on the synthesis of scytonemin, a UV-screening pigment, in a cyanobacterium (*Chroococcidiopsis* sp.). *Arch. Microbiol.* **2002**, *177*, 322–331. [CrossRef] [PubMed]

29. Fleming, E.D.; Castenholz, R.W. Effects of periodic desiccation on the synthesis of the UV-screening compound, scytonemin, in cyanobacteria. *Environ. Microbiol.* **2007**, *9*, 1448–1455. [CrossRef] [PubMed]

30. Soule, T.; Garcia-Pichel, F.; Stout, V. Gene expression patterns associated with the biosynthesis of the sunscreen scytonemin in *Nostoc punctiforme* ATCC 29133 in response to UVA radiation. *J. Bacteriol.* **2009**, *191*, 4639–4646. [CrossRef] [PubMed]

31. Janssen, J.; Soule, T. Gene expression of a two-component regulatory system associated with sunscreen biosynthesis in the cyanobacterium *Nostoc punctiforme* ATCC 29133. *FEMS Microbiol. Lett.* **2016**, *363*, fnv235. [CrossRef]

32. Soule, T.; Gao, Q.; Stout, V.; Garcia-Pichel, F. The global response of *Nostoc punctiforme* ATCC 29133 to UVA stress, assessed in a temporal DNA microarray study. *Photochem. Photobiol.* **2013**, *89*, 415–423. [CrossRef] [PubMed]

33. Soule, T.; Ferreira, D.; Lothamer, J.; Garcia-Pichel, F. The independent and shared transcriptomic response to UVA, UVB and oxidative stress in the cyanobacterium *Nostoc punctiforme* ATCC 29133. *Photochem. Photobiol.* **2021**, *97*, 1063–1071. [CrossRef] [PubMed]

34. Allen, M.; Arnon, D.I. Studies on nitrogen-fixing blue-green algae. I. Growth and nitrogen-fixation by *Anabaena cylindrica*. *Plant Physiol.* **1955**, *30*, 366–372. [CrossRef] [PubMed]

35. Campbell, E.L.; Summers, M.L.; Christman, H.; Martin, M.E.; Meeks, J.C. Global gene expression patterns of *Nostoc punctiforme* in steady-state dinitrogen-grown heterocyst-containing cultures at single time points during the differentiation of akinetes and hormogonia. *J. Bacteriol.* **2007**, *186*, 5247–5256. [CrossRef] [PubMed]

36. Sorrels, C.M.; Proteau, P.J.; Gerwick, W.H. Organization, evolution, and expression analysis of the biosynthetic gene cluster for scytonemin, a cyanobacterial UV-absorbing pigment. *Appl. Environ. Microbiol.* **2009**, *75*, 4861–4869. [CrossRef] [PubMed]

37. Pfaffl, M.W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **2001**, *29*, 2002–2007. [CrossRef] [PubMed]

Article

Host Cyanobacteria Killing by Novel Lytic Cyanophage YongM: A Protein Profiling Analysis

Shanshan Zhang [1,2], **Baohua Zhao** [1], **Jing Li** [2], **Xiaofei Song** [2], **Yigang Tong** [2] **and Wenlin An** [2,3,*]

1   College of Life Science, Hebei Normal University, Shijiazhuang 050024, China;
    zhangshanshanhbtu@163.com (S.Z.); zhaobaohua@hebtu.edu.cn (B.Z.)
2   College of Life Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China;
    frostli@163.com (J.L.); 13261297988@163.com (X.S.); tongyigang@mail.buct.edu.cn (Y.T.)
3   Department of Scientific Research Management, National Vaccine and Serum Institute, Beijing 100176, China
*   Correspondence: anwlin@163.com; Tel.: +86-010-52245228

Abstract: Cyanobacteria are autotrophic prokaryotes that can proliferate robustly in eutrophic waters through photosynthesis. This can lead to outbreaks of lake "water blooms", which result in water quality reduction and environmental pollution that seriously affect fisheries and aquaculture. The use of cyanophages to control the growth of cyanobacteria is an important strategy to tackle annual cyanobacterial blooms. YongM is a novel lytic cyanophage with a broad host spectrum and high efficiency in killing its host, cyanobacteria FACHB-596. However, changes in cyanophage protein profile during infestation and killing of the host remains unknown. To characterize the proteins and its regulation networks involved in the killing of host cyanobacteria by YongM and evaluate whether this strain YongM could be used as a chassis for further engineering to be a powerful tool in dealing with cyanobacterial blooms, we herein applied 4D label-free high-throughput quantitative proteomics to analyze differentially expressed proteins (DEPs) involved in cyanobacteria host response infected 1 and 8 h with YongM cyanophage. Metabolic pathways, such as photosynthesis, photosynthesis-antennal protein, oxidative phosphorylation, ribosome, carbon fixation, and glycolysis/glycol-isomerization were significantly altered in the infested host, whereas DEPs were associated with the metabolic processes of photosynthesis, precursor metabolites, energy production, and organic nitrogen compounds. Among these DEPs, key proteins involved in YongM-host interaction may be photosystem I P700 chlorophyll-a apolipoprotein, carbon dioxide concentration mechanism protein, cytochrome B, and some YongM infection lysis-related enzymes. Our results provide comprehensive information of protein profiles during the invasion and killing of host cyanobacteria by its cyanophage, which may shed light on future design and manipulation of artificial cyanophages against water blooms.

Keywords: cyanophage; *Nostoc* sp.; label-free quantitative proteomics; photosynthesis; substance metabolism; energy metabolism

1. Introduction

Water eutrophication is a global water pollution problem that causes massive production of cyanobacteria [1,2]. The latter can lead to a cyanobacteria bloom, causing huge economic and environmental losses [3–5]. Cyanophages are a group of viruses that specifically infect prokaryotic cyanobacteria [6–9], playing a very important role in the regulation of the population and influencing evolution [10]. These viruses are biological control agents and have great potential for controlling harmful blooms of cyanobacteria [11]. However, most reported cyanophages have strong host specificity, a narrow cyanobacteria-killing spectrum, and a long lysis cycle [12,13]. For example, MA-LMM01, isolated from Lake Mikata, Japan [14], could only infect a toxic strain of *Microcystis aeruginosa* and had a long incubation period [15,16]. The *Podoviridae* cyanophages Pf-WMP3 [17] and Pf-WMP4 [18]

can infect the same host cyanobacteria, but with a narrow algaecidal spectrum. Infection experiments have shown that S-LBS1 takes about four days to induce cyanobacterial lysis, and only in strain TCC793 of *Synechococcus* sp. [19]. In addition, newly discovered cyanophages in the last 2–3 years, such as PA-SR01, only lysed *Pseudanabaena* sp. KCZY-C8 [20]. Mic1 was only infective to its host, *Microcystis wesenbergii* FACHB-1339 [21]. Cyanophages S-H68 and S-B68, isolated from Bohai Sea waters, had incubation periods as long as 41 and 49 h, respectively [22,23]. Cyanobacterial blooms are usually formed by mixed outbreaks of multiple cyanobacteria with a rapid reproduction rate [24,25]. Therefore, application of existing cyanophages to treat cyanobacterial blooms is limited [26,27].

Cyanophage infection of cyanobacteria involves multiple proteins, and conventional methods cannot systematically analyze their mechanisms of action in a high throughput fashion. The quantitative proteomics system is a powerful approach in investigating the pivotal proteins involved in the cyanophage's infection of host cyanobacteria and deciphering the cyanobacteria-killing mechanism of cyanophages. Studies on *Prochlorococcus* sp. MED4 and its cyanophage P-SSP7 have shown that infection is followed by complex gene expression dynamics, but little is known about the initial cyanophage-host interactions [28]. The infection of the cyanophage S-SM1 with *polycoccus* WH8102 has been studied with an isotope-labeled proteomic approach to determine the uptake of extracellular nitrogen and its use in biosynthesis of viral and host proteins [29]. The effect on photosynthesis of the cyanophage P-TIM68 carrying PSI and PSII-related genes during infection of the *Prochlorococcus* host MIT9515 in tropical oceans showed that P-TIM68 may maintain or even manipulate host photosynthesis [30]. Genomics, proteomics, and transcriptomics have been used to trace Shine–Dalgarno sequence differences in *Synechocystis* sp., *Microcystis aeruginosa*, cyanophage, tobacco chloroplasts, and *Arabidopsis chloroplasts* [31]. However, off-label quantitative proteomics have not been used to investigate the mechanism of cyanophage infestation of host cyanobacteria. Also, further research is needed on the kinetics, phagocytosis pattern, and mechanism of interaction between the cyanophage and host [32].

The newly isolated *Myoviridae* cyanophage from water samples of Kunming Dianchi Lake, YongM, has efficient lysis characteristics and can rapidly infest and lyse its host *Nostoc* sp. FACHB-596 within 24 h. YongM was able to lyse 18 of 33 tested cyanobacteria strains, indicating a broad spectrum of lysis hosts. The lysis characteristics of YongM have been studied by exposure experiments and absorbance measurements. The color change of the host solution was not obvious 1 h after YongM infestation, but it turned yellow after 8 h, representing the lysis phase. This lysis efficiency is more rapid and efficient than any strain reported so far and it can be potentially applied to control cyanobacterial blooms. Therefore, the host cyanobacteria 1 and 8 h after YongM infestation were used for quantitative proteomic comparison. DEPs in the host were screened and the expression characteristics were bioinformatically analyzed. GO analysis revealed that after infection, most of the DEPs were associated with photosynthesis, light response, precursor metabolites and energy production, and metabolic processes of organic nitrogen compounds. KEGG pathway enrichment results indicated that metabolic pathways, such as photosynthesis, photosynthesis-antenna protein, oxidative phosphorylation, ribosome, carbon fixation, and glycolysis/glycoisomerization were significantly altered in the infested host cyanobacteria. In addition, there were several key, differentially expressed proteins, such as photosystem I P700 chlorophyll-a apolipoprotein, carbon dioxide concentration mechanism protein, and cytochrome B. These may serve as hotspots for the study of the YongM-host interaction mechanism. They can also provide genetic resources and a theoretical basis for future cyanophage transformation to produce efficient and broad-spectrum cyanophages, and to understand their interaction with the hosts.

2. Materials and Methods

2.1. Materials

The strain of host cyanobacteria FACHB-596 was obtained from a freshwater cyanobacteria culture bank at the Institute of Hydrobiology, Wuhan Academy of Sciences, China. Cyanophage YongM was isolated from water samples of the Dianchi Lake in Kunming, Yunnan. It was preserved in the Center for General Microbiology of Microbial Culture Collection Management Committee, Conservation Number: CGMCC No.18383.

BG-11 medium was purchased from Qingdao Hope Bio-Technology Co., Ltd, Qingdao, China. (product number: HB8793). Its ingredients include sodium nitrate, potassium hydrogen phosphate trihydrate, magnesium sulfate heptahydrate, calcium chloride dihydrate, citric acid, ammonium ferric citrate, EDTA, sodium carbonate, boric acid, manganese chloride monohydrate, zinc sulfate heptahydrate, copper sulfate pentahydrate, sodium molybdate dihydrate, and cobalt nitrate hexahydrate. A GXZ intelligent light incubator was purchased from Ningbo Jiangnan Instrument Factory, Ningbo, China (model number: GXZ-280B).

2.2. Separation, Purification, and Suspension Preparation

Samples of surface water were collected from Yunnan Dianchi (latitude: 24.964616, longitude: 102.666303) and centrifuged at $12,000 \times g$ for 20 min at 4 °C. The supernatant was filtered through 0.45 µm and 0.22 µm pore-sized nitrocellulose membranes (Merck Millipore Ltd., Shanghai, China). The filtrate was added to a volume ten times larger of FACHB-596 cyanobacteria in logarithmic growth solution. The solution was mixed and incubated until yellowing was evident. The procedure was repeated three times. The produced lysate was then serially diluted ten-fold with BG-11 medium and used in double-layer agar plate spread experiments. After the growth of phagocytic spots were observed, several individual spots were dug up and suspended in 5 mL of FACHB-596 growth suspension in logarithmic phase. The above phagocytic spot experiment was repeated after yellowing, until phagocytic spots of uniform shape and size were formed on the plate. The purified lysate was centrifuged at $5000 \times g$ for 10 min at 4 °C, producing a suspension of purified YongM in the supernatant.

2.3. Genome Enrichment and Extraction

High titer cyanophage-cyanobacteria culture lysate (30 mL) was centrifuged at $6000 \times g$ for 20 min, and the supernatant was sequentially filtered through 0.45 µm and 0.22 µm pore size nitrocellulose membranes. The filtrate was centrifuged in a density gradient at 4 °C for 1 h at $35,000 \times g$ using 20% (w/v) and 40% (w/v) sucrose solution freshly prepared within a week. After discarding the supernatant, the precipitate was resuspended with 200 µL of PBS (0.01 M, PH7.4) and the cyanophage genome was extracted using the Roche High Pure Viral Nucleic Acid Kit (Roche, Basel, Switzerland).

2.4. Infection and Sample Collection

The host culture, in logarithmic phase, was inoculated with a YongM suspension at an optimal MOI = 0.1 (experimental group T) whereas another sample was inoculated only with BG-11 medium (control group C). Each group was placed in a light incubator at 25 °C with a light intensity of 2000 lx and exposed to 12 h light-dark cycles. After 1 h and 8 h of inoculation, the cultures in both groups were gently shaken and a sample (3 mL) was taken immediately from each group. These samples were centrifuged at $6000 \times g$ for 15 min at 4 °C. The sediments were washed twice with PBS and stored at −80 °C for protein extraction and proteome sequencing.

2.5. Protein Extraction and Digestion

For extraction, lysis buffer (500 µL, 8 mM urea, 30 mM HEPES, 1 mM PMSF, 2 mM EDTA and 10 mM DTT) was added to alga cells, followed by a 5 min ultrasonic ice bath treatment (pulse on 2 s, pulse off 3 s, power 230 W). The supernatant was centrifuged at

$2000 \times g$ for 30 min. DTT was added until the final concentration was 10 mM. After the samples were incubated at 56 °C for 1 h, IAM was quickly added to a final concentration of 55 mM and the solution was incubated in the dark for 1 h. The protein concentration was quantified by the Bradford method.

For digestion, 40 µg protein from each sample was centrifuged in ultrafiltration tubes with 3 K MCWO at $14{,}000 \times g$ (4 °C, 40 min) and the supernatant was discarded. The pellet was solubilized with 200 µL of 50 mM NH_4HCO_3 and was centrifuged at 4 °C for 40 min at $14{,}000 \times g$, discarding the supernatant. The process was repeated twice. Trypsin (1 µg/µL) was added to the samples at a protein substrate to enzyme ratio of 30. The samples were incubated in water at 37 °C for 24 h. The digested samples were lyophilized and solubilized with 25 mM NH_4HCO_3 (30 µL per tube).

2.6. Proteome Sequencing

Desalted peptide mixtures were loaded onto a Acclaim PePmap C18-reversed phase column (75 µm × 2 cm, 3 µm, 100 Å thermo scientific) and separated with a reversed phase C18 column (75 µm × 10 cm, 5 µm, 300 Å, Agela Technologies, Tianjin, China) mounted on a Dionex ultimate 3000 nano LC system. Peptides were eluted using a gradient of 5–80% (*v/v*) acetonitrile in 0.1% formic acid over 45min at a flow rate of 300 nL min^{-1} combined with a Q Exactive mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA). The eluates were directly entered using Q—Exactive MS (Thermo Fisher Scientific, Waltham, MA, USA), setting in positive ion mode and data-dependent manner with full MS scan from 350–2000 *m/z*, full scan resolution at 70,000, MS/MS scan resolution at 17500.MS/MS, scan with minimum signal threshold 1×10^5, isolation width at 2 Da. To evaluate the performance of this mass spectrometry on the Label-free samples, two dd-MS2 acquisition modes was used, AGC target 1×10^5, normalized collision energy (NCE) was systemically examined 30, stepped 27%.

2.7. Data Analysis

The free software MaxQuant (version 1.6.0.1) (developed by Jürgen Cox & Matthias Mann in Max-Planck Institute for Biochemistry, Martinsried, Germany) was used for mass spectrometry data acquisition and quantitative processing. MSstats ANOVA in R software was used to assess the significance of differences. DEPs were defined as those with fold change (FC) greater than or equal to 1.2, or less than 0.833, between experimental and control groups by *t*-test $p < 0.05$. The PCA analysis, volcanic maps, heat map of DEPs clustering, and associated variance data were analyzed by GraphPad Prism (8.0.2) and R Package (V.2.0.3). To determine the biological role of DEPs, we used Gene Ontology (GO) analysis with the Gene Ontology database and the annotation and classification of proteins according to the three aspects of GO analysis, namely biological processes, cellular components, and molecular functions. The Kyoto Encyclopedia of Genes and Genomes database (KEGG) was used to identify DEP-related metabolic and signal transduction pathways.

3. Results and Analysis

3.1. Effect of YongM Infection on Host Growth

YongM was added to log-phase host cyanobacteria FACHB-596 cells and was incubated under a circadian rhythm (alternate 12 h dark and light periods). The color change of the cyanobacteria solution is not obvious after 1 h, where YongM is still in the latent phase, but the solution turned yellow after 8 h (Figure 1), indicating that the phagosome was in lysis phase. The susceptibility of 33 cyanobacterial strains to YongM is shown in Table 1.
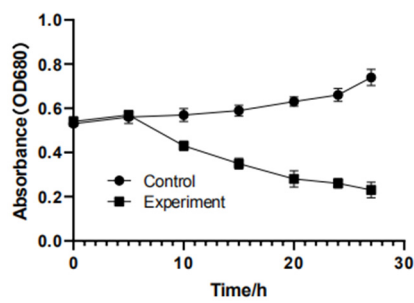
(**a**)   (**b**)

Figure 1. Growth of host cyanobacteria FACHB-596 after cyanophage YongM infection. (**a**) Host cyanobacteria FACHB-596 absorbance versus infection time by YongM; (**b**) Phenotypic changes of YongM-infected host cyanobacteria FACHB-596 cells after 1 h (**left**) and 8 h (**right**).

Table 1. Host range analysis of YongM against 33 cyanobacterial strains.

Orders	Family	Species	Strains	Susceptibility	Origin
Chroococcales	*Microcystaceae*	*Microcystis aeruginosa*	FACHB-905	+	China
			FACHB-925	−	Australia
			FACHB-942	−	China
			FACHB-469	+	France
			FACHB-924	−	Australia
			FACHB-1326	−	China
			FACHB-912	−	China
		M. wesenbergii	FACHB-908	+	China
			FACHB-1112	−	China
			FACHB-1317	−	China
			FACHB-1318	−	China
			FACHB-929	+	Japan
		M. viridis	FACHB-979	+	Japan
		M. Flos-aquae	FACHB-1028	+	China
		Microcystis sp.	FACHB-915	+	France
		M. elabens	FACHB-916	−	Japan
		M. Panniformis	FACHB-1757	−	China
		M. ichthyoblabe Kutz	FACHB-1294	−	China
			FACHB-1409	+	China
	Chroococcacea	*Chroococcus* sp.	FACHB-193	+	China
Nostocales	*Aphanizomenonaceae*	*Aphanizomenon flos-aquae*	FACHB-1039	−	China
			FACHB-1040	+	China
		Anabaena flos-aquae	FACHB-245	+	USA
		Anabaena sp.	FACHB-418	+	France
		Dolichospermum flos-aquae	FACHB-1255	+	China
	Nostocaceae	*Nostoc* sp.	FACHB-596	+	China
Oscillatoriale	*Microcoleaceae*	*Planktothrix agardhii*	FACHB-1166	−	China
			FACHB-920	+	Japan
		Planktothricoides raciborskii	FACHB-881	+	China
	Oscillatoriaceae	*Oscillatoria planctonica*	FACHB-708	+	China

Table 1. *Cont.*

Orders	Family	Species	Strains	Susceptibility	Origin
Hormogonales	*Scytonemataceae*	*Plectonema*	FACHB-402	−	USA
			FACHB-240	−	USA
Synechococcales	*Synechococcaceae*	*Synechococcus* sp.	FACHB-805	+	Australia

Note: "+" suspective "−" unsuspective.

3.2. Altered FACHB-596 Protein Expression Profile Caused by YongM Infection

Logarithmic stage host cyanobacteria FACHB-596 infested with YongM after 1 h and 8 h were used as experimental groups (T1 and T8, respectively), whereas cells not exposed to the cyanophage were used as control (C1 and C8, respectively). Proteomic and bioinformatic analyses showed up to 573 proteins differentially expressed after cyanophage infection. Principal component analysis (PCA) of these DEPs (Figure 2a) showed a small separation between the samples in the same group, but a large separation for samples belonging to different groups. This indicates that the experiments were reproducible, and the data was reliable. This also suggests that there are protein expression changes between experimental and control groups.

(**a**)

(**b**)

(**c**)

(**d**)
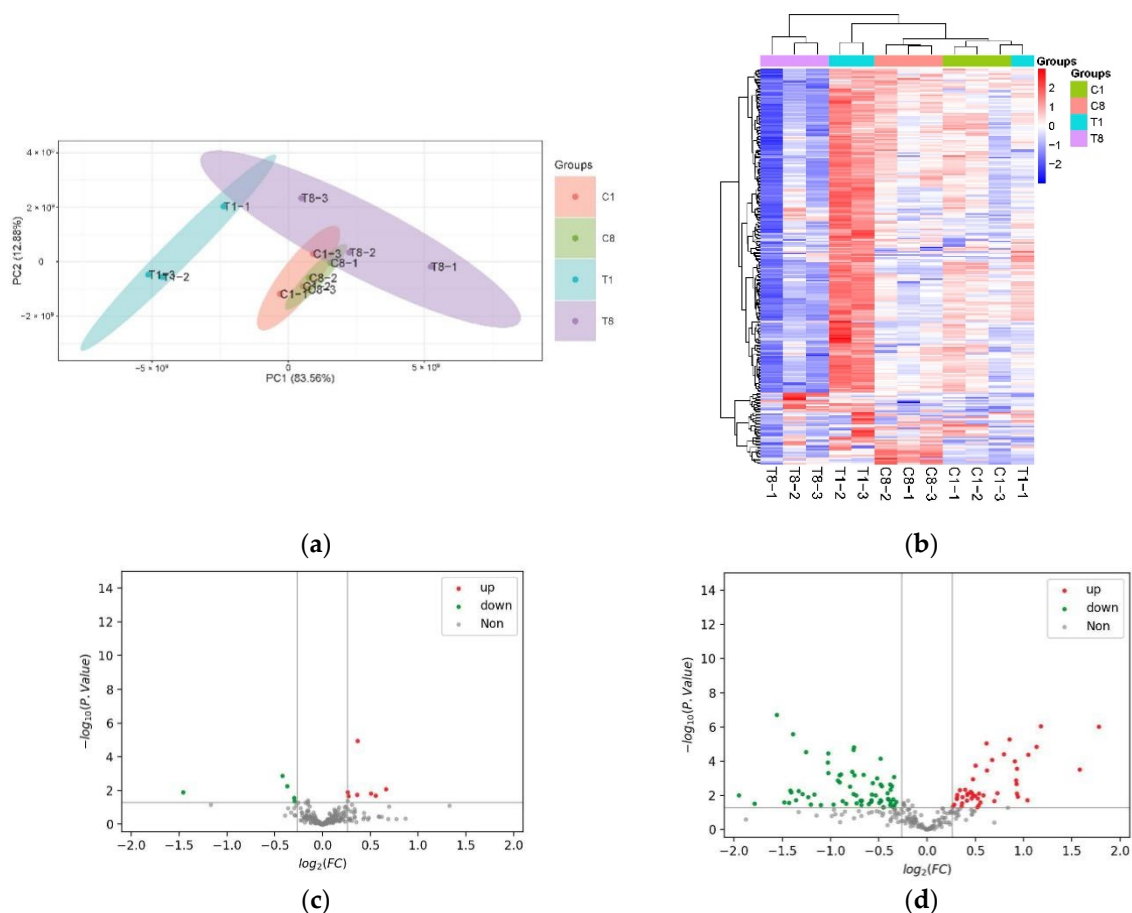
Figure 2. Changes in protein expression in host cyanobacteria caused by YongM infection. (**a**) PCA of DEPs; (**b**) heat map; (**c**,**d**) volcanic maps of DEPS after 1 h infection (**c**) and 8 h infection (**d**).

To visualize the changes in the FACHB-596 cyanobacteria cells altered proteome caused by YongM infection, DEPs were clustered and heatmaps were drawn (Figure 2b). A total of 12 DEPs (5 down-regulated and 7 up-regulated) were significantly altered after 1 h of infection compared to the control (Figure 2c). The most down-regulated protein was the 50S ribosomal protein L23 (A0A1Z4KLJ5_ANAVA), whereas the most up-regulated

protein was the ATP-dependent zinc metalloproteinase FtsH (Q8YXF2_NOSS1). After 8 h, 112 DEPs (67 down-regulated and 45 up-regulated) were altered (Figure 2d). The most down-regulated protein being peroxidase 2 family protein/glutaminase (Q8YWR3_NOSS1) and the most up-regulated protein being photosystem I P700 chlorophyll-a apolipoprotein A1 (A0A1Z4M5R2_MICDP).

3.3. Functional Analysis of DEPs

The number of DEPs individually expressed during 1 h and 8 h after infection in the YongM-infected and uninfected host cells were 127, 74, 74 and 43, respectively. The top 3 statistics of DEPs (Table 2) show that after 1 h of infection, only the membrane-bound, ATP-dependent zinc metalloproteinase FtsH, generally conserved in prokaryotes, was significantly up-regulated. This protein degrades proteins with low thermodynamic stability and lacks unfolding enzyme activity. This suggests that it may be part of a prokaryotic self-protection mechanism to check whether proteins are folded correctly. Ribosomal protein S14, one of the major constituent proteins of small subunit 30 S, is significantly up-regulated in bacteria, cyanobacteria, and plants, indicating enhanced translational activity and high protein synthesis. An additional All4042 protein of unknown function was also significantly up-regulated. The significantly down-regulated outer membrane efflux protein (Alr2887 protein) forms a trimeric channel that can export a variety of substrates in Gram-negative bacteria. After 8 h of infection, photosystem I P700 chlorophyll-a apolipoprotein A1 was significantly up-regulated. This is a membrane protein complex that uses light energy to mediate the transfer of electrons from plastid proteins to ferric oxide reducing proteins and is primarily involved in cyanobacterial photosynthesis. Carbon dioxide concentration mechanism protein is present in a number of autotrophic and non-autotrophic multiple polyhedral shell proteins (CcmK), therefore it may be related to CO_2 utilization. A significantly down-regulated protein was the late competence development protein (A0A1Z4IAM0) that produces an adaptive response to external stimuli and is required for cellular uptake of exogenous DNA from the environment leading to transforming capacity. The latter is prevalent in bacteria and may be involved in gene transfer. The significantly down-regulated photosystem I reaction center subunit IV, present on the matrix side of the cystoid membrane, can form complexes with ferricoxigenin and ferricoxigenin oxidoreductase at the photosystem I reaction center and is associated with cyanobacterial photosynthesis. The phycocyanin β subunit, a phycobilisome protein, collects light energy through water-soluble complexes of the phycobilisome, which are attached to the outer surface of the cystoid membrane and can transfer absorbed energy to the photosynthetic reaction center with >95% efficiency.

Table 2. The three most significant DEPs in host cyanobacteria induced by YongM infection of FACHB-596 cells. ↑: significantly up-regulated; ↓: significantly down-regulated.

Accession	Protein Name	*p* Value	Adj. *p*-Value	T1:C1	T8:C8
Q8YXF2	ATP-dependent zinc metalloproteinase FtsH	0.008218777	0.52091175	1.5848247↑	
Q8YPZ8	All4042 protein	0.020302562	0.55631214	1.4674468↑	
A0A6P0S0D5	30S ribosomal protein S14	0.014660408	0.53405772	1.4211440↑	
Q8YT39	Alr2887 protein	0.005383913	0.45763262	0.7740167↓	
Q8YTI1	All2736 protein	0.001327486	0.16925451	0.7467759↓	
A0A1Z4KLJ5	50Sribosomal protein L23	0.01222975	0.52091175	0.3643173↓	
A0A1Z4M5R2	Photosystem I P700 chlorophyll-a apolipoprotein A1	0.000000967	0.0000825		3.42825728↑
Q8YYI1	Carbon Dioxide Concentration Mechanism protein	0.000305335	0.00390828		2.98863046↑
Q8YTI1	All2736 protein	0.000000887	0.0000825		2.26210112↑
A0A1Z4IAM0	Unknown functional protein	0.000000195	0.0000499		0.33966838↓
A0A1Z4KN79	Argininosuccinate synthase	0.030938777	0.07920326		0.28915421↓
Q8YWR3	Peroxidase 2 family of proteins/glutaminase	0.009777517	0.04120970		0.25880121↓

In summary, after 8 h of infection, host cells showed a weakened defense against exogenous nucleic acid. Expression of argininosuccinate synthase was reduced, resulting in the inhibition of the penultimate step in arginine biosynthesis. This makes it impossible to form argininosuccinate, AMP, and pyrophosphate from citrulline and aspartate.

3.4. GO Analysis of DEPs

GO analysis revealed that the main biological process of DEPs after 1 h of infection was protein metabolism. Cellular components were mainly intrinsic to the membrane, and molecular functions were mainly long- and medium-chain fatty acid-coenzyme ligase activity and fatty acid ligase activity (Figure 3a). After 8 h of infection, the main biological processes of DEPs were precursor metabolites and energy production, photosynthesis, and light reaction. The cellular components were mainly cells and their cell membrane parts. The molecular functions were mainly related to electron transfer activity (Figure 3b). In addition, DEPs were also involved in carbon fixation in photosynthesis, amino acid synthesis, and biosynthesis of antibiotics.
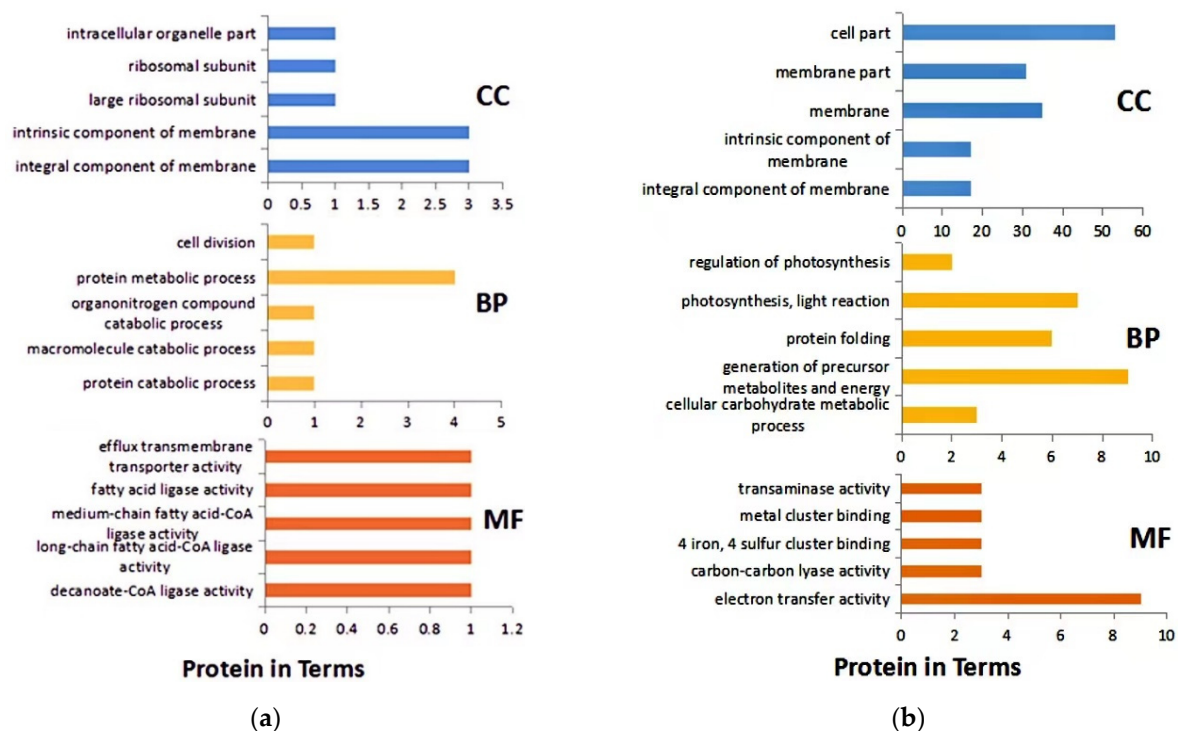
Figure 3. GO analysis of DEPs. (**a**) after 1 h infection; (**b**) after 8 h infection. BP: Biological process; CC: Cell composition; MF: Molecular function.

3.5. KEGG Pathway Analysis of DEPs

KEGG Pathway analysis is the analysis of information networks connecting known molecular interactions, such as metabolic pathways, complexes, and biochemical reactions. The KEGG Pathway was used to analyze DEPs through the MaxQuant Cloud Platform for Integrative Analysis of Genomic Data Collection. Four metabolic pathways, involving 7 DEPs, were significantly enriched 1 h after infections, and 4 of them showed elevated DEPs, accounting for 57.1% (Figure 4a). A total of 41 pathways involving 141 DEPs were enriched 8 h after infection, and six metabolic pathways were significantly enriched with DEPs common to all of them. For example, the photosynthesis-related protein (A0A5Q0GEQ2) was involved in both photosynthesis-related and oxidative phosphorylation metabolic pathways (Figure 4b). Six metabolic pathways were significantly enriched after 1 h and 8 h containing 39 DEPs (Table 3). Among the DEPs found in all pathways, 116 DEPs down-regulated (82.3%), suggesting that 8 h after infection most protein expression in the

host cyanobacteria was inhibited, resulting in a significant decrease in metabolic activity and a weakening of the vital processes. Enriched pathways (Figure 4) included those directly related to cyanobacteria pigment synthesis and energy supply, such as ribosome metabolism, carbon fixation, oxidative phosphorylation, energy metabolism, photosynthesis, and secondary metabolites (e.g., phenylpropanoids and dicarboxylic acids).

(**a**)   (**b**)
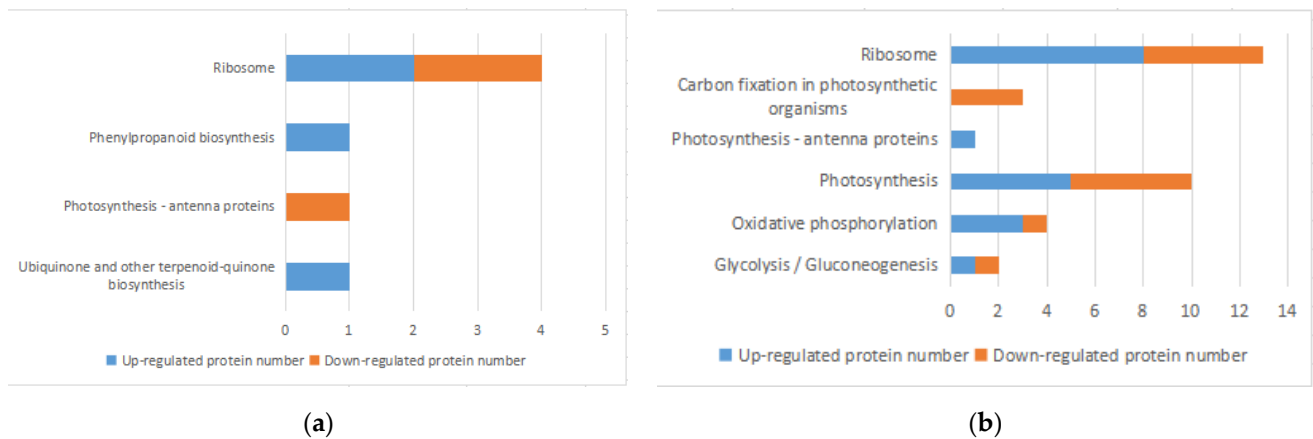
Figure 4. Enrichment of KEGG pathway in DEPs. (**a**,**b**) KEGG enrichment histogram of DEPs after 1 h infection (**a**) and after 8 h infection (**b**).

Table 3. List of DEPs involved in metabolic pathways after infection of host cyanobacteria FACHB-596 cells by YongM.

KEGG Pathway	Protein Name	Protein Description	Upward/Downward
Photosynthesis-antennal proteins	ApcB	Allophycocyanin β subunit	Down
	CpcB	Phycocyanin associated rod junction protein	Up
Ribosomal	L23	Ribosomal large subunit protein L23	Down
	L2	Ribosomal large subunit protein L2	Down
	S14	Ribosomal small subunit protein S14	Up
	L35	Ribosomal large subunit protein L35	Up
	L3	Ribosomal large subunit protein L3	Down
	L4	Ribosomal large subunit protein L4	Down
	L22	Ribosomal large subunit protein L22	Down
	L29	Ribosomal large subunit protein L29	Up
	L5	Ribosomal large subunit protein L5	Down
	S5	Ribosomal small subunit protein S5	Up
	L13	Ribosomal large subunit protein L13	Up
	L1	Ribosomal large subunit protein L1	Up
	L35	Ribosomal large subunit protein L35	Up
	S6	Ribosomal small subunit protein S6	Down
	L21	Ribosomal large subunit protein L21	Up
	S16	Ribosomal small subunit protein S16	Up
	S1	Ribosomal small subunit protein S1	Up
Oxidative phosphorylation	NdhK	NAD (P) H-quinone oxidoreductase subunit K(EC:7.1.1.2)	Down
	alpha	F-type H+/Na+ transporter ATPase subunit α (EC:7.1.2.2 7.2.2.1)	Up
	b	ATPase B subunit	Up
	OSCP	F type H~+ transport ATPase subunit	Up

Table 3. *Cont.*

KEGG Pathway	Protein Name	Protein Description	Upward/Downward
Photosynthesis	Psb-A	Photosystem II P680 reaction center D1 protein (EC:1.10.3.9)	Down
	Psb-O	Photosystem II oxygen evolution enhancer protein 1	Down
	Psb-U	Photosystem IIPsbU protein	Down
	Psb-V	Photosystem II cytochrome C550	Down
	Psb27	Photosystem IIPsb27 protein	Up
	PsaA	Photosystem I P700 chlorophyll-a apolipoprotein A1	Up
	PetA	Acocytochrome F	Up
	PetE	plastocyanin	Down
	alpha	F-type H+/Na+ transporter ATPase subunit α (EC:7.1.2.2 7.2.2.1)	Up
	b	ATPase B subunit	Up
Carbon fixation	2.7.9.1	Pyruvate, orthophosphate dikinase	Down
	4.1.1.39	Ribulose bisphosphate carboxylase large chain	Down
	2.7.2.3	phosphoglyceric kinase	Down
	2.7.9.1	Pyruvate, orthophosphate dikinase	Down
Glycolysis and gluconeogenesis	2.7.2.3	phosphoglyceric kinase	Down
	2.3.1.12	Pyruvate dehydrogenase E2 component (dihydrolipamidoacetyltransferase)	Up

3.5.1. Analysis of Photosynthetic DEPs

The phycobilisomes consist of multiple photosynthetic light-harvesting proteins (i.e., photosynthesis-associated haptoglobin) attached to the outer surface of the cystoid membrane [33]. Cyanobacteria collects light energy through multiple light-harvesting proteins in this complex and transfers the absorbed energy to photosynthetic reaction centers with an efficiency greater than 95%. Among these light-harvesting proteins, phycocyanin is the core component of the water-soluble complex, receiving and transferring energy to the chlorophyll of the cystoid membrane [34]. When YongM infested the host for 1 h, the expression of phycocyanin in the host was reduced, the light-harvesting ability was weakened whereas photosynthesis was inhibited. In contrast, after 8 h YongM infiltration, we observed elevated phycocyanin-related rod-linked proteins, which contributed to light uptake and increased photosynthetic efficiency.

The PSII oxygen precipitation complex (OEC) consists of five subunits, PsbO, PsbP, PsbQ, PsbU, and PsbV [35], which can be photolyzed by water to provide protons for PSI. Its P680 reaction center contains chlorophyll a, which photolyzes water and produces ATP using a proton pump [36]. The P700 reaction center of PSI contains chlorophyll, which absorbs electrons and associated hydrogen delivered from PSII to reduce NADP+ to NADPH. Both ATP and NADPH are then used in light-independent dark reactions, culminating in the conversion of carbon dioxide to glucose and the release of oxygen as a byproduct. PSII is a multisubunit protein-pigment complex in which the core, chlorophyll, and β-carotene bind mainly to the haptoglobin CP43 (PsbC) and CP47 (PsbB), which transmit excitation energy to the reaction center proteins PsbA and PsbD, combining all redox-active cofactors involved in energy conversion [37,38]. After 8 h of infection, the subunits responsible for water photolysis, PsbO, PsbP, PsbQ, and PetE related to photosynthetic electron transport were all down regulated, whereas the photoresponsive phase was suppressed. The up-regulated proteins were Psb27, a highly conserved component of photosystem II consisting of four helices [39], PsaA, a membrane-integrated protein involved in constituting PSI [40] and the cytochrome complex PetA [41], a protein associated with senescence whose up-regulation may indicate that photosynthesis was almost stagnant, and cells were in a state of senescence.

In addition to the above pathway, all DEPs in the carbon fixation pathway were down-regulated after 8 h; phosphoglycerate kinase (PGK), present in all living organisms,

catalyzes the interconversion of ATP and ADP and is highly conserved throughout evolution, converting 1,3-diphosphoglycerate to 3-phosphoglycerate in the second step of glycolysis and forming a molecule of ATP [42]. In most cells, this reaction is essential for ATP production in aerobic bacteria, fermentation in anaerobic bacteria, and carbon fixation in plants. Ribulose diphosphate carboxylase large subunit (RbcL) is the macromolecular subcomplex site of Rubsico, and the RbcX protein has been identified as possessing a chaperone-like protein function, as it contributes to the correct assembly of RbcL and RbcS subunits during Rubsico biosynthesis [43]. This metabolic pathway is complemented by down-regulated pyruvate phosphate dikinase, which catalyzes the reversible conversion of ATP to AMP, pyrophosphate, and phosphoenolpyruvate (PEP). These two enzymes are essential for carbon fixation in cyanobacteria cells, resulting in an inadequate energy supply mechanism during carbon fixation and blocked organic matter synthesis in the host cyanobacteria.

3.5.2. Analysis of Energy Metabolism DEPs

After 8 h of infection, expression of ubiquinone oxidoreductase, an enzyme that catalyzes the transfer of electrons from NADH to ubiquinone and is associated with proton translocation across membranes, was reduced. This is the largest and most complex enzyme in the respiratory chain. In cyanobacteria, the three subunits NdhM, NdhN, and NdhO of the NDH-1 complex influence the electron flow around PSI by providing additional ATP for plastids and photosynthesis in cyanobacteria [44]. The NADH dehydrogenase I complex transfers electrons through FMN and Fe-sulfur centers to quinones in the respiratory and photosynthetic chains and couples redox reactions to proton transport, resulting in proton redox energy conserved as a proton gradient [45,46]. Also, among the upregulated expression proteins, alpha is transmembrane ATPases, which are membrane-bound enzyme complexes/ion transport proteins that use ATP hydrolysis to drive proton translocation across membranes. There are also transmembrane ATPases that work in reverse, using energy in the proton gradient and transmembrane ion flow across the ATPase proton channel to drive ATP synthesis. b is F-ATPase, also known as ATP synthase, a type of transmembrane ATPase that hydrolyzes ATP in bacteria to generate a proton gradient [47,48]. The above suggests that infection by YongM results in the restriction of proton transport during oxidative phosphorylation of the host.

In addition, two DEPs, one up-regulated and another down-regulated, were involved in the glycolytic and gluconeogenic metabolic pathways after 8 h of infection. The up-regulated protein is the 2-oxoacid dehydrogenase acyltransferase, which catalyzes the overall conversion of α-ketoacid to acyl-CoA and CO_2 [49], whereas the down-regulated one is PGK, which is involved in the carbon fixation pathway that is also down-regulated. This protein could catalyze the interconversion of ATP and ADP. Thus, the infestation of YongM may prevent host glycolysis and gluconeogenesis pathways from proceeding normally by regulating the overall conversion of α-keto acid to acyl-CoA and carbon dioxide and the expression of catalytic enzymes that convert ATP and ADP.

3.5.3. Analysis of Translation Function DEPs

Ribosomes are organelles present in all organisms that use mRNA as a template for the targeted synthesis of genetic information carried by genes. About 1/3 of the ribosome mass consists of proteins that are named small subunit proteins (S1 to S31) and large subunit proteins (L1 to L44), based on their ribosomal subunit size. Many ribosomal proteins, especially those of the large subunit, have long finger-like protrusions that extend into the rRNA core and stabilize their structure. In the large subunits, about 1/3 of the 23S rRNA nucleotides are at least in Van der Waals contact with the protein, while L22 interacts with all six structural domains of the 23S rRNA. The proteins S4 and S7 of the starting 16S rRNA assembly are located at the junctions of five and four RNA helices, respectively, and can be used to organize and stabilize the tertiary structure of the rRNA [50,51].

After 1 h of exposure, we observed a decrease in protein expression of ribosomal large subunit proteins L25 and L2, which are known to bind 23S rRNA and have peptidyl transferase activity in *E. coli*. Among the proteins with elevated expression, S14 is one of the proteins from the small subunit of the ribosome. This protein is required for the assembly of 30S particles in *E. coli* and may also be responsible for determining the conformation of 16S rRNA at the A site [52]. Cyanobacteria are Gram-negative bacteria, and the function of its ribosomal proteins is similar to that of *E. coli*. After 8 h of infection, the number of proteins with elevated expression increased significantly, e.g., L1, the largest protein from the ribosomal subunit, with RNA binding site highly conserved and RNA chaperone activity. Compared with 1 h, differentially expressed ribosomal proteins were significantly increased after 8 h. Five large subunit proteins associated with ribosomal rRNA core stability were up-regulated (L29, L13, L1, L35, L21) which facilitate the stable translation of mRNA and sustained protein synthesis. In addition, the expression of the S5 protein, which is closely related to the function of 30S small subunit assembly and can reduce the rate of translation errors, was also up regulated. We speculate that this is consistent with the high vital activity of the cyanophage after 8 h, when they are in the stage of cleavage.

3.6. Structural and Functional Proteins of YongM

Amplification, concentration, and mass spectrometry (LC-MS/MS) analysis of YongM were performed to further validate the results obtained. By delineating and comparing the functional modules of the predicted Open Reading Frames (ORFs) and the actual identified proteins by mass spectrometry (Figure 5b,c), we found that 32 proteins (33.34% of the predicted ORFs) out of 93 predicted proteins (Figure 5a) in the YongM genome could be identified by mass spectrometry proteomics. These included 14 structural proteins, 5 functional proteins, and 13 unknown functional proteins (Table 4), most of which were encoded by the identified structural genes. The molecular weights of the identified proteins were very close to the predicted values. In addition, we identified two lysozymes (Anabaena phage Elbi) and a single lysozyme (Nostoc phage N1) by protein profiling, encoded by the predicted ORFs at position 41 and 86 of the YongM genome, with sizes of 99 kDa and 12.65 kDa, respectively. Four genes were related to DNA anabolism: thymidylate kinase, DNA polymerase, nucleic acid endonuclease, and alkaline phosphatase, which may contribute to the rapid infection and proliferation of YongM [53]. Therefore, we speculate that the lysis efficiency of YongM is inextricably linked to its own dual lytic enzymes and DNA synthesis-related proteins. In addition, the ability of YongM to lyse its host within 12 h and its relatively wide host range may also be attributed to its multiple restriction endonucleases and methylesterases. YongM may serve as a model for future studies on the interaction between these enzymes and freshwater cyanobacteria.

Table 4. Protein profile identification results of YongM.

No.	Location	Description	Score	Coverage
1	20651..21748	putative major capsid protein (Nostoc phage A1)	10,458	86%
2	23869..25389	tail sheath protein (Nostoc phage A1)	4372	46%
3	57421..58554	tail collar protein (Nostoc phage A1)	2265	24%
4	25487..25996	tail tube protein (Nostoc phage A1)	2101	91%
5	18162..19766	hypothetical protein (Nostoc phage A1)	1811	62%
6	29345..32047	lysozyme (Anabaena phage Elbi)	1151	30%
7	5775..7658	alkaline phosphatase D family protein (Salinivenus lutea)	1132	35%
8	26714..28780	tail protein (Nostoc phage A1)	1119	34%
9	60572..61747	baseplate J tail protein (Nostoc phage N1)	1048	44%
10	5256..5762	Capsid fiber protein	1037	61%
11	58577..59965	tail fiber protein (Nostoc phage N1)	995	42%
12	21893..22378	hypothetical protein (Nostoc phage A1)	925	49%
13	19898..20644	putative outer membrane protein (Nostoc phage A1)	912	49%
14	61784..62131	lysozyme (Nostoc phage N1)	672	73%
15	4580..5233	hypothetical protein (Nostoc phage A1)	453	24%

Table 4. *Cont.*

| No.	Location	Description	Score	Coverage
16	32813..33661	hypothetical protein (Nostoc phage N1)	376	37%
17	8358..9086	hypothetical protein (Nostoc phage A1)	288	20%
18	59997..60572	tail collar protein (Nostoc phage A1)	198	30%
19	23110..23832	hypothetical protein (Nostoc phage A1)	169	25%
20	7738..8361	hypothetical protein (Nostoc phage A1)	156	40%
21	62202..62531	hypothetical protein	101	64%
22	32074..32541	hypothetical protein	100	26%
23	63660..64472	Baseplate structural protein	71	11%
24	22375..22701	putative tail-component	66	19%
25	22706..23113	hypothetical protein	57	18%
26	15042..15680	hypothetical protein (Nostoc phage A1)	29	6%
27	42619..43221	hypothetical protein (Microcystis phage Me-ZS1)	29	6%
28	55094..57100	DNA polymerase delta catalytic subunit	19	1%
29	3..722	Thymidylate kinase (EC 2.7.4.9)	19	7%
30	20661..26102	tail length tape-measure protein (Microcystis phage Me-ZS1)	16	2%
31	41382..41699	hypothetical protein (Microcystis phage Me-ZS1)	14	6%
32	26900..27886	tail tube protein (Microcystis phage Me-ZS1)	13	4%
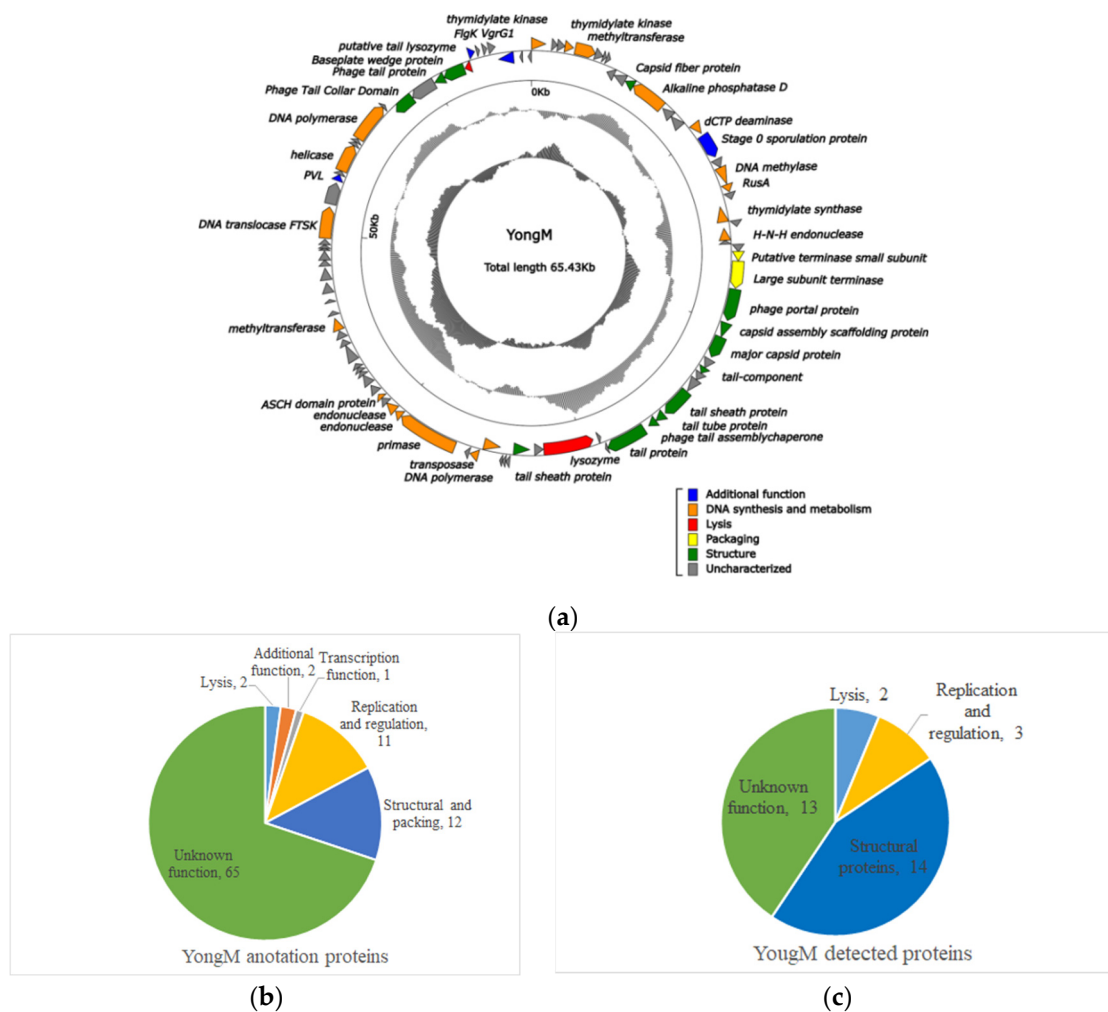
Figure 5. Comparison of predicted ORFs of YongM with proteins identified by proteomics. (**a**) Genome map of YongM; (**b**) heat map of DEPs due to cyanophage YongM infestation of host cyanobacteria FACHB-596 cells; (**c**) protein function/metabolic pathway and proportion of YongM detected proteins.

4. Discussion and Conclusions

Since the discovery and isolation of cyanophages, their biological functions and ecological importance in aquatic ecosystems have received great attention, especially in the control of harmful cyanobacterial blooms. Indeed, cyanophages have become important focal points in environmental science and virology. In recent years, molecular biology methods have been used to study the genetic diversity and evolutionary history of cyanophages [54,55]. For example, transcriptomic and metabolomic analyses can be performed on lytic cyanophage-infected host cells. However, gene expression can also be regulated at the post-translational level, therefore proteomic analyses are essential to fully understand the cellular response to specific conditions and to track the fate of each protein [56]. However, the systems involved in the regulation of protein expression levels are very complex, producing a very dynamic proteome. Thus, technical limitations have left most of the viral proteome unresolved, except in a few cases of classic examples [57]. Recent improvements in instrumentation, and the development of highly sensitive analytical methods, have facilitated advances in viral proteomics, but most studies are still focused on the biology of eukaryotic viruses.

The present study systematically analyzed for the first-time host FACHB-596 DEPs, and their involvement in metabolic pathways, during YongM infestation. This was done at the protein level, using a non-standard quantitative proteomics approach (Figure 6). One hour after infection, expression of cyanobacterial proteins with light-collecting activity was down-regulated, photosynthesis was inhibited and resistance of the host to the exogenous genome was reduced. This may create the ideal conditions for YongM infection and gene replication. The decrease in the expression of ribosomal large subunit proteins L25 and L2 reduced the binding of L2 protein to 23S rRNA and inhibited its peptidyl transferase activity. The host ribosome function was impaired. After 8 h of infestation, the expression of phycocyanin-associated rod-linked proteins and large subunit proteins associated with ribosomal rRNA core stability were upregulated, facilitating light energy uptake and the stable translation of mRNA and sustained protein synthesis. In addition, the expression of S5 protein, closely related to the function of 30S small subunit assembly and able to reduce the rate of translation errors, was also up regulated. This may be consistent with the high vital activity of the cyanophage after 8 h, when it was in the ascending stage of lysis. YongM may induce the ageing of the host in addition to initiate dual lytic enzyme cleavage for rapid passive cleavage and death of the host. Indeed, we found that the expression of the senescence-related protein-cytochrome complex PetA was up-regulated after 8 h of infection. In the carbon fixation pathway, the expression of all DEPs was down regulated. The expression of phosphoglycerate kinase and pyruvate phosphate dikinase was reduced, resulting in an insufficient energy supply for carbon fixation in the host, blocking organic matter synthesis and affecting vital activities. In addition, abnormal expression of the host's proton transport-related proteins and catalytic enzymes in the oxidative phosphory-lation, the glycolysis/gluconeogenesis pathway, and the interconversion of ATP and ADP prevented the host's metabolic pathways from proceeding normally.

In conclusion, these results provide systematic information on protein profiling during the invasion and killing of host cyanobacteria by cyanophages. The identification of significant DEPs and cyanophage infestation, cleavage-related effector enzymes in the host cyanobacteria after infestation may provide insights into the design and manipulation of artificial phages against water blooms, as well as genetic resources and stimulate theoretical studies on phage-host interactions.
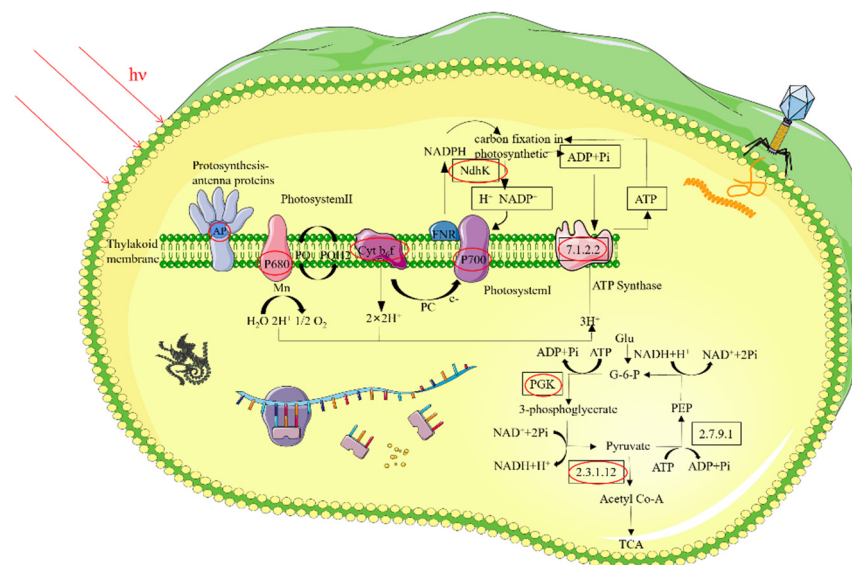
Figure 6. Schematic diagram of key proteins and pathways involved in the lysis of cyanobacteria FACHB-596 cells caused by YongM infestation.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gilbert, P.M. Eutrophication, harmful algae and biodiversity—Challenging paradigms in a world of complex nutrient changes. *Mar. Pollut. Bull.* **2017**, *124*, 591–606. [CrossRef]
2. Hu, X.; Zhang, R.; Ye, J.; Xu, X.; Zhang, Y.; Wu, C. Monitoring and research of microcystins and environmental factors in a typical artificial freshwater aquaculture pond. *Environ. Sci. Pollut. Res.* **2018**, *25*, 5921–5933. [CrossRef]
3. Harke, M.J.; Steffen, M.M.; Gobler, C.J.; Otten, T.G.; Wilhelm, S.W.; Wood, S.A.; Paerl, H.W. A review of the global ecology, genomics, and biogeography of the toxic cyanobacterium, Microcystis spp. *Harmful Algae* **2016**, *54*, 4–20. [CrossRef]
4. Chen, L.; Chen, J.; Zhang, X.; Xie, P. A review of reproductive toxicity of microcystins. *J. Hazard. Mater.* **2016**, *301*, 381–399. [CrossRef]
5. Suttle, C. Viruses in the sea. *Nature* **2005**, *437*, 356–361. [CrossRef]
6. Safferman, R.S.; Cannon, R.E.; Desjardins, P.R.; Gromov, B.V.; Haselkorn, R.; Sherman, L.A.; Shilo, M. Classification and nomenclature of viruses of cyanobacteria. *Intervirology* **1983**, *19*, 61–66. [CrossRef]
7. Mirzaei, M.K.; Eriksson, H.; Kasuga, K.; Haggård-Ljungquist, E.; Nilsson, A.S. Genomic, Proteomic, Morphological, and Phylogenetic Analyses of vB_EcoP_SU10, a Podoviridae Phage with C3 Morphology. *PLoS ONE* **2014**, *10*, e116294. [CrossRef]
8. Safferman, R.S.; Morris, M.E. Algal virus:isolation. *Science* **1963**, *140*, 679–680. [CrossRef]
9. Brussaard, C.; Wilhelm, S.W.; Thingstad, F.; Weinbauer, M.G.; Bratbak, G.; Heldal, M.; Kimmance, S.A.; Middelboe, M.; Nagasaki, K.; Paul, J.H.; et al. Global-scale processes with a nanoscale drive: The role of marine viruses. *ISME J.* **2008**, *2*, 575–578. [CrossRef]
10. Roux, S.; Brum, J.R.; Dutilh, B.E.; Sunagawa, S.; Duhaime, M.B.; Loy, A.; Poulos, B.T.; Solonenko, N.; Lara, E.; Poulain, J.; et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses Ocean microbes drive biogeochemical cycling on a global scale. *Nature* **2016**, *537*, 689–693. [CrossRef]
11. Lin, W.; Li, D.F.; Sun, Z.T.; Tong, Y.; Yan, X.; Wang, C.; Zhang, X.; Pei, G. A novel freshwater cyanophage vB_MelS-Me-ZS1 infecting bloom-forming cyanobacterium Microcystis elabens. *Mol. Biol. Rep.* **2020**, *47*, 7979–7989. [CrossRef] [PubMed]

12. Jonge, P.A.; Nobrega, F.L.; Brouns, S.J.J.; Dutilh, B.E. Molecular and Evolutionary Determinants of Bacteriophage Host Range. *Trends Microbiol.* **2019**, *27*, 51–63. [CrossRef] [PubMed]

13. Silpe, J.E.; Bassler, B.L. A Host-Produced Quorum-Sensing Autoinducer Controls a Phage Lysis-Lysogeny Decision. *Cell* **2019**, *176*, 268–280. [CrossRef] [PubMed]

14. Yoshida, T.; Nagasaki, K.; Takashima, Y.; Shirai, Y.; Tomaru, Y.; Takao, Y.; Sakamoto, S.; Hiroishi, S.; Ogata, H. Ma-LMM01 infecting toxic Microcystis aeruginosa illuminates diverse cyanophage genome strategies. *J. Bacteriol.* **2008**, *190*, 1762–1772. [CrossRef]

15. Xia, H.; Li, T.; Deng, F.; Hu, Z. Freshwater cyanophages. *Virol. Sin.* **2013**, *28*, 253–259. [CrossRef]

16. Yoshida-Takashima, Y.; Yoshida, M.; Ogata, H.; Nagasaki, K.; Hiroishi, S.; Yoshida, T. Cyanophage infection in the bloom-forming cyanobacteria Microcystis aeruginosa in surface freshwater. *Microbes Environ.* **2012**, *27*, 350–355. [CrossRef] [PubMed]

17. Liu, X.; Kong, S.; Shi, M.; Fu, L. Genomic Analysis of Freshwater Cyanophage Pf-WMP3 Infecting Cyanobacterium Phormidium foveolarum: The Conserved Elements for a Phage. *Microb. Ecol.* **2008**, *56*, 671–680. [CrossRef]

18. Liu, X.; Shi, M.; Kong, S.; Gao, Y.; An, C. Cyanophage Pf-WMP4, a T7-like phage infecting the freshwater cyanobacterium Phormidium foveolarum: Complete genome sequence and DNA translocation. *Virology* **2007**, *366*, 28–39. [CrossRef]

19. Zhong, K.X.; Suttle, C.A.; Baudoux, A.C.; Derelle, E.; Colombet, J.; Cho, A.; Caleta, J.; Six, C.; Jacquet, S. A New Freshwater Cyanosiphovirus Harboring Integrase. *Front. Microbiol.* **2018**, *9*, 2204. [CrossRef]

20. Zhang, D.; You, F.; He, Y.; Shu, H.; Gin, Y. Isolation and Characterization of the First Freshwater Cyanophage Infecting Pseudanabaena. *J. Virol.* **2020**, *94*, e00682-20. [CrossRef]

21. Yang, F.; Jin, H.; Wang, X.; Li, Q.; Zhang, J.; Cui, N.; Jiang, Y.; Chen, Y.; Wu, Q.; Zhou, C.; et al. Genomic Analysis of Mic1 Reveals a Novel Freshwater Long-Tailed Cyanophage. *Front. Microbiol.* **2020**, *11*, 484. [CrossRef] [PubMed]

22. Xue, C.; Liu, X.; Wang, Q.; Lin, T.; Wang, M.; Liu, Q.; Shao, H.; Jiang, Y. The isolation and genome sequencing of a novel cyanophage S-H68 from the Bohai Sea, China. *Mar. Genom.* **2020**, *53*, 100739. [CrossRef] [PubMed]

23. Huang, L.; Liu, Q.; Liu, X.; Wang, Q.; Jiang, Y. Isolation and Complete Genome Sequence of a Novel Cyanophage S-B68. *Curr. Microbiol.* **2020**, *77*, 2385–2390. [CrossRef]

24. Sukenik, A.; Kaplan, A. Cyanobacterial Harmful Algal Blooms in Aquatic Ecosystems: A Comprehensive Outlook on Current and Emerging Mitigation and Control Approaches. *Microorganisms* **2021**, *9*, 1472. [CrossRef] [PubMed]

25. Paerl, H.W.; Fulton, R.S.; Moisander, P.H.; Dyble, J. Harmful freshwater algal blooms, with an emphasis on cyanobacteria. *Sci. World J.* **2001**, *1*, 76–113. [CrossRef]

26. Ulius, S.; Mazur-Marzec, H.; Vitonyt, I.; Kvederaviciute, K.; Holmfeldt, K. Insights into cyanophage-mediated dynamics of nodularin and other non-ribosomal peptides in Nodularia spumigena. *Harmful Algae* **2018**, *78*, 69–74.

27. Eigemann, F.; Schwartke, M.; Schulz-Vogt, H. Niche separation of Baltic Sea cyanobacteria during bloom events by species interactions and autecological preferences. *Harmful Algae* **2018**, *72*, 65–73. [CrossRef]

28. Murata, K.; Zhang, Q.; Galaz-Montoya, J.G.; Fu, C.; Coleman, M.L.; Osburne, M.S.; Schmid, M.F.; Sullivan, M.B.; Chisholm, S.W.; Chiu, W. Visualizing Adsorption of Cyanophage P-SSP7 onto Marine Prochlorococcus. *Sci. Rep.* **2017**, *7*, 44176. [CrossRef]

29. Waldbauer, J.R.; Coleman, M.L.; Rizzo, A.I.; Campbell, K.L.; Zhang, L. Nitrogen sourcing during viral infection of marine cyanobacteria. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 15590–15595. [CrossRef]

30. Fridman, S.; Flores-Uribe, J.; Larom, S.; Alalouf, O.; Liran, O.; Yacoby, I.; Béjà, O. A myovirus encoding both photosystem I and II proteins enhances cyclic electron flow in infected Prochlorococcus cells. *Nat. Microbiol.* **2017**, *2*, 1350–1357. [CrossRef]

31. Wei, Y.; Xia, X. Unique Shine-Dalgarno Sequences in Cyanobacteria and Chloroplasts Reveal Evolutionary Differences in Their Translation Initiation. *Genome Biol. Evol.* **2019**, *11*, 3194–3206. [CrossRef] [PubMed]

32. Šulčius, S.; Šimoliūnas, E.; Alzbutas, G.; Gasiūnas, G.; Jauniškis, V.; Kuznecova, J.; Miettinen, S.; Nilsson, E.; Meškys, R.; Roine, E.; et al. Genomic Characterization of Cyanophage vB_AphaS-CL131 Infecting Filamentous Diazotrophic Cyanobacterium Aphanizomenon flos-aquae Reveals Novel Insights into Virus-Bacterium Interactions. *Appl. Environ. Microbiol.* **2019**, *85*, e01311-18.

33. Maccoll, R. Allophycocyanin and energy transfer. *Biochim. Biophys. Acta (BBA)-Bioenerg.* **2004**, *1657*, 73–81. [CrossRef]

34. Jiang, T.; Zhang, J.P.; Chang, W.R.; Liang, D.C. Crystal structure of R-phycocyanin and possible energy transfer pathways in the phycobilisome. *Biophys. J.* **2001**, *81*, 1171–1179. [CrossRef]

35. Thornton, L.E.; Ohkawa, H.; Roose, J.L.; Kashino, Y.; Keren, N.; Pakrasi, H.B. Homologs of plant PsbP and PsbQ proteins are necessary for regulation of photosystem ii activity in the cyanobacterium Synechocystis 6803. *Plant Cell* **2004**, *16*, 2164–2175. [CrossRef]

36. Shi, L.X.; Schröder, W.P. The low molecular mass subunits of the photosynthetic supracomplex, photosystem II. *Biochim. Biophys Acta* **2004**, *1608*, 75–96. [CrossRef]

37. Kamiya, N.; Shen, J.R. Crystal structure of oxygen-evolving photosystem II from Thermosynechococcus vulcanus at 3.7-A resolution. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 98–103. [CrossRef]

38. Raymond, J.; Blankenship, R.E. The evolutionary development of the protein complement of photosystem 2. *Biochim. Biophys. Acta (BBA)-Bioenerg.* **2004**, *1655*, 133–139. [CrossRef]

39. Huang, G.; Xiao, Y.; Pi, X.; Zhao, L.; Zhu, Q.; Wang, W.; Kuang, T.; Han, G.; Sui, S.; Shen, J. Supplementary Information for Structural insights into a dimeric Psb27-photosystem II complex from a cyanobacterium Thermosynechococcus vulcanus.pdf. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2018053118. [CrossRef]

40. Webber, A.N.; Malkin, R. Photosystem I reaction-centre proteins contain leucine zipper motifs. A proposed role in dimer formation. *FEBS Lett.* **1990**, *264*, 1–4. [CrossRef]

41. Panavas, T.; Pikula, A.; Reid, P.D.; Rubinstein, B.; Walker, E.L. Identification of senescence-associated genes from daylily petals. *Plant Mol. Biol.* **1999**, *40*, 237–248. [CrossRef]

42. Watson, H.C.; Walker, N.P.; Shaw, P.J.; Bryant, T.N.; Wendell, P.L.; Fothergill, L.A.; Perkins, R.E.; Conroy, S.C.; Dobson, M.J.; Tuite, M.F. Sequence and structure of yeast phosphoglycerate kinase. *EMBO J.* **1982**, *1*, 1635–1640. [CrossRef]

43. Tarnawski, M.; Krzywda, S.; Bialek, W.; Jaskolski, M.; Szxzepaniak, A. Structure of the RuBisCO chaperone RbcX from the thermophilic cyanobacterium Thermosynechococcus elongatus. *Acta Cryst. Sect. F Struct. Biol. Cryst. Commun.* **2011**, *67*, 851–857. [CrossRef]

44. Battchikova, N.; Zhang, P.; Rudd, S.; Ogawa, T.; Aro, E.M. Identification of NdhL and Ssl1690 (NdhO) in NDH-1L and NDH-1M complexes of Synechocystis sp. PCC 6803. *J. Biol. Chem.* **2005**, *280*, 2587–2595. [CrossRef]

45. Rumeau, D.; Bécuwe-Linka, N.; Beyly, A.; Louwagie, M.; Garin, J.; Peltier, G. New subunits NDH-M, -N, and -O, encoded by nuclear genes, are essential for plastid Ndh complex functioning in higher plants. *Plant Cell* **2005**, *17*, 219–232. [CrossRef]

46. Zhang, P.; Battchikova, N.; Paakkarinen, V.; Katoh, H.; Iwai, M.; Ikeuchi, M.; Pakrasi, H.; Ogawa, T.; Aro, E.M. Isolation, subunit composition and interaction of the NDH-1 complexes from Thermosynechococcus elongatus BP-1. *Biochem. J.* **2005**, *390*, 513–520. [CrossRef]

47. Wilkens, S.; Zhang, Z.; Zheng, Y. A structural model of the vacuolar ATPase from transmission electron microscopy. *Micron* **2005**, *36*, 109–126. [CrossRef]

48. Leyva, J.; Bianchet, M.A.; Amzel, L.M. Understanding ATP synthesis: Structure and mechanism of the F1-ATPase (Review). *Mol. Membr. Biol.* **2003**, *20*, 27–33. [CrossRef]

49. Mattevi, A.; Obmolova, G.; Kalk, K.H.; Westphal, A.H.; Kok, A.D.; Hol, W.G. Refined Crystal Structure of the Catalytic Domain of Dihydrolipoyl Transacetylase (E2p) from Azotobacter vinelandii at 2·6 Å Resolution. *J. Mol. Biol.* **1993**, *230*, 1183–1199. [CrossRef]

50. Maguire, B.A.; Zimmermann, R.A. The Ribosome in Focus. *Cell* **2001**, *104*, 813–816. [CrossRef]

51. Sanyal, S.C.; Liljas, A. The end of the beginning: Structural studies of ribosomal proteins. *Curr. Opin. Struct. Biol.* **2000**, *10*, 633–636. [CrossRef]

52. Chan, Y.L.; Suzuki, K.; Olvera, J.; Wool, I.G. Zinc finger-like motifs in rat ribosomal proteins S27 and S29. *Nucleic Acids Res.* **1993**, *21*, 649–655. [CrossRef]

53. Hertveldt, K.; Lavigne, R.; Pleteneva, E.; Sernova, N.; Kurochkina, L.; Korchevskii, R.; Robben, J.; Mesyanzhinov, V.; Krylov, V.N.; Volckaert, G. Genome comparison of pseudomonas aeruginosa large phages. *J. Mol. Biol.* **2005**, *354*, 536–545. [CrossRef]

54. Ma, Y.; Allen, L.Z.; Palenik, B. Diversity and genome dynamics of marine cyanophages using metagenomic analyses. *Environ. Microbiol. Rep.* **2015**, *6*, 583–594. [CrossRef]

55. Hevroni, G.; Enav, H.; Rohwer, F.; Beja, O. Diversity of viral photosystem-I psaA genes. *ISME J.* **2015**, *9*, 1892–1898. [CrossRef]

56. Lemay, M.L.; Otto, A.; Maa, S.; Plate, K.; Becher, D.; Moineau, S. Investigating Lactococcus lactis MG1363 Response to Phage p2 Infection at the Proteome Level. *Mol. Cell. Proteom.* **2019**, *18*, 704–714. [CrossRef]

57. Gao, E.; Huang, Y.; Ning, D. Metabolic Genes within Cyanophage Genomes: Implications for Diversity and Evolution. *Genes* **2016**, *7*, 80. [CrossRef]