microorganisms

Special Issue Reprint

Bioinformatics and Omic Data Analysis in Microbial Research

Edited by
Douglas D. Rhoads

mdpi.com/journal/microorganisms

MDPI

Bioinformatics and Omic Data Analysis in Microbial Research

Bioinformatics and Omic Data Analysis in Microbial Research

Guest Editor

Douglas D. Rhoads

MDPI

Basel • Beijing • Wuhan • Barcelona • Belgrade • Novi Sad • Cluj • Manchester

Guest Editor
Douglas D. Rhoads
Department of Biological
Sciences
University of Arkansas
Fayetteville, NC
USA

Editorial Office
MDPI AG
Grosspeteranlage 5
4052 Basel, Switzerland

This is a reprint of the Special Issue, published open access by the journal *Microorganisms* (ISSN 2076-2607), freely accessible at: https://www.mdpi.com/si/199202.

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

Lastname, A.A.; Lastname, B.B. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

ISBN 978-3-7258-9255-6 (Hbk)
ISBN 978-3-7258-9256-3 (PDF)
https://doi.org/10.3390/books978-3-7258-9256-3

© 2026 by the authors. Articles in this reprint are Open Access and distributed under the Creative Commons Attribution (CC BY) license. The reprint as a whole is distributed by MDPI under the terms and conditions of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) license (https://creativecommons.org/licenses/by-nc-nd/4.0/).

Contents

About the Editor . . . . . **vii**

L. Caroline House, Amer Hasan, Andi Asnayanti, Adnan A. K. Alrubaye, Jeff Pummill and Douglas Rhoads
Phylogenomic Analyses of Three Distinct Lineages Uniting *Staphylococcus cohnii* and *Staphylococcus urealyticus* from Diverse Hosts
Reprinted from: *Microorganisms* **2024**, *12*, 1549,
https://doi.org/10.3390/microorganisms12081549 . . . . . **1**

Giulia Gatti, Francesca Taddei, Anna Marzucco, Maria Sofia Montanari, Giorgio Dirani, Silvia Zannoli, et al.
Isolation and Genomic Analysis of a Case of *Staphylococcus argenteus* ST2250 Related to Sepsis in Italy
Reprinted from: *Microorganisms* **2024**, *12*, 1485,
https://doi.org/10.3390/microorganisms12071485 . . . . . **17**

Sandeep Kondakala, Sunghyun Yoon, Soumana Daddy Gaoh, Ohgew Kweon, Seong-Jae Kim and Mark E. Hart
Directional and Strain-Specific Interaction Between *Lactobacillus plantarum* and *Staphylococcus aureus*
Reprinted from: *Microorganisms* **2024**, *12*, 2432,
https://doi.org/10.3390/microorganisms12122432 . . . . . **26**

Athish Ram Das, Nisha Pillai, Bindu Nanduri, Michael J. Rothrock Jr. and Mahalingam Ramkumar
Exploring Pathogen Presence Prediction in Pastured Poultry Farms through Transformer-Based Models and Attention Mechanism Explainability
Reprinted from: *Microorganisms* **2024**, *12*, 1274,
https://doi.org/10.3390/microorganisms12071274 . . . . . **43**

Mark Cannon, Gustavo Ferrer, Mari Tesch and Matthew Schipma
Whole-Genome Deep Sequencing of the Healthy Adult Nasal Microbiome
Reprinted from: *Microorganisms* **2024**, *12*, 1407,
https://doi.org/10.3390/microorganisms12071407 . . . . . **61**

Monique A. Felix, Jing Han, Bijay K. Khajanchi, Yasser M. Sanad, Shaohua Zhao and Steven L. Foley
Salmonella enterica serovar Schwarzengrund: Distribution, Virulence, and Antimicrobial Resistance
Reprinted from: *Microorganisms* **2025**, *13*, 92,
https://doi.org/10.3390/microorganisms13010092 . . . . . **75**

Karen E. Beenken, Mara J. Campbell, Stephanie D. Byrum, Rick D. Edmondson, Samuel G. Mackintosh, Alan J. Tackett and Mark S. Smeltzer
Staphylococcus aureus Proteins Implicated in the Reduced Virulence of *sarA* and *sarA/agr* Mutants in Osteomyelitis
Reprinted from: *Microorganisms* **2025**, *13*, 181,
https://doi.org/10.3390/microorganisms13010181 . . . . . **88**

Alonso R. Poma Ticona, Karita C. R. Santos, Heber E. Ramirez-Arua, Roberto Castellanos, Jéssica Pinheiro Silva, Pedro R. Vieira Hamann, et al.
Functional Characterization, Genome Assembly, and Annotation of *Geobacillus* sp. G4 Isolated from a Geothermal Field in Tacna, Peru
Reprinted from: *Microorganisms* **2025**, *13*, 1374,
https://doi.org/10.3390/microorganisms13061374 . . . . . **107**

Pedro Sepúlveda-Rebolledo, Carolina González-Rosales, Mark Dopson, Ernesto Pérez-Rueda, David S. Holmes and Jorge H. Valdés
Comparative Genomics of Sigma Factors in *Acidithiobacillia* Sheds Light into the Transcriptional Regulatory Networks Involved in Biogeochemical Dynamics in Extreme Acidic Environments
Reprinted from: *Microorganisms* **2025**, *13*, 1199,
https://doi.org/10.3390/microorganisms13061199 . . . . . **130**

Yunzhu Chen, Zhuangzhuang Gao, Yan Yang, Qiang Liu, Lijuan Jiang, Jingzhen Chen, et al.
Revealing the Existence of Diverse Strategies for Phosphorus Solubilization and Acquisition in Plant-Growth Promoting *Streptomyces misionensis* SwB1
Reprinted from: *Microorganisms* **2025**, *13*, 378,
https://doi.org/10.3390/microorganisms13020378 . . . . . **147**

About the Editor

Douglas D. Rhoads

Douglas D. Rhoads, University of Arkansas, has been on the faculty of this institution since 1990. His research interests include molecular genetics and genomics of eukaryotes and prokaryotes, including muti-gene traits, and evolution of gene/genome structure. His current research is a collaborative effort focused on applying bioinformatics and molecular genomics to improve broiler production; he is primarily focused on bacterial diseases. Collaborations are also underway on developing genomes for non-model species including snakes, spiders, and scorpions.

microorganisms

MDPI

Article

Phylogenomic Analyses of Three Distinct Lineages Uniting *Staphylococcus cohnii* and *Staphylococcus urealyticus* from Diverse Hosts

L. Caroline House [1,2], Amer Hasan [2,3], Andi Asnayanti [2], Adnan A. K. Alrubaye [2,4], Jeff Pummill [2,5] and Douglas Rhoads [2,*]

1 John Brown University, Siloam Springs, AR 72761, USA; carolinelouisa77@gmail.com
2 Cell and Molecular Biology Pogram, University of Arkansas, Fayetteville, AR 72703, USA; amer.a@covm.uobaghdad.edu.iq (A.H.); aasnayan@uark.edu (A.A.); aakhalaf@uark.edu (A.A.K.A.); jpummil@uark.edu (J.P.)
3 Department of Veterinary Public Health, College of Veterinary Medicine, University of Baghdad, Baghdad P.O. Box 1417, Iraq
4 Department of Poultry Science, University of Arkansas, Fayetteville, AR 72703, USA
5 Arkansas High Performance Computing Center, University of Arkansas, Fayetteville, AR 72703, USA
* Correspondence: drhoads@uark.edu

Abstract: We sequenced and assembled genomes for 17 isolates of *Staphylococcus cohnii* isolated from osteomyelitis lesions in young broilers from two separate experiments where we induced lameness using a hybrid wire-litter flooring system. Whole genome comparisons using three different methods support a close relationship of genomes from both *S. cohnii* and *Staphylococcus urealyticus*. The data support three different lineages, which we designated as Lineage 1, Lineage 2, and Lineage 3, uniting these two species within an evolving complex. We present evidence for horizontal transfer between lineages of genomic regions from 50–440 kbp. The transfer of a 186 kbp region from Lineage 1 to Lineage 2 appears to have generated Lineage 3. Human-associated isolates appear to be limited to Lineages 2 and 3 but Lineage 2 appears to contain a higher number of human pathogenic isolates. The chicken isolates from our lameness trials included genomically diverse isolates from both Lineage 1 and 2, and isolates from both lineages were obtained from osteomyelitis lesions of individual birds. Our results expand the diversity of Staphylococci associated with osteomyelitis in poultry and suggest a high diversity in the microbiome of day-old chicks. Our data also support a reevaluation and unification of the taxonomic classifications of *S. cohnii* and *S. urealyticus*.

Keywords: osteomyelitis; horizontal transfer; genome evolution; *Staphylococcus*; bioinformatics

1. Introduction

The Staphylococcus genus of Gram-positive bacteria comprises species identified as normal flora and opportunistic pathogens, isolated from a wide variety of animals [1–4]. Some Staphylococcus species have been associated with specific hosts. Our research group has been exploring the conditions leading to lameness outbreaks in the broiler industry. Lameness typically affects 1–2% of broilers, with Bacterial Chondronecrosis with Osteomyelitis (BCO) contributing about half of these cases [5–7]. However, flocks can experience lameness outbreaks that spread through a facility, affecting 10–15% of the flock, where the majority of cases are BCO. One of the primary incidents that can contribute to an outbreak is a significant stress event (heat, respiratory infection, etc.) and we have employed growth on a suspended wire-floor as our primary source of stress [6–9]. Historically, *Staphylococcus aureus* has been most commonly associated with BCO and systemic infections in chickens [10–18]. More recently, *Staphylococcus agnetis* has also been found infecting poultry [19–21] Previously, *S. agnetis* had been primarily associated with sub-clinical mastitis in cattle [22–24], but has also now been reported in skin infections

in camelids [25], cattle [26], and poultry [20,27]. We have characterized isolates of *S. agnetis* obtained from femoral and tibial head necrosis lesions from broilers reared on wire-flooring [19]. One particular isolate is highly virulent when administered in a single dose to young broilers, and can spread from infected birds through the air to induce lameness in other birds in the same facility [28]. Whole genome comparisons show that the chicken isolates are closely related to cattle isolates [21]. Pangenome analyses suggest that changes in host and tissue tropism of *S. agnetis* may be driven by horizontal transfer of mobile genomic elements from *S. aureus* [29]. In both *S. aureus* and *S. agnetis*, host preference appears to be facilitated by horizontal gene transfer [16,21,29,30].

Recently, we have developed a new hybrid system for inducing lameness where birds are raised on wire flooring to induce lameness and the infection spreads to other chickens in the same facility raised on standard litter flooring [31]. This system best models the infectious spread of BCO in broiler facilities. Additionally, we do not control the source of the infection, which can derive from the environment or the bird's own microbiome. Over the course of two lameness trials in Fall 2020 and Spring 2021 using the hybrid wire/litter floor model, we recovered 77 distinct cultures from femoral or tibial head necrosis lesions, 35 of which were identified as *Staphylococcus cohnii*. This particular species has been primarily associated with infections in mammals, and more specifically humans [32–34], with very few reported cases of pathogenesis in poultry [35]. *S. cohnii* has been described as a species complex containing three recognizable clades [32,33], including two subspecies (*S. cohnii* subsp. *cohnii* and *S. cohnii* subsp. *urealyticus*). Whole genome analyses have led some to propose splitting the complex into two species [33] Given the paucity of publications, or genome assemblies, for *S. cohnii* infecting chickens, and the high-frequency of isolation in two successive BCO challenge experiments, we chose to examine the genomic diversity of our BCO isolates of *S. cohnii*. Standard microbiological techniques suggested a diverse population, and whole genome comparisons expand the available genomes for elucidating the evolution of this taxonomic group and our understanding of the phylogenetic evolution of pathogenesis and host tropism. Our results also contribute to a better understanding of the etiology of BCO outbreaks in the broiler industry.

2. Materials and Methods

2.1. Media

Media included: CHROMagar Orientation (CO) and CHROMagar *Staphylococcus* (CS) (DRG International, Inc., Springfield, NJ, USA) and tryptic soy broth (TSB) (Difco Laboratories, Franklin Lakes, NJ, USA). Difco bacteriological agar was added at 1.5% for solidification, when necessary.

2.2. S. cohnii Isolates

S. cohnii isolates were part of a collection of isolates obtained from swabs of femoral head necrosis lesions from Experiment 29 (fall 2020) and Experiment 30 (spring 2021). Details of the animal work have been published elsewhere [31]. Briefly, there were two pens of broiler breeder male chicks raised on wire flooring to induce lameness. Separately in the same facility, broiler breeder males were raised on standard litter flooring. Around 31 days of age, lameness developed in the birds on wire-flooring and the infection spreads to the birds on litter flooring. At 56 days of age, birds were euthanized and the femurs aseptically exposed for diagnosis of femoral head necrosis. Sterile swabs were used to sample the lesions then streaked on CO plates. Colony morphology and color were used for initial discrimination. Individual colonies were streak purified on new CO plates and PCR sequenced for the 16S v1-5 regions for species identification [19,36]. Cultures were stored at −80 °C in 40% glycerol 50% TSB, and at 4 °C on TSB agar slants.

PCR-ready DNA was extracted using NaOH on 20 µL aliquots from overnight cultures followed by 4-fold dilution as described [37]. Species and lineage specific qPCR (quantitative PCR) was in triplicates of 20 µL in 1x Taq Buffer (50 mM Tris pH 8.3, 1.25 mM $MgCl_2$, 300 ng/µL BSA), 0.2 mM dNTPs, 0.25 µM primers, 1 X EvaGreen® Dye (Biotium, Fremont,

CA, USA), 2 µL of extracted DNA, and 4 U Taq Polymerase in 96 well plates in a CFX96 Touch Real-Time PCR Detection System (Bio-Rad Laboratories, Inc., Hercules, CA, USA). Cycle parameters were: 90 °C for 45 s, 10 cycles of 90 °C for 15 s, anneal temperature (see below) for 15 s, and 72 °C for 60 s; followed by 30 cycles with the same parameters with a plate read after each cycle. The products were then subjected to high resolution melt (HRM) analysis to verify that any amplification signal was not artifactual, consisting of 72 °C for 180 s, 90 °C for 60 s, 70 °C for 120 s, followed by a 70 °C to 90 °C melt with steps of 0.1 °C for 5 s and plate read. The amplification and melt curves were analyzed using CFX Manager v 3.1 (Bio-Rad Laboratories). Staphylococcal species verification targeted the pfbA gene [38]; for *S. cohnii* the primers were forward: 5′-GCACTCGAATKGCATCTTTAGAC-3′; reverse: 5′-GGGCAGATACACGYGCGATTC-3′; with anneal temperature 60 °C. *S. cohnii* Lineage 1 specific primers (forward: 5′-CTCGTACAACTTAAACCTAGTATAG-3′; reverse: 5′-GTTTAATTTGTGTGTGTAAATCTTTTTCGT-3′) targeted the gene for der, a GTPase that plays an essential role in the late steps of ribosome biogenesis. Lineage 2 specific primers (forward: 5′-CACATATGGCATTAATCGTTACATTGA-3′; reverse: 5′-CTTTACTAATTAAAAGTTGATKAACAAAATGG-3′) targeted the gene for Nat, an arylamine N-acetyltransferase. Anneal temperature for lineage-specific primers was 59 °C.

2.3. Genome Sequence and Assembly

Total DNA was isolated by the method of Dyer and Iandolo [39] as modified [21]. Briefly, overnight 2 mL TSB cultures were diluted into 20 mL of fresh TSB and grown 2 h at 37 °C to log phase before treatment with lysostaphin, followed by standard SDS-lysis, organic extraction, RNase/protease digestion, and ethanol precipitation. The final purification was performed using NanoSep 100 k Omega spin cups (Stock#: OD100C33, Pall Corporation, Port Washington, NY, USA) according to the manufacturer's instructions. Purified total bacterial DNAs were submitted to SeqCenter (Pittsburgh, PA, USA) for 2 × 151 paired-end sequencing on an Illumina NextSeq 2000 (Illumina Corp., San Diego, CA, USA). Sequence data were uploaded to BV-BRC [40] for processing through the Trim Galore pipeline and Unicycler genome assembly. Whole-genome sequence reads and assemblies were submitted to NCBI under BioProject PRJNA1126273, containing BioSamples SAMN41936528 through SAMN41936546.

2.4. Bioinformatics

NCBI genome assemblies were downloaded 10 January 2024, using genome_updater v0.6.3 (https://github.com/pirovc/genome_updater, accessed on 10 January 2024). Phylogenomic trees of genome assemblies were produced based on kmer comparisons using PopPUNK v2.6.0 [41], based on core genome single nucleotide polymorphisms (SNPs) using ParSNP v1.7.4 [42], based on average nucleotide identity using pyANI v 0.2.12 [43], and based on percentage of conserved proteins using POCP-nf v 2.3.2 [44]. Newick trees were midpoint rooted using Archeopteryx v 0.9928 beta [45] and rendered in MicroReact [46]. Proksee was used for genome comparisons, FastANI analyses, and mobile element identification [47] based on the mobnileOG-db [48]. BLAST searches were employed at NCBI or a local installation of version 2.13.0 [49]. Assemblies were annotated using Prokka v1.14.5 [50]. Pan and core gene partitioning was performed using Roary v3.13.0 [51] and Pirate [52]. Scoary v1.6.16 [53] was used for pan-genome-wide association of output from Roary. Maast v1.0.8 [54] was used to generate a vcf containing genotype data for core genome SNPs for each genome, which were then processed further in Microsoft Office 365 Excel to calculate and plot SNP frequencies with respect to isolate traits. SeqKit [55] was used to convert fasta files to and from tab-delimited files, and to extract specific polypeptide sequences from the Prokka annotation files. Microsoft Excel was used to curate tab delimited sequences to filter for evident orthologs based on amino and terminal sequences, and for full-length orthologs. Excel was also used to analyze and summarize BLAST and SNP data. Variant identification in DNA or polypeptide sequences was based on alignment

using Clustal Omega in MegAlignPro (Lasergene v17; DNAStar, Inc.; Madison, WI, USA) with variant tables exported to Excel for filtering, sorting, tabulation and graphing.

3. Results

3.1. Phylogenomic Analyses Identify Three Lineages

The 35 presumptive isolates of *S. cohnii* from Experiments 29 (Fall 2020) and 30 (Spring 2021) had been assigned to species based on PCR sequencing of the 16S V1-V5 region [19,36]. Species was verified by qPCR-HRM (high resolution melt) with primers specific for the *S. cohnii* pfbA gene [38], then each isolate was scored for color on CO and CS chromogenic media. Four different color types, designated A through D, were identified (Table 1). We submitted 18 genomic DNAs representing all four color types for Illumina sequencing. Unicycler assemblies of the trimmed sequence reads ranged from 20–40 contigs, with a consistent size of 2.58–2.70 Mbp (Table S1), and GC percentage ranging from 32.31 to 32.53. These values agree with those for most *S. cohnii* assemblies in NCBI.

Table 1. *Staphylococcus cohnii* isolates from Experiment (Exp) 29 and 30. For each Isolate Number is provided the source Bird identifier, whether whole genome sequence (WGS) was performed, the Lineage as determined from WGS or qPCR-HRM, and Colony Color on CHROMagar Orientaion (CO) or CHROMagar Staphylococcus (CS); l—lightgreen/lime, p—pink, t—turquoise, w—white.

Isolate Number	Exp	Bird	WGS	Lineage	Colony Color	
					CO	CS
1637	29	10A	yes	1	p	l
1649	29	10A	yes	1	t	t
1662	29	10B		1	p	l
1667	29	10B	yes	2	w	l
1670	29	10B		1	t	t
1638	29	11A	yes	1	p	l
1650	29	11A		1	t	t
1651	29	11B		1	t	t
1652	29	12B		1	t	t
1653	29	13A		1	t	t
1654	29	13B		1	t	t
1663	29	13B	yes	1	p	l
1655	29	14A		1	t	t
1656	29	14B		1	t	t
1657	29	2A		1	t	t
1658	29	2A	yes	1	p	l
1659	29	3B	yes	1	p	l
1664	29	3B	yes	2	w	l
1644	29	5B		1	t	t
1645	29	6A	yes	1	t	t
1636	29	6B		1	p	l
1646	29	6B		1	t	t
1661	29	9A	yes	1	p	l
1665	29	9A	yes	2	w	l
1669	29	9A		1	t	t
1639	29	9B		1	p	l
1648	29	9B		1	t	t
1666	29	9B	yes	2	w	l
1697	30	2	yes	2	t	t
1725	30	3	yes	2	t	t
1711	30	4	yes	2	w	l
1701	30	6	yes	1	t	t
1728	30	8	yes	2	p	l
1714	30	10	yes	2	p	l
1715	30	10	yes	2	p	l

We downloaded all 96 assemblies from NCBI for *S. cohnii* and all 16 assemblies for *Staphylococcus urealyticus* (sometimes mis-labeled as *S. ureilyticus*). Lavecchi et al. [33] proposed that *S. cohnii* and *S. urealyticus* actually represent a single species complex. Our initial evaluation of the assemblies showed that 108 of the 112 assemblies from NCBI were of sufficient quality and sequence relatedness. Four assemblies were excluded because they were either misidentified for species, duplicate assemblies, or bad assemblies of mixed species (Table S1). We used four different methods to examine the phylogenomic relationships of the 108 NCBI assemblies together with our 18 new assemblies. Figure 1 shows the PopPUNK tree based on core genome kmer comparisons. Figure S1 is the ParSNP tree based on core genome SNPs. Figure S2 is a neighbor joining tree of the pyANI pair-wise average nucleotide identity (ANI) distances. Figure S3 is a neighbor-joining tree based on distances calculated from Percentage of Conserved Proteins (POCP). PopPUNK, ParSNP, and pyANI produce very similar trees showing three distinct lineages. These results are similar to the findings of Lavecchia et al. based on core genome multilocus alignments for 65 assemblies. Our Lineage 1, 2, and 3 correspond to their clades B, A2, and A1, respectively. The *S. urealyticus* assemblies (circles) are restricted to Lineage 2. BLASTn alignments of the 16S V1-V5 regions show that all 124 assemblies have the same 16S rDNA genes with a few exceptions for polymorphisms, likely attributable to assembly artifacts. Five isolates, DE0361, DE0360, SE4-2, SE4-1, and SE3-10, consistently appear intermediate between Lineage 3 and the other two lineages in all three trees. Isolate hu-01 appears intermediate only in the ParSNP and pyANI trees. Excluding the five intermediate isolates, the average ANI distance from Lineage 3 to either Lineage 1 or 2 is less than the distance from Lineage 1 to Lineage 2 (Table 2). The genetic distance between lineages is 8 to 10 times the genetic distance within each lineage. Thus, Lineages 1 and 2 are the most different with Lineage 3 in-between. The five intermediate isolates (m) are closer to Lineage 3 and slightly closer to Lineage 2 than Lineage 1.

Table 2. Average Nucleotide Identity (±SD) between the isolates of the three lineages plus the intermediate (m) isolates.

Lineage	n	1	2	3	m
1	56	0.99012 ± 0.00771	0.92023 ± 0.00566	0.92526 ± 0.00333	0.94591 ± 0.00196
2	40		0.99290 ± 0.00591	0.96082 ± 0.00171	0.94935 ± 0.00025
3	23			0.99558 ± 0.00368	0.97667 ± 0.00216
m	5				0.99996 ± 0.00004

POCP is a generally accepted method for genus and species delineation [56]. The POCP based tree in Figure S3 does not show three lineages and the topology is quite distinct from the three other methods. Particularly striking is the placement on long branches of *S. urealyticus* 3636 and *S. cohnii* SNUC-4556. The assembly for isolate 3636 comprises 2.5 Mb in 1291 contigs with an N50 of only 3.6 kbp. Thus, the overall small contig size likely results in fewer and shorter predicted polypeptides, and thus a much lower POCP value. The assembly for isolate SNUC_4556 consists of 1461 contigs spanning 3.2 Mb with a N50 of 8.4 kb. The genome length suggests contamination and the large number of short contigs would also compromise the POCP value. The application of POCP for comparisons of a large number of draft assemblies appears to explain the radically different phylogeny in Figure S3 and argues for only applying POCP to complete genome assemblies. Table 3 presents POCP and ANI values comparing the nine complete or chromosome-level assemblies (four each from lineages 1 and 3, and one from Lineage 2) along with four complete genomes for *Staphylococcus nepalensis* as the closest relative to *S. cohnii/S. urealyticus* [57]. POCP values within *S. nepalensis* or within lineages range from 90–95 and between lineages from 88–89, while each lineage is more distant at 84–85 compared to *S. nepalensis*. ANI values present the same picture of each lineage being approximately the same distance from *S. nepalensis* with less distance between lineages.

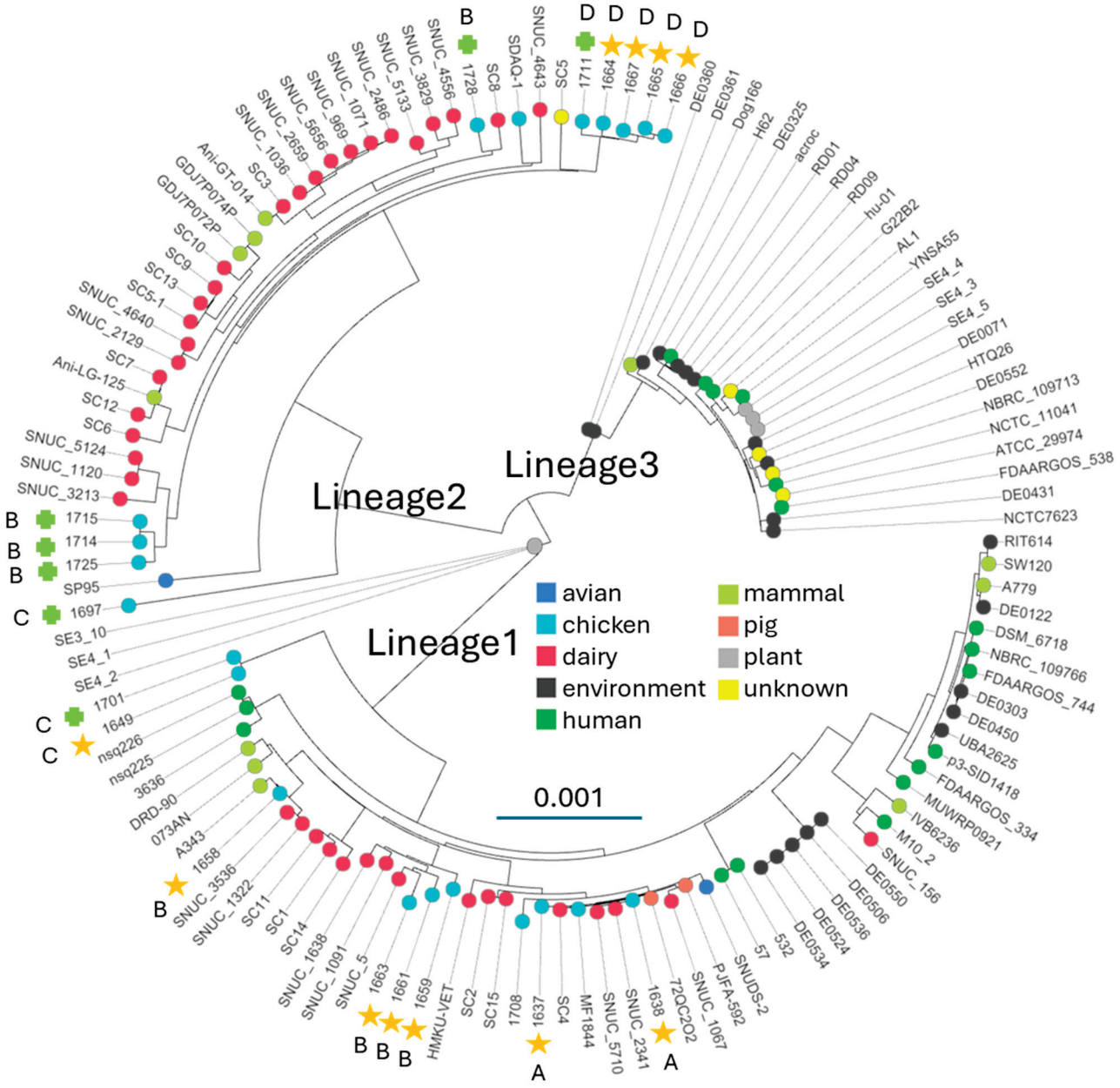


Figure 1. PopPUNK phylogenomic tree of genomes from 126 isolates of *Staphylococcus cohnii* (squares) and *Staphylococcus urealyticus* (*circles*). Strain names are the leaves, with nodes colored according to host (legend) as determined from NCBI Biosample entries. Chicken isolates from Experiment 29 (gold star) and Experiment 30 (green plus) are indicated along with color group (A, B, C, D) based on growth on CHROMagar media (Table 1). The branches for the three lineages described in the text are indicated.

FastANI comparisons of the nine complete genomes along with those from *S. nepalensis* indicate that the genomes in all three lineages and both species are largely syntenic (Figure S4). However, the average genome size is larger for *S. nepalensis* at 2.9 Mbp versus *S. cohnii*/*S. urealyticus* at 2.5 Mbp. The analyses are complicated by not all assemblies starting at the same point or even in the same orientation. For example, the FDAARGOS_538 genome was best compared as a reverse complement, while the FDAARGOS_334 genome is circularly permuted relative to the representatives from lineages 2 and 3 and *S. nepalensis*.

Table 3. Percentage of Conserved Proteins (POCP) and Average Nucleotide Identity (ANI) values (±SD) for complete genomes for *Staphylococcus nepalensis* (Sn) and the three lineages of *Staphylococcus cohnii* (1, 2, 3) discussed in the text. Values for Lineage 2 compared to Lineage 2 are not applicable (NA) as there is only one complete genome for that lineage.

		Sn	**1**	**2**	**3**
POCP	n	4	4	1	4
	Sn	93.9 ± 1.8	84.2 ± 1.4	85.4 ± 1.6	84.0 ± 0.7
	1		90.4 ± 2.3	88.5 ± 1.5	89.2 ± 1.6
	2			NA	88.2 ± 0.6
	3				95.3 ± 2.0
ANI	Sn	0.997 ± 0.0015	0.857 ± 0.0007	0.879 ± 0.0459	0.859 ± 0.0012
	1		0.990 ± 0.0036	0.919 ± 0.0003	0.925 ± 0.0015
	2			NA	0.961 ± 0.0004
	3				0.997 ± 0.0017

3.2. Lineage Association with Host and Virulence

Examination of the hosts and disease status of the isolates in each lineage suggest that Lineage 3 may be largely non-pathogenic. The lineage contains a number of environmental (black) and plant (gray) isolates. Unfortunately, many of the isolates from mammals and humans have NCBI biosample data that are largely inconclusive for disease status. Human isolates (green) appear to be restricted to Lineage 1 and 3. Summaries of the biosample data for the human-associated isolates in Lineage 3 include:

- YNSA55: China, human, survey of S. aureus, no disease state recorded
- NCTC11041: Homo sapiens skin, no disease
- FDAARGOS-538: Homo sapiens, clinical isolate
- H62: air, environmental
- Acroc: human skin, no disease state recorded
- RD01, RD04, RD09: Bacteria were isolated from multiple surfaces in washrooms on a university campus in Alberta, Canada.
- Hu-01: China Homo sapiens, skin swab, no disease
- G22B2: Homo sapiens, gall bladder

G22B2 and FDAARGOS-538 may be the only Lineage 3 isolates from diseased samples. The human-associated isolates in Lineage 1 include:

- 57: Homo sapiens, blood; opportunistic pathogen isolated from human patient
- 532: Homo sapiens, catheter; opportunistic pathogen isolated from human patient
- MUWRP0921: human UTI
- FDAARGOS_334: Homo sapiens, clinical isolate
- P3-SID1418: human skin, toe web space
- FDAARGOS_744: Homo sapiens, clinical isolate
- NBRC_109766: unknown
- RIT614: smartphone
- Nsq225: human, linezolid-resistant clinical isolate
- Nsq226: human, linezolid-resistant clinical isolate
- 3636: human oral infection

Therefore Lineage 1 appears to be more clearly associated with human pathogenesis. None of the 40 isolates in Lineage 2 are human isolates. The dairy isolates (red) are found only in Lineage 1 and 2, as are our new chicken isolates (light blue). The other mammalian isolates (light green; not human and not dairy) can be found in Lineages 2 and 3, with the single mammal isolate in Lineage 3 isolated from a dog ear canal, and therefore perhaps not a pathogen. This is all consistent with pathogenesis being prevalent in Lineage 1 and 2 and not in Lineage 3. The five or six intermediate isolates are also not associated with a disease state in the biosample entries.

3.3. Horizontal Transfer between Lineages

Overall frequencies of core genome SNPs for all genomes in the three lineages were determined using Maast with FDAARGOS_538, a complete assembly from Lineage 3, as the reference. SNPs were categorized as intergenic (n = 39,360), non-wobble (n = 174,466), or wobble (n = 23,029) based on the NCBI annotation for the FDAARGOS_538 genome. SNP frequencies for each of the three codon categories were plotted for all three lineages based on chromosomal position in FDAARGOS_538. Figure 2 shows the plots for SNP frequency differences for non-wobble SNPs for pair-wise comparisons of the three lineages (panel A) and for each of the lineages compared to the five intermediate isolates (designated m in Table S1). Plots for the other two SNP categories show very similar patterns. SNP frequency differences were averaged over a 40 SNP sliding window to smooth the graphs. Gaps in the plots represent mobile elements (transposons, prophage, chromosomal islands) and tRNA/rDNA clusters. The frequency difference plots shows that Lineage 1 genomes have a high SNP frequency difference from both Lineage 2 and Lineage 3 throughout the chromosome except for a 186 kbp region from 1977 to 2162 (region 4 in panel B) spanning 184 protein encoding genes. For that region, Lineage 1 and Lineage 3 have similar SNP frequencies and a high SNP frequency difference from Lineage 2. Thus, Lineage 3 appears to have arisen from Lineage 2 by the replacement of this chromosomal region with the orthologous region from Lineage 1. KEGG analysis of EC numbers and inspection of the annotations did not identify a particular pathway or virulence function for these 184 genes. Figure 2 panel B shows the intermediate isolates also contain region 4 from Lineage 1 but also appear to have acquired region 2 (482 to 837 kbp) from Lineage 1. There are also region 1 (42-to 482 kbp) and region 3 (837 to 866 kbp) that appear to be moderately divergent from all of Lineage 1, 2, and 3. BLASTn queries of the NCBI microbial genomes database (i.e., 19 May 2024 query of draft and complete genomes in Staphylococcaceae excluding both *S. cohnii* and *S. urealyticus/ureilyticus*) with subsequences from regions 1 and 3 did not identify any potential source of this region in other Staphylococcus species.

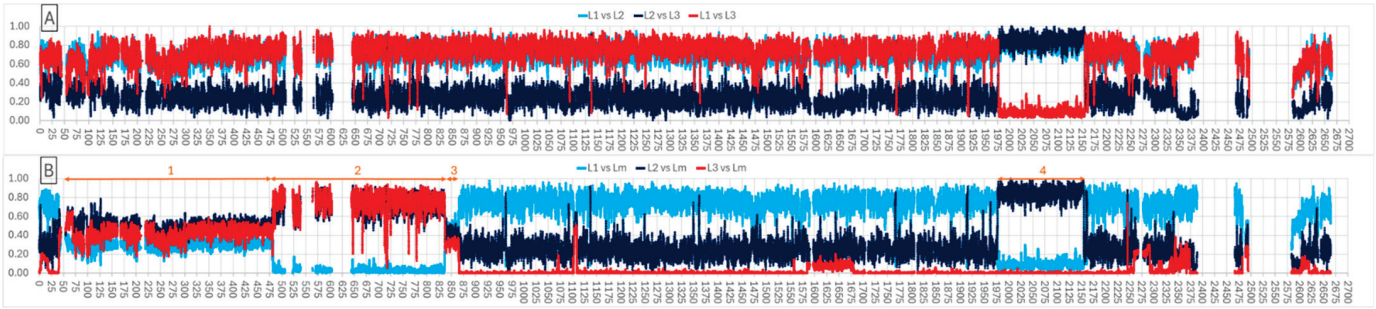


Figure 2. Core genome SNP frequency plots comparing frequencies of non-wobble SNPs in three lineages of *Staphylococcus cohnii* and *Staphylococcus urealyticus*. FDAARGOS_538 was the reference for SNP annotation and chromosomal position (X axis in kbp). Panel A shows the difference in SNP frequencies between Lineage 1 and Lineage 2 (L1 vs. L2; light blue), Lineage 2 and Lineage 3 (L2 vs. L3; dark blue), and Lineage 1 and Lineage 3 (L1 vs. L3; red). Panel B plots the non-wobble SNP frequency differences for the three Lineages relative to the five intermediate isolates (Lm) categorized as m in Table S1. Lineage 1 and m (L1 vs. Lm; light blue), Lineage 2 and m (L2 vs. Lm; dark blue), and Lineage 3 and m (L3 vs. Lm; red). Regions 1 through 4 discussed in the text are delimited in Panel B.

3.4. Lineage Specific Polypeptides

Lavecchia et al. [33] previously identified 24 "clade-specific" genes based on 66 genomes of *S. cohnii* and *S. urealyticus.* With an expanded set of 126 genomes, we readdressed this question. The Prokka generated predicted polypeptides (.faa) files for all 126 genomes were catenated and used to make a blast protein database, which was analyzed by BLASTp queries with the .faa file for reference genome from each lineage: Lineages 1 FDAAR-

GOS_334, Lineage 2 SDAQ-1, and Lineage 3 FDAARGOS_538. BLASTp results were filtered for pident*qcov $\geq$ 8100 (i.e., average of pident and qcov $\geq$ 90), then evaluated for hits for all of the isolates in each lineage. This identified 36 polypeptides present in all genomes from Lineage 1 and missing in one, or both, of the other lineages (Table S2). Seven, possibly eight, of the polypeptides missing from Lineage 3 but present in all Lineage 2 genomes are clustered from PEG (protein encoded gene) 540 to 570, including genes for Nitrate metabolism, nitrogen/oxygen sensors, production of uroporphyrinogen, and a helix-turn-helix regulatory protein.

There are six polypeptides that are unique to a single lineage. In Lineage 1, there are three polypeptides not found in Lineage 2 or 3, a 321 residue RNA binding protein, and a 38 residue hypothetical protein, although BLASTp scoring for a 38 residue oligo peptide can be questionable. There is only one polypeptide uniquely present in Lineage 2; a 67 residue hypothetical protein. Lineage 3 has three unique polypeptides; a 67 residue DUF1146 domain containing protein, a 221 residue metallo-beta-lactamase domain-containing protein, and a 42 residue IS family transposase. Inspection of the annotations reveals no obvious virulence determinants associated with the radiation of Lineages 1 and 2 from Lineage 3. Where Lavecchia et al. identified 24 clade-specific polypeptides, we identified 36 lineage-specific genes. Reciprocal BLASTp analyses of their 24 vs. the 37 from this work identify only three possible overlaps. They identified a 50 residue hypothetical protein in clade A1 (equivalent to our Lineage 3) that corresponds with a 38 residue IS1182 transposase in our data. They identified a 126 residue LytTR family transcriptional regulator in clade A2 (our Lineage 2) that is related (pident = 32; qcov = 62) to a 152 residue LytTR family transcriptional regulator that we found present in all of Lineage 1, some of Lineage 2, and none of Lineage 3. They identified a 281 residue DUF4352 domain-containing protein in clade B that is highly related (pident = 99; qcov 100) to a 321 residue RNA binding protein uniquely present in Lineage 1 (equivalent to their clade B). Although both groups used BLASTp to identify clade/lineage-specific polypeptides, the results are not concordant and may relate to annotations based on incomplete assemblies and the number of assemblies utilized in each analysis.

The Prokka annotations were processed further through Roary to identify core and accessory genomes. The Roary output was then analyzed using Scoary to identify accessory genes positively or negatively associated with specific traits. Traits included host (i.e., human, mammal, animal, chicken) and lineage. However, there is always concern with assigning host from NCBI Biosample data. For example, isolates from washroom sinks, or from human skin swabs, could be from any source (i.e., plants, soil) or they could be valid human associated strains. The Roary-Scoary output identified no host-specific accessory genes (100% for both Specificity and Sensitivity). Roary identified 1017 core genes (99–100% of strains) with an additional 216 soft core genes ($\geq$95%). Shell genes ($\geq$15% and <95%) numbered 23,274, with 10,269 cloud genes (>0% and <15%). Conversely, Pirate uses a variable percentage identity threshold to partition clusters into pan and core genome with 2076 genes in $\geq$95% of isolates; 446 in $\geq$25% and <95%; with 4791 < 25%.

3.5. Lineage Diagnostic qPCR

The Roary-Scoary pipeline did identify lineage specific genes (Table 4). Lineage 1 appeared to have the highest number of lineage-specific genes with 84, even though there were more genomes (Table 2) included from that lineage. Further inspection of these lineage-specific genes reveals that most actually have orthologs in the other lineages where the orthologs differ by enough SNPs and indels that Roary identifies them as different genes when run with the default settings. Based on the Scoary output, we selected the Arylamine-N-acetyltransferase (nat) and GTPase (der) for design of lineage specific PCR primers. The primers targeted multiple SNP positions specific to the ortholog of a particular lineage. DerF1 + R1 was specific for Lineage 1, NatF1 + R1 specific for Lineage 3, and NatF1 + R1 specific for Lineage 2. These primers were employed in qPCR-HRM to diagnosis the remaining 19 *S. cohnii* isolates from Experiment 29 and 30 for which we did not have

genome assemblies (Table 1). The qPCR-HRM results were clear for 18 of the 19 additional isolates but isolate 1645 was positive for DerF1 + R1 and NatF2 + R2, suggesting it was both Lineage 1 and Lineage 2. Therefore, we sequenced and assembled the genome for 1645 and based on PopPUNK analysis 1645 clusters with 1701 and 1649, which appear as an early branch of Lineage 1 (Figure 1). The Prokka annotation, as well as the Roary-Scoary analyses, identify 1649 and 1701 as containing both the Lineage 1 versions of der and nat. ANI computed distances for 1645 to 1701 is 0.9893; 1645 to 1649 is 0.9890; and 1649 to 1701 is 0.9990. The highest value for any of these three to the rest of Lineage 1 is less than 0.97. Prokka annotated contigs containing the Nat gene from 1645 (contig3 263 kbp), 1649 (contig2 427 kbp), and 1701 (contig3 340 kbp) were aligned in MegAlignPro along with a sequence variation plot (Figure S5). The contigs differ for annotation and sequence up to the RlmH gene near 23.9 kbp in contig3 from 1645. The three contigs then appear syntenic through the end of all three contigs (approximately 239 kbp); however, for the last approximately 50 kbp, the 1645 contig has an abundant number of SNPs and indels relative to the contigs from 1649 and 1701. The sequence variation begins two genes upstream of the nat gene. The alignments and sequence variation are most consistent with 1645 having acquired an approximately 50 kbp region, including the nat gene, from a Lineage 2 isolate. Whether this was by transformation, conjugation, or transduction is not known but is additional evidence of horizontal transfer between lineages. Further investigation would necessitate complete chromosome level assemblies of these three isolates for more extensive comparisons.

Table 4. Scoary output for genes differentially represented in *Staphylococcus cohnii* genomes corresponding to particular traits. Gene counts are for those showing statistical differences for the phenotype (Bonferroni corrected $p < 0.05$), for those with positive association (Odds Ratio > 4.4), negative association (Odds Ratio < 0.24), and those with 100% Specificity (presence in all isolates with the trait), and Sensitivity (present only in isolates with the trait).

Trait	Bonferroni $p < 0.05$	Gene Count Odds Ratio		100% Specificity and Sensitivity
		>4.4	<0.24	
Lineage 1	1756	857	899	84
Lineage 2	1716	867	849	14
Lineage 3	1311	707	604	13
Human	333	195	138	0
Mammal	7	0	7	0
Avian	7	7	0	0
Chicken	17	17	0	0
Animal	166	37	129	0

With the diagnosis of lineage for all 35 *S. cohnii* isolates, we determined that colony color on CO and CS media was not always a good indicator of lineage (Table 1). Although pink on CO and turquoise on CS, which was exclusively Lineage 1 (7 of 7), and white on CO and light green on CS was exclusively Lineage 2 (5 of 5). For pink on CO and light green on CS, two were Lineage 1 and three were Lineage 2. For turquoise on both CO and CS, sixteen isolates were Lineage 1 and two were Lineage 2.

3.6. S. cohnii Plasmids

There are sixteen plasmids annotated in the five complete assemblies (five in FDAARGOS_334, one in FDAARGOS_538, six in FDAARGOS_744, two in SDAQ-1, and two in SNUDS2). These were used as BLASTn queries to screen for highly related sequences in all the assemblies. The work was primarily focused on plasmid association with particular lineages or hosts (Table S3). Our subjective measure of prevalence was 50% or greater for the maximum BLASTn bitscore for the plasmid with itself. The three largest (29–37 kbp) plasmids in FDAARGOS_334 are partially related to each other, as well as to the 34 kbp

plasmid in FDAARGOS_538, the four largest (36–53 kbp) plasmids in FDAARGOS_744, plasmid 2 in SDAQ-1, and plasmid 1 in SNUDS-2. However, Clustal Omega alignments of these eight plasmids show that the relationships are piecemeal and not shared across all the plasmids (Figure S6). Related sequences based on BLASTn bitscores can be found in other assemblies but are limited to 19 or fewer genomes depending on the plasmid used in the query. There were no host-specific or lineage-specific-associations for any of the 16 plasmids.

3.7. Genetic Diversity in BCO Infections

Both Experiment 29 and Experiment 30 yielded a high degree of genetic diversity in *S. cohnii* genomes, as both lineages were isolated from each experiment (Figure 1; Experiment 29 gold star and Experiment 30 green plus). In Experiment 29, there were four closely related Lineage 2 isolates and six Lineage 1 isolates from four different radiations. For Experiment 30, there were isolates closely related to some of the diverse isolates in Experiment 29 (Figure 1, see 1711 and 1701) but wholly new genotypes were also present in Experiment 30 (e.g., 1697, 1714, and 1715). When we assigned lineage using qPCR-HRM for the rest of the isolates, we found Lineage 1 predominated in Experiment 29 (24 of 28 isolates) while Lineage 2 predominated in Experiment 30 (6 of 7 isolates). The genetic diversity observed within experiments also occurs within individual birds. Isolates 1659 and 1664 were from bird 3B in Experiment 29 and map to Lineage 1 and 2, respectively. Similarly, 1661 and 1665 are from bird 9A in Experiment 29 and map to Lineage 1 and 2, respectively. Further, Bird 10A in Experiment 29 yielded isolates 1637 and 1649, which are both Lineage 1 but in quite distinct radiations within that Lineage (Figure 1).

4. Discussion

Classic bacteriological methods (e.g., colony morphology, enzyme profiles, sugar fermentation, fatty-acid profiles, DNA-DNA hybridization) led others to propose that *S. cohnii* from humans and primates consisted of two distinguishable sub species, *S. cohnii* subsp. *cohnii* and *S. cohnii* subsp. *urealyticus* (see [32]). More recently, based on alignments of 1468 core genes from 66 genomes, others suggested that the *S. cohnii* complex (*S. cohnii*, *S. cohnii* subsp. *cohnii*, and *S. cohnii* subsp. *urealyticus*) comprises at least three different clades [33]. Their clades, A1, A2, and B, correspond to our lineages, 3, 2, and 1, respectively. The tree structure produced by Lavecchia et al. suggested a closer alliance between A1 and A2, which is in agreement with our ANI data showing Lineage 2 and Lineage 3 are closer to each other than either is to Lineage 1 (Table 2). Using three different methods for whole genome comparison, we derived the same basic topology of three lineages with a few intermediate isolates (Figures 1, S1 and S2). Typically, species delineation uses a threshold of genomic relatedness of 95–96% [43,58,59]. Table 2 shows average ANI values greater than 0.96 between Lineage 2 and 3, and just below 0.93 from both Lineage 2 and 3 to Lineage 1. Thus, it would appear that Lineage 1 might be considered as one species and Lineage 2 and 3 a different, single species. However, there are intermediate isolates we grouped into "m", which have ANI values when compared to Lineages 1, 2, and 3 of 0.94, 0.95, and 0.98, respectively. Therefore, drawing strict species boundaries would be problematic. Further, as we found for isolate 1645, genomic exchange appears to be occurring between Lineage 1 and 2, and even Lineage 3 appears to have possibly arisen by transfer of a 186 kbp region from a Lineage 1 genome to replace the orthologous region in a Lineage 2 genome (Figure 2). Lineage m isolates also appear to be distinct from both Lineage 1, 2, and 3 for additional regions in the genome. Therefore, a single species complex may be more appropriate. With further investigation and production of additional complete genome assemblies, it may be possible to identify additional genomic transfers between the three lineages. The intermediate isolates we classified as "m" might be particularly appropriate genomes to examine. Transferred regions indicated from our data include: (a) at least 50 kbp acquired from Lineage 2 in isolate 1645, (b) 186 kbp acquired by a Lineage 2 genome from Lineage 1 to originate Lineage 3, (c) the 440 kbp region 1 and 29 kbp region 3 (Figure 2B)

acquired by a Lineage 3 genome from an unknown source in the intermediate m genomes, and (d) 355 kbp region 2 (Figure 2B) identified in the intermediate m genomes acquired from a Lineage 1 genome. Given the large sizes, these regions are most likely acquired through conjugative transfer, rather than transduction or direct uptake of environmental DNA. Prophage mobilization of distant chromosomal elements has been documented in *Staphylococcus aureus* [60] but the transfers are limited to phage head capacity. Given that pathogenesis of staphylococci is often associated with biofilm formation we speculate that these horizontal transfers may be promulgated within biofilms.

Given the evidence of genetic exchanges between the lineages and the >92% relatedness measurements, our data is most supportive of designating all isolates of *S. urealyticus* and *S. cohnii* as members of a single species complex, similar to the complex comprising *Escherichia coli* and Shigella. Whole genome-based examination of *E. coli*/*Shigella* complex identifies fourteen phylogroups with four Shigella species clustered within *E. coli* [61,62]. Further pan-genome analyses of *E. coli* genomes suggested that major differences were associated with small mobile genome islands [63] or small sets of pan-genes [64]. There was no evidence for transfer of large chromosomal regions as we have detected in *S. cohnii* isolate 1645, Lineage 3, and the intermediate m genomes.

In our BCO induction experiments, source birds are raised on wire-flooring to stress the birds inducing lameness [7,8,28,31,36,65]. The infection spreads through the air to target birds raised on standard litter flooring, inducing BCO-lameness in 40–50% of the target birds. The necrotic lesions in the proximal femoral and tibial heads yield multiple different species even within a single bird. The original source of these bacteria could be the environment (the specific barn, farm personnel, insects, etc.). However, the more likely source is the inherent microbiome of the birds on wire-flooring. For both experiment 29 and 30, the same facility was used with no intervening flocks and the facility was cleaned out and thoroughly disinfected (water lines, floors and equipment) with dilute bleach between experiments. The source of the one-day-old chicks for both experiments was a commercial hatchery but the eggs were likely from different breeder flocks. The range of bacterial species identified from the BCO lesions differed between Experiment 29 and Experiment 30, but in both experiments the most numerous species was *S. cohnii*. Subsequent experiments performed in the same facility have detected far fewer isolates of *S. cohnii*. Since all the experiments were in the same research facility and sourced from the same hatchery, the breadth and genomic diversity of the bacteria present in the BCO lesions appears to vary with the batch of chicks and likely the particular breeder flock producing the eggs. Thus, our data from the genomic diversity of *S. cohnii* isolates suggests that even for 1-day old chicks they must have a rather complex and diverse microbiome.

We have developed inexpensive and rapid methods for environmental sampling for species identification and quantification [37]. It will be important to determine the levels of different Staphylococci in the air and feces during our BCO challenge experiments to understand the dynamics of the spread of pathogens (e.g., *S. aureus*, *S. cohnii*, *E. coli*, *Salmonella* sp.). Further examination must also be made of the microbiome of chicks at hatch to determine the contributions of the parent flock vs. the hatchery environment.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/microorganisms12081549/s1, Figure S1. ParSNP phylogenomic tree of 126 genomes of isolates of *Staphylococcus cohnii* (squares) and *Staphylococcus urealyticus* (*circles*) with FDAARGOS_538 as reference, based on 82,870 core genome SNPs. Strain names are the leaves, with nodes colored according to host as determined from NCBI Biosample entries. The branches for the three lineages discussed in the text are indicated. Figure S2. pyANI phylogenomic tree of 126 genomes of isolates of *Staphylococcus cohnii* (squares) and *Staphylococcus urealyticus* (*circles*). Strain names are the leaves, with nodes colored according to host as determined from NCBI Biosample entries. The branches for the three lineages discussed in the text are indicated. Figure S3. Percentage of Conserved Proteins phylogenomic tree of 126 genomes of isolates of *Staphylococcus cohnii* (squares) and *Staphylococcus urealyticus* (*circles*) isolates. Strain names are the leaves, with nodes colored according to host as determined from NCBI Biosample entries. Figure S4. FastANI alignments from

ProkSee server of representative complete genomes of *Staphylococcus nepalensis* and the three lineages of *Staphylococcus cohnii*. Genomes are L1: Lineage 1, FDAARGOS_334; L2: Lineage 2, SDAQ-1; L3: Lineage 3, FDAARGOS_538 reverse complement; and Sn: *S. nepalensis*, JS11. For the top panel the central genome is Lineage 1 compared to Lineage 3 above and Lineage 2 below. The other panels are each lineage compared to *S. nepalensis*. Approximate genomic distance is indicated at the bottom along with the heat map for the connecting lines. Figure S5. Graphical alignments of contigs containing Nat gene from three closely related *Staphylococcus cohnii* isolates from chickens. The upper panel is the graphical view of the Mauve alignment in MegAlignPro for the indicated contigs from isolates 1645, 1649, and 1701. For each contig sequence blocks are represented by colors. Below each contig graphic are the sequence ruler showing position in the contig, with the prokka annotation indicated by directional arrows. The bottom panel is an aligned histogram of SNP counts per 1000 bp in isolate 1645 relative to 1649 and 1701, based on the variants report for the orange sequence block region from MegAlignPro. The histogram shows significant variation for the first app. 23 kbp and the last app. 50 kbp of the aligned orange sequence block. Figure S6. Clustal Omega alignments between *Staphylococcus cohnii* plasmids identified as sharing sequence identity in BLAST searches. Alignments were graphically depicted in MegAlignPro. Table S1. Genome assemblies for Species Staphylococcus cohnii (SC), Staphylococcus urealyticus (SU), and Staphylococcus nepalensis (Sn) used for these analyses. Newick Strain is the strain designation compatible with Newick tree formatting. BioSample, BioProject, Assembly, Level, Size, Contig count, Host/Source are from NCBI or this study. Lineage assignments are provided based on three software packages (PopPUNK, ParSNP, or pyANI) or from Lavecchia et al. [33]). Notes indicates footnotes for those assemblies excluded from the analyses. Table S2. Polypeptides missing in one or more Staphylococcus cohnii lineages. BLASTp was used to search a database of all predicted Prokka polypeptides from the genome annotations. The database was queried with polypeptides sequences for FDAARGOS_334, SDAQ-1 and FDAARGOS_538, as reference genomes for Lineage 1, 2, and 3, respectively. Minimum hit score was Pident*Qcov $\geq$ 8100. For each Lineage as query the Annotation and Source are listed from Prokka (P) unless Prokka predicted a hypothetical protein in which case the annotation from UniProt.org (U) is listed. PEG in Reference is the protein encoded gene number, Length of the polypeptide, and EC number, are for the reference genome annotation. Best BLASTp is the accession from searches of NR at NCBI. Missing From indicates the lineages the polypeptide is Not In, the number of genomes the polypeptide is missing from in each of the three Lineages (L1, L2, L3) with the total count of genomes in parentheses. Homolog in Lavecchia lists the NCBI accession for homologs from Lavecchia et al. [33]. Table S3. Prevalence of Staphylococcus cohnii plasmids in 126 genome assemblies. Plasmids are numbered and size in kbp listed from each of five complete genomes. Number of assemblies with >50% maximum BLASTn bitscore lists the counts from each of the three Lineages (L1, L2 or L3 and Total) for assemblies producing BLASTn hits totalling more than 50% of the maximum bitscore (i.e., the plasmid with itself).

Author Contributions: D.R. wrote the original draft, all authors reviewed and edited the final draft, D.R. was responsible for the final draft. A.H., A.A. and A.A.K.A. performed all animal work and the initial identification of the isolates. L.C.H., A.A. and D.R. performed the genome sequencing. L.C.H., J.P. and D.R. were responsible for the bioinformatics. D.R. conceived and directed the research. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by funding to D.R. and A.A.K.A. from the Arkansas Biosciences Institute, a component of the Arkansas Tobacco Settlement. L.C.H. was supported by the Arkansas INBRE program, with a grant from the National Institute of General Medical Sciences, (NIGMS), P20 GM103429 from the National Institutes of Health. The funders had no role in study design, data collection, analysis, decision to publish, or preparation of the manuscript.

Data Availability Statement: All genomes analyzed in this research are available from the National Center for Biotechnology Information, National Library of Medicine (https://www.ncbi.nlm.nih.gov).

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

Research Assurance: All animal work received prior approval from the University of Arkansas Institutional Animal Care and Use Committee (#21066, #21085 and #21085-1).

References

1. Becker, K.; Heilmann, C.; Peters, G. Coagulase-negative Staphylococci. *Clin. Microbiol. Rev.* **2014**, *27*, 870. [CrossRef] [PubMed]
2. Cheung, G.Y.C.; Otto, M. Virulence Mechanisms of Staphylococcal Animal Pathogens. *Int. J. Mol. Sci.* **2023**, *24*, 14587. [CrossRef] [PubMed]
3. Cox, H.U.; Newman, S.S.; Roy, A.F.; Hoskins, J.D. Species of Staphylococcus isolated from animal infections. *Cornell Vet.* **1984**, *74*, 124–135. [PubMed]
4. Mathema, B.; Mediavilla, J.; Chen, L.; Kreiswirth, B. Evolution and taxonomy of staphylococci. In *Staphylococci in Human Disease*, 2nd ed.; Crossley, K., Jefferson, K., Archer, G., Folwer, V., Eds.; John Wiley & Sons: Oxford, UK, 2009; pp. 31–64.
5. Siegel, P.B.; Barger, K.; Siewerdt, F. Limb Health in Broiler Breeding: History Using Genetics to Improve Welfare. *J. Appl. Poult. Res.* **2019**, *28*, 785–790. [CrossRef]
6. Wideman, R.F.; Prisby, R.D. Bone circulatory disturbances in the development of spontaneous bacterial chondronecrosis with osteomyelitis: A translational model for the pathogenesis of femoral head necrosis. *Front. Endocrinol.* **2013**, *3*, 183. [CrossRef]
7. Wideman, R.F. Bacterial chondronecrosis with osteomyelitis and lameness in broilers: A review. *Poult. Sci.* **2016**, *95*, 325–344. [CrossRef] [PubMed]
8. Wideman, R.F.; Hamal, K.R.; Stark, J.M.; Blankenship, J.; Lester, H.; Mitchell, K.N.; Lorenzoni, G.; Pevzner, I. A wire-flooring model for inducing lameness in broilers: Evaluation of probiotics as a prophylactic treatment. *Poult. Sci.* **2012**, *91*, 870–883. [CrossRef]
9. Gilley, A.D.; Lester, H.; Pevzner, I.Y.; Anthony, N.B.; Wideman, R.F. Evaluating portable wire-flooring models for inducing bacterial chondronecrosis with osteomyelitis in broilers. *Poult. Sci.* **2014**, *93*, 1354–1367. [CrossRef] [PubMed]
10. Harry, E.G. The Characteristics of *Staphylococcus aureus* Isolated from Cases of Staphylococcosis in Poultry. *Res. Vet. Sci.* **1967**, *8*, 479–489. [CrossRef]
11. Kibenge, F.S.; Wilcox, G.; Pass, D. Pathogenicity of four strains of Staphylococci isolated from chickens with clinical tenosynovitis. *Avian Pathol.* **1983**, *12*, 213–220. [CrossRef]
12. Mutalib, A.; Riddell, C.; Osborne, A.D. Studies on the pathogenesis of Staphylococcal osteomyelitis in chickens. II. Role of the respiratory tract as a route of infection. *Avian Dis.* **1983**, *27*, 157–160. [CrossRef] [PubMed]
13. Mutalib, A.; Riddell, C.; Osborne, A.D. Studies on the Pathogenesis of Staphylococcal Osteomyelitis in Chickens. I. Effect of Stress on Experimentally Induced Osteomyelitis. *Avian Dis.* **1983**, *27*, 141–156. [CrossRef]
14. McNamee, P.T.; Smyth, J.A. Bacterial chondronecrosis with osteomyelitis ('femoral head necrosis') of broiler chickens: A review. *Avian Pathol.* **2000**, *29*, 477–495. [CrossRef]
15. Butterworth, A.; Reeves, N.A.; Harbour, D.; Werrett, G.; Kestin, S.C. Molecular typing of strains of *Staphylococcus aureus* isolated from bone and joint lesions in lame broilers by random amplification of polymorphic DNA. *Poult. Sci.* **2001**, *80*, 1339–1343. [CrossRef] [PubMed]
16. Lowder, B.V.; Guinane, C.M.; Ben Zakour, N.L.; Weinert, L.A.; Conway-Morris, A.; Cartwright, R.A.; Simpson, A.J.; Rambaut, A.; Nübel, U.; Fitzgerald, J.R. Recent human-to-poultry host jump, adaptation, and pandemic spread of *Staphylococcus aureus*. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 19545–19550. [CrossRef] [PubMed]
17. Ekesi, N.S.; Dolka, B.; Alrubaye, A.; Rhoads, D. Analysis of Genomes of Bacterial Isolates from Lameness Outbreaks in Broilers. *Poult. Sci.* **2021**, *100*, 101148. [CrossRef]
18. Fitzgerald, J.R. Livestock-associated *Staphylococcus aureus*: Origin, evolution and public health threat. *Trends Microbiol.* **2012**, *20*, 192–198. [CrossRef]
19. Al-Rubaye, A.A.K.; Couger, M.B.; Ojha, S.; Pummill, J.F.; Koon, J.A., II; Wideman, R.F., Jr.; Rhoads, D.D. Genome analysis of *Staphylococcus agnetis*, an agent of lameness in broiler chickens. *PLoS ONE* **2015**, *10*, e0143336. [CrossRef]
20. Poulsen, L.L.; Thøfner, I.; Bisgaard, M.; Olsen, R.H.; Christensen, J.P.; Christensen, H. *Staphylococcus agnetis*, a potential pathogen in broiler breeders. *Vet. Microbiol.* **2017**, *212*, 1–6. [CrossRef]
21. Shwani, A.; Adkins, P.R.F.; Ekesi, N.S.; Alrubaye, A.; Calcutt, M.J.; Middleton, J.R.; Rhoads, D.D. Whole genome comparisons of *Staphylococcus agnetis* isolates from cattle and chickens. *Appl. Environ. Microbiol.* **2020**, *86*, e00484-20. [CrossRef]
22. Adkins, P.R.F.; Middleton, J.R.; Calcutt, M.J.; Stewart, G.C.; Fox, L.K. Species identification and strain typing of *Staphylococcus agnetis* and *Staphylococcus hyicus* isolates from bovine milk by use of a novel multiplex PCR assay and pulsed-field gel electrophoresis. *J. Clin. Microbiol.* **2017**, *55*, 1778–1788. [CrossRef] [PubMed]
23. Adkins, P.R.F.; Dufour, S.; Spain, J.N.; Calcutt, M.J.; Reilly, T.J.; Stewart, G.C.; Middleton, J.R. Molecular characterization of non-aureus *Staphylococcus* spp. from heifer intramammary infections and body sites. *J. Dairy Sci.* **2018**, *101*, 5388–5403. [CrossRef] [PubMed]
24. Adkins, P.R.F.; Dufour, S.; Spain, J.N.; Calcutt, M.J.; Reilly, T.J.; Stewart, G.C.; Middleton, J.R. Cross-sectional study to identify staphylococcal species isolated from teat and inguinal skin of different-aged dairy heifers. *J. Dairy Sci.* **2018**, *101*, 3213–3225. [CrossRef] [PubMed]
25. Akarsu, H.; Liljander, A.; Younan, M.; Brodard, I.; Overesch, G.; Glücks, I.; Labroussaa, F.; Kuhnert, P.; Perreten, V.; Monecke, S.; et al. Genomic Characterization and Antimicrobial Susceptibility of Dromedary-Associated *Staphylococcaceae* from the Horn of Africa. *Appl. Environ. Microbiol.* **2022**, *88*, e01146-22. [CrossRef] [PubMed]

26. Naseem, M.N.; Turni, C.; Gilbert, R.; Raza, A.; Allavena, R.; McGowan, M.; Constantinoiu, C.; Ong, C.T.; Tabor, A.E.; James, P. Role of *Staphylococcus agnetis* and *Staphylococcus hyicus* in the Pathogenesis of Buffalo Fly Skin Lesions in Cattle. *Microbiol. Spectr.* **2022**, *10*, e00873-22. [CrossRef] [PubMed]
27. Ayala, D.I.; Grum, D.S.; Evans, N.P.; Russo, K.N.; Kimminau, E.A.; Trible, B.R.; Lahoti, M.M.; Novak, C.L.; Karnezos, T.P. Identification and characterization of the causative agents of Focal Ulcerative Dermatitis in commercial laying hens. *Front. Vet. Sci.* **2023**, *10*, 1110573. [CrossRef] [PubMed]
28. Alrubaye, A.; Ekesi, N.S.; Hasan, A.; Koltes, D.A.; Wideman, R., Jr.; Rhoads, D. Chondronecrosis with osteomyelitis in broilers: Further defining a bacterial challenge model using standard litter flooring and protection with probiotics. *Poult. Sci.* **2020**, *99*, 6474–6480. [CrossRef] [PubMed]
29. Rhoads, D.D.; Pummill, J.; Ekesi, N.S.; Alrubaye, A.A.K. Horizontal transfer of probable chicken-pathogenicity chromosomal islands between *Staphylococcus aureus* and *Staphylococcus agnetis*. *PLoS ONE* **2023**, *18*, e0283914. [CrossRef] [PubMed]
30. Richardson, E.J.; Bacigalupe, R.; Harrison, E.M.; Weinert, L.A.; Lycett, S.; Vrieling, M.; Robb, K.; Hoskisson, P.A.; Holden, M.T.G.; Feil, E.J.; et al. Gene exchange drives the ecological success of a multi-host bacterial pathogen. *Nat. Ecol. Evol.* **2018**, *2*, 1468–1478. [CrossRef]
31. Asnayanti, A.; Do, A.D.T.; Alharbi, K.; Alrubaye, A. Inducing experimental bacterial chondronecrosis with osteomyelitis lameness in broiler chickens using aerosol transmission model. *Poult. Sci.* **2024**, *103*, 103460. [CrossRef]
32. Kloos, W.E.; Wolfshohl, J.F. *Staphylococcus cohnii* Subspecies: *Staphylococcus cohnii* subsp. *cohnii* subsp. nov. and *Staphylococcus cohnii* subsp. *urealyticum* subsp. nov. *Int. J. Syst. Evol. Microbiol.* **1991**, *41*, 284–289. [CrossRef]
33. Lavecchia, A.; Chiara, M.; De Virgilio, C.; Manzari, C.; Pazzani, C.; Horner, D.; Pesole, G.; Placido, A. Comparative Genomics Suggests a Taxonomic Revision of the *Staphylococcus cohnii* Species Complex. *Genome Biol. Evol.* **2021**, *13*, evab020. [CrossRef]
34. Soldera, J.; Nedel, W.L.; Cardoso, P.R.C.; d'Azevedo, P.A. Bacteremia due to *Staphylococcus cohnii* ssp. *urealyticus* caused by infected pressure ulcer: Case report and review of the literature. *Sao Paulo Med. J.* **2013**, *131*, 59–61.
35. Marek, A.; Stepień-Pyśniak, D.; Pyzik, E.; Adaszek, Ł.; Wilczyński, J.; Winiarczyk, S. Occurrence and characterization of Staphylococcus bacteria isolated from poultry in Western Poland. *Berl. Münchener Tierärztliche Wochenschr.* **2016**, *129*, 147–152.
36. Al-Rubaye, A.A.K.; Ekesi, N.S.; Zaki, S.; Emami, N.K.; Wideman, R.F.; Rhoads, D.D. Chondronecrosis with osteomyelitis in broilers: Further defining a bacterial challenge model using the wire flooring model. *Poult. Sci.* **2017**, *96*, 332–340. [CrossRef]
37. Shwani, A.; Zuo, B.; Alrubaye, A.; Zhao, J.; Rhoads, D.D. A Simple, Inexpensive Alkaline Method for Bacterial DNA Extraction from Environmental Samples for PCR Surveillance and Microbiome Analyses. *Appl. Sci.* **2024**, *14*, 141. [CrossRef]
38. Shwani, A. Bacterial Chondronecrosis with Osteomyelitis in Broilers: Genomics, Phylogenomics, and Methods to Detect Specific Pathogens during Outbreaks. Ph.D. Dissertation, University of Arkansas, Fayetteville, AR, USA, 2022.
39. Dyer, D.W.; Iandolo, J.J. Rapid isolation of DNA from *Staphylococcus aureus*. *Appl. Environ. Microbiol.* **1983**, *46*, 283–285. [CrossRef]
40. Olson, R.D.; Assaf, R.; Brettin, T.; Conrad, N.; Cucinell, C.; Davis, J.J.; Dempsey, D.M.; Dickerman, A.; Dietrich, E.M.; Kenyon, R.W.; et al. Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): A resource combining PATRIC, IRD and ViPR. *Nucleic Acid Res.* **2022**, *51*, D678–D689. [CrossRef]
41. Lees, J.A.; Harris, S.R.; Tonkin-Hill, G.; Gladstone, R.A.; Lo, S.W.; Weiser, J.N.; Corander, J.; Bentley, S.D.; Croucher, N.J. Fast and flexible bacterial genomic epidemiology with PopPUNK. *Genome Res.* **2019**, *29*, 304–316. [CrossRef]
42. Treangen, T.J.; Ondov, B.D.; Koren, S.; Phillippy, A.M. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* **2014**, *15*, 524. [CrossRef]
43. Pritchard, L.; Glover, R.H.; Humphris, S.; Elphinstone, J.G.; Toth, I.K. Genomics and taxonomy in diagnostics for food security: Soft-rotting enterobacterial plant pathogens. *Anal. Methods* **2016**, *8*, 12–24. [CrossRef]
44. Hölzer, M. POCP-nf: An automatic Nextflow pipeline for calculating the percentage of conserved proteins in bacterial taxonomy. *Bioinformatics* **2024**, *40*, btae175. [CrossRef]
45. Han, M.V.; Zmasek, C.M. phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinform.* **2009**, *10*, 356. [CrossRef]
46. Argimón, S.; Abudahab, K.; Goater, R.J.E.; Fedosejev, A.; Bhai, J.; Glasner, C.; Feil, E.J.; Holden, M.T.G.; Yeats, C.A.; Grundmann, H.; et al. Microreact: Visualizing and sharing data for genomic epidemiology and phylogeography. *Microb. Genom.* **2016**, *2*, e000093. [CrossRef] [PubMed]
47. Grant, J.R.; Enns, E.; Marinier, E.; Mandal, A.; Herman, E.K.; Chen, C.Y.; Graham, M.; Van Domselaar, G.; Stothard, P. Proksee: In-depth characterization and visualization of bacterial genomes. *Nucleic Acids Res.* **2023**, *51*, W484–W492. [CrossRef]
48. Brown, C.L.; Mullet, J.; Hindi, F.; Stoll, J.E.; Gupta, S.; Choi, M.; Keenum, I.; Vikesland, P.; Pruden, A.; Zhang, L. mobileOG-db: A Manually Curated Database of Protein Families Mediating the Life Cycle of Bacterial Mobile Genetic Elements. *Appl. Environ. Microbiol.* **2022**, *88*, e0099122. [CrossRef]
49. Madden, T. BLAST® Command Line Applications User Manual. 2008. Available online: https://www.ncbi.nlm.nih.gov/books/NBK279691/ (accessed on 25 September 2020).
50. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [CrossRef] [PubMed]
51. Page, A.J.; Cummins, C.A.; Hunt, M.; Wong, V.K.; Reuter, S.; Holden, M.T.G.; Fookes, M.; Falush, D.; Keane, J.A.; Parkhill, J. Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **2015**, *31*, 3691–3693. [CrossRef]
52. Bayliss, S.C.; Thorpe, H.A.; Coyle, N.M.; Sheppard, S.K.; Feil, E.J. PIRATE: A fast and scalable pangenomics toolbox for clustering diverged orthologues in bacteria. *GigaScience* **2019**, *8*, giz119. [CrossRef]

53. Brynildsrud, O.; Bohlin, J.; Scheffer, L.; Eldholm, V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol.* **2016**, *17*, 238. [CrossRef]
54. Shi, Z.J.; Nayfach, S.; Pollard, K.S. Maast: Genotyping thousands of microbial strains efficiently. *Genome Biol.* **2023**, *24*, 186. [CrossRef] [PubMed]
55. Shen, W.; Le, S.; Li, Y.; Hu, F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS ONE* **2016**, *11*, e0163962. [CrossRef] [PubMed]
56. Lalucat, J.; Mulet, M.; Gomila, M.; García-Valdés, E. Genomics in Bacterial Taxonomy: Impact on the Genus Pseudomonas. *Genes* **2020**, *11*, 139. [CrossRef] [PubMed]
57. Graña-Miraglia, L.; Arreguín-Pérez, C.; López-Leal, G.; Muñoz, A.; Pérez-Oseguera, A.; Miranda-Miranda, E.; Cossío-Bayúgar, R.; Castillo-Ramírez, S. Phylogenomics picks out the par excellence markers for species phylogeny in the genus *Staphylococcus*. *PeerJ* **2018**, *6*, e5839. [CrossRef]
58. Varghese, N.J.; Mukherjee, S.; Ivanova, N.; Konstantinidis, K.T.; Mavrommatis, K.; Kyrpides, N.C.; Pati, A. Microbial species delineation using whole genome sequences. *Nucleic Acid Res.* **2015**, *43*, 6761–6771. [CrossRef] [PubMed]
59. 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]
60. Chen, J.; Quiles-Puchalt, N.; Chiang, Y.N.; Bacigalupe, R.; Fillol-Salom, A.; Chee, M.S.J.; Fitzgerald, J.R.; Penadés, J.R. Genome hypermobility by lateral transduction. *Science* **2018**, *362*, 207–212. [CrossRef] [PubMed]
61. Lan, R.; Reeves, P.R. *Escherichia coli* in disguise: Molecular origins of Shigella. *Microbes Infect.* **2002**, *4*, 1125–1132. [CrossRef] [PubMed]
62. Abram, K.; Udaondo, Z.; Bleker, C.; Wanchai, V.; Wassenaar, T.M.; Robeson, M.S.; Ussery, D.W. Mash-based analyses of *Escherichia coli* genomes reveal 14 distinct phylogroups. *Commun. Biol.* **2021**, *4*, 117. [CrossRef]
63. Yang, T.; Gao, F. High-quality pan-genome of *Escherichia coli* generated by excluding confounding and highly similar strains reveals an association between unique gene clusters and genomic islands. *Brief. Bioinform.* **2022**, *23*, bbac283. [CrossRef]
64. Gonzalez-Alba, J.M.; Baquero, F.; Cantón, R.; Galán, J.C. Stratified reconstruction of ancestral *Escherichia coli* diversification. *BMC Genom.* **2019**, *20*, 936. [CrossRef] [PubMed]
65. Alrubaye, A.A.K.; Ekesi, N.S.; Hasan, A.; Elkins, E.; Ojha, S.; Zaki, S.; Dridi, S.; Wideman, R.F.; Rebollo, M.A.; Rhoads, D.D. Chondronecrosis with Osteomyelitis in Broilers: Further Defining Lameness-Inducing Models with Wire or Litter Flooring, to Evaluate Protection with Organic Trace Minerals. *Poult. Sci.* **2020**, *99*, 5422–5429. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

microorganisms

MDPI

Communication

Isolation and Genomic Analysis of a Case of *Staphylococcus argenteus* ST2250 Related to Sepsis in Italy

Giulia Gatti [1,*], Francesca Taddei [2], Anna Marzucco [2], Maria Sofia Montanari [2], Giorgio Dirani [2], Silvia Zannoli [2], Laura Grumiro [2], Martina Brandolini [1], Claudia Colosimo [1], Laura Dionisi [1], Ludovica Ingletto [1], Alessandra Mistral De Pascali [1], Alessandra Scagliarini [1], Vittorio Sambri [1,2] and Monica Cricca [1,2]

[1] Department of Medical and Surgical Sciences—DIMEC, Alma Mater Studiorum, University of Bologna, 40138 Bologna, Italy; martina.brandolini3@unibo.it (M.B.); claudia.colosimo2@unibo.it (C.C.); laura.dionisi2@unibo.it (L.D.); ludovica.ingletto@unibo.it (L.I.); alessandra.depascal3@unibo.it (A.M.D.P.); alessand.scagliarini@unibo.it (A.S.); vittorio.sambri@unibo.it (V.S.); monica.cricca3@unibo.it (M.C.)

[2] Unit of Microbiology, The Great Romagna Hub Laboratory, 47522 Cesena, Italy; francesca.taddei@auslromagna.it (F.T.); anna.marzucco@auslromagna.it (A.M.); sofi.monta.msm@gmail.com (M.S.M.); giorgio.dirani@auslromagna.it (G.D.); silvia.zannoli@auslromagna.it (S.Z.); laura.grumiro@auslromagna.it (L.G.)

* Correspondence: giulia.gatti12@unibo.it

Abstract: *Staphylococcus argenteus*, identified in 2006, represents a challenging case of bacterial taxonomic identification because of its high similarity to *Staphylococcus aureus*. In this context, neither mass spectrometry (MS) nor 16S gene analysis cannot precisely reveal the difference between the two species. In our study, the sensitivity to antibiotics of *S. argenteus* isolated from blood culture was tested, and the investigation of the bacterial genome was performed by Multi-Locus Sequence Typing (MLST) and Whole-Genome Next-Generation Sequencing (WG-NGS). The pathogen was identified as ST2250 and presented perfectly matched resistance genes, namely *aph(3′)-III*, *mgrA*, and *sepA*, whereas the virulence gene detected was *scn*. Two plasmids were found: the pSAS plasmid, belonging to the family of Inc18, and plasmid pN315, belonging to the Rep3 group. The epidemiological distribution and the spread of *S. argenteus* infection are scarcely documented, particularly when associated with sepsis. Therefore, a correct taxonomy identification, antibiogram, and resistance gene analysis may help in acquiring knowledge about this bacterium and implement its detection and treatment.

Keywords: *Staphylococcus argenteus*; NGS; bloodstream infections; *Staphylococcus aureus*; genomic analysis

1. Introduction

In 2006, a new presumable lineage of *Staphylococcus aureus*, assigned to Clonal Complex 75 (CC75), was isolated from samples from individuals belonging to indigenous communities in the Northern Territory, Australia. Multi-Locus Sequence Typing (MLST) proved this was a very interesting and challenging case study in bacterial systematics, so much so that CC75 was renamed *Staphylococcus argenteus* [1,2]. The phenotypic difference between *S. aureus* and *S. argenteus* is a non-pigmented appearance of colonies that is due to the lack of the crtOPQMN operon, essential for the synthesis of staphyloxanthin, a carotenoid pigment [3]. Through genomic investigation, *S. argenteus* can be considered an ancestor of *S. aureus*, with which it shares about 95% of its Average Nucleotide Identity (ANI), although 16S gene results are not distinguishable between the two. A difference can be detected by studying the *nuc*, *gap*, *rpoB*, *sodA*, *tuf*, and *hsp60* genes or performing MLST [4,5]. Worldwide, *S. argenteus* is concentrated in three "hotspot" regions, namely Southeast Asia, Australia, and the Amazon, but has also been reported in New Zealand, Fiji, America, and, in rare cases, in Europe [5,6]. In subtropical areas, *S. argenteus* is considered an emerging pathogen among patients with community-onset staphylococcal infections; linked to this

phenomenon, *S. argenteus* seems to have recently also been found in healthcare-associated infections among recently hospitalized patients [7]. Besides its spread, the clinical implication of *S. argenteus* relates to the presence of genes encoding staphylococcal enterotoxins, hemolysin, and Panton–Valentine leucocidin [8]. Additionally, public health systems have been warned about the antimicrobial resistance genes reported in *S. argenteus* isolates, since penicillin (*blaZ*-positive) and methicillin (mostly *mecA*-positive)-resistant bacteria similar to methicillin-resistant *S. aureus* (MRSA) have been described in detail [9,10]. Although most *S. argenteus* cases have been found to be sensitive to methicillin, some isolates among patients in Aboriginal communities in Northern Australia affected by impetiginous lesions have been described as resistant to this antimicrobial agent [7,8]. In view of healthcare-associated infections, the acquisition of resistance to methicillin by *S. argenteus* draws attention to high antibiotic pressure and its further implications in public health [7]. *S. argenteus* is known to be a sporadic pathogen causing skin and soft tissue infections, bone infections, and sepsis [11,12], cases in which it can be misdiagnosed as *S. aureus* bacteremia. In light of this, the clinical conditions of patients with *S. argenteus* bacteremia can worsen if they develop lower respiratory tract infections, which increase mortality compared to methicillin-sensitive *Staphylococcus aureus* (MSSA) infections [7,13]. A plausible explanation for this enhanced trend relates to the acquisition of species-specific *S. argenteus* virulence factors [7]. In this study, we describe the first *S. argenteus* case isolated at the Greater Romagna Hub Laboratory from a blood culture, which, in our knowledge, represents the first case of *S. argenteus* bacteremia described in Italy.

2. Materials and Methods

2.1. Sampling

The blood sample was collected in an emergency room at Riccione and delivered to the Greater Romagna Hub Laboratory in BACT/ALERT® culture media (bioMérieux, Marcy l'Etoile, France). The sample was incubated for about 22 h until positivity was declared. The sample was processed according to the diagnostic routine of the laboratory.

2.2. Identification and Antibiogram

Then, the blood culture was seeded on Columbia + 5% sheep blood, COS (bioMérieux, Marcy l'Etoile, France). The identification of the isolated colonies was performed using Vitek® MS (bioMérieux, Marcy l'Etoile, France) and resulted in *S. argenteus* identification; since it was the first case detected in our laboratory, the isolate was tested for antimicrobial sensitivity and with an NGS approach. The Antibiotic Sensitivity Test (AST) was performed according to EUCAST's breakpoint guidelines [14] calculated using MicroScan WalkAway (Beckman Coulter Inc., Brea, CA, USA). The isolated colony presented a peculiar feature: it resembled a *S. aureus* colony, but without pigmentation. For this reason, the colony was selected for an in-depth investigation.

2.3. Sequencing

To confirm the mass spectrometry spectrum, the whole bacterial genome was investigated. The sample was extracted using the automated system of Maxwell® CSC Pathogen Total Nucleic Acid Kit (Promega Corporation, Madison, WI, USA). The library was prepared using the DNA Prep kit (Illumina Inc., San Diego, CA, USA) according to the manufacturer's instructions and sequenced on MiSeq platform (Illumina Inc., San Diego, CA, USA). The raw sequence FastQ files were deposited in the Sequence Reads Archive (SRA) (BioProject ID: PRJNA1032758).

2.4. Bioinformatical Analysis and Confirmation

Forward and reverse reads were controlled using FastQC v.0.12.1 (Babraham Institute, Cambridge, UK), trimmed using Trimmomatic v.0.38.1 by using default parameters, and de novo assembled using SPAdes v.3.13.0 with default parameters. The confirmation of pathogen identification was performed by using a ribosomal MLST (rMLST) on 53 target genes, and then the MLST profile was traced using Ridom SeqSphere+ v.9.0 with 7 target genes described in the literature (*arcC*, *aroE*, *glpF*, *gmk*, *pta*, *tpi,* and *yqiL*). The presence of genes involved in resistance mechanisms was investigated using the ResFinder v.4.1 and Comprehensive Antibiotic Resistance Database (CARD v.3.2.8) databases. Additionally, the PlasmidFinder v.2.0 database was queried to search for the presence of plasmids, and the Virulence Factor Gene Database to screen for additional virulence factors. The mobile genetic elements (MGEs) were investigated by the MobileElementFinder database v.1.0.3.

Once the profile of the resistance genes was obtained, phenotypic resistance was confirmed for the corresponding antibiotic by performing ETEST® (bioMérieux, Marcy l'Etoile, France) and MicroScan WalkAway in triplicate. To complete the phenotypic resistance pattern, sensitivity to fosfomycin was evaluated by the agar dilution method AD Fosfomycin 0.25–256 assay (Liofilchem srl, Roseto degli Abruzzi, Italy). Resistance was declared according to EUCAST's guidelines [14]. To better understand the evolution similarity of the isolate, 15 complete *S. argenteus* genomes were downloaded from the NCBI database and a phylogenetic tree was inferred in MEGA11 [15] using a UPGMA method and the evolutionary distances were computed using the Maximum Composite Likelihood method. The tree was visualized using FigTree v.1.4.4 (accessed on 10 July 2024: http://tree.bio.ed.ac.uk/software/figtree/) and was midpoint-rooted.

2.5. S. argenteus MSHR1132

The *S. argenteus* genome (NCBI RefSeq Assembly: GCF_000236925.1) was downloaded from NCBI and included in the analysis. The sequence belonged to the MSHR1132 strain and was deposited in November 2011 by The Wellcome Trust Sanger Institute. There are 2689 annotated genes, including 2533 protein-coding genes. The bacterium was isolated from a blood culture of a woman affected by necrotizing fasciitis and the genome was classified in an ST 1850 [2].

3. Results

3.1. rMLST and MLST

The rMLST resulted in rST 4649, typically associated with the *S. argenteus* genomic profile. According to the MLST, the genome sequence is classified in an ST 2250 and target genes attributed to its alleles (Table 1). According to the genomes deposited in the public database PubMLST (https://pubmlst.org/, accessed on 16 July 2024), 111 isolates were recorded with rST 4649 and 3 with ST 2250.

Table 1. Results of MLST of Staphylococcus argenteus isolated from blood culture at the Greater Romagna Hub Laboratory. The 7 target genes were classified, and the sequence type was determined.

Genes	Allele
arcC	151
aroE	325
glpF	215
gmk	34
pta	175
tpi	180
yqiL	169
ST *	2250

* ST = sequence type.

3.2. Antibiogram

The antibiogram performed in microdilution showed bacterial sensitivity to amikacin (≤4 mg/mL), ceftaroline (0.25 mg/mL), ciprofloxacin (≤0.25 mg/mL), clindamycin (≤0.12 mg/mL), daptomycin (≤0.25 mg/mL), erythromycin (≤0.5 mg/mL), gentamycin (≤1 mg/mL), levofloxacin (≤0.5 mg/mL), linezolid (≤2 mg/mL), minocycline (≤1 mg/mL), moxifloxacin (≤0.25 mg/mL), oxacillin (≤0.25 mg/mL), penicillin (≤0.03 mg/mL), teicoplanin (≤1 mg/mL), tetracycline (≤1 mg/mL), tigecycline (≤0.25 mg/mL), tobramycin (≤1 mg/mL), trimethoprim and sulfamethoxazole (≤1/19 mg/mL), and vancomycin (1 mg/mL). Therefore, the isolate did not present any resistance to the tested antibiotics. In addition, the agar dilution method for the detection of fosfomycin resistance resulted in an MIC of 8 mg/mL; therefore, the isolate was sensitive to the antibiotic according to EUCAST's guidelines (Table 2).

Table 2. The antibiogram results of Staphylococcus argenteus. The isolate was sensitive to 20 antibiotics tested.

Antibiotic	MIC * (mg/mL)	Result
Amikacin	≤4	Sensitive
Ceftaroline	0.25	Sensitive
Ciprofloxacin	≤0.25	Sensitive
Clindamycin	≤0.12	Sensitive
Daptomycin	≤0.25	Sensitive
Erythromycin	≤0.5	Sensitive
Gentamycin	≤1	Sensitive
Levofloxacin	≤0.5	Sensitive
Linezolid	≤2	Sensitive
Minocycline	≤1	Sensitive
Moxifloxacin	≤0.25	Sensitive
Oxacillin	≤0.25	Sensitive
Penicillin	≤0.03	Sensitive
Teicoplanin	≤1	Sensitive
Tetracycline	≤1	Sensitive
Tigecycline	≤0.25	Sensitive
Tobramycin	≤1	Sensitive
Trimethoprim/Sulfamethoxazole	≤1/19	Sensitive
Vancomycin	1	Sensitive
Fosfomycin	8	Sensitive

* Minimum Inhibitory Concentration.

3.3. Resistance Genes, Plasmids, and Virulence Genes

At a genomic level, the resistance gene identified by both the queried databases (ResFinder and CARD) had a perfect match in terms of identity and sequence length with *aph(3′)-III* (aph(3′)-III_M26832), which can inactivate aminoglycoside antibiotics such as amikacin, butirosin, kanamycin, neomycin, lividomycin, paromomycin, and ribostamycin. The CARD database found two other additional perfectly matched genes: *mgrA* (inducing resistance to fluoroquinolones, cephalosporins, penams, tetracyclines, peptide antibiotics, and disinfecting and antiseptic agents) and *sepA* (inducing resistance to disinfecting and antiseptic agents). Other sequences that are probably responsible for resistance mechanisms were detected, even though there was no perfect match: *blaZ* (ResFinder and CARD) (96.8% identity and 100% length), *arlR* (98.63% identity and 100% length), *norC* (97.4% identity and 100% length), *Staphylococcus aureus fosB* (97.12 identity and 100% length), *Staphylococcus aureus lmrS* (90.44% identity and 100% length), *mepR* (92.81% identity and 100% length), *sdrM* (89.04% identity and 100% length), and the *vanT* gene in the *vanG* cluster (34.23% identity and 55.48% length) (CARD). Given the presence of *blaZ*, resistance to penicillin was confirmed, with an MIC of 0.125 mg/mL. Indeed, according to EUCAST's guidelines, the isolate presented resistance to penicillin.

Also, to better understand the isolate's resistance and invasiveness capability, the following plasmids were investigated: the pSAS plasmid (reference sequence BX571858; 20,652 base pairs, bp) with the *rep16* gene and belonging to the family of Inc18 was found, with an identity and length match of 100% for Coding Sequence 8. In addition, the plasmid pN315 (reference sequence AP003139; 24,653 bp) with the *rep5a* gene and belonging to the plasmid family Rep3 was detected, with a length match of 100% and an identity percentage of 99.88% (one single-nucleotide polymorphism) in the Coding Sequence of SAP001. Regarding the virulence genes, a perfect match in length and sequence was declared with the *scn* gene, a staphylococcal complement inhibitor, and a 99.6% identity match was found with the *sak* gene encoding staphylokinase. Just one MGE was found: ISSau5 (NC_002952.2), belonging to the IS30 family, an insertion sequence of 1136 nucleotides. The evolutionary relation of the 16 genomes was calculated and the original tree, with a total of 2,834,926 positions, was analyzed. The isolate's genome is most likely related to the *S. argenteus* MSHR1132 strain (Figure 1).

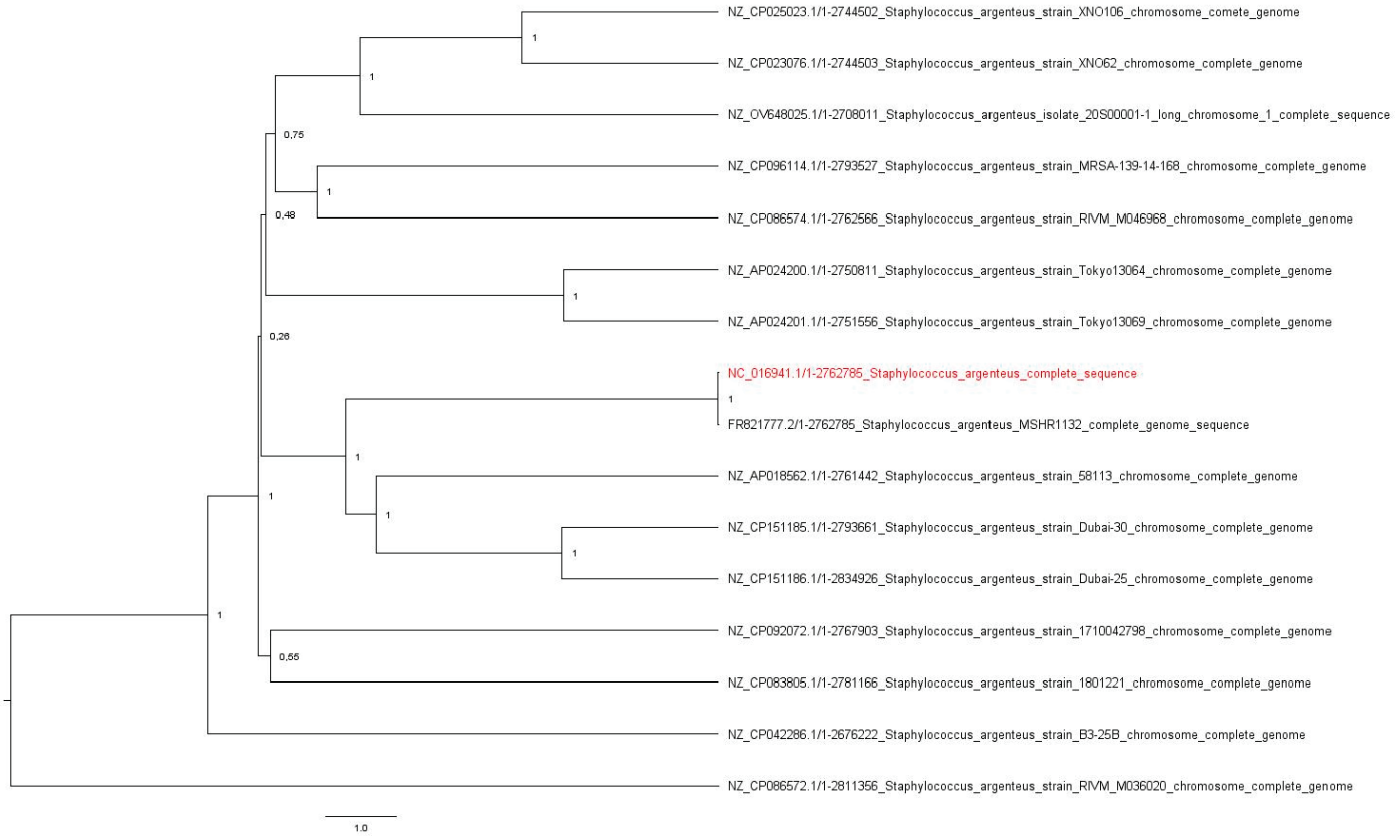


Figure 1. The original tree resulting from the UPGMA analysis. The evolutionary distances among 16 genomes were calculated with the Composite Likelihood Method using Mega 11 v.11.0.13. The tree was midpoint rooted using FigTree v.1.4.4. The isolate's genome is highlighted in red.

Similarity with the *S. aureus* reference genome (GCF_000013425.1) was calculated using Average Nucleotide Identity and was 87% (Figure 2).

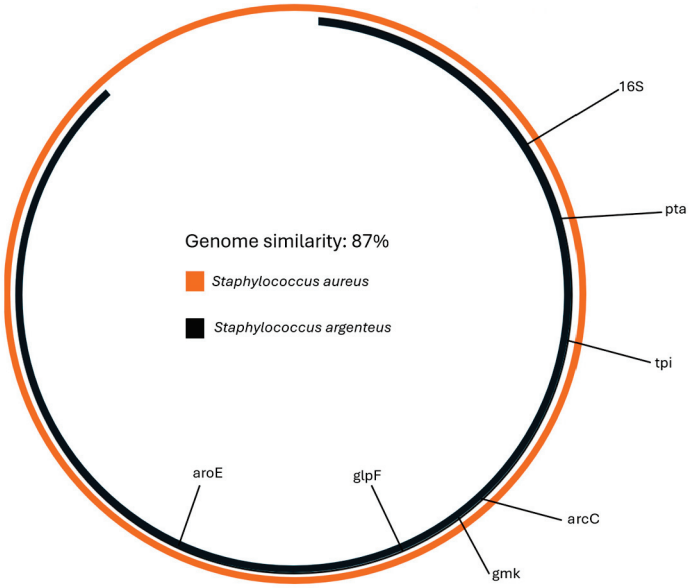


Figure 2. The similarity between S. aureus and the genome of the isolate was 87%. The similarity was calculated with Average Nucleotide Identity analysis.

4. Discussion

Since its first isolation, *S. argenteus* has been documented as a probable blood stream contaminant; in particular, ST2250 was found to cause sepsis [16,17]. In the Great Romagna Hub Laboratory, a positive blood culture isolate was classified as *S. argenteus* with the use of mass spectrometry; this was then confirmed with ANI analysis and MLST. The presence of resistance genes was investigated through the ResFinder and CARD databases: the two were concordant on the presence of *aph(3′)-III* (aph(3′)-III_M26832), a gene encoding aminoglycoside-3′′-O-phosphoryltransferase III and responsible for the inactivation of aminoglycoside antibiotics by catalyzing acetyltransferase and phosphotransferase reactions [18]. *mgrA*, only found in CARD, controls over 300 genes involved in antibiotic resistance and virulence, including an upregulation of exotoxins and capsular polysaccharide [19]. Moreover, *sepA* has been described in staphylococci encoding efflux pumps related to biocide resistance [20]. The other detected sequences did not present a perfect match with the genes deposited in the databases, even though their presence can be investigated in more depth. Particularly, the knowledge of the role of *blaZ,* found in the genome of *S. argenteus,* in regulating penicillin resistance can be extended to the identification of the gene's repressors (*blaI*) and regulators (*blaR1*) [21]. Particularly, genomic analysis can reveal a resistance profile which may be missed or misdiagnosed due to technical errors during the laboratory workflow. This suggests that a collaboration between molecular biology techniques and standard wet-lab methodology can trace the pathogen resistance trend more precisely. In addition, the presence of the *vanT* gene, grouped in the *vanG* cluster, can be related to a peculiar pattern of vancomycin resistance [22,23]. Indeed, the antibiogram result obtained for vancomycin was considered dubious, and, according to EUCAST's guidelines [14], which suggest a breakpoint of 2 mg/L, the MIC that was found was reduced, at 1 mg/L. Other investigations are needed to confirm the presence of the *vanT* sequence and to understand its impact on MIC. A crucial role in the acquisition of antimicrobial resistance is played by plasmids; indeed, ST2250 has been proven to possess the plasmid p2_1801221, with *rep5a* and *rep16* involved as determinants in penicillin resistance due to carrying the *blaZ* gene [9]. In *S. argenteus*, the *blaZ* gene is more likely linked to plasmid acquisition than to chromosomal origin. The study of Aung et al., according to Olsen's classification, divided plasmid-related *blaZ* in *S. argenteus* into P1 and P2 subgroups; this subdivision may describe a multiple origin of *blaZ* [24]. In particular, the Inc18 plasmid family is commonly isolated from bacteria that cause nosocomial infections and encodes resistance to different antimicrobial agents such as macrolides, lincosamides, streptogramins,

and vancomycin. Among staphylococci, Inc18 plasmids carrying the *vanA* gene are responsible for vancomycin resistance acquired from enterococci. This acquisition can likely be correlated with poor hygiene conditions, and in this setting, the rare vancomycin-resistant *S. aureus* is generally also resistant to methicillin and derives its *vanA*-mediated resistance from enterococci. Nevertheless, vancomycin resistance incidence in MRSA was revealed to be a serious threat not only for the treatment of *S. aureus* infections but also for all *S. aureus* complex-related bacterial species [25]. In the same manner, the *scn* virulence gene, due to its role as an immune evasion marker, may be linked to animal-to-human transmission and is generally found in MSSA [26]. The presence of the *scn* gene raises the need to take strict hygiene controls in animal-related and animal-derived products [27]. According to the One Health epidemiological approach, the investigation and the understanding of the resistance and virulence genes acquired by *S. argenteus* may be relevant not only to the analysis of their distribution but also to the prevention and treatment of the human healthcare-associated matter in question. More precisely, bloodstream infections may be fatal for patients and indeed a precise identification of the infective agent, its virulence, and resistance are necessary both for the monitoring of the spread and for the control of the patient's infection. The single MGE found in the sequence was previously described in MRSA252, a hospital-acquired pathogen [28]. Considering the increase in sepsis incidence in recent years and the variety of its causative agents, a new frontier of research opens in correlating the infection with gut microbiota dysbiosis [29]. The alteration of gastrointestinal mucosa integrity allows for the infiltration of various entities, including bacteria [30]. Together with other commensal microorganisms, *S. aureus* may be found to be an intestinal colonizer pathogen that enters the bloodstream and causes diffuse infections [31]. Due to its large similarity with *S. aureus*, it can be assumed that *S. argenteus* could present the same outcome, but other studies are needed to confirm its presence in the gut microbiota. On the other hand, new evidence has been reported in the literature about *S. argenteus* isolation from blood cultures and its severity for patients [12]. Due to the scarcity of information, its antibiotic resistance and geographical distribution need to be traced and monitored. In Australia, in 2023, *S. argenteus* ST 2250 was isolated in a patient without prior exposure and was found to be resistant to daptomycin [32]. Similarly, in Korea, *S. argenteus* ST 8342 and 8343 lead to severe clinical outcomes in both community- and healthcare-acquired infections [33]. Therefore, studying the epidemiological trend of *S. argenteus* by using molecular assays or NGS helps not only in characterizing the pathogen but also in preventing patient complications [13].

5. Conclusions

The assortment of human pathogens requires continuous monitoring of novel deleterious mutations and causative agents to prevent severe sequelae. Investigating and understanding the phenotypic and genotypic characteristics of uncommon infections is significant in order to organize an efficient diagnostic workflow and administer the proper treatment. In this scenario, as far as the authors know, this is the first documented case of *S. argenteus* causing sepsis in Italy.

Author Contributions: Conceptualization, G.G. and F.T.; methodology, G.G. and M.C.; software, G.G.; validation, F.T., A.M. and M.S.M.; formal analysis, G.D. and S.Z.; investigation, L.G., L.D. and M.B.; resources, V.S.; data curation, G.G.; writing—original draft preparation, G.G.; writing—review and editing, A.M.D.P. and M.C.; visualization, A.M.D.P., C.C. and L.I.; supervision, M.C.; project administration, G.G.; funding acquisition, V.S., A.S. and M.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by EU funding within the NextGenerationEU-MUR PNRR Extended Partnership initiative on Emerging Infectious Diseases (Project no. PE00000007, INF-ACT). The Ph.D. scholarship of Giulia Gatti was funded by the European Union—NextGenerationEU through the Italian Ministry of University and Research under PNRR—Mission 4 Component 2, Investment 3.3 "Partnerships extended to universities, research centers, companies und funding of basic research projects" D.M. 352/2021-CUP J33C2200133009.

Institutional Review Board Statement: Ethical approval or informed consent were not required because this study has been performed using exclusively anonymized, leftover samples derived from routine diagnostic procedures. Anonymization was achieved by using the current procedure (AVR-PPC P09, rev.2) checked by the local ethics board.

Data Availability Statement: The original data presented in the study are openly available in BioProject. Accession Number: PRJNA1032758.

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

References

1. McDonald, M.; Dougall, A.; Holt, D.; Huygens, F.; Oppedisano, F.; Giffard, P.M.; Inman-Bamber, J.; Stephens, A.J.; Towers, R.; Carapetis, J.R.; et al. Use of a single-nucleotide polymorphism genotyping system to demonstrate the unique epidemiology of methicillin-resistant *Staphylococcus aureus* in remote aboriginal communities. *J. Clin. Microbiol.* **2006**, *44*, 3720–3727. [CrossRef] [PubMed] [PubMed Central]
2. Holt, D.C.; Holden, M.T.; Tong, S.Y.; Castillo-Ramirez, S.; Clarke, L.; Quail, M.A.; Currie, B.J.; Parkhill, J.; Bentley, S.D.; Feil, E.J.; et al. A very early-branching *Staphylococcus aureus* lineage lacking the carotenoid pigment staphyloxanthin. *Genome Biol. Evol.* **2011**, *3*, 881–895. [CrossRef] [PubMed] [PubMed Central]
3. Rossi, B.F.; Bonsaglia, É.C.R.; Castilho, I.G.; Dantas, S.T.A.; Langoni, H.; Pantoja, J.C.F.; Júnior, A.F.; Gonçalves, J.L.; Santos, M.V.; Mota, R.A.; et al. First investigation of *Staphylococcus argenteus* in a Brazilian collections of S. aureus isolated from bovine mastitis. *BMC Vet. Res.* **2020**, *16*, 252. [CrossRef] [PubMed] [PubMed Central]
4. Eshaghi, A.; Bommersbach, C.; Zittermann, S.; Burnham, C.A.; Patel, R.; Schuetz, A.N.; Patel, S.N.; Kus, J.V. Phenotypic and Genomic Profiling of *Staphylococcus argenteus* in Canada and the United States and Recommendations for Clinical Result Reporting. *J. Clin. Microbiol.* **2021**, *59*, e02470-20. [CrossRef] [PubMed] [PubMed Central]
5. Becker, K.; Schaumburg, F.; Kearns, A.; Larsen, A.R.; Lindsay, J.A.; Skov, R.L.; Westh, H. Implications of identifying the recently defined members of the *Staphylococcus aureus* complex *S. argenteus* and *S. schweitzeri*: A position paper of members of the ESCMID Study Group for Staphylococci and Staphylococcal Diseases (ESGS). *Clin. Microbiol. Infect.* **2019**, *25*, 1064–1070. [CrossRef] [PubMed]
6. Wu, S.; Pang, R.; Huang, J.; Zhang, F.; Cai, Z.; Zhang, J.; Chen, M.; Xue, L.; Gu, Q.; Wang, J.; et al. Evolutionary Divergence of the Novel Staphylococcal Species *Staphylococcus argenteus*. *Front. Microbiol.* **2021**, *12*, 769642. [CrossRef] [PubMed] [PubMed Central]
7. Chen, S.Y.; Lee, H.; Wang, X.M.; Lee, T.F.; Liao, C.H.; Teng, L.J.; Hsueh, P.R. High mortality impact of *Staphylococcus argenteus* on patients with community-onset staphylococcal bacteraemia. *Int. J. Antimicrob. Agents* **2018**, *52*, 747–753. [CrossRef] [PubMed]
8. Bayer, A.S.; Tong, S.Y.C. Case Commentary: Daptomycin Resistance in *Staphylococcus argenteus*-from Northern Australia to San Francisco. *Antimicrob. Agents Chemother.* **2020**, *64*, e01502-20. [CrossRef] [PubMed] [PubMed Central]
9. Shittu, A.O.; Layer-Nicolaou, F.; Strommenger, B.; Nguyen, M.T.; Bletz, S.; Mellmann, A.; Schaumburg, F. First Report of a Methicillin-Resistant, High-Level Mupirocin-Resistant *Staphylococcus argenteus*. *Front. Cell. Infect. Microbiol.* **2022**, *12*, 860163. [CrossRef] [PubMed] [PubMed Central]
10. Kitagawa, H.; Ohge, H.; Hisatsune, J.; Masuda, K.; Aziz, F.; Hara, T.; Kuroo, Y.; Sugai, M. Low incidence of *Staphylococcus argenteus* bacteremia in Hiroshima, Japan. *J. Infect. Chemother.* **2020**, *26*, 140–143. [CrossRef] [PubMed]
11. Jauneikaite, E.; Pichon, B.; Mosavie, M.; Fallowfield, J.L.; Davey, T.; Thorpe, N.; Nelstrop, A.; Sriskandan, S.; Lamb, L.E. *Staphylococcus argenteus* transmission among healthy Royal Marines: A molecular epidemiology case-study. *J. Infect.* **2021**, *83*, 550–553. [CrossRef] [PubMed]
12. Zhan, Y.; Tang, R.; Zhang, Y.; Li, X.; Fan, Y.; Wang, J.; Wang, H.; Lu, Z. Fatal bacteremia caused by *Staphylococcus argenteus*: A case report. *Medicine* **2023**, *102*, e35866. [CrossRef] [PubMed] [PubMed Central]
13. Chen, J.H.K.; Leung, H.Y.; Wong, C.M.C.; Yuen, K.Y.; Cheng, V.C.C. Prevalence and Characteristics of Invasive *Staphylococcus argenteus* among Patients with Bacteremia in Hong Kong. *Microorganisms* **2023**, *11*, 2435. [CrossRef] [PubMed] [PubMed Central]
14. EUCAST—European Committee on Antimicrobal Susceptibility Testing. Available online: https://www.eucast.org/clinical_breakpoints (accessed on 24 October 2023).
15. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027. [CrossRef] [PubMed]
16. Dupieux, C.; Blondé, R.; Bouchiat, C.; Meugnier, H.; Bes, M.; Laurent, S.; Vandenesch, F.; Laurent, F.; Tristan, A. Community-acquired infections due to *Staphylococcus argenteus* lineage isolates harbouring the Panton-Valentine leucocidin, France, 2014. *Euro Surveill.* **2015**, *20*, 21154. [CrossRef] [PubMed]
17. Hsu, J.C.; Wan, T.W.; Lee, H.; Wang, X.M.; Lin, Y.T.; Jung, C.J.; Lee, T.F.; Hsueh, P.R.; Teng, L.J. Heterogeneity of Molecular Characteristics among *Staphylococcus argenteus* Clinical Isolates (ST2250, ST2793, ST1223, and ST2198) in Northern Taiwan. *Microorganisms* **2020**, *8*, 1157. [CrossRef] [PubMed] [PubMed Central]
18. Chandrakanth, R.K.; Raju, S.; Patil, S.A. Aminoglycoside-resistance mechanisms in multidrug-resistant *Staphylococcus aureus* clinical isolates. *Curr. Microbiol.* **2008**, *56*, 558–562. [CrossRef] [PubMed]

19. Lee, J.; Carda-Diéguez, M.; Žiemytė, M.; Vreugde, S.; Cooksley, C.; Crosby, H.A.; Horswill, A.R.; Mira, A.; Zilm, P.S.; Kidd, S.P. Functional mgrA Influences Genetic Changes within a *Staphylococcus aureus* Cell Population over Time. *J. Bacteriol.* **2022**, *204*, e0013822. [CrossRef] [PubMed] [PubMed Central]
20. Conceição, T.; Coelho, C.; de Lencastre, H.; Aires-de-Sousa, M. High Prevalence of Biocide Resistance Determinants in *Staphylococcus aureus* Isolates from Three African Countries. *Antimicrob. Agents Chemother.* **2015**, *60*, 678–681. [CrossRef] [PubMed] [PubMed Central]
21. Olsen, J.E.; Christensen, H.; Aarestrup, F.M. Diversity and evolution of blaZ from *Staphylococcus aureus* and coagulase-negative staphylococci. *J. Antimicrob. Chemother.* **2006**, *57*, 450–460. [CrossRef] [PubMed]
22. Arias, C.A.; Weisner, J.; Blackburn, J.M.; Reynolds, P.E. Serine and alanine racemase activities of VanT: A protein necessary for vancomycin resistance in *Enterococcus gallinarum* BM4174. *Microbiology* **2000**, *146 Pt 7*, 1727–1734. [CrossRef] [PubMed]
23. Peltier, J.; Courtin, P.; El Meouche, I.; Catel-Ferreira, M.; Chapot-Chartier, M.P.; Lemée, L.; Pons, J.L. Genomic and expression analysis of the vanG-like gene cluster of Clostridium difficile. *Microbiology* **2013**, *159 Pt 7*, 1510–1520. [CrossRef] [PubMed]
24. Aung, M.S.; Urushibara, N.; Kawaguchiya, M.; Hirose, M.; Ike, M.; Ito, M.; Kobayashi, N. Distribution of Virulence Factors and Resistance Determinants in Three Genotypes of *Staphylococcus argenteus* Clinical Isolates in Japan. *Pathogens* **2021**, *10*, 163. [CrossRef] [PubMed] [PubMed Central]
25. Kohler, V.; Vaishampayan, A.; Grohmann, E. Broad-host-range Inc18 plasmids: Occurrence, spread and transfer mechanisms. *Plasmid* **2018**, *99*, 11–21. [CrossRef] [PubMed]
26. Yu, S.; Zhou, Y.; Feng, D.; Jiang, Q.; Li, T.; Jiang, G.; Zhou, Z.; Li, H. Whole genome sequence-based characterization of virulence and antimicrobial resistance gene profiles of *Staphylococcus aureus* isolated from food poisoning incidents in eastern China. *Front. Microbiol.* **2023**, *14*, 1225472. [CrossRef] [PubMed] [PubMed Central]
27. Mama, O.M.; Gómez, P.; Ruiz-Ripa, L.; Gómez-Sanz, E.; Zarazaga, M.; Torres, C. Antimicrobial Resistance, Virulence, and Genetic Lineages of Staphylococci from Horses Destined for Human Consumption: High Detection of *S. aureus* Isolates of Lineage ST1640 and Those Carrying the *lukPQ* Gene. *Animals* **2019**, *9*, 900. [CrossRef] [PubMed] [PubMed Central]
28. Holden, M.T.G.; Feil, E.J.; Lindsay, J.A.; Peacock, S.J.; Day, N.P.J.; Enright, M.C.; Foster, T.J.; Moore, C.E.; Hurst, L.; Atkin, R.; et al. Complete Genomes of Two Clinical Staphylococcus Aureus Strains: Evidence for the Rapid Evolution of Virulence and Drug Resistance. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 9786–9791. [CrossRef] [PubMed]
29. Yang, S.; Guo, J.; Kong, Z.; Deng, M.; Da, J.; Lin, X.; Peng, S.; Fu, J.; Luo, T.; Ma, J.; et al. Causal Effects of Gut Microbiota on Sepsis and Sepsis-Related Death: Insights from Genome-Wide Mendelian Randomization, Single-Cell RNA, Bulk RNA Sequencing, and Network Pharmacology. *J. Transl. Med.* **2024**, *22*, 10. [CrossRef]
30. Assimakopoulos, S.F.; Triantos, C.; Thomopoulos, K.; Fligou, F.; Maroulis, I.; Marangos, M.; Gogos, C.A. Gut-Origin Sepsis in the Critically Ill Patient: Pathophysiology and Treatment. *Infection* **2018**, *46*, 751–760. [CrossRef]
31. Liu, A.; Garrett, S.; Hong, W.; Zhang, J. Staphylococcus Aureus Infections and Human Intestinal Microbiota. *Pathogens* **2024**, *13*, 276. [CrossRef]
32. Lim, C.; Mowlaboccus, S.; Daley, D.A.; Shoby, P.; Coombs, G.W. Daptomycin Non-Susceptible Staphylococcus Argenteus Isolated from a Patient without Prior Antibiotic Exposure. *Int. J. Antimicrob. Agents* **2024**, *63*, 107046. [CrossRef] [PubMed]
33. Lee, M.; Choi, Y.; Choi, S.J.; Moon, S.M.; Kim, E.S.; Kim, H.B.; Ahn, S.; Lee, H.; Kim, J.; Shin, D.W.; et al. *Staphylococcus argenteus* Bacteremia in the Republic of Korea. *Microbiol. Spectr.* **2024**, *12*, e02798-23. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

microorganisms

MDPI

Article

Directional and Strain-Specific Interaction Between *Lactobacillus plantarum* and *Staphylococcus aureus*

Sandeep Kondakala [1,†], Sunghyun Yoon [1,†], Soumana Daddy Gaoh [1], Ohgew Kweon [1], Seong-Jae Kim [1] and Mark E. Hart [1,2,*]

[1] Division of Microbiology, National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, AR 72079, USA; sandeep.kondakala@fda.hhs.gov (S.K.); sunghyun.yoon@fda.hhs.gov (S.Y.); soumana.daddy-gaoh@fda.hhs.gov (S.D.G.); oh-gew.kweon@fda.hhs.gov (O.K.); seong-jae.kim@fda.hhs.gov (S.-J.K.)

[2] Department of Microbiology and Immunology, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA

* Correspondence: mark.hart@fda.hhs.gov

† These authors contributed equally to this work.

Abstract: The interaction between *Lactobacillus plantarum* and *Staphylococcus aureus* strains FRI-1169 and MN8, two original isolated strains from menstrual toxic shock syndrome (mTSS) cases, is a key focus for developing non-antibiotic strategies to control *S. aureus*-related infections. While the antagonistic effects of *Lactobacilli* species on *S. aureus* through mechanisms like organic acid and bacteriocin production are known, the molecular dynamics of these interactions remain underexplored. This study employs a proteomic approach to analyze the interactions between *L. plantarum* WCFS1 and *S. aureus* strains, FRI-1169 and MN8, during co-culture. We profiled differentially expressed proteins (DEPs) found in the spent media and cytosols of both bacteria, revealing distinct directional and strain-specific responses. The findings demonstrate that *L. plantarum* exerts a more pronounced effect on *S. aureus*, with more DEPs and upregulated proteins, while *S. aureus* showed fewer DEPs and more downregulated proteins. These strain-specific interactions highlight the complex metabolic and regulatory adjustments between these bacterial species. This research provides valuable insights into the molecular mechanisms of *Lactobacillus-S. aureus* antagonism and underscores the potential of proteomic analysis as a powerful tool for studying bacterial dynamics in co-culture systems.

Keywords: antagonism; *Lactobacillus plantarum*; non-antibiotic biocontrol agent; proteomics; *Staphylococcus aureus* infection

1. Introduction

Staphylococcus aureus strains FRI-1169 and MN8 were originally isolated from menstrual toxic shock syndrome (mTSS) cases in the late 1970s [1–4]. These strains have become significant models for studying TSS due to their unique toxin production profiles [5]. While both strains produce toxic shock syndrome toxin-1 (TSST), *S. aureus* MN8, classified as a USA200 pulsed-field gel electrophoresis (PFGE) type, is known for being a low producer of α-toxin due to a nonsense mutation within the open reading frame of α-toxin, whereas *S. aureus* FRI-1169, though not part of the USA200 clonal type, is recognized for its high α-toxin production [5]. Despite the extensive knowledge of these strains, there remains a need for comprehensive genotypic and phenotypic evaluations, particularly concerning their roles in mTSS. The incidence of mTSS in the United States is rare, approximately 1 case/100,000 among menstrual age women (13–54 years of age) [6,7]. However, the fact that 5–10% of cases end in death warrants further efforts to find alternative approaches other than antibiotic therapy that can be developed for prevention and treatment before an event such as an epidemic with an antibiotic-resistant, TSST-1 producing strain of *S. aureus* occurs [8,9]. This becomes even more critical in light of the potential re-emergence of

S. aureus mTSS clonal groups and the rise of antibiotic-resistant, TSST-1-producing strains of *S. aureus* [10–12].

The interaction between *Lactobacillus* species and *S. aureus* has attracted significant interest for its potential in non-antibiotic control of *S. aureus*-related infections, including toxic shock syndrome (TSS) and other serious conditions. *Lactobacillus* species, valued for their probiotic benefits, exhibit antagonistic effects on *S. aureus* through a variety of mechanisms, including the production of organic acids, bacteriocins, and hydrogen peroxide, as well as competition for nutrients [13,14]. Despite these findings, the precise molecular mechanisms underlying these interactions, especially at the proteomic level, are not yet fully understood. Uncovering these molecular dynamics is essential for the further development of *Lactobacillus* species as a biotherapeutic means for infection control.

Proteomic analysis has emerged as a crucial tool for elucidating bacterial interactions, providing detailed insights into cellular responses and metabolic adjustments. Previous research has often focused on isolated factors, such as the production of organic acids or bacteriocins, without exploring the broader, overall protein response as the result of interactions between *Lactobacillus* and *S. aureus*. Our group demonstrated that acid production by *Lactobacillus* species inhibits *S. aureus* MN8 but did not fully investigate the molecular mechanisms involved [15]. Additionally, while *L. plantarum* has well-established antagonistic properties, its specific effects on different *S. aureus* strains have not been extensively characterized. This study seeks to fill this gap by using proteomic data to explore interactions between *L. plantarum* and two *S. aureus* strains, FRI-1169 and MN8, with a particular focus on differentially expressed proteins (DEPs) and their functional roles in bacterial dynamics. This study utilizes a co-culture system that physically separates *S. aureus* from *L. plantarum* during growth and, thus, allows spent media produced by one species to interact with the other species without cell-to-cell contact. In addition, we were also able to isolate cells of each species and determine cytosolic differences as a result of this interaction.

In this study, we present the first comprehensive proteomic analysis of interactions between *L. plantarum* and *S. aureus*. By profiling DEPs in the spent media and cytosols of both *L. plantarum* and the two *S. aureus* strains during co-culture, this research sheds light on the metabolic and regulatory pathway changes in response to each other. The findings presented enhance our understanding of *Lactobacillus*-*S. aureus* antagonism and demonstrate the utility of proteomic analysis as a powerful tool for studying complex bacterial dynamics in co-culture systems.

2. Materials and Methods

2.1. Bacterial Strains, Strain Maintenance, and Growth Conditions

Staphylococcus aureus clinical strains FRI-1169 (1980 WI; NCH 240) and MN8 (1980 MN; NCH 245) were isolated from cases of mTSS and were provided by P.D. Fey at the University of Nebraska Medical Center and were confirmed to produce toxic shock syndrome toxin-1 (TSST-1) by western analysis. Strains were maintained as frozen (−80 °C) stocks in brain heart infusion (BHI; Difco Laboratories, Detroit, MI, USA) broth containing 25% (*w*/*v*) glycerol and streaked for isolation on either tryptic soy broth (TSB; Difco Laboratories) or TSB containing 1.5% agar. *Lactobacillus plantarum* WCFS1 (HLH 10) was obtained from Wageningen Center for Food Sciences, Wageningen, Netherlands, and maintained as a frozen stock in Lactobacilli deMan, Rogosa and Sharpe (MRS) broth (Difco Laboratories) containing 10–15% (*w*/*v*) glycerol and streaked for isolation on MRS containing 1.5% agar. While *Lactobacillus* species, namely *L. crispatus*, *L. gasseri*, *L. iners*, and *L. jensenii* are the predominant members of this genus found in most vaginal tracts of healthy women [16–18], *L. plantarum* can be found in the urogenital tract of women but its numbers appear to be low [16,17,19], thus leaving questions as to its role as a member of the vaginal microbiota. Nevertheless, our decision to utilize *L. plantarum* WCFS1 [20] for most of our studies thus far, was made primarily due to convenience. *Lactobacillus plantarum* WCFS1 was the first of the genus *Lactobacillus* to have its genome completely sequenced and annotated [21,22],

thus facilitating the molecular studies to isolate and clone important *L. plantarum* genetic determinants for the purpose of cloning the lysostaphin gene (*end*) into *L. plantarum* WCFS1 [20].

2.2. Co-Culture and Sample Preparation

Cells of *L. plantarum* WCFS1 and *S. aureus* strains FRI-1169 and MN8 were cultured overnight (15–18 h) in MRS and TSB media, respectively, using a flask to a medium volume of 2.5. A 1 mL aliquot of each overnight culture was transferred to a sterile microcentrifuge tube, washed twice with 1 mL of sterile, deionized water, and once with 1 mL of buffered, Hart modified genital tract secretion (HmGTS) medium [23]. The medium was made without the inclusion of bovine serum albumin to facilitate bacterial protein identification by mass spectrophotometry. The cells were then suspended in 1 mL of HmGTS media, and optical density (OD) readings were determined for each culture spectrophotometrically at 550 nm. Optical density values were used to dilute cultures to approximately 1×10^7 colony-forming units (cfu)/mL using a previously determined standard curve of cfu as a function of OD. Two ml of *S. aureus* cell suspensions were then aseptically pipetted into each well of a six-well BD Falcon cell culture plate (Becton Dickinson, Franklin Lakes, NJ, USA). Cup-shaped inserts with 0.4 μm filters were aseptically placed into each well, and 2 mL of *L. plantarum* cell suspensions were aseptically pipetted into each insert. Control samples consisted of 2 mL cell suspensions of either *S. aureus* or *L. plantarum* aseptically pipetted into both the well and insert. Plates were incubated under stationary, aerobic conditions at 37 °C prior to harvesting both well and insert contents at 8 and 24 h of growth. Viable cell counts were determined using a portion of each of the cell suspensions at 0, 8, and 24 h using 10-fold serial dilutions and plating on mannitol salt agar (MSA), which is selective for *S. aureus*, and Rogosa agar (RA), which is selective for *L. plantarum*. The remaining volume of each cell suspension was centrifuged (12,000× *g*, for 10 min at 4 °C) to separate spent media from cells. Spent media were filter-sterilized into 15 mL polypropylene, screw-capped, conical centrifuge tubes using 0.2 μm syringe filters. Portions of each filter-sterilized spent medium were pipetted into 4 mL capacity, Centricon tubes (Centricon Ultracel YM-3 filters with 3000 MWCO, Millipore Corporation, Bedford, ME, USA) and centrifuged (7500× *g*) at 4 °C for 1 h. The retentates were recovered and stored at –80 °C until needed. Cell pellets from the first centrifugation were washed one time in one ml of TE (10 mM and 1 mM EDTA) buffer (pH 8.0), suspended in 0.4 mL of TE, and transferred to FastPrep blue tubes (Qbiogene, Irvine, CA, USA) containing acid-washed and RNase free, 0.1 mm silica beads and lysed using the FP120 reciprocator (Qbiogene) with a setting of 6 m/second for 40 s. Cell lysates were immediately cooled on ice for 15 min and then subjected to centrifugation (10,000× *g* for 1 min at 4 °C). Supernatants were recovered and stored at –80 °C.

2.3. Protein Quantification and Sample Quality Assessment

The protein concentration for each sample was determined using the bicinchoninic acid (BCA) assay (Pierce Biotechnology, Waltham, MA, USA) following the manufacturer's instructions. The quality of the processed samples was assessed using NuPAGE Tris-Bis mini gels (4–12% gradient) according to the manufacturer's instructions (Invitrogen Life Technologies, Carlsbad CA, USA).

2.4. Protein Digestion and Peptide Preparation

An aliquot of the protein extract (50 μg) was reduced with 12 mM dithiothreitol (DTT) at room temperature for 1 h. This was followed by alkylation with 15 mM iodoacetamide (IAA) at room temperature for 1 h. Trypsin (sequencing grade, Promega, Madison, WI, USA) was added at a 1:20, enzyme to substrate dilution, and the digestion was carried out overnight at 37 °C. The digested peptides were acidified with 1% formic acid and desalted using C18 solid-phase extraction columns (Waters, Milford, MA, USA). Peptides were then dried in a vacuum concentrator and stored at –80 °C until mass spectrometry analysis.

2.5. Mass Spectrometry Analysis

For all samples, 1 μg was analyzed by nano LC/MS with a Waters M-class HPLC system interfaced with an Orbitrap Exploris 480 Mass Spectrometer (ThermoFisher, Walthma, MA, USA). For generation of the DIA Chromatogram Library, six gas-phase fractions (GPF) injections were acquired for six ranges: 396 to 502, 496 to 602, 596 to 702, 696 to 802, 796 to 902, and 896 to 1002. Sequentially, full scan MS data (60,000 FWHM resolution) was followed by 26 × 4 m/z precursor isolation windows, another full scan and 26 × 4 m/z windows staggered by 2 m/z. The products were acquired at 30,000 FWHM resolution. The automatic gain control (AGC) target was set to 1×10^6 for both full MS and product ion data. The maximum ion inject time (IIT) was set to 50 ms for full MS and "dynamic" mode for products with nine data points required across the peak and the NCE was set to 30.

For sample analysis, full scan MS data (60,000 FWHM resolution) from m/z 385 to 1015 was followed by 61 × 10 m/z precursor isolation windows, another full scan from m/z 385 to 1015 was followed by 61 × 10 m/z windows staggered by 5 m/z. The products were acquired at 15,000 FWHM resolution. The maximum ion inject time (IIT) was set to 50 ms for full MS and "dynamic" mode for products with nine data points required across the peak and the NCE was set to 30. An injection of the pool was included at the start and end of the batch.

Data were processed using Scaffold for data-independent acquisition (Scaffold DIA 3.2.1; Proteome Software) which incorporated the following steps: raw MS files were converted to mzML (ProteoWizard 3.0.19254), including deconvolution of staggered windows and final conversion to DIA format. The alignment was based on retention times and used to create the custom DIA-based chromatogram libraries (ELIB); we searched data using the Prosit library for data-dependent acquisition (DLIB) and the chromatogram/reference library. Filtering of the database was performed using Percolator at 1% peptide false discovery rate (FDR). Peak areas for detected peptides were calculated using Encyclopedia (0.9.6). For each peptide, the five highest-quality fragment ions were selected for quantification. The data was searched with the following parameters: Enzyme: Trypsin, Databases: Uniprot *L. plantarum* and *S. aureus*, Fixed Modification: Carbamidomethyl (C); Precursor Mass Tolerance, Fragment Mass Tolerance and Library Fragment Tolerance were set at 10 ppm, Peptide and Protein FDR set at 0.01, Peptide Length: 6–30 AA, Maximum Missed Cleavages and Minimum Peptides set to 1 and Peptide charge set at 2–3.

2.6. Bioinformatic Analysis

Differentially expressed proteins (DEPs) were identified by filtering the proteomic data with a fold ratio of ≥ 2.0 based on normalized protein intensity values. The identified differentially expressed proteins (DEPs) were analyzed for Cluster of Orthologous Groups (COG) for functional categories, and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (http://www.genome.jp/kegg/ accessed on 7 August 2024) was used for pathway mapping [24]. In this study, the hypergeometric test was used to estimate the probability of finding a specific number of proteins associated with a particular function or pathway within the sample, taking into account the total number of proteins with that function in the entire population and the size of the sample. Figures and statistical analyses were generated using an in-house Python (3.10.0) script.

3. Results and Discussion

By enhancing our understanding of the mechanisms underlying *Lactobacillus–S. aureus* interactions, new avenues for the biocontrol of *S. aureus* using non-antibiotic methods that do not harm bacterial ecosystems are being explored [25–27]. The interaction between *L. plantarum* and *S. aureus* has been extensively studied to understand their dynamics [15,20,28]. Functional genomic approaches, such as transcriptomic and proteomic analyses, are considered powerful tools for investigating *Lactobacillus–S. aureus* interactions in terms of antagonistic mechanisms of *Lactobacillus* and cellular responses of *S. aureus*. To our knowledge, this study is the first to demonstrate that these bacterial interactions can

be effectively analyzed through proteome analysis. We profiled the proteins in the spent media and cytosols of *S. aureus* strains FRI-1169 and MN8, as well as *L. plantarum* WCFS1, during co-culture, focusing on the proteins that were differentially expressed. Identifying these DEPs not only deepens our understanding of the specific interactions between these bacterial species but also underscores the potential of proteome analysis as a method for studying bacterial dynamics.

3.1. Interaction between L. plantarum WCFS1 and S. aureus FRI-1169

3.1.1. Comparative Analysis of Proteins from *L. plantarum* WCFS1 During Co-Culture with *S. aureus* FRI-1169

To assess the response of *L. plantarum* WCFS1 during co-culture with *S. aureus* FRI-1169, we examined the protein expression profiles in both the spent medium and cellular cytosol of *L. plantarum* WCFS1 (Figure 1). Compared to control cultures, a total of 354 proteins were identified in the spent medium at 8 and 24 h, with 231 and 241 proteins differentially expressed at these time points, respectively. In the cytosol, 1328 proteins were identified, of which 160 and 485 proteins were differentially expressed at 8 and 24 h. Notably, many of the DEPs were upregulated, suggesting an active metabolic and cellular response by *L. plantarum* to *S. aureus*. This upregulation may reflect mechanisms that allow *L. plantarum* to enhance its antagonistic activity or adapt to environmental changes in the co-culture, highlighting the dynamic nature of bacterial interactions and the potential of *L. plantarum* to reinforce metabolic pathways to either sustain its own growth or inhibit *S. aureus*.

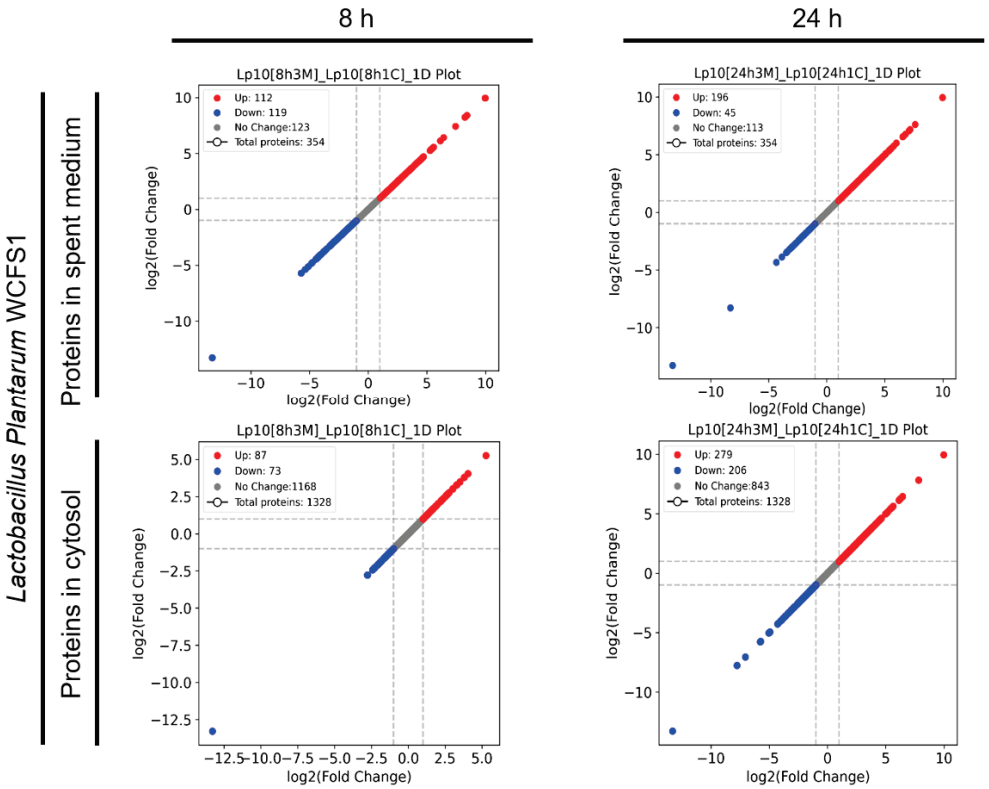


Figure 1. A comparative analysis of protein expression profiles in the spent medium and cytosol of *L. plantarum* WCFS1 during co-culture with *S. aureus* FRI-1169 at 8 h and 24 h.

To further elucidate the biological processes and pathways involved in the interaction between *L. plantarum* WCFS1 and *S. aureus* FRI-1169, COG, and KEGG enrichment analyses were performed on the identified proteins from *L. plantarum* WCFS1. The COG enrichment analysis of DEPs from both the spent medium and cytosol at 8 and 24 h highlighted several key functional categories (Figure 2A). In the COG enrichment analysis of the DEPs in the spent medium, prominent among these were "translation, ribosomal structure, and

biogenesis (J)", "nucleotide transport and metabolism (F)", "amino acid transport and metabolism (E)", and "cell wall/membrane/envelope biogenesis (M)", indicating active engagement by *L. plantarum* in essential metabolic and structural processes throughout the co-culture period. On the other hand, in the COG enrichment analysis of the DEPs in the cytosol, at 8 h, the prominent pathways of *L. plantarum* primarily were nutrient acquisition and metabolic activity, with significant enrichment in "amino acid transport and metabolism (E)" and "coenzyme transport and metabolism (H)". By 24 h, the metabolism shifted toward processes related to gene regulation and genomic stability, with enrichment observed in "transcription (K)" and "replication, recombination, and repair (L)". Additionally, consistent enrichment in "carbohydrate transport and metabolism (G)" throughout the experiment, along with later shifts toward "lipid transport and metabolism (I)" and "energy production and conversion (C)", reflect the ongoing metabolic adjustments made by *L. plantarum* WCFS1 in response to co-culture with *S. aureus* FRI-1169. These observations suggest that *L. plantarum* dynamically regulates its metabolic and structural processes, initially prioritizing nutrient acquisition and later focusing on cellular maintenance and genomic integrity as the co-culture progresses. This indicates a well-coordinated response that could allow *L. plantarum* to adapt to the environmental conditions presented by the presence of *S. aureus*.

The KEGG enrichment analysis of proteins from the spent medium of *Lactobacillus plantarum* WCFS1 during co-culture with *Staphylococcus aureus* FRI-1169 revealed key metabolic and biosynthetic activities at both 8 and 24 h (Figure 2B). In the spent medium, at 8 h, the prominently enriched pathways included "metabolic pathways", "aminoacyl-tRNA biosynthesis", and "biosynthesis of secondary metabolites", suggesting that *L. plantarum* was actively engaged in core metabolic and biosynthetic processes early in the co-culture. By 24 h, these pathways remained highly enriched, with additional enrichment observed in pathways such as "amino sugar and nucleotide sugar metabolism" and "two-component system", reflecting ongoing metabolic activities and the emergence of regulatory responses as the co-culture progressed.

In the cytosol, *L. plantarum* WCFS1 exhibited significant enrichment at 8 h in pathways related to "phenylalanine, tyrosine, and tryptophan biosynthesis", "streptomycin biosynthesis", and "inositol phosphate metabolism", indicating robust early biosynthetic activity. Furthermore, enrichment in pathways like "quorum sensing" and "peroxisome" pointed to broader regulatory mechanisms being activated in response to *S. aureus*. By 24 h, the enrichment of the top 20 pathways increased significantly, suggesting a more dynamic and adaptive metabolic response in the later stages of co-culture. These results highlight how *L. plantarum* adjusts its metabolic and regulatory pathways over time, likely in response to interactions with *S. aureus*, underscoring the complexity of the bacterial interactions.

To analyze the cellular response between *L. plantarum* WCFS1 and *S. aureus* FRI-1169 and between *L. plantarum* WCFS1 and *S. aureus* MN8, we examined the expression of proteins related to antibiotic resistance, heavy metal resistance, and virulence factors (Supplementary Tables S1 and S2). These functions are often used to assess bacterial status in unfavorable environments, indicating how bacteria adapt, survive, and potentially thrive despite stressors. Additionally, other proteins with notable upregulation were analyzed for their roles during co-culture. Understanding these protein expressions would help elucidate the mechanisms employed by each bacterium in response to the challenging environment.

In the genome of *L. plantarum* WCFS1, 29 heavy metal resistance genes were identified, with 20 proteins detected, 10 of which were upregulated. Proteins like CopY/TcrY and CopB may regulate internal conditions [28], while YieF and ChrR are associated with oxidative stress management [29], although these responses likely reflect regular metabolic activities rather than stress responses. Among the 37 identified virulence factors, 24 proteins were detected, with 8 showing upregulation. The Clp proteins are involved in maintaining protein quality [30–32].

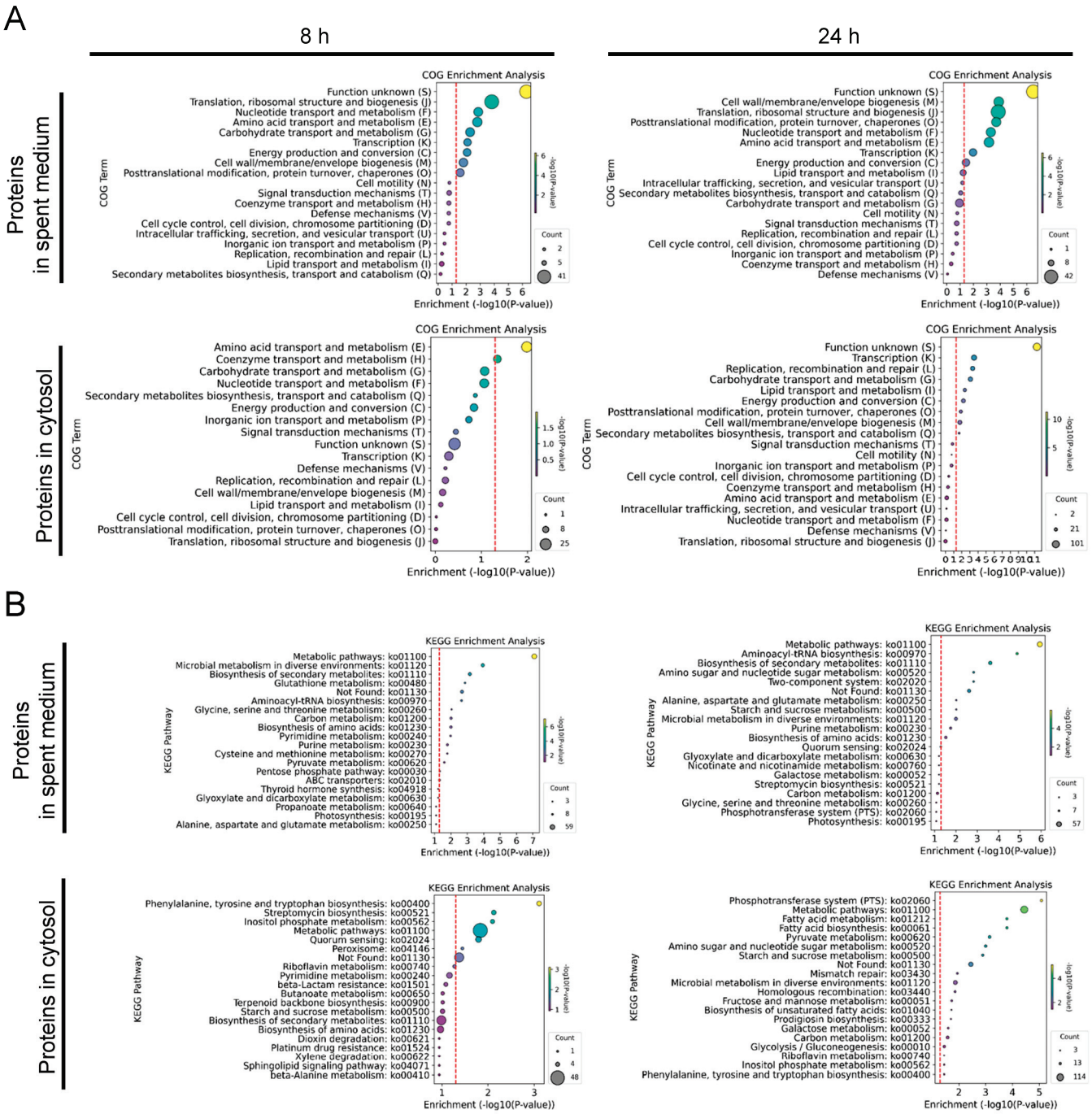


Figure 2. Functional and pathway analysis of proteins of *L. plantarum* WCFS1 during co-culture with *S. aureus* FRI-1169. COG enrichment analysis of DEPs in the spent medium and cytosol of *L. plantarum* WCFS1 at 8 and 24 h (**A**). KEGG enrichment analysis of proteins of *L. plantarum* WCFS1 at 8 and 24 h (**B**). The size of each circle represents the number of proteins. The hue of the circles symbolizes −log10(*p*-value) values. The red dashed lines represent the value of −log10(*p*-value = 0.05).

3.1.2. Comparative Analysis of Proteins from *S. aureus* FRI-1169 During Co-culture with *L. plantarum* WCFS1

The growth inhibition of *S. aureus* FRI-1169 by *L. plantarum* WCFS1 was supported by proteomic evidence. To investigate the response of *S. aureus* FRI-1169 during co-culture with *L. plantarum* WCFS1, the protein expression profiles from both the spent medium and cellular cytosol were analyzed (Figure 3). At 8 and 24 h, 451 proteins were identified in

the spent medium, with 242 and 118 proteins differentially expressed at these respective time points. Similarly, 1515 cytosolic proteins were identified, with differential regulation observed in 95 proteins at 8 h and 87 at 24 h.

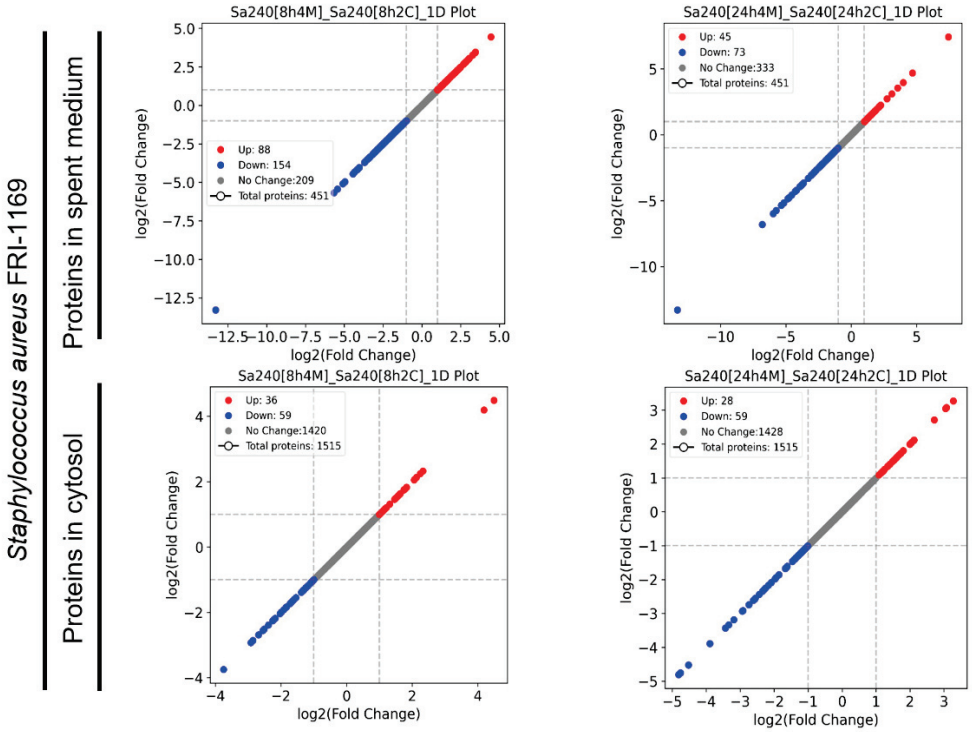


Figure 3. A comparative analysis of protein expression profiles in the spent medium and cytosol of *S. aureus* FRI-1169 during co-culture *L. plantarum* WCFS1 at 8 h and 24 h.

COG enrichment analysis of DEPs from the spent medium revealed dynamic changes over time (Figure 4A). At 8 h, there was significant enrichment in "translation, ribosomal structure and biogenesis (J)", "inorganic ion transport and metabolism (P)", and "defense mechanisms (V)", suggesting that *S. aureus* was mounting early responses focused on protein synthesis, ion regulation, and managing stress, likely in reaction to the acidic environment created by *L. plantarum*. By 24 h, the emphasis shifted towards categories like "transcription (K)", "amino acid transport and metabolism (E)", and "coenzyme transport and metabolism (H)", indicating an adaptive response involving gene regulation and nutrient metabolism as *S. aureus* adjusted to the co-culture. In the cytosol, no significant enrichment was detected.

KEGG enrichment analysis of DEPs in the spent medium at 8 h showed significant enrichment in pathways related to protein synthesis and general metabolism, including "metabolic pathways", "ribosome", and "aminoacyl-tRNA biosynthesis" (Figure 4B). Other pathways, such as "ABC transporters" and "glycerolipid metabolism", were enriched, indicating adaptations in nutrient uptake and lipid processing. By 24 h, continued enrichment in pathways such as "methane metabolism" and "RNA degradation" highlighted sustained metabolic activity (Figure 4B). In the cytosol, at 8 h, enrichment in pathways like "glycerolipid metabolism" and "two-component system" pointed to early metabolic adaptations. Pathways like "pantothenate and CoA biosynthesis" and "arginine biosynthesis" suggest ongoing nutrient metabolism and homeostasis, particularly in response to the acidic conditions imposed by *L. plantarum*. By 24 h, amino acid, and lipid metabolism remained central, with significant enrichment in "fatty acid degradation" and "arginine biosynthesis", reflecting further adaptation and stress resistance mechanisms.

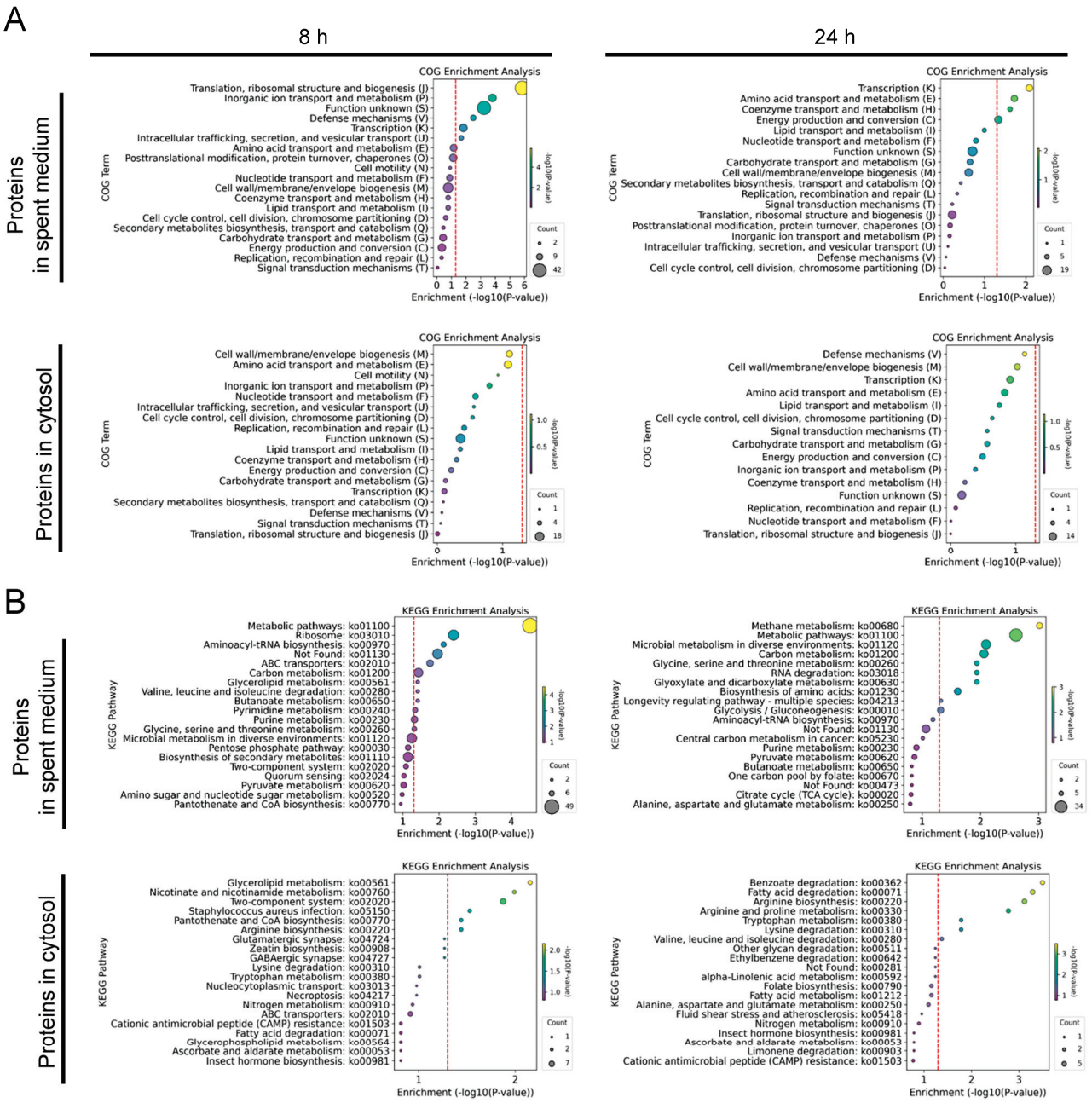


Figure 4. Functional and pathway analysis of proteins of *S. aureus* FRI-1169 during co-culture with *L. plantarum* WCFS1. COG enrichment analysis of DEPs in the spent medium and the cytosol of *S. aureus* FRI-1169 at 8 and 24 h (**A**). KEGG enrichment analysis of proteins of *S. aureus* FRI-1169 at 8 and 24 h (**B**). The size of each circle represents the number of proteins. The hue of the circles symbolizes −log10(*p*-value) values. The red dashed lines represent the value of −log10(*p*-value = 0.05).

Notably, *S. aureus* FRI-1169 exhibited upregulation of proteins related to antibiotic resistance and stress response. Among the 23 antibiotic resistance genes identified, 12 corresponding proteins were detected during co-culture, including three upregulated proteins: MgrA, RpoB2, and VanR. MgrA, a global transcriptional regulator, modulates resistance mechanisms and is in involved with virulence gene regulation [33–35]. RpoB2, a subunit of RNA polymerase, likely plays a role in regulating gene expression under these stress conditions [36], while VanR, a regulator of vancomycin resistance, may help maintain cell wall integrity [37]. Additionally, out of the 123 identified virulence factors, 76 corresponding proteins were detected, with 16 upregulated, including ClpB, ClpC, GroL, and others

involved in stress responses and protein quality control [38,39]. These proteins likely help *S. aureus* manage the harsh conditions of the co-culture, ensuring its survival despite the antagonistic effects of *L. plantarum*.

3.2. Interaction Between L. plantarum WCFS1 and S. aureus MN8

3.2.1. Comparative Analysis of Proteins from *L. plantarum* WCFS1 During Co-Culture with *S. aureus* MN8

To evaluate the response of *L. plantarum* WCFS1 when co-cultured with *S. aureus* MN8, we analyzed the protein expression profiles from both the spent medium and cytosol (Figure 5). At 8 and 24 h, 380 and 381 proteins were identified in the spent medium, respectively, with 94 and 144 proteins showing differential expression at these time points. In the cytosol, 1189 proteins were detected, with 64 proteins differentially expressed at 8 h and 316 at 24 h. This differential regulation reflects an adaptive response by *L. plantarum* in co-culture with *S. aureus* MN8, including shifts in metabolic processes and cellular activities over time.

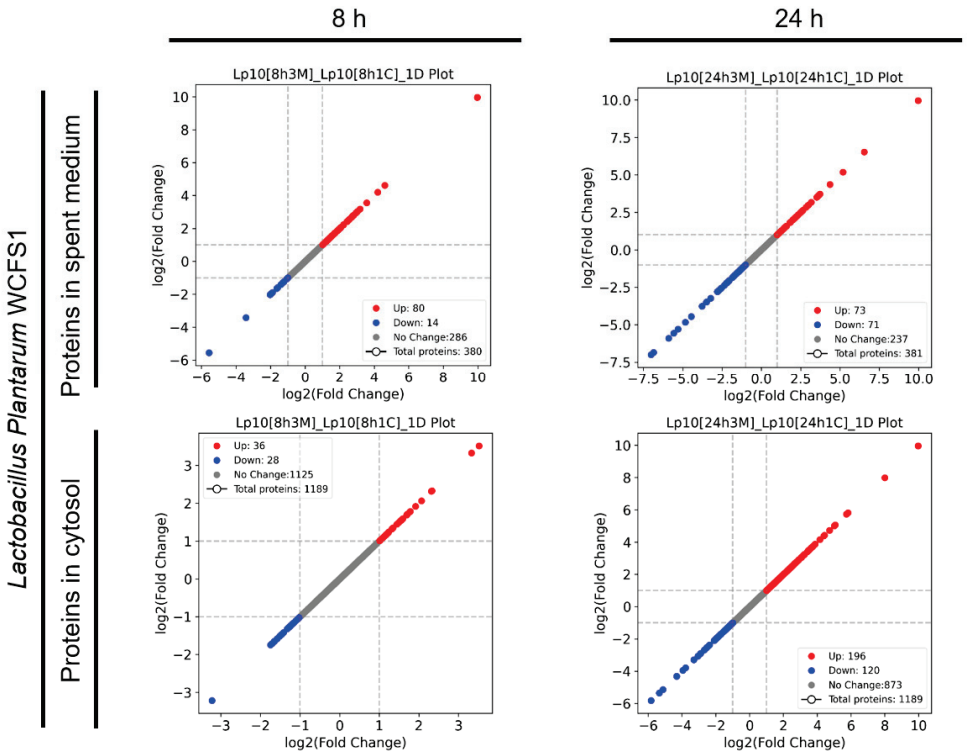


Figure 5. A comparative analysis of protein expression profiles in the spent medium and cytosol of *L. plantarum* WCFS1 during co-culture with *S. aureus* MN8 at 8 h and 24 h.

COG enrichment analysis of the DEPs revealed dynamic functional responses in *L. plantarum* during co-culture (Figure 6A). At 8 h, key categories such as "amino acid transport and metabolism (E)" and "nucleotide transport and metabolism (F)" were enriched, indicating early metabolic activity. By 24 h, there was a notable shift toward "translation, ribosomal structure, and biogenesis (J)", suggesting an increased focus on protein synthesis as the co-culture progressed. In the cytosol, early enrichment at 8 h in categories such as "amino acid transport and metabolism (E)" and "signal transduction mechanisms (T)" indicated active nutrient acquisition and cellular signaling. By 24 h, the focus shifted to "transcription (K)" and "cell wall/membrane/envelope biogenesis (M)", reflecting gene regulation and structural functions in response to the ongoing interaction.

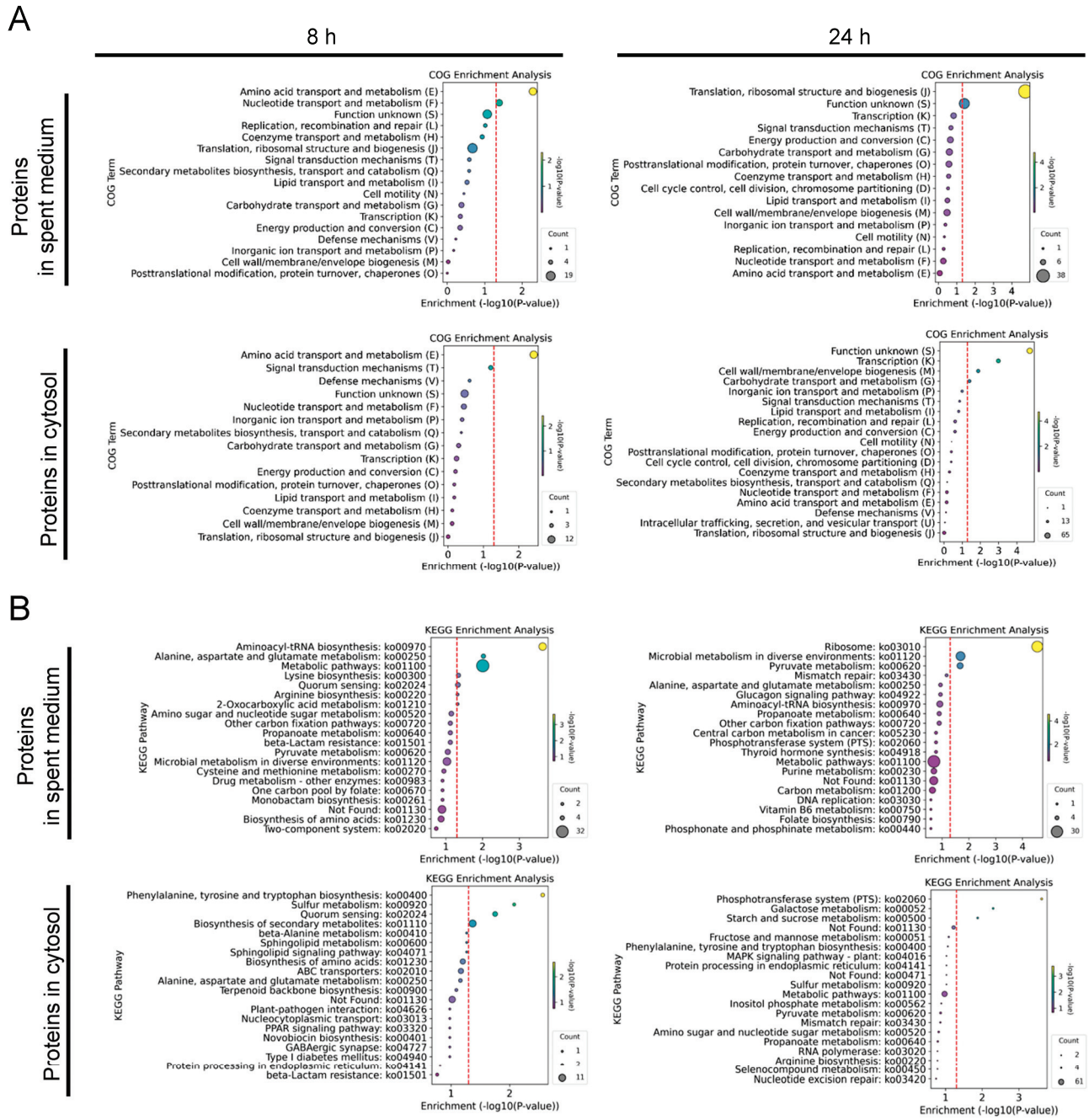


Figure 6. Functional and pathway analysis of proteins from *L. plantarum* WCFS1 during co-culture with *S. aureus* MN8. COG enrichment analysis of DEPs in the spent medium and cytosol of *L. plantarum* WCFS1 at 8 and 24 h (**A**). KEGG enrichment analysis of proteins from *L. plantarum* WCFS1 at 8 and 24 h (**B**). The size of each circle represents the number of proteins. The hue of the circles symbolizes −log10(*p*-value) values. The red dashed lines represent the value of −log10(*p*-value = 0.05).

KEGG enrichment analysis of the DEPs further highlighted the metabolic adjustments made by *L. plantarum* (Figure 6B). At 8 h, pathways like "aminoacyl-tRNA biosynthesis", "alanine, aspartate, and glutamate metabolism", and "lysine biosynthesis" were enriched, pointing to active protein synthesis and amino acid metabolism. By 24 h, pathways involved in "ribosome" and "pyruvate metabolism" were prominent, reflecting continued protein synthesis and energy production. In the cytosol, early enrichment in "phenylalanine, tyrosine, and tryptophan biosynthesis" and "sulfur metabolism" suggested normal

metabolic activity, while by 24 h, enrichment in the "phosphotransferase system (PTS)" and "starch and sucrose metabolism" indicated ongoing carbohydrate metabolism.

Out of the 29 heavy metal resistance genes of *L. plantarum* WCFS1, 18 corresponding proteins were detected during co-culture with *S. aureus* MN8, with 5 upregulated, including GalE, RecG, and Rrp2. Although no heavy metal stress was present, the upregulation suggests cellular processes related to environmental adaptation, with GalE maintaining cellular functions [39] and RecG and Rrp2 supporting cellular maintenance [40,41]. Among 37 virulence factors identified, 24 corresponding proteins were detected, with 8, such as ClpE, ClpB, Tuf, GroL, PpiB, RfbB, TagH, and KatA, upregulated [42].

3.2.2. Comparative Analysis of Proteins from *S. aureus* MN8 During Co-Culture with *L. plantarum* WCFS1

To evaluate how *S. aureus* MN8 responds to co-culture with *L. plantarum* WCFS1, the protein expression profiles in both the spent medium and cellular cytosol were analyzed. As shown in Figure 7, protein expression in the spent medium of *S. aureus* MN8 at 8 and 24 h was compared to that in the control culture. A total of 591 proteins were identified at both time points, with 147 proteins differentially expressed at 8 h and 31 proteins at 24 h. In the cytosol, 1442 proteins were identified, with differential regulation observed in 43 proteins at 8 h and 158 proteins at 24 h.

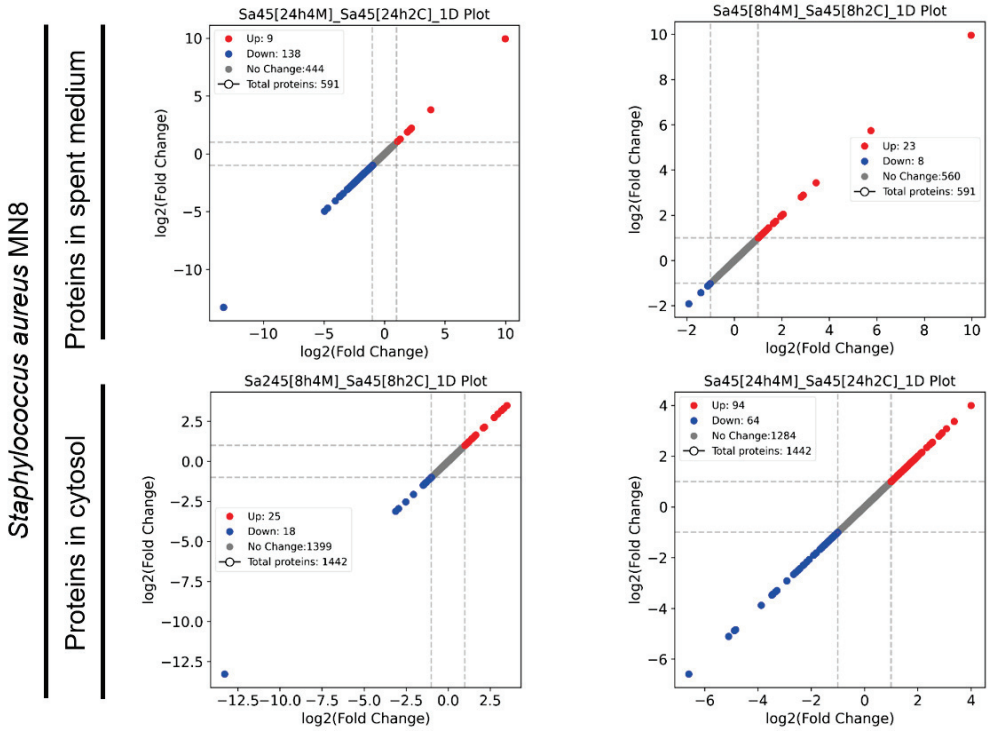


Figure 7. A comparative analysis of protein expression profiles in the spent medium and cytosol of *S. aureus* MN8 during co-culture with *L. plantarum* WCFS1at 8 h and 24 h.

COG enrichment analysis (Figure 8A) revealed that at 24 h, DEPs in the spent medium of *S. aureus* MN8 showed enrichment in categories such as "energy production and conversion", "lipid transport and metabolism (I)", and "amino acid transport and metabolism". These enrichments indicate active metabolic processes essential for cellular adaptation to environmental changes, potentially including adjustments to pH fluctuations caused by *L. plantarum*. In the cytosolic proteins, *S. aureus* MN8 exhibited early metabolic activity at 8 h, with enrichment in "amino acid transport and metabolism", "secondary metabolites biosynthesis, transport, and catabolism (Q)", and "coenzyme transport and metabolism

(H)". By 24 h, the focus remained on "amino acid transport and metabolism", suggesting continued engagement in essential metabolic processes.

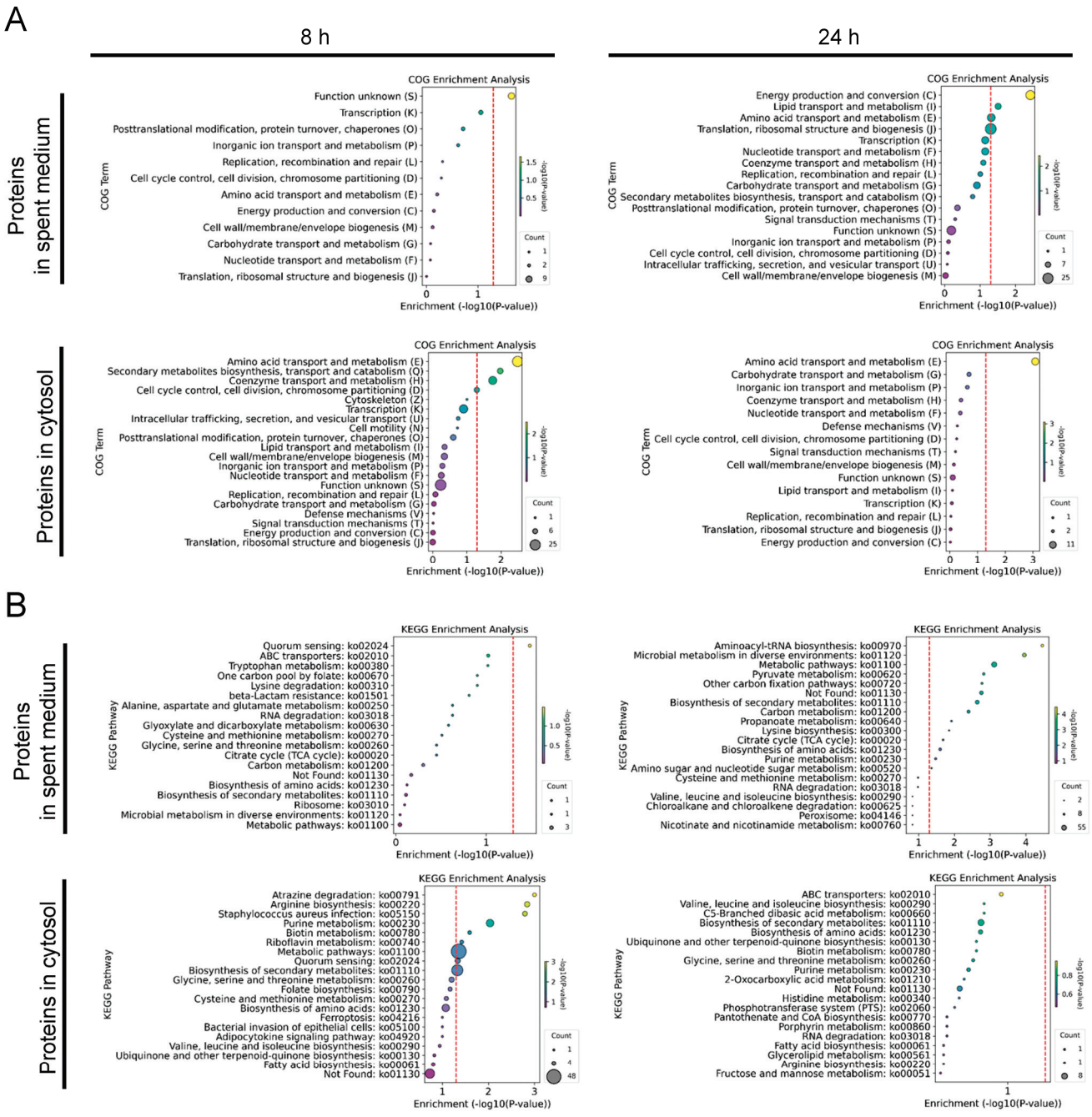


Figure 8. Functional and pathway analysis of proteins of *S. aureus* MN8 during co-culture with *L. plantarum* WCFS1. COG enrichment analysis of DEPs in the spent medium and the cytosol of *S. aureus* MN8 at 8 and 24 h (**A**). KEGG enrichment analysis of proteins of *S. aureus* MN8 at 8 and 24 h (**B**). The size of each circle represents the number of proteins. The hue of the circles symbolizes −log10(*p*-value) values. The red dashed lines represent the value of −log10(*p*-value = 0.05).

KEGG enrichment analysis of the DEPs (Figure 8B) highlighted key pathways in the spent medium at 8 h, including significant enrichment in "quorum sensing", indicating that bacterial communication was actively occurring early in the co-culture. By 24 h, pathways such as "aminoacyl-tRNA biosynthesis", "microbial metabolism in diverse environments", and "metabolic pathways" were enriched, reflecting continued protein synthesis, metabolic

activity, and microbial interactions. Further enrichment in "pyruvate metabolism", "biosynthesis of secondary metabolites", and "citrate cycle (TCA cycle)" suggested ongoing energy production and carbon metabolism as *S. aureus* MN8 adapted to the co-culture environment. In the cytosol, enrichment at 8 h in pathways like "atrazine degradation", "arginine biosynthesis", and "*Staphylococcus* aureus infection" pointed to roles in nitrogen processing, amino acid biosynthesis, and virulence mechanisms. Enrichment in "purine metabolism", "riboflavin metabolism", and "quorum sensing" at 8 h indicated nucleotide metabolism and bacterial communication during the early stages of interaction.

In the genome of *S. aureus* MN8, 23 antibiotic resistance genes, 36 heavy metal resistance genes, and 116 virulence factor genes were identified, with several corresponding proteins detected during co-culture with *L. plantarum* WCFS1. Among these, the virulence factors SdrF, UreC, UreG, and FabZ were upregulated. UreC and UreG, components of the urease system, likely played a role in maintaining pH by converting urea into ammonia, helping *S. aureus* neutralize the acidic environment induced by *L. plantarum* [43].

3.3. Differential Expression of Proteins from L. plantarum WCFS1 During Co-Culture

3.3.1. Protein Expression Profiles of *L. plantarum* WCFS1 During Co-Culture

We assessed whether the protein expression profiles of *L. plantarum* WCFS1 are consistent when co-cultured with two different strains of *Staphylococcus aureus*—FRI-1169 and MN8 (Supplementary Table S3). The Venn diagram (Figure 9) illustrates the comparison of differentially expressed proteins between the two co-cultures. In the spent medium, 41 and 67 proteins (25.7%) were differentially expressed in the co-cultures with *S. aureus* FRI-1169 and MN8, respectively, while 314 proteins were expressed in common. In the cell lysate, 173 and 34 cytosolic proteins (15.1%) were differentially expressed, with 1155 proteins shared between the two conditions. These results demonstrate that the protein expression profiles of *L. plantarum* WCFS1 are distinctly influenced depending on the *S. aureus* strain present in the co-culture, reflecting the specific interactions and physiological adjustments in response to each strain.

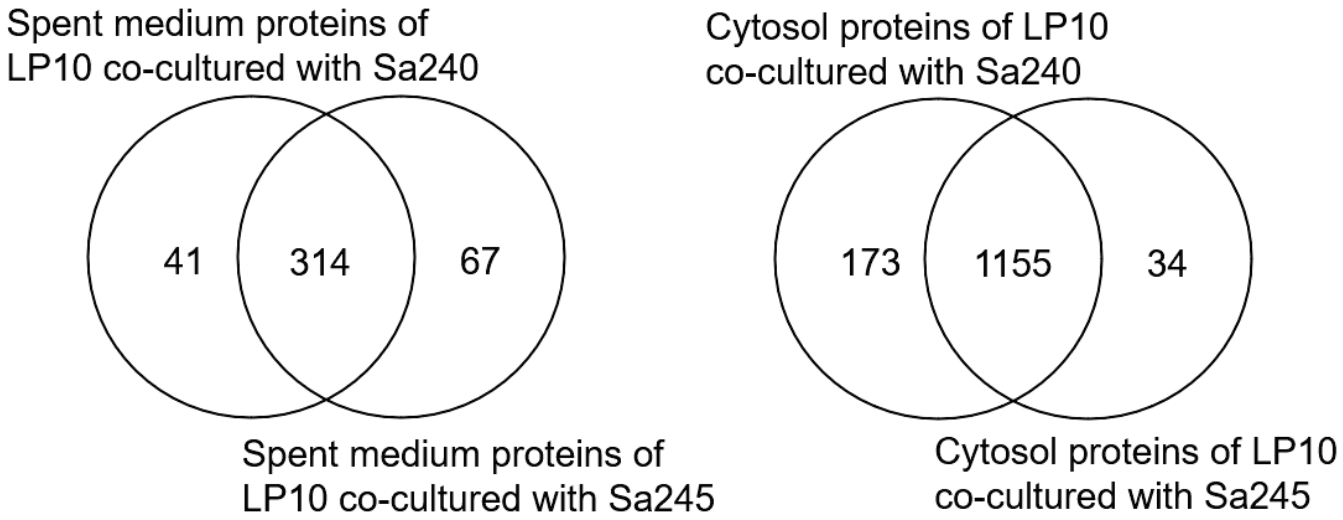


Figure 9. Venn diagram analysis of DEPs in the spent medium (**left**) and the cytosol (**right**) of *L. plantarum* WCFS1 during co-culture with *S. aureus* FRI-1169 and *S. aureus* MN8.

3.3.2. Functional Analysis of Differentially Expressed Proteins from *L. plantarum* WCFS1 During Co-Culture with *S. aureus* FRI-1169 and *S. aureus* MN8

To further investigate the impact of co-culture on *L. plantarum* WCFS1, we analyzed the functional roles of differentially expressed proteins in both the spent medium and cytosol (Supplementary Table S3). These proteins offer insight into *L. plantarum*'s physiological state and its potential influence on *S. aureus*. In the spent medium from the co-culture

with *S. aureus* FRI-1169, proteins such as ArgR (involved in nitrogen regulation), CelC (carbohydrate metabolism), and OppC (peptide transport) were identified, suggesting that *L. plantarum* releases proteins associated with nutrient acquisition. Proteins like HslO (a heat shock protein) and SrtA (involved in cell wall synthesis) hint at interactions affecting protein stability and cell wall integrity in *S. aureus* FRI-1169. In contrast, during co-culture with *S. aureus* MN8, proteins like VanR (related to cell wall stability) and MvaA (involved in isoprenoid biosynthesis) were differentially expressed, indicating a potential impact on *S. aureus* MN8′s membrane integrity. Additionally, proteins like ClpC (protein quality control) and FolP (folate biosynthesis) suggest effects on stress responses and metabolic functions, while the upregulation of GalE (carbohydrate metabolism) points to shifts in nutrient utilization. In the cytosol, *L. plantarum* WCFS1 exhibited the expression of proteins like CopR (involved in copper resistance) and CydA (linked to respiratory processes) during co-culture with *S. aureus* FRI-1169, suggesting a focus on oxidative stress management and metabolic balance. During co-culture with *S. aureus* MN8, proteins such as ArgH (involved in arginine biosynthesis) and MsrA (responsible for repairing oxidative damage) were differentially expressed, indicating adaptive responses related to amino acid metabolism and oxidative stress repair. These findings underscore the distinct metabolic and stress responses by *L. plantarum* to the two different strains of *S. aureus*, highlighting its ability to modulate its physiological processes to adapt to varying co-culture conditions.

4. Conclusions

This study provides the first comprehensive proteomic analysis of interactions between *Lactobacillus plantarum* WCFS1 and *Staphylococcus aureus* strains FRI-1169 and MN8, indicating a directional and strain-specific antagonism. *L. plantarum* exhibited a stronger impact on *S. aureus*, with more differentially expressed and upregulated proteins, while *S. aureus* strains responded with fewer DEPs and more downregulated proteins. The results demonstrate that *L. plantarum* influences key metabolic, regulatory, and stress response pathways in *S. aureus*, highlighting its potential as a non-antibiotic biocontrol agent. Furthermore, strain-specific differences in *S. aureus* responses suggest complex bacterial dynamics during co-culture. These findings underscore the effectiveness of proteomic analysis in informing new hypotheses and finding ways to manage *S. aureus* infections without the use of antibiotics or actually enhance antibiotic efficacy.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/microorganisms12122432/s1. Supplemental Table S1. Lists of proteins related to antibiotic resistance, heavy metal resistance, and virulence factors identified in the spent medium (Sheets 1 to 3) and cytosol (Sheets 4 to 6) of *L. plantarum* WCFS1 and *S. aureus* FRI-1169 (Sheets 7 to 9 and 10 to 12, respectively) during co-culture with each other. Supplemental Table S2. Lists of proteins related to antibiotic resistance, heavy metal resistance, and virulence factors identified in the spent medium (Sheets 1 to 3) and cytosol (Sheets 4 to 6) of *L. plantarum* WCFS1 and *S. aureus* MN8 (Sheets 7 to 9 and 10 to 12, respectively) during co-culture with each other. Supplemental Table S3. A list of proteins of *L. plantarum* WCFS1 during co-culture. In the spent medium, a total of 41 proteins (Sheet 1) and 67 proteins (Sheet 2) were differentially expressed in co-culture with *S. aureus* FRI-1169 and MN8, respectively, with 314 proteins expressed in common (Sheet 3). In the cytosol, 173 proteins (Sheet 4) and 34 proteins (Sheet 5) were differentially expressed in co-culture with *S. aureus* FRI-1169 and MN8, respectively, with 1155 proteins expressed in common (Sheet 6).

Author Contributions: S.K., S.Y. and M.E.H. were responsible for experimentation and draft formulation; S.D.G. and O.K. were responsible for data analysis, bioinformatics and formulation of result presentation; S.-J.K. and M.E.H. were responsible for draft review and finalization of manuscript. M.E.H. conceived and directed the research and was responsible for the final draft. All authors have read and agreed to the published version of the manuscript.

Funding: This project was supported by research protocol, E-07735.01 awarded to M.E.H. by NCTR and by postdoctoral fellowships awarded to S.K., S.Y. and S. D.-G. by Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. Food and Drug Administration.

Data Availability Statement: The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Acknowledgments: The authors are grateful to Barnette, McKenzie, and Schnackenberg for their valuable suggestions and for reviewing the manuscript. This presentation reflects the views of the authors and does not necessarily reflect those of the U.S. Food and Drug Administration. Any mention of commercial products is for clarification only and is not intended as approval, endorsement, or recommendation.

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

References

1. Bergdoll, M.S.; Crass, B.A.; Reiser, R.F.; Robbins, R.N.; Davis, J.P. A new staphylococcal enterotoxin, enterotoxin F, associated with toxic-shock-syndrome *Staphylococcus aureus* isolates. *Lancet* **1981**, *1*, 1017–1021. [CrossRef]
2. Blomster-Hautamaa, D.; Schlievert, P.M. Preparation of toxic shock syndrome toxin-1. *Methods Enzymol.* **1988**, *165*, 37–43. [CrossRef] [PubMed]
3. Lee, A.C.; Crass, B.A.; Bergdoll, M.S. Investigation by syringe method of effect of tampons on production in vitro of toxic shock syndrome toxin 1 by *Staphylococcus aureus*. *J. Clin. Microbiol.* **1987**, *25*, 87–90. [CrossRef] [PubMed]
4. Schlievert, P.M.; Kelly, J.A. Clindamycin-induced suppression of toxic-shock syndrome--associated exotoxin production. *J. Infect. Dis.* **1984**, *149*, 471. [CrossRef] [PubMed]
5. Spaulding, A.; Satterwhite, E.; Lin, Y.-C.; Chuang-Smith, O.; Frank, K.; Merriman, J.; Schaefers, M.; Yarwood, J.; Peterson, M.; Schlievert, P. Comparison of *Staphylococcus aureus* strains for ability to cause infective endocarditis and lethal sepsis in rabbits. *Front. Cell. Infect. Microbiol.* **2012**, *2*, 18. [CrossRef]
6. Centers for Disease Control and Prevention (CDC). Summary of Notifiable Infectious Diseases and Conditions—United States, 2015. *Morb. Mortal. Wkly. Rep.* **2017**, *64*, 1–143. [CrossRef]
7. Hajjeh, R.A.; Reingold, A.; Weil, A.; Shutt, K.; Schuchat, A.; Perkins, B.A. Toxic shock syndrome in the United States: Surveillance update, 1979-1996. *Emerg. Infect. Dis.* **1999**, *5*, 807–810. [CrossRef]
8. Parsonnet, J.; Hansmann, M.A.; Delaney, M.L.; Modern, P.A.; Dubois, A.M.; Wieland-Alter, W.; Wissemann, K.W.; Wild, J.E.; Jones, M.B.; Seymour, J.L.; et al. Prevalence of toxic shock syndrome toxin 1-producing *Staphylococcus aureus* and the presence of antibodies to this superantigen in menstruating women. *J. Clin. Microbiol.* **2005**, *43*, 4628–4634. [CrossRef]
9. Todd, J.; Fishaut, M.; Kapral, F.; Welch, T. Toxic-shock syndrome associated with phage-group-I Staphylococci. *Lancet* **1978**, *2*, 1116–1118. [CrossRef]
10. Gaventa, S.; Reingold, A.L.; Hightower, A.W.; Broome, C.V.; Schwartz, B.; Hoppe, C.; Harwell, J.; Lefkowitz, L.K.; Makintubee, S.; Cundiff, D.R.; et al. Active surveillance for toxic shock syndrome in the United States, 1986. *Rev. Infect. Dis.* **1989**, *11* (Suppl. S1), S28–S34. [CrossRef]
11. DeVries, A.S.; Lesher, L.; Schlievert, P.M.; Rogers, T.; Villaume, L.G.; Danila, R.; Lynfield, R. Staphylococcal toxic shock syndrome 2000-2006: Epidemiology, clinical features, and molecular characteristics. *PLoS ONE* **2011**, *6*, e22997. [CrossRef] [PubMed]
12. Schlievert, P.M.; Tripp, T.J.; Peterson, M.L. Reemergence of staphylococcal toxic shock syndrome in Minneapolis-St. Paul, Minnesota, during the 2000-2003 surveillance period. *J. Clin. Microbiol.* **2004**, *42*, 2875–2876. [CrossRef] [PubMed]
13. Sikorska, H.; Smoragiewicz, W. Role of probiotics in the prevention and treatment of meticillin-resistant *Staphylococcus aureus* infections. *Int. J. Antimicrob. Agents* **2013**, *42*, 475–481. [CrossRef] [PubMed]
14. Carvalho, F.M.; Mergulhao, F.J.M.; Gomes, L.C. Using Lactobacilli to fight *Escherichia coli* and *Staphylococcus aureus* biofilms on urinary tract devices. *Antibiotics* **2021**, *10*, 1525. [CrossRef]
15. Elkins, C.A.; Munoz, M.E.; Mullis, L.B.; Stingley, R.L.; Hart, M.E. *Lactobacillus*-mediated inhibition of clinical toxic shock syndrome *Staphylococcus aureus* strains and its relation to acid and peroxide production. *Anaerobe* **2008**, *14*, 261–267. [CrossRef]
16. Gustafsson, R.J.; Ahrné, S.; Jeppsson, B.; Benoni, C.; Olsson, C.; Stjernquist, M.; Ohlsson, B. The *Lactobacillus* flora in vagina and rectum of fertile and postmenopausal healthy Swedish women. *BMC Women's Health* **2011**, *11*, 17. [CrossRef]
17. Ravel, J.; Gajer, P.; Abdo, Z.; Schneider, G.M.; Koenig, S.S.K.; McCulle, S.L.; Karlebach, S.; Gorle, R.; Russell, J.; Tacket, C.O.; et al. Vaginal microbiome of reproductive-age women. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 4680–4687. [CrossRef]
18. Van de Wijgert, J.H.H.M.; Borgdorff, H.; Verhelst, R.; Crucitti, T.; Francis, S.; Verstraelen, H.; Jespers, V. The vaginal microbiota: What have we learned after a decade of molecular characterization? *PLoS ONE* **2014**, *9*, e105998. [CrossRef]
19. Chaban, B.; Links, M.G.; Jayaprakash, T.P.; Wagner, E.C.; Bourque, D.K.; Lohn, Z.; Albert, A.Y.K.; van Schalkwyk, J.; Reid, G.; Hemmingsen, S.M.; et al. Characterization of the vaginal microbiota of healthy Canadian women through the menstrual cycle. *Microbiome* **2014**, *2*, 23. [CrossRef]
20. Liu, H.; Gao, Y.; Yu, L.-R.; Jones, R.C.; Elkins, C.A.; Hart, M.E. Inhibition of *Staphylococcus aureus* by lysostaphin-expressing *Lactobacillus plantarum* WCFS1 in a modified genital tract secretion medium. *Appl. Environ. Microbiol.* **2011**, *77*, 8500–8508. [CrossRef]
21. Kleerebezem, M.; Boekhorst, J.; van Kranenburg, R.; Molenaar, D.; Kuipers, O.P.; Leer, R.; Tarchini, R.; Peters, S.A.; Sandbrink, H.M.; Fiers, M.W.; et al. Complete genome sequence of *Lactobacillus plantarum* WCFS1. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 1990–1995. [CrossRef]

22. Siezen, R.J.; Francke, C.; Renckens, B.; Boekhorst, J.; Weis, M.; Kleerebezem, M.; van Hijum, S.A. Complete resequencing and reannotation of the *Lactobacillus plantarum* WCFS1 genome. *J. Bacteriol.* **2012**, *194*, 195–196. [CrossRef]
23. Stingley, R.L.; Liu, H.; Mullis, L.B.; Elkins, C.A.; Hart, M.E. *Staphylococcus aureus* toxic shock syndrome toxin-1 (TSST-1) production and *Lactobacillus* species growth in a defined medium simulating vaginal secretions. *J. Microbiol. Methods* **2014**, *106c*, 57–66. [CrossRef]
24. Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [CrossRef]
25. Redondo-Lopez, V.; Cook, R.L.; Sobel, J.D. Emerging role of lactobacilli in the control and maintenance of the vaginal bacterial microflora. *Rev. Infect. Dis.* **1990**, *12*, 856–872. [CrossRef]
26. Reid, G.; Dols, J.; Miller, W. Targeting the vaginal microbiota with probiotics as a means to counteract infections. *Curr. Opin. Clin. Nutr. Metab. Care* **2009**, *12*, 583–587. [CrossRef]
27. Reid, G.; Jass, J.; Sebulsky, M.T.; McCormick, J.K. Potential uses of probiotics in clinical practice. *Clin. Microbiol. Rev.* **2003**, *16*, 658–672. [CrossRef]
28. Li, H.; Jia, M.; Qi, Q.; Wang, Q. Engineered probiotic *Lactobacillus plantarum* WCSF I for monitoring and treatment of *Staphylococcus aureus* infection. *Microbiol. Spectr.* **2023**, *11*, e0182923. [CrossRef]
29. Wang, Y.; Li, D.; Chitrakar, B.; Zhang, X.; Zhang, N.; Liu, C.; Li, Y.; Wang, M.; Tian, H.; Li, C. Copper inhibits postacidification of yogurt and affects its flavor: A study based on the *cop* operon. *J. Dairy Sci.* **2023**, *106*, 897–911. [CrossRef]
30. Lin, W.H.; Chien, C.C.; Ou, J.H.; Yu, Y.L.; Chen, S.C.; Kao, C.M. Cleanup of Cr(VI)-polluted groundwater using immobilized bacterial consortia via bioreduction mechanisms. *J. Environ. Manag.* **2023**, *339*, 117947. [CrossRef]
31. Ingmer, H.; Vogensen, F.K.; Hammer, K.; Kilstrup, M. Disruption and analysis of the *clpB*, *clpC*, and *clpE* genes in *Lactococcus lactis*: ClpE, a new Clp family in gram-positive bacteria. *J. Bacteriol.* **1999**, *181*, 2075–2083. [CrossRef]
32. Cao, C.; Wang, J.; Liu, Y.; Kwok, L.Y.; Zhang, H.; Zhang, W. Adaptation of *Lactobacillus plantarum* to ampicillin involves mechanisms that maintain protein homeostasis. *mSystems* **2020**, *5*, e00853-19. [CrossRef]
33. Ingavale, S.; van Wamel, W.; Luong, T.T.; Lee, C.Y.; Cheung, A.L. Rat/MgrA, a Regulator of autolysis, is a regulator of virulence genes in *Staphylococcus aureus*. *Infect. Immun.* **2005**, *73*, 1423–1431. [CrossRef]
34. Truong-Bolduc, Q.C.; Hooper, D.C. Phosphorylation of MgrA and its effect on expression of the NorA and NorB efflux pumps of *Staphylococcus aureus*. *J. Bacteriol.* **2010**, *192*, 2525–2534. [CrossRef]
35. Luong, T.T.; Dunman, P.M.; Murphy, E.; Projan, S.J.; Lee, C.Y. Transcription profiling of the *mgrA* regulon in *Staphylococcus aureus*. *J. Bacteriol.* **2006**, *188*, 1899–1910. [CrossRef]
36. Campbell, E.A.; Korzheva, N.; Mustaev, A.; Murakami, K.; Nair, S.; Goldfarb, A.; Darst, S.A. Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell* **2001**, *104*, 901–912. [CrossRef]
37. Perichon, B.; Courvalin, P. VanA-type vancomycin-resistant *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* **2009**, *53*, 4580–4587. [CrossRef]
38. Qoronfleh, M.W.; Bortner, C.A.; Schwartzberg, P.; Wilkinson, B.J. Enhanced levels of *Staphylococcus aureus* stress protein GroEL and DnaK homologs early in infection of human epithelial cells. *Infect. Immun.* **1998**, *66*, 3024–3027. [CrossRef]
39. Frees, D.; Gerth, U.; Ingmer, H. Clp chaperones and proteases are central in stress survival, virulence and antibiotic resistance of *Staphylococcus aureus*. *Int. J. Med. Microbiol.* **2014**, *304*, 142–149. [CrossRef]
40. Tsai, Y.K.; Lin, T.H. Sequence, organization, transcription and regulation of lactose and galactose operons in *Lactobacillus rhamnosus* TCELL-1. *J. Appl. Microbiol.* **2006**, *100*, 446–459. [CrossRef]
41. Licandro-Seraut, H.; Scornec, H.; Pedron, T.; Cavin, J.F.; Sansonetti, P.J. Functional genomics of *Lactobacillus casei* establishment in the gut. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E3101–E3109. [CrossRef] [PubMed]
42. Morel-Deville, F.; Fauvel, F.; Morel, P. Two-component signal-transducing systems involved in stress responses and vancomycin susceptibility in *Lactobacillus sakei*. *Microbiology* **1998**, *144 Pt 10*, 2873–2883. [CrossRef] [PubMed]
43. Zhou, C.; Bhinderwala, F.; Lehman, M.K.; Thomas, V.C.; Chaudhari, S.S.; Yamada, K.J.; Foster, K.W.; Powers, R.; Kielian, T.; Fey, P.D. Urease is an essential component of the acid response network of *Staphylococcus aureus* and is required for a persistent murine kidney infection. *PLoS Pathog.* **2019**, *15*, e1007538. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

microorganisms

MDPI

Article

Exploring Pathogen Presence Prediction in Pastured Poultry Farms through Transformer-Based Models and Attention Mechanism Explainability

Athish Ram Das [1], Nisha Pillai [2], Bindu Nanduri [1], Michael J. Rothrock, Jr. [3] and Mahalingam Ramkumar [2,*]

[1] Department of Comparative Biomedical Sciences, College of Veterinary Medicine, Mississippi State University, Starkville, MS 39762, USA; ar2903@msstate.edu (A.R.D.); bnanduri@cvm.msstate.edu (B.N.)
[2] Department of Computer Science and Engineering, Mississippi State University, Starkville, MS 39762, USA; pillai@cse.msstate.edu
[3] Egg Safety and Quality Research Unit, USDA-ARS U.S. National Poultry Research Center, Athens, GA 30605, USA; michael.rothrock@usda.gov
* Correspondence: ramkumar@cse.msstate.edu

Abstract: In this study, we explore how transformer models, which are known for their attention mechanisms, can improve pathogen prediction in pastured poultry farming. By combining farm management practices with microbiome data, our model outperforms traditional prediction methods in terms of the F1 score—an evaluation metric for model performance—thus fulfilling an essential need in predictive microbiology. Additionally, the emphasis is on making our model's predictions explainable. We introduce a novel approach for identifying feature importance using the model's attention matrix and the PageRank algorithm, offering insights that enhance our comprehension of established techniques such as DeepLIFT. Our results showcase the efficacy of transformer models in pathogen prediction for food safety and mark a noteworthy contribution to the progress of explainable AI within the biomedical sciences. This study sheds light on the impact of effective farm management practices and highlights the importance of technological advancements in ensuring food safety.

Keywords: food safety; pathogen; transformer; PageRank

1. Introduction

The growing concern over *Salmonella*, *Listeria*, and *Campylobacter* in poultry presents a pressing public health issue, as these pathogens lead to significant food-borne illnesses globally [1–3]. Addressing the challenge of detecting and managing these bacteria in pastured poultry farms calls for innovative strategies to ensure food safety and protect public health. Our paper introduces a machine-learning-based approach to improving the detection and control of pathogens in poultry production environments. Additionally, we present a new method in the field of explainable AI [4], offering a clear insight into the complex decision-making processes of advanced transformer models.

At the heart of our research is the microbiome—the diverse community of microorganisms inhabiting specific environments such as the gut of poultry. These microbial populations are pivotal in determining the health and disease susceptibility of their hosts. Some microbiota clusters can affect pathogen levels either by maintaining a balance that inhibits pathogenic growth or by fostering conditions conducive to an increase in pathogens [5]. Understanding these complex interactions is essential for effective pathogen risk management.

Recent research in the field of food safety and microbiome has focused on predictive microbiology [6,7] and understanding the decision-making processes of machine learning models. Standard methods for assessing feature importance, such as SHAP [8] and DeepLIFT [9], are commonly used across various datasets. However, these methods often overlook interactions within the microbiota, which can significantly influence the outcomes of these models. Consequently, simple feature importance analyses may not fully

explain the underlying biological processes. By employing transformer models, which are renowned for their ability to detect complex data patterns through attention mechanisms [10], we aim to identify microbiota clusters that indicate the presence or absence of pathogens. These models are particularly adept at uncovering data patterns and dependencies [10], making them well suited for analyzing microbiome data relationships that impact pathogen prevalence.

This study not only aims for improved pathogen prediction accuracy but also highlights the critical role of model explainability in biomedical applications. Explainability ensures the trustworthiness, transparency, and reliability of AI predictions for both poultry farmers and biologists. Our method analyzes the model's attention matrix and applies the PageRank algorithm [11] to clarify feature significance. Furthermore, we extend explainable AI techniques by applying spectral clustering for data cluster analysis and transforming attention matrices into adjacency matrices for graph-based visualizations, improving our interpretive capabilities for AI decision making.

While exploring attention weights in transformer models to demystify their "black-box" nature is not entirely new [12–14], we recognize the substantial potential of these weights. Although past studies suggest that attention weights should not be directly equated with explanations [15], the more recent literature suggests that these values can indeed be utilized for explanation [16,17]. Our approach also does not view them as direct interpretations or explanations; instead, we consider attention values as a step towards explaining a model.

By proposing the pathogen prediction method for pastured poultry farming, our paper contributes to the fields of predictive microbiology and food safety. It also paves the way for future research in explainable AI within biomedical sciences, using attention matrices to unveil the reasoning behind model decisions. The broader implications of our findings offer crucial insights into farm management practices that can drive the establishment of an optimal host microbiome that does not support food-borne pathogens and underscores the importance of novel analytical approaches in promoting food safety.

In summary, the primary objective of our study is to develop machine learning models with enhanced explainability for food safety. This research describes highly accurate machine learning models utilizing transformer architectures and introduces a novel model explanation technique that complements existing interpretation methods and shows significant promise for future applications.

2. Dataset

In the introductory section, it was highlighted that our research methodology was tested by combining two distinct datasets: the Farm Management Practices dataset and the Microbiome dataset, both of which were sourced from pastured poultry farms. These datasets were acquired from the foundational work of Hwang et al. and Rothrock et al. [18,19].

2.1. Farm Management Practices Dataset

This dataset is based on a longitudinal study conducted from March 2014 to November 2017, which covered 42 broiler flocks across 11 pastured poultry farms in the southeastern United States with flock sizes varying between 25 and 1500 birds. The data encompassed both pre-harvest (Feces and Soil) and post-harvest (Ceca, WCR-F, and WCR-P) samples. Documented by Xinran Xu et al. in 2021 [20], the study collected data on poultry farms using movable pens, which were shifted to new pastures daily. The configuration, number, and application of temporary fencing around the pens varied across the farms.

The data collection encompassed 40 key farm practice variables across the lifecycle of a flock, totaling 160 variables, including metadata. The dataset includes variables for the presence of *Salmonella*, *Listeria*, and *Campylobacter*, which are used as the target variables for our study.

2.2. Microbiome Dataset

To generate the Microbiome dataset, 16S rRNA gene high-throughput Illumina sequencing was applied to analyze temporal pre-harvest samples, including feces and soil, from the same 41 pastured poultry flocks. This sequencing facilitated the determination of the relative abundance of operational taxonomic units (OTUs). The unique genera identified within the OTUs served as machine learning predictors for assessing the prevalence of food-borne pathogens—*Salmonella*, *Campylobacter*, and *Listeria*—at various growth stages of poultry, categorized as START (2–4 weeks old), MID (5–7 weeks old), and END (8–11 weeks old).

Microbiome Analysis

DNA Extraction: DNA was extracted from samples according to a semi-automated hybrid DNA extraction protocol [21]. An enzymatic method based on the QIAamp DNA Stool Mini Kit (QIAGEN, Valencia, CA, USA) was combined with a mechanical method with the FastDNA Spin Kit for Feces (MP Biomedicals, Solon, OH, USA). A QIAcube Robotic Workstation was used to purify DNA using the DNA Stool-Human Stool-Pathogen Detection Protocol. Using a Take3 plate and the Synergy H4 multimode plate reader (BioTek, Winooski, VT, USA), the DNA concentration of each sample was determined spectrophotometrically after purification.

Illumina MiSeq Library Analysis: A dataset from the Earth Microbiome Project Laboratory at the U.S. Department of Energy, Argonne National Laboratory (Argonne, IL, USA) was used for library construction and sequencing. The hypervariable V4 domain of the bacterial 16S rRNA gene was amplified using the F515 (5′-CACGGTCGKCGGCGCCATT-3′) and R806 (5′-GGACTACHVGGGTWTCT AAT-3′) primer set, with each primer containing Illumina adapter regions (Illumina, Inc., San Diego, CA, USA) and the reverse primer containing the Golay barcodes to facilitate multiplexing [22]. The Illumina MiSeq platform was used to obtain raw reads. The QIIME v1.9.1 pipeline (Quantitative Insights Into Microbial Ecology) generated and processed 3,297,242 raw sequence reads [23]. R1 reads were filtered for quality, and libraries were split according to the Golay barcode sequences (split library *fastq.py* script, default parameters). With the usearch option [24] and the *pick _otus.py* script (*-m* usearch, all other parameters were set to the default), sequences were chimera checked against the gold.fa database (http://drive5.com/uchime/gold.fa, accessed on 20 May 2024) and clustered based on their sequence similarity (97%) into operational taxonomic units (OTUs). For each OTU, a representative sequence was selected using the *pick_rep_set.py* script (utilizing the most abundant method for picking, all other parameters were set to the default) and used for taxonomic assignment with UCLUST and the Greengenes 13_8 database [25] using *assign_taxonomy.py* (default parameters). Using PyNAST (http://pynast.sourceforge.net, accessed on 10 December 2014) [26], sequences were aligned (*align_seqs.py* script, default parameters) and filtered (*filter_alignment.py*). Following this, a phylogenetic tree was generated using *make_phylogeny.py* (using the default settings and FastTree).

In the context of machine learning, the dataset incorporates input features derived from two distinct sources: the Farm Management Practices dataset and the Microbiome dataset. Specifically, the Farm Management Practices dataset contributes 157 input features, while the Microbiome dataset provides 1823 input features. The model targets three outcome variables, corresponding to the prevalence of three food-borne pathogens: *Salmonella*, *Campylobacter*, and *Listeria*.

3. Method

The methodology of this study is structured into three primary stages: data preprocessing, model architecture design and training, and model decoding. While these stages are commonly found in any machine learning project, our approach necessitates further elaboration due to the specific adaptations made to accommodate the transformer model.

3.1. Data Pre-Processing

In addressing the farm management data, which comprise both categorical and continuous variables, a critical step in our data pre-processing involved encoding these variables into fixed-dimension vectors. This encoding is essential for facilitating the computation within the attention mechanism of the transformer model. Drawing insights from the work of Xin Huang [27], we observed that traditionally, attention mechanisms in the context of tabular data primarily accommodated categorical features, leaving continuous features to bypass the attention blocks.

To effectively incorporate continuous features into the attention mechanism, we developed a new encoding technique named "scalar-to-vector encoding". The details of this method are provided below.

Given a scalar value S, a maximum value max, a minimum value min, and a target vector dimension n, we first define the bin size as follows:

$$binsize = \frac{max - min}{n - 1}$$

The bin boundaries spanning the complete range of our data are set according to this $binsize$. The scalar value S is then mapped to a corresponding n-dimensional vector V where all elements are initially set to zero. To embed S into V, we identify the immediate lower L and upper U bin boundaries that S falls between. The vector V is then updated at indices corresponding to L and U as follows:

$$V_L = \frac{U - S}{U - L}$$

$$V_U = \frac{S - L}{U - L}$$

All other elements in V remain at zero, ensuring that V represents the scalar value S with respect to its position within the specified range.

This method allows continuous variables to be transformed into a fixed-dimensional vector format, making them compatible for processing through the attention mechanism. This extends the capabilities of the transformer model to accommodate scalar continuous values as well. Our design builds upon the foundational principles outlined by Xin Huang [27], expanding its utility to include continuous data alongside categorical inputs.

3.2. Model

Utilizing the scalar-to-vector encoding mentioned in the previous subsection, we merged categorical features from the Farm Management Practices dataset and continuous features from both the Farm Management Practices dataset and the Microbiome dataset, preparing them for introduction into the transformer model, as shown in Figure 1. Adopting the transformer encoder architecture outlined in "Attention is All You Need" by Vaswani et al. (2017) [10], we applied it to predict the binary presence of pathogens.

3.3. Model Decoding

In the field of model explanation, Shapley Additive Explanations (SHAP) [8] and Deep Learning Important Features (DeepLIFT) [9] are established methods. Our motivation for creating a new model explainability approach stems from two factors. Firstly, the performance of transformers across different data types has been better, a benefit linked to their attention mechanism. This mechanism implies that analyzing the attention matrix can unlock various possibilities in Explainable AI.

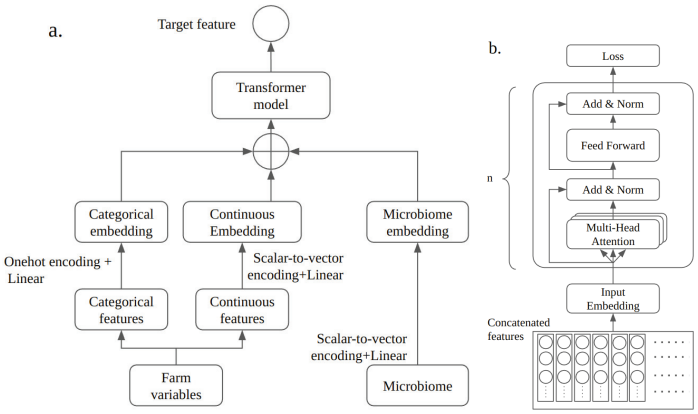


Figure 1. Illustration of the architecture of the transformer model highlighting the integration of both categorical (via one-hot encoding) and continuous (via scalar-to-vector encoding) features into the transformer's encoder blocks. Panel (**a**) shows the encoding processes, and Panel (**b**) shows the transformer model's encoder architecture.

Our transformer model features multiple layers of multi-head attention networks in series, with each layer's output feeding into the subsequent layer's input. For our purposes in explainable AI and to keep the complexity manageable, only the attention weights from the first block were used for further analysis. Thus, minor modifications in the formula for computing attention [15] allow us to derive the attention matrices for our analysis; these are represented as follows:

$$A(Q_{l1}, K_{l1}, V_{l1}) = softmax(\frac{Q_{l1}K_{l1}^T}{\sqrt{d_k}})V_{l1}$$

Here, A represents the attention value, and Q, K, and V denote the Query, Key, and Value, respectively, with the subscript $l1$ signifying that these components are exclusively derived from the first layer. This approach is used to construct the attention matrix essential for our model decoding and is utilized for three distinct analyses.

Feature Importance with PageRank: We initially applied this approach to calculate feature importance, a common practice in decoding black-box models. PageRank is an algorithm used by Google Search to rank web pages in their search engine results [11]. It measures the importance of website pages by counting the number and quality of links to a page to determine a rough estimate of a website's importance. Analogously, by using the PageRank algorithm, the significance of a feature can be calculated based on its attention weights.

The PageRank vector, PR, was determined by solving the following eigenvalue problem:

$$PR = \alpha APR + \frac{1-\alpha}{N}1$$

where

- PR is the vector of PageRank values for all features;
- α is the damping factor;
- A is the attention matrix, where element A_{ij} denotes the softmax attention weight of feature i to feature j;
- N is the total number of features;
- 1 is a vector with all elements equal to 1.

Spectral Clustering to Identify the Signature of Microbiota: The attention mechanism simplifies the process of identifying feature clusters. These clusters, which are feature groups affecting the model's output together, are more directly observable in the attention matrix. The importance of identifying such clusters is particularly relevant in our data, as

recognizing microbiota clusters is significant in microbiology. Spectral clustering utilizes the spectral features of attention weights to achieve this goal.

Spectral clustering [28] is a technique used in machine learning and data mining to identify clusters in data based on the spectrum (eigenvalues) of the similarity matrix of the data. It works by transforming the data into a lower-dimensional space in which clusters are more apparent and can be easily identified using traditional clustering techniques such as K-means [29] or HAC (Hierarchical Agglomerative Clustering) [30], the latter of which we used in our study.

Our method involved constructing a similarity matrix, S, from the data and then deriving the Laplacian matrix, L, from S. HAC was then performed in the space spanned by the eigenvectors of L corresponding to its smallest eigenvalues.

Additionally, these attention matrices can be mapped out in a graph structure to visualize the interactions among the identified OTUs in the microbiota, which is an interesting research question in microbial ecology.

4. Experiments and Results

Our investigation focused on two primary questions: whether a prediction model using a transformer architecture, as depicted in Figure 1, outperforms multi-layer perceptron (MLP) models and whether the attention matrix can offer valuable insights into explainable AI. To address the first question, we carried out extensive testing across several models using our dataset, applying a grid search to fine-tune the hyperparameters. The findings from these tests were used in evaluating the Tab-transformer model's performance relative to that of MLP models. For the second question, we extracted and analyzed the attention matrix from our model, and a detailed discussion of this analysis is presented in the following sections.

4.1. Transformer Model Evaluation

For model evaluation, we conducted a series of detailed experiments involving various combinations of epochs, learning rates, test sizes, dropout values, attention layers, and linear layers. These combinations were applied to predict pathogen presence in the following different contexts:

- Pre-harvest *Salmonella* samples;
- Pre-harvest *Listeria* samples;
- Pre-harvest *CampyCapetown* samples;
- Post-harvest *Salmonella* samples;
- Post-harvest *Listeria* samples;
- Post-harvest *CampyCapetown* samples.

We executed the experiments across the following four distinct model architectures to gauge their effectiveness:

- Multi-layer perceptron (MLP) - Farm Management Practice Variables.
- Multi-head transformers without scalar-to-vector embedding—Farm Management Practice variables.
- Multi-head transformers with scalar-to-vector embedding—Farm Management Practice variables.
- Multi-head transformers with scalar-to-vector embedding—Farm Management Practice variables and Microbiome.

This approach allowed us to compare the efficiency and accuracy of each architecture under similar experimental conditions.

The results of these experiments are shown in Table 1. The table presents the F1 scores for all the models tested on our dataset, demonstrating the gradual enhancement in performance as the models evolve. Our final architecture incorporates a multi-head transformer with scalar-to-vector encoding for continuous variables. The dataset used for this final model combines poultry management variables and microbiome data.

The results of these experiments demonstrate that the performance of the transformer models is better than that of the multi-layer perceptrons. Moreover, the model gives much better predictions of pathogen presence when the farm management data are combined with the Microbiome dataset. With these results, we can infer that the attention mechanism is working in our model, which is essentially the core of transformer models, and we were ready to further investigate the attention matrix.

Table 1. F1 scores for all the models tested on the dataset.

	Pre Harvest			**Post Harvest**		
	Salmonella	***Listeria***	***Campy***	***Salmonella***	***Listeria***	***Campy***
MLP	0.79	0.67	0.84	0.78	0.87	0.95
Multi-Head Transformer w/o scalar-to-vector embedding	0.79	0.72	0.84	0.78	0.92	0.96
Multi-Head Transformer w/ scalar-to-vector embedding	0.78	0.71	0.84	0.83	0.91	0.97
Multi-Head Transformer w/ scalar-to-vector embedding	**0.86**	**0.79**	**0.86**	**0.89**	**0.92**	**0.97**

4.2. PageRank Results and Evaluation

As described in the Methods section, we extracted the attention weights from the first multi-head attention block's self-attention heads for all test scenarios. We then applied the PageRank algorithm to allocate scores to each feature, with the aggregate of these scores equaling 1 amongst all 1823 microbiota features and around 60 farm variable features. The top PageRank-valued features were identified for validation. The top 10 microbiota features identified by the method for post-harvest *Salmonella* are shown in Figure 2 as an example. Similar tables were generated for pre-harvest and post-harvest *Salmonella*, *Listeria*, and *CampyCapetown*.

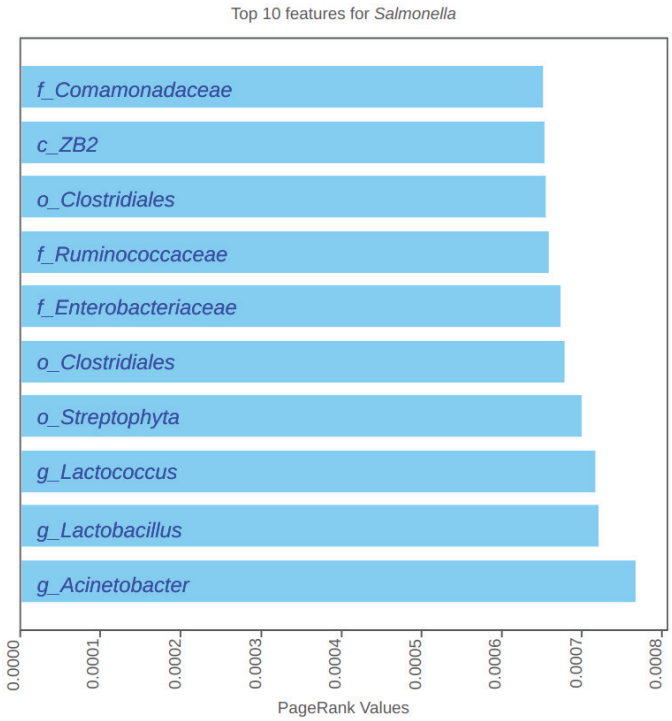


Figure 2. Top 10 most important features recognized by PageRank for post-harvest *Salmonella*.

The outcomes were verified through both qualitative and quantitative means. For the qualitative evaluation, we cross-referenced the top microbiota features identified by our approach with the existing literature to confirm their associations with probiotic or pathogenic characteristics, as discussed in detail in Section 5. For quantitative evaluation, we used DeepLIFT to rank all features in the dataset and then compared the top 100 features identified by both our method and DeepLIFT. This comparison revealed a significant concurrence between the two methods, with approximately 35% of the top features being recognized by both. The Venn diagram in Figure 3 illustrates some of these results. For a comprehensive view of the findings, please refer to the Appendix A.

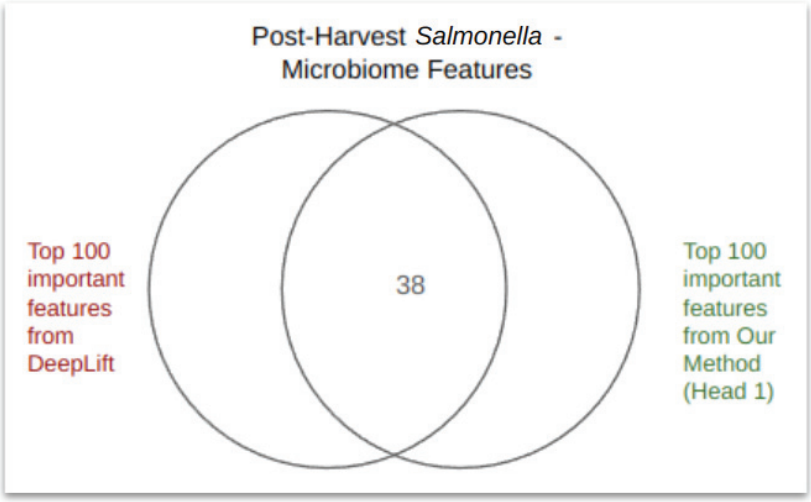


Figure 3. Comparison of the DeepLIFT and PageRank results for post-harvest *Salmonella*. The top 100 features from a total of 1824 were chosen to construct the Venn diagram, which reveals a notable level of agreement between the two methods.

4.3. Spectral Clustering Results

To determine microbiota clusters within the six experimental scenarios (Section 4.1), we utilized attention matrices. The selection of the number of clusters, a critical hyperparameter in our clustering approach, necessitated a methodical decision-making process. We opted for Hierarchical Agglomerative Clustering (HAC) over K-means due to the clarity provided by HAC's dendrograms in determining the optimal number of clusters. For each scenario, a dendrogram was generated to establish the number of clusters. Subsequently, we applied HAC with average linkage for feature classification into clusters. With this approach, we were able to identify clusters composed of microbes with similar ecological properties. An example of two clusters identified through this method is shown in Table 2. In one of the clusters, the majority of species, including *Actinobacteria*, *Acidobacteria*, *Bacteroidetes*, *Rhodoplanes*, *Bacillus*, *Myxococcales*, and *Candidatus Nitrososphaera*, are non-pathogenic and beneficial microorganisms primarily distributed in soil and water environments. These groups play crucial roles in ecosystem functioning, such as in nutrient cycling and organic matter decomposition. In contrast, the *Lactobacillales* found in the other cluster are typically associated with environments rich in carbohydrates, including dairy products, fermented foods, the gastrointestinal tract of humans and animals, and plant surfaces.

Table 2. Evaluation of two clusters identified through HAC. Microbes usually found in soil (cluster B/2) and gut (cluster A/1) are accurately identified and grouped separately.

Bacteria	Cluster
g__Rummeliibacillus	2
g__Salinicoccus	2
g__Enterococcus	2
g__Streptococcus	2
f__Enterobacteriaceae;g__	2
g__Candidatus Nitrososphaera	0
o__iii1-15;f__;g__	0
o__RB41;f__;g__	0
f__Ellin6075;g__	0
f__Gaiellaceae;g__	0
o__Solirubrobacterales;f__;g__	0
f__Solirubrobacteraceae;g__	0
f__Cytophagaceae;g__	0
o__Sphingobacteriales;f__;g__	0
f__Chitinophagaceae;g__	0
g__Bacillus	0
g__Rhodoplanes	0
o__Myxococcales;f__;g__	0
g__DA101	0

5. Discussion

This research aims to make powerful machine learning models more interpretable and transparent when analyzing complex biological data. By combining techniques to explain these models, using effective transformer neural networks, and applying algorithms such as PageRank, we can gain deeper insights into fundamental biological processes. This approach has wide applications to other animal production systems beyond the pastured poultry production system used in this study. It can also be applied to drug discovery, analyzing evolutionary relationships between species, mining agricultural/biomedical text data, understanding gene regulatory networks, and predicting protein structures. Making these advanced models more interpretable will help advance our understanding of biology and lead to new applications that can improve animal and human health.

This section presents a discussion of our findings and related research, focusing on microbiota features that have been identified as significantly associated with the presence of food-borne pathogens. *Salmonella* and *Campylobacter* are among the most prevalent pathogens associated with poultry and are leading causes of bacterial food-borne illness [31]. Taxonomically, *Salmonella* belongs to the phylum *Proteobacteria*, while *Campylobacter* belongs to the phylum *Epsilonbacteraeota* [32]. Listeria, another food-borne pathogen, is a concern not only in poultry but also in a wide range of other foods, such as dairy and ready-to-eat products, and it belongs to the phylum *Firmicutes* [33]. However, the presence and significance of *Listeria* in poultry production, particularly in pastured systems, is more associated with the environment and processing facilities than with the birds themselves.

5.1. Role of Firmicutes

Our research finds *Firmicutes* to be one of the major influencing factors for the prevalence of *Salmonella*, *Listeria*, and *Campylobacter* in both pre-harvest (feces, soil) samples and post-harvest (post-processing) samples. While *Firmicutes* is a large phylum of bacteria that includes many non-pathogenic and beneficial organisms, the environmental dynamics of pastured poultry systems can affect the prevalence of various pathogens, including those not classified under *Firmicutes* [34]. Some beneficial *Firmicutes* in the gut flora of pastured

poultry may help in competitively excluding pathogenic bacteria by competing for nutrients and attachment sites in the gastrointestinal tract [35]. Studies show that inoculation with *S. Enteritidis* [36] resulted in significant positive correlations with *Firmicutes*, notably affecting the relative abundance of 18 genera.

We found that the family *Bacillaceae*, belonging to the phylum *Firmicutes* [32], influenced the prevalence of *Salmonella*, *Listeria*, and *Campylobacter* in our experiments. The ability of *Bacillus* strains to produce a wide array of antimicrobial peptides (AMPs) and bacteriocins is crucial in their antagonistic effects against enteropathogenic bacteria in the gastrointestinal tract [37]. Specifically, the growth of *Listeria* is inhibited in contaminated environments or products [38]. *Bacillus subtilis PS-216* showed effective inhibition against *Campylobacter jejuni* under microaerobic conditions, demonstrating its potential as a probiotic that could integrate into the chicken intestinal microbiome and combat *campylobacteriosis* [39]. Bacillus strains can produce extracellular polysaccharides, vitamins, and exoenzymes that support the growth of beneficial microbiota, contributing to a healthier gut environment [40,41] and potentially reducing the colonization of pathogens by strengthening the birds' natural defenses against infections and possibly reducing pathogen shedding.

Our results show that *Rummeliibacillus*, a Gram-positive rod-shaped bacterium, is associated with *Listeria* and *Campylobacter*. Factors such as soil composition and microbial diversity can either facilitate or limit the survival of *Listeria* [42], indicating that microbial competition, potentially including competition from *Rummeliibacillus*, could influence the prevalence of *Listeria*. Lactic acid bacteria (LAB) [43], including *Lactococcus* [44] and *Lactobacillus* species, are known for their role in producing fermented foods and for their ability to inhibit the growth of pathogenic bacteria through the production of lactic acid, bacteriocins, and AMPs. Acidification of the environment due to lactic acid generation by LAB inhibits the growth of [45,46] *Listeria* and *Salmonella*. *Lactobacillus* species have been extensively studied for their probiotic properties in poultry [47], demonstrating benefits such as reduced *Salmonella* contamination (https://today.uconn.edu/2021/06/probiotic-intervention-to-prevent-salmonella-infection-in-poultry/ accessed on 8 April 2024), improved growth performance, immune enhancement, gut microbe sustainability, and contributions to health. *Lactobacillus* cultures or bacteriocins could be used in rinses or coatings for poultry meat post-processing to reduce surface contamination by *Salmonella* [48].

Our results indicate the influence of *Lysinibacillus* species on the presence of *Salmonella* and that of the *Planococcaceae* family on the presence of *Listeria* in the pre-harvest phase, and this suggests a multifaceted approach involving the management of animal waste, the monitoring and treatment of irrigation water, and practices to reduce contamination in food production environments. The role of soil-dwelling or fecal bacteria, such as *Lysinibacillus* [49], in influencing these processes, directly or indirectly, through effects on microbial communities remains a critical area for further research. Another influential genus in our findings, *Solibacillus*, a Gram-positive, rod-shaped, spore-forming bacteria, could potentially compete with *Salmonella* for nutrients in the soil, alter soil microbial community composition, and limit the latter's ability to proliferate.

Anoxybacillus is a genus of thermophilic [50], facultatively anaerobic bacteria within the *Firmicutes* phylum that is known for its ability to thrive at high temperatures and for its presence in diverse environments, including hot springs and dairy products. Detecting *Anoxybacillus* in post-harvest environments indicates that higher temperatures or specific nutrient availability could impact the survival or proliferation of pathogens such as *Campylobacter*.

The class *Clostridia*, part of the *Firmicutes* phylum, includes significant food-borne pathogens that impact poultry and can pose risks to human health [51]. Our results indicate that *Clostridia* plays a critical role in the prevalence of *Salmonella* and *Campylobacter* in soil and fecal samples. Conversely, the presence of *Salmonella* and *Listeria* was associated *Clostridia* species in post-processed meat samples. By contributing to the fermentation

process and production of short-chain fatty acids (SCFAs) in the gut [52], The *Ruminococcaceae* family, part of the *Clostridia* class, might help create an intestinal environment that is less favorable for *Salmonella* and *Campylobacter* proliferation. The association between the *Syntrophomonas* genus and *Listeria* contamination in our results suggests a potential role for this group of anaerobic, syntrophic bacteria in modulating the prevalence of *Listeria* in the post-processing environment, possibly due to the syntrophic degradation of butyrate in environmental and industrial processes limiting organic waste that can support pathogens such as *Listeria* [53].

5.2. Role of Proteobacteria

The phylum *Proteobacteria* represents a vast and diverse group of Gram-negative bacteria that are classified into various classes. *Acinetobacter* can be found in soil and animal (including human) feces, though they are not a predominant component of the gut flora [54], so they could indicate exposure to the environment or the consumption of contaminated food or water [55–57]. In soil, these bacteria are key players in breaking down organic substances and nutrient recycling [58].

Our previous studies have underscored the potential role of other microbial species, such as *Acinetobacter*, in the ecology of these pathogens. Specifically, our analysis indicates that *Acinetobacter* played a critical role in pre-harvest (fecal and soil) samples where *Salmonella* was detected. Detection of *Salmonella* and *Acinetobacter* from poultry operations in Washington state was reported [59]. Understanding the nature of this relationship and whether it is synergistic or antagonistic could inform the development of targeted interventions and more effective pathogen control strategies on farms.

5.3. Clustering to Uncover the Microbial Community Structure

ClusterA

Rummeliibacillus is a part of the *Firmicutes* phylum, and these Gram-positive bacteria are known to form endospores, which allow them to survive in harsh environmental conditions. ***Streptococcus*** is another member of the *Firmicutes* phylum; these are Gram-positive cocci known for their role in both health (as part of the normal microbiota) and disease (causing various infections). *Enterococcus*, *Streptococcus*, and members of the *Enterobacteriaceae* family are primarily associated with the gastrointestinal tract of animals and humans [60], playing roles that range from benign colonization to causing serious infections. ***Salinicoccus*** is a genus of Gram-positive, halophilic bacteria that belong to the family *Staphylococcaceae* [61]. These bacteria are typically found in environments with high salt concentrations [62], such as salt lakes, saline soils, and salted food products, and they have potential applications in biotechnology, including the biodegradation of pollutants in saline conditions and the production of enzymes and other bioactive compounds. ***Enterococcus***, including species such as *Enterococcus faecalis*, is a genus of bacteria that are part of the natural microbiota of the human gastrointestinal tract [63] but can also be found in soil, water, food, and decaying vegetation [64]. In soil, *Enterococcus* species may be introduced through the application of animal manure as fertilizer, contributing to the microbial diversity of agricultural soils.

ClusterB

All of the groups in this cluster are ubiquitous in the environment and are found in a wide range of habitats from soil and water to extreme environments such as hot springs (*Crenarchaeota*) and acidic mines (***Acidobacteria***). This distribution underlines their adaptability and the vast diversity of metabolic strategies that they have evolved to exploit different ecological niches. They play crucial roles in their respective ecosystems and are involved in nutrient cycling, decomposing organic matter, and, in some cases, forming symbiotic relationships with plants [65] (e.g., certain ***Alphaproteobacteria***, such as *Rhizobia*) or animals. Their metabolic diversity allows them to perform a variety of biochemical processes critical to Earth's biogeochemical cycles, such as carbon and nitrogen cycling.

6. Conclusions

In this study, we set out to investigate two main questions: firstly, whether incorporating a transformer model into our unique framework would significantly enhance our ability to predict pathogen presence in poultry production environments; secondly, whether we could introduce a novel method for determining feature importance using the attention matrix and PageRank. Although our final model showed modest improvements in predictive performance over previous models applied to our dataset, this improvement was attributed more to the inclusion of microbiome data than to the transformer model itself. However, the primary goal of implementing the transformer model was to facilitate our second objective, which indeed yielded promising results. The method of computing feature importance through the attention matrix and PageRank aligned well with DeepLIFT's findings and was validated as biologically relevant. Additionally, our study introduced an effective method for identifying microbiota signatures using the same attention matrices, yielding meaningful outcomes. Consequently, we view our contribution as a positive step forward in the field of explainable AI, and we anticipate that it will inspire further research in this direction.

Author Contributions: Conceptualization, A.R.D. and N.P.; methodology, A.R.D. and M.J.R.J.; software, A.R.D.; validation, A.R.D.; formal analysis, A.R.D.; investigation, A.R.D. and M.J.R.J.; resources, M.J.R.J. and B.N.; data curation, M.J.R.J. and A.R.D.; writing—original draft preparation, A.R.D. and N.P.; writing—review and editing, A.R.D., B.N., N.P., M.J.R.J. and M.R.; visualization, A.R.D.; supervision, M.J.R.J. and B.N.; project administration, B.N. and M.J.R.J.; funding acquisition, B.N. and M.J.R.J. All authors have read and agreed to the published version of the manuscript.

Funding: The dataset used in this study was provided by the Agricultural Research Service USDA CRIS Project "Reduction of Invasive Salmonella enterica in Poultry through Genomics Phenomics and Field Investigations of Small Multi-Species Farm Environments" #6040-320000-011-00-D. This work was supported by the USDA-ARS NACA Project "Developing Detection and Modeling Tools for the Geospatial and Environmental Epidemiology of Animal Disease" #58-6064-3-017.

Data Availability Statement: Restrictions apply to the availability of these data. Data were obtained from the Agricultural Research Service, USDA, and are available from Dr. Rothrock with the permission of the USDA.

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

Appendix A

Appendix A.1. Transformer Model Performance

The performance of the prediction models tested on our dataset through the evolution of our models is shown in Figure A1.

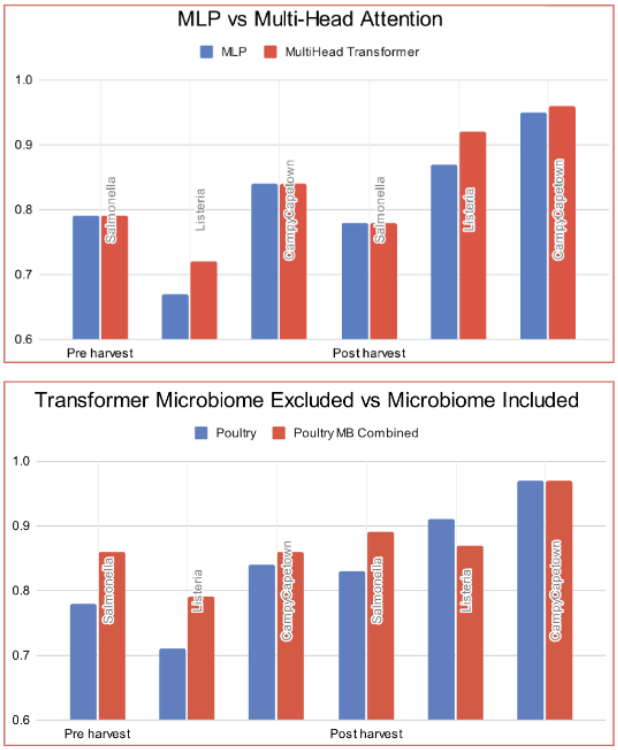


Figure A1. Comparison of the average F1 scores on the test set for predictions made by the different models.

Appendix A.2. PageRank Results

The top 10 features identified with the PageRank method for different scenarios are shown in Figures A2 and A3. Figures A4 and A5 show the agreement in results for both DeepLIFT and our method across different scenarios for the microbiota features. A similar agreement was also found in the variables of theFarm Management Practices dataset, and this is shown in Figure A6.

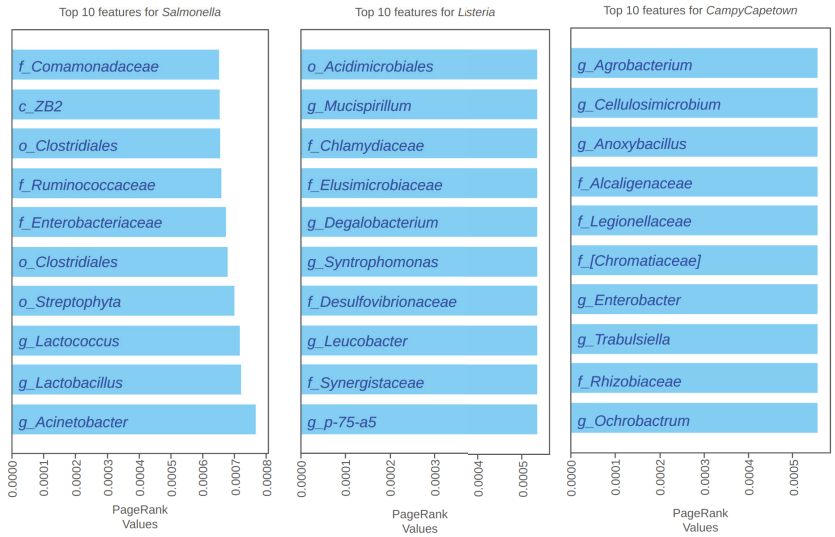


Figure A2. Top 10 features identified with PageRank for post-harvest *Salmonella*, *Listeria*, and *CampyCapetown*.

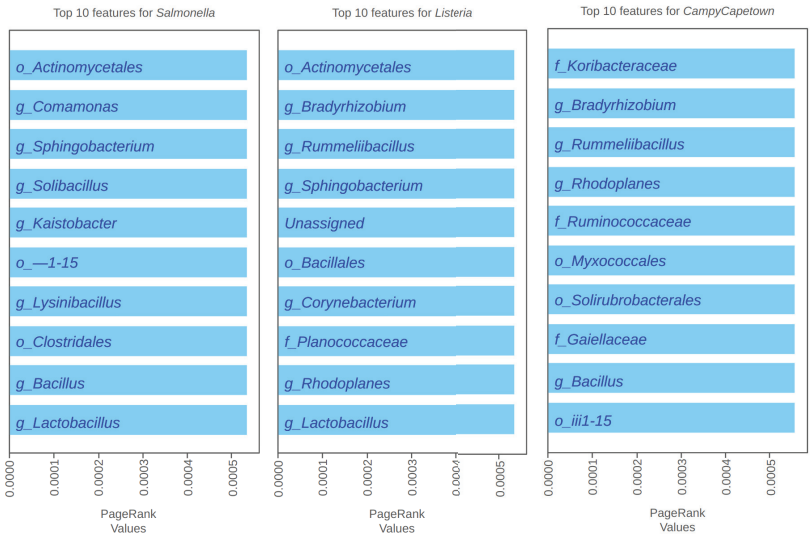


Figure A3. Top 10 features identified with PageRank for pre-harvest *Salmonella*, *Listeria*, and *Campy-Capetown*.

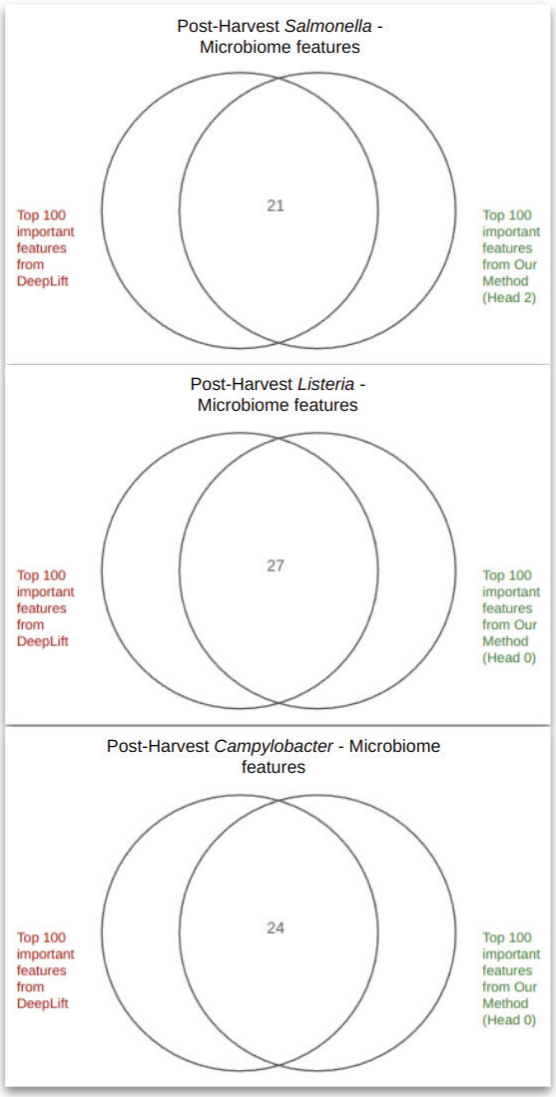


Figure A4. Comparison of DeepLIFT and PageRank. The feature importance identification across different post-harvest scenarios shows a notable agreement.

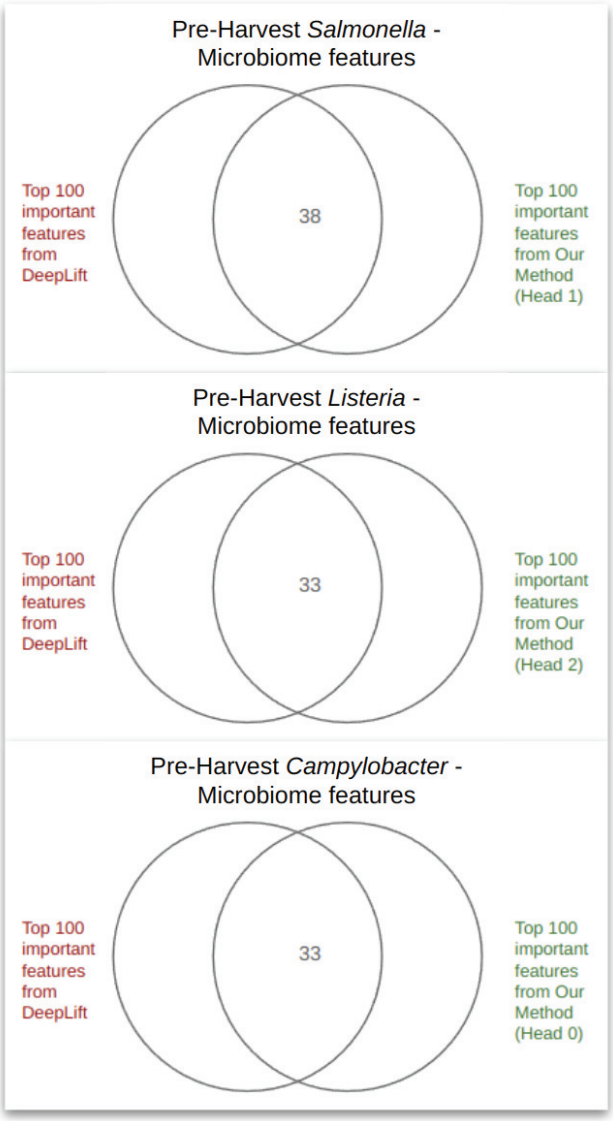


Figure A5. Comparison of DeepLIFT and PageRank. The feature importance identification across different pre-harvest scenarios shows a notable agreement.

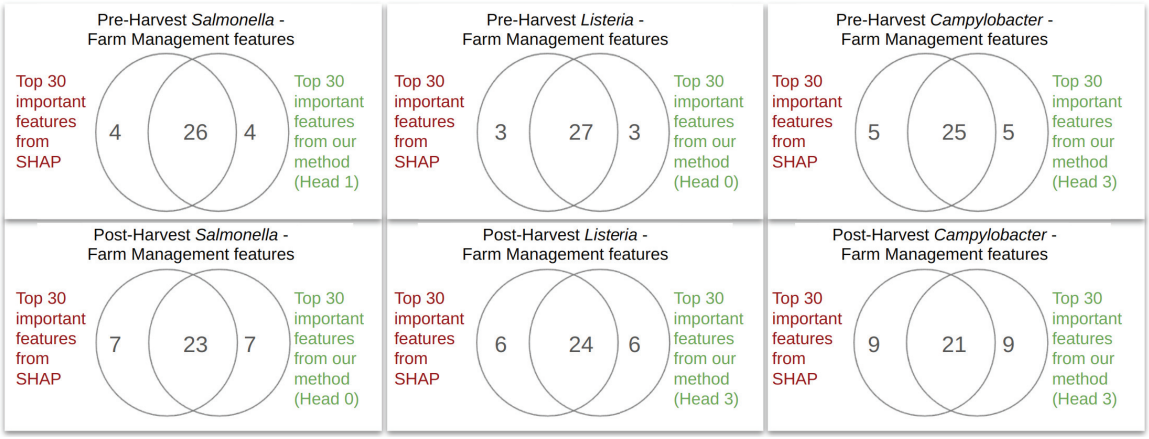


Figure A6. Comparison of DeepLIFT and PageRank. The feature importance identification across all scenarios for farm management features shows a notable agreement.

References

1. Rothrock, M.J.; Davis, M.L.; Locatelli, A.; Bodie, A.; McIntosh, T.G.; Donaldson, J.R.; Ricke, S.C. Listeria Occurrence in Poultry Flocks: Detection and Potential Implications. *Front. Vet. Sci.* **2017**, *4*, 125. [CrossRef]
2. DuPont, H.L. The growing threat of foodborne bacterial enteropathogens of animal origin. *Clin. Infect. Dis.* **2007**, *45*, 1353–1361. [CrossRef]
3. Authority, European Food Safety. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2017. *EFSa J.* **2018**, *16*.
4. Adadi, A.; Berrada, M. Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access* **2018**, *6*, 52138–52160. [CrossRef]
5. Aruwa, C.E.; Pillay, C.; Nyaga, M.M.; Sabiu, S. Poultry gut health–microbiome functions, environmental impacts, microbiome engineering and advancements in characterization technologies. *J. Anim. Sci. Biotechnol.* **2021**, *12*, 119. [CrossRef] [PubMed]
6. Asnicar, F.; Thomas, A.M.; Passerini, A.; Waldron, L.; Segata, N. Machine learning for microbiologists. *Nat. Rev. Microbiol.* **2024**, *22*, 191–205. [CrossRef]
7. Malakar, S.; Sutaoney, P.; Madhyastha, H.; Shah, K.; Chauhan, N.S.; Banerjee, P. Understanding gut microbiome-based machine learning platforms: A review on therapeutic approaches using deep learning. *Chem. Biol. Drug Des.* **2024**, *103*, e14505. [CrossRef]
8. Lundberg, S.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. *arXiv* **2017**, arXiv:1705.07874v2.
9. Shrikumar, A.; Greenside, P.; Kundaje, A. Learning Important Features Through Propagating Activation. Differences *arXiv* **2019**, arXiv:1704.02685.
10. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.u.; Polosukhin, I. Attention is All you Need. In *Advances in Neural Information Processing Systems*; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: San Jose, CA, USA, 2017; Volume 30.
11. Page, L.; Brin, S.; Motwani, R.; Winograd, T. The PageRank Citation Ranking: Bringing Order to the Web. In Proceedings of the The Web Conference, Toronto, ON, Canada, 11–14 May 1999.
12. Xu, K.; Ba, J.; Kiros, R.; Cho, K.; Courville, A.; Salakhudinov, R.; Zemel, R.; Bengio, Y. Show, attend and tell: Neural image caption generation with visual attention. In Proceedings of the International Conference on Machine Learning, PMLR, Lille, France, 6 July–11 July 2015; pp. 2048–2057.
13. Choi, E.; Bahadori, M.T.; Sun, J.; Kulas, J.; Schuetz, A.; Stewart, W. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. *Adv. Neural Inf. Process. Syst.* **2016**, *29*.
14. Lei, T.; Barzilay, R.; Jaakkola, T. Rationalizing neural predictions. *arXiv* **2016**, arXiv:1606.04155.
15. Jain, S.; Wallace, B.C. Attention is not explanation. *arXiv* **2019**, arXiv:1902.10186.
16. Liu, G.; Zhang, J.; Chan, A.B.; Hsiao, J.H. Human attention-guided explainable artificial intelligence for computer vision models. *Neural Networks* **2024**, *177*, 106392. [CrossRef] [PubMed]
17. Kotipalli, B. The Role of Attention Mechanisms in Enhancing Transparency and Interpretability of Neural Network Models in Explainable AI. Master's Thesis, Harrisburg University of Science and Technology, Harrisburg, PA, USA, 2024.
18. Hwang, D.; Rothrock, M.J.; Pang, H.; Dev Kumar, G.; Mishra, A. Farm management practices that affect the prevalence of Salmonella in pastured poultry farms. *LWT* **2020**, *127*, 109423. [CrossRef]
19. Rothrock Jr, M.J.; Locatelli, A.; Feye, K.M.; Caudill, A.J.; Guard, J.; Hiett, K.; Ricke, S.C. A microbiomic analysis of a pasture-raised broiler flock elucidates foodborne pathogen ecology along the farm-to-fork continuum. *Front. Vet. Sci.* **2019**, *6*, 260. [CrossRef]
20. Xu, X.; Rothrock, M.J.; Mohan, A.; Kumar, G.D.; Mishra, A. Using farm management practices to predict Campylobacter prevalence in pastured poultry farms. *Poult. Sci.* **2021**, *100*, 101122. [CrossRef]
21. Rothrock Jr, M.J.; Hiett, K.L.; Gamble, J.; Caudill, A.C.; Cicconi-Hogan, K.M.; Caporaso, J.G. A hybrid DNA extraction method for the qualitative and quantitative assessment of bacterial communities from poultry production samples. *J. Vis. Exp. JoVE* **2014**, *94*, e52161.
22. Caporaso, J.G.; Lauber, C.L.; Walters, W.A.; Berg-Lyons, D.; Lozupone, C.A.; Turnbaugh, P.J.; Fierer, N.; Knight, R. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 4516–4522. [CrossRef]
23. Caporaso, J.G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F.D.; Costello, E.K.; Fierer, N.; Peña, A.G.; Goodrich, J.K.; Gordon, J.I.; et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **2010**, *7*, 335–336. [CrossRef]
24. Edgar, R.C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **2010**, *26*, 2460–2461. [CrossRef]
25. DeSantis, T.Z.; Hugenholtz, P.; Larsen, N.; Rojas, M.; Brodie, E.L.; Keller, K.; Huber, T.; Dalevi, D.; Hu, P.; Andersen, G.L. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **2006**, *72*, 5069–5072. [CrossRef]
26. Caporaso, J.G.; Bittinger, K.; Bushman, F.D.; DeSantis, T.Z.; Andersen, G.L.; Knight, R. PyNAST: A flexible tool for aligning sequences to a template alignment. *Bioinformatics* **2010**, *26*, 266–267. [CrossRef] [PubMed]
27. Huang, X.; Khetan, A.; Cvitkovic, M.; Karnin, Z.S. TabTransformer: Tabular Data Modeling Using Contextual Embeddings. *arXiv* **2020**, arXiv:2012.06678.
28. Von Luxburg, U. A tutorial on spectral clustering. *Stat. Comput.* **2007**, *17*, 395–416. [CrossRef]
29. MacQueen, J. Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Oakland, CA, USA, 21 June 1967; Volume 1, pp. 281–297.

30. Jain, A.K.; Murty, M.N.; Flynn, P.J. Data clustering: A review. *ACM Comput. Surv. (CSUR)* **1999**, *31*, 264–323. [CrossRef]
31. Hoffmann, S.A.; Maculloch, B.; Batz, M. Economic Burden of Major Foodborne Illnesses Acquired in the United States, 2015. Available online: https://www.ers.usda.gov/webdocs/publications/43984/52807_eib140.pdf?v=2344.4 (accessed on 1 June 2024).
32. Vos, P.; Garrity, G.; Jones, D.; Krieg, N.R.; Ludwig, W.; Rainey, F.A.; Schleifer, K.H.; Whitman, W.B. *Bergey's Manual of Systematic Bacteriology: Volume 3: The Firmicutes*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2011; Volume 3.
33. Doyle, M.P.; Diez-Gonzalez, F.; Hill, C. *Food Microbiology: Fundamentals and Frontiers*; John Wiley & Sons: Hoboken, NJ, USA, 2020.
34. Muyyarikkandy, M.S.; Parzygnat, J.; Thakur, S. Uncovering changes in microbiome profiles across commercial and backyard poultry farming systems. *Microbiol. Spectr.* **2023**, *11*, e01682-23. [CrossRef] [PubMed]
35. Leser, T.D.; Mølbak, L. Better living through microbial action: The benefits of the mammalian gastrointestinal microbiota on the host. *Environ. Microbiol.* **2009**, *11*, 2194–2206. [CrossRef] [PubMed]
36. Liu, L.; Lin, L.; Zheng, L.; Tang, H.; Fan, X.; Xue, N.; Li, M.; Liu, M.; Li, X. Cecal microbiome profile altered by Salmonella enterica, serovar Enteritidis inoculation in chicken. *Gut Pathog.* **2018**, *10*, 34. [CrossRef]
37. Zhu, J.; Chen, Y.; Imre, K.; Arslan-Acaroz, D.; Istanbullugil, F.R.; Fang, Y.; Ros, G.; Zhu, K.; Acaroz, U. Mechanisms of probiotic Bacillus against enteric bacterial infections. *One Health Adv.* **2023**, *1*, 21. [CrossRef]
38. Caulier, S.; Nannan, C.; Gillis, A.; Licciardi, F.; Bragard, C.; Mahillon, J. Overview of the antimicrobial compounds produced by members of the Bacillus subtilis group. *Front. Microbiol.* **2019**, *10*, 435128. [CrossRef]
39. Šimunović, K.; Stefanic, P.; Klančnik, A.; Erega, A.; Mandic Mulec, I.; Možina, S.S. Bacillus subtilis PS-216 antagonistic activities against Campylobacter jejuni NCTC 11168 are modulated by temperature, oxygen, and growth medium. *Microorganisms* **2022**, *10*, 289. [CrossRef] [PubMed]
40. Luise, D.; Bosi, P.; Raff, L.; Amatucci, L.; Virdis, S.; Trevisi, P. *Bacillus* spp. probiotic strains as a potential tool for limiting the use of antibiotics, and improving the growth and health of pigs and chickens. *Front. Microbiol.* **2022**, *13*, 801827. [CrossRef] [PubMed]
41. Mazanko, M.S.; Popov, I.V.; Prazdnova, E.V.; Refeld, A.G.; Bren, A.B.; Zelenkova, G.A.; Chistyakov, V.A.; Algburi, A.; Weeks, R.M.; Ermakov, A.M.; et al. Beneficial effects of spore-forming Bacillus probiotic bacteria isolated from poultry microbiota on broilers' health, growth performance, and immune system. *Front. Vet. Sci.* **2022**, *9*, 877360. [CrossRef] [PubMed]
42. Vivant, A.L.; Garmyn, D.; Piveteau, P. Listeria monocytogenes, a down-to-earth pathogen. *Front. Cell. Infect. Microbiol.* **2013**, *3*, 87. [CrossRef] [PubMed]
43. Anjana, A.; Tiwari, S.K. Bacteriocin-producing probiotic lactic acid bacteria in controlling dysbiosis of the gut microbiota. *Front. Cell. Infect. Microbiol.* **2022**, *12*, 851140.
44. Zhang, G.; Raheem, A.; Gao, X.; Zhang, J.; Shi, L.; Wang, M.; Li, M.; Yin, Y.; Li, S.; Cui, X.; et al. Cytoprotective Effects of Lactobacilli on Mouse Epithelial Cells during Salmonella Infection. *Fermentation* **2022**, *8*, 101. [CrossRef]
45. Webb, L.; Ma, L.; Lu, X. Impact of lactic acid bacteria on the control of Listeria monocytogenes in ready-to-eat foods. *Food Qual. Saf.* **2022**, *6*, fyac045. [CrossRef]
46. Rushdy, A.A.; Gomaa, E.Z. Antimicrobial compounds produced by probiotic Lactobacillus brevis isolated from dairy products. *Ann. Microbiol.* **2013**, *63*, 81–90. [CrossRef]
47. Kadam, J.H.; Pawar, R.S.; Din, M.F.M.; Zambare, V. Advances on Probiotics Utilization in Poultry Health and Nutrition. In *Advances in Probiotics for Health and Nutrition*; IntechOpen: London, UK, 2023.
48. Hafez, H.M. Poultry meat and food safety: Pre–and post-harvest approaches to reduce foodborne pathogens. *World's Poult. Sci. J.* **1999**, *55*, 269–280. [CrossRef]
49. Naureen, Z.; Rehman, N.U.; Hussain, H.; Hussain, J.; Gilani, S.A.; Al Housni, S.K.; Mabood, F.; Khan, A.L.; Farooq, S.; Abbas, G.; et al. Exploring the potentials of Lysinibacillus sphaericus ZA9 for plant growth promotion and biocontrol activities against phytopathogenic fungi. *Front. Microbiol.* **2017**, *8*, 1477. [CrossRef]
50. Saw, J.H.; Mountain, B.W.; Feng, L.; Omelchenko, M.V.; Hou, S.; Saito, J.A.; Stott, M.B.; Li, D.; Zhao, G.; Wu, J.; et al. Encapsulated in silica: Genome, proteome and physiology of the thermophilic bacterium Anoxybacillus flavithermus WK1. *Genome Biol.* **2008**, *9*, 1–16. [CrossRef] [PubMed]
51. Mak, P.H.; Rehman, M.A.; Kiarie, E.G.; Topp, E.; Diarra, M.S. Production systems and important antimicrobial resistant-pathogenic bacteria in poultry: A review. *J. Anim. Sci. Biotechnol.* **2022**, *13*, 148. [CrossRef] [PubMed]
52. Sanjorjo, R.A.; Tseten, T.; Kang, M.K.; Kwon, M.; Kim, S.W. In Pursuit of Understanding the Rumen Microbiome. *Fermentation* **2023**, *9*, 114. [CrossRef]
53. Fusco, W.; Lorenzo, M.B.; Cintoni, M.; Porcari, S.; Rinninella, E.; Kaitsas, F.; Lener, E.; Mele, M.C.; Gasbarrini, A.; Collado, M.C.; et al. Short-Chain Fatty-Acid-Producing Bacteria: Key Components of the Human Gut Microbiota. *Nutrients* **2023**, *15*, 2211. [CrossRef] [PubMed]
54. Zhao, Y.; Wei, H.M.; Yuan, J.L.; Xu, L.; Sun, J.Q. A comprehensive genomic analysis provides insights on the high environmental adaptability of Acinetobacter strains. *Front. Microbiol.* **2023**, *14*, 1177951. [CrossRef] [PubMed]
55. Davies, A.R.; Board, R.; Board, R. *Microbiology of Meat and Poultry*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1998.
56. Joshi, S.G.; Litake, G.M. Acinetobacter baumannii: An emerging pathogenic threat to public health. *World J. Clin. Infect. Dis.* **2013**, *3*, 25–36. [CrossRef]

57. Samtiya, M.; Matthews, K.R.; Dhewa, T.; Puniya, A.K. Antimicrobial resistance in the food chain: Trends, mechanisms, pathways, and possible regulation strategies. *Foods* **2022**, *11*, 2966. [CrossRef] [PubMed]
58. Calderon, R.B.; Jeong, C.; Ku, H.H.; Coghill, L.M.; Ju, Y.J.; Kim, N.; Ham, J.H. Changes in the microbial community in soybean plots treated with biochar and poultry litter. *Agronomy* **2021**, *11*, 1428. [CrossRef]
59. Shah, D.H.; Board, M.M.; Crespo, R.; Guard, J.; Paul, N.C.; Faux, C. The occurrence of Salmonella, extended-spectrum β-lactamase producing Escherichia coli and carbapenem resistant non-fermenting Gram-negative bacteria in a backyard poultry flock environment. *Zoonoses Public Health* **2020**, *67*, 742–753. [CrossRef]
60. Herrera, P.; Kwon, Y.M.; Ricke, S.C. Ecology and pathogenicity of gastrointestinal Streptococcus bovis. *Anaerobe* **2009**, *15*, 44–54. [CrossRef]
61. Hyun, D.W.; Whon, T.W.; Cho, Y.J.; Chun, J.; Kim, M.S.; Jung, M.J.; Shin, N.R.; Kim, J.Y.; Kim, P.S.; Yun, J.H.; et al. Genome sequence of the moderately halophilic bacterium Salinicoccus carnicancri type strain Crm T (=DSM 23852 T). *Stand. Genom. Sci.* **2013**, *8*, 255–263. [CrossRef] [PubMed]
62. Cycil, L.M.; DasSarma, S.; Pecher, W.; McDonald, R.; AbdulSalam, M.; Hasan, F. Metagenomic insights into the diversity of halophilic microorganisms indigenous to the Karak Salt Mine, Pakistan. *Front. Microbiol.* **2020**, *11*, 1567. [CrossRef] [PubMed]
63. Klein, G. Taxonomy, ecology and antibiotic resistance of enterococci from food and the gastro-intestinal tract. *Int. J. Food Microbiol.* **2003**, *88*, 123–131. [CrossRef] [PubMed]
64. Micallef, S.A.; Goldstein, R.E.R.; George, A.; Ewing, L.; Tall, B.D.; Boyer, M.S.; Joseph, S.W.; Sapkota, A.R. Diversity, distribution and antibiotic resistance of *Enterococcus* spp. recovered from tomatoes, leaves, water and soil on US Mid-Atlantic farms. *Food Microbiol.* **2013**, *36*, 465–474. [CrossRef]
65. Bao, Y.; Dolfing, J.; Guo, Z.; Chen, R.; Wu, M.; Li, Z.; Lin, X.; Feng, Y. Important ecophysiological roles of non-dominant Actinobacteria in plant residue decomposition, especially in less fertile soils. *Microbiome* **2021**, *9*, 84. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

microorganisms

MDPI

Article

Whole-Genome Deep Sequencing of the Healthy Adult Nasal Microbiome

Mark Cannon [1,2,*], Gustavo Ferrer [3], Mari Tesch [3] and Matthew Schipma [4]

[1] Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL 60611, USA
[2] Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA
[3] Aventura Hospital Pulmonary and Critical Care Fellowship, Aventura, FL 33180, USA; gferrer@pulmonary-institute.com (G.F.); research@drferrerbiopharma.com (M.T.)
[4] QDSC, NUSeq Core, Northwestern University, Chicago, IL 60611, USA; m-schipma@northwestern.edu
* Correspondence: drmarkcannon@outlook.com

Abstract: This study aimed to determine shifts in microbial populations regarding richness and diversity from the daily use of a popular over-the-counter nasal spray. In addition, the finding of nasal commensal bacterial species that overlap with the oral microbiome may prove to be potential probiotics for the "gateway microbiomes". Nasal swab samples were obtained before and after using the most popular over-the-counter (OTC) nasal spray in 10 participants aged 18–48. All participants were healthy volunteers with no significant medical histories. The participants were randomly assigned a number by randomizing software and consisted of five men and five women. The sampling consisted of placing a nasal swab atraumatically into the nasal cavity. The samples were preserved and sent to Northwestern University Sequencing Center for whole-genome deep sequencing. After 21 days of OTC nasal spray use twice daily, the participants returned for further nasal microbiome sampling. The microbial analysis included all bacteria, archaea, viruses, molds, and yeasts via deep sequencing for species analysis. The Northwestern University Sequencing Center utilized artificial intelligence analysis to determine shifts in species and strains following nasal spray use that resulted in changes in diversity and richness.

Keywords: deep sequencing; nasal microbiome; gateway microbiome; diversity; richness; probiotics; nasal spray

1. Introduction

The number of patients with an allergic response to their environment is increasing and is now more than 40% in numerous populations in the United States and Europe [1]. Allergic rhinitis refers to an allergic response that occurs in the nose [2]. The number of patients suffering from allergic rhinitis, commonly known as hay fever, in the United States is approximately 30% [3]. Allergic rhinitis contributes to lost or unproductive time at work and school, sleep problems, and reduced participation in outdoor activities. The ability to control asthma and allergic rhinitis in individuals with asthma has been linked to the control of allergic rhinitis [4].

According to published studies, most individuals with asthma develop rhinitis [5]. In addition, the presence of allergic rhinitis (seasonal or perennial) significantly increases the likelihood of asthma; up to 40% of people with allergic rhinitis have or will have asthma [6]. It is also essential to define how breathing is affected in patients with obstructive sleep apnea (OSA) [7]. Breathing starts at the nose, and while OSA is characterized by the collapse of the muscles of the oropharyngeal airways, nasal obstruction and OSA are usually co-existing conditions that worsen each individual condition [8–10]. Therapies to improve airflow through the nose in compromised patients significantly reduce daytime and nighttime symptoms [11]. The nose is responsible for almost 50% of the resistance when transporting air from the nose to the lungs and plays an essential role in humidification,

heating, and air filtration [12]. The tissues inside the nose, known as the nasal mucosa, are dynamic organs regulated by the autonomic nervous system [13]. Periodic nasal congestion and decongestion are termed the "nasal cycle" [14]. In patients with a permanent one-sided nasal obstruction, the nasal cycle can make breathing air into the body difficult [15].

Nasal sprays are used by a large percentage of the population, especially with recent news reports demonstrating the inhibitory action of xylitol nasal sprays and other ingredients on the attachment (then endocytosis) of SARS-CoV-2 virus to nasal cells [16–18]. In addition, xylitol has also been reported to inhibit the oncogenesis of oral cells, making xylitol nasal sprays even more desirable for this population [19]. However, no study has analyzed the potential shift in the nasal microbiome after using xylitol nasal spray.

Nasal and lung microbiomes may be affected by the gut microbiome [20]. Formula-fed infants are at an increased risk of infections [21]. Owing to the crosstalk between the mucosal systems of the gastrointestinal and respiratory tracts, adding synbiotics (prebiotics and probiotics) to infant formulas may prevent infections even at distant sites [22,23]. In a study on infants who were born full-term and weaned from breast milk, they were randomized to either prebiotic formula (fructo- and galactooligosaccharides) or the same prebiotic formula along with *Lactobacillus paracasei* ssp. *paracasei* F19 (synbiotics) from 1 to 6 months of age. Synbiotic feeding led to reduced Klebsiella, enrichment of bifidobacteria, and increases in microbial degradation metabolites implicated in immune signaling and in the gut–lung and gut–skin axes [23]. With the synbiotic effects on gut microbiota development, *L. fermentum* PCC and *L. reuteri* RC-14 were more resistant to gastric conditions, and their survival rate was further improved in the presence of 5 out of 10 tested pectins. Additionally, two pectins positively affected the viability of the less resistant *L. rhamnosus LGG* and *L. paracasei* F-19 [24].

Since the advent of high-throughput sequencing, PCR-amplified 16S sequences have typically been clustered based on their similarity to generate operational taxonomic units (OTUs), and representative OTU sequences have been compared with reference databases to infer the likely taxonomy. However, convenient and powerful usage of 16S rRNA has necessitated certain assumptions, such as the now historical assumption that sequences of >95% identity represent the same genus. In contrast, sequences with >97% identity represent the same species [25,26].

Objective:

The principal objective of this study was to discover the normal complete nasal microbiome after using an over-the-counter nasal spray, which, as of this date, has yet to be completed. In addition, more information is needed on the microbiome shifts that occur in response to xylitol nasal spray. Before the widespread use of nasal sprays, what may be considered normal should be determined. Shifts in microbial diversity and richness should be regarded as likely.

2. Materials and Methods

This study performed a nasal swab test before and after using the most popular OTC nasal spray in 10 participants aged 18–48. All participants had non-significant medical histories with no history of allergies or medication use that could affect the nasal microbiome. The participants consisted of five men and five women who were randomly assigned a number using randomizing software. The participants signed an informed consent form (submitted to the IRB and accepted), and the principal investigator kept all personal protected information in a hard copy safely locked in a secure cabinet.

An e-mail newsletter recruited participants from all patients of the principal investigator, and recruiting posters were placed in the reception area of the PI office. No inducements or compensation were provided to the participants. The study protocol, informed consent, medical history, and sampling are discussed separately. The samples were preserved and shipped by FedEx to Northwestern University's NUSeq Center for Whole Genome Deep Sequencing. After 21 days of OTC nasal spray use twice daily, the participants returned for another sampling of their nasal microbiome, which FedEx then shipped to NUSeq for

deep sequencing. The participants were instructed to use the spray before dismissal during the first visit. The spray was obtained as samples normally distributed to clinics at no charge and was already in the principal investigator's inventory. These samples were often administered to pediatric patients diagnosed with sleep apnea.

Sampling consisted of placing a nasal swab atraumatically into the anterior nasal cavity, similar to self-performed sampling for COVID-19 tests. No injury was reported, and only slight discomfort, if any, was reported. The swab was placed approximately one inch into the nares, rotated for 15 s, and then placed into the other nares for another rotation of 15 s. The swabs were placed in a preservative and sealed safely. Biosafety envelopes were used for shipping according to previous protocols. After the second sampling, patients were excluded from the study. All personal information was destroyed after two years.

Xlear Sinus Care is an OTC xylitol solution. Xylitol is a natural sugar (pentose aldose) that inhibits biofilm formation [27]. Pure xylitol is a white crystalline substance in many fruits and vegetables [28]. According to the manufacturer, the Xlear solution cleanses, hydrates, dries, and irritates tissues. Xlear products use a patented xylitol solution that inhibits bacteria and pulls moisture into the nasal cavity. According to recently published research, xylitol also inhibits the viral invasion of cells [17].

Ingredients:

- Purified water;
- Xylitol;
- Saline;
- Grapefruit seed extract.

Study Data:

Microbiome Analysis. The quality of the reads in the FASTQ format was evaluated using FastQC (v1.12.1). Reads were trimmed to remove Illumina adapters from the 3' ends using Cutadapt (3.7) [29]. Paired-end trimmed reads were aligned using Kraken2 (2.1.2) with default parameters, except for using 12 threads instead of the default value of 1 [30]. The standard Kraken2 database includes archaea, bacterial, viral, plasmid, and human genomes. Taxa were determined using the standard database, which was loaded using kraken2-build. Alignments that could not be classified into known taxa were excluded. Taxon abundance estimates were performed using Bracken (2.7) with default parameters, except for using 12 threads instead of the default value of 1 [31]. Normalization and differential expression were calculated using DESeq2 (1.14.1) [32]. The libraries were constructed using Nextera kits according to the manufacturer's instructions (Illumina, San Diego, CA, USA). The samples were sequenced paired-end 150 base pairs on the NovaSeq 6000 instrument (Illumina, San Diego, CA, USA). The total sequencing depth was nearly 2.3 billion reads with an average of 47 million reads per sample. Default parameters were used to improve standardization and remain within NUseq Center protocols. Data reside at the NUSeq Center at Northwestern University, and the nasal samples are preserved for future additional analysis.

3. Results

The analysis determined 2558 taxa with greater abundances of the following top taxa: *Xanthobacter autotrophicus*, *Streptomyces* sp. *ICC1*, *Streptomyces armeniacus*, *Streptococcus thermophilus*, *Streptococcus sanguinis*, *Streptococcus pyogenes*, *Streptococcus oralis*, *Staphylococcus epidermidis*, *Staphylococcus aureus*, *Salmonella enterica*, *Rothia aeria*, *Ralstonia pickettii*, *Pseudomonas aeruginosa*, *Peptoniphilus harei*, *Lawsonella clevelandensis*, *Klebsiella pneumoniae*, *Finegoldia magna*, *Escherichia coli*, *Enterococcus faecium*, *Dolosigranulum pigrum*, *Delftia lacustris*, *Cutibacterium granulosum*, *Cutibacterium acnes*, *Corynebacterium tuberculostearicum*, *Corynebacterium striatum*, *Corynebacterium segmentosum*, *Corynebacterium propinquum*, *Corynebacterium macginleyi*, *Corynebacterium kefirresidentii*, *Corynebacterium glutamicum*, *Corynebacterium diphtheriae*, *Burkholderia dolosa*, *Brevibacillus brevis*, *Bartonella krasnovii*, *Anaerococcus prevotii*, *Aeromonas caviae*, and *[Haemophilus] ducreyi*.

The routine use of xylitol nasal spray changed the relative abundance of ten taxa (Table 1). This result is not surprising because recently published research has demonstrated that even a few species being changed have a cascading effect throughout the microbiome [23]. Xylitol should reduce the number of pathogenic bacteria, which would also affect the pathogens' synergistic co-pathogens [33–35]. Reports have previously been published that the oral microbiome has a "downstream effect" on the gut microbiome [36,37]. It would not be surprising if the nasal gateway microbiome did not have the same effect, with a reduction in pathogens of the anterior nasal passageway strain/species shifting the respiratory microbiome.

Akkermansia muciniphila
Rhodococcus qingshengii
Streptomyces sp. *ICC1*
Streptomyces armeniacus
Actinomyces sp. *oral taxon 414*
Klebsiella oxytoca
Acinetobacter ursingii
Klebsiella variicola
Enterobacter hormaechei
Brevundimonas diminuta

Table 1. The relative abundance of these ten taxa was significantly different after using the xylitol nasal spray. Four species decreased, and six increased in abundance. *Rhodococcus qingshengii* (soil organism with anti-fungal properties) significantly increased, but *Akkermansia muciniphilia* substantially decreased. *Brevundimonas diminuta* was also decreased.

Microbe	log2FoldChange	lfcSE	Stat	*p* Value
Akkermansia muciniphila	−21.42441822	2.967913092	−7.218681128	5.25×10^{-13}
Rhodococcus qingshengii	20.77410089	2.97113001	6.991986491	2.71×10^{-12}
Streptomyces sp. *ICC1*	6.922848082	1.548527233	4.470601442	7.80×10^{-6}
Streptomyces armeniacus	−3.191007111	0.919334063	−3.470998453	0.000518527
Actinomyces sp. *oral*	−3.771603944	1.615989861	−2.333927976	0.019599494
Klebsiella oxytoca	5.272476299	2.286510228	2.305905408	0.021115918
Acinetobacter ursingii	6.534219931	2.968012555	2.201547268	0.027697304
Klebsiella variicola	1.349165754	0.630129643	2.141092342	0.032266592
Enterobacter hormaechei	2.117319005	1.007575689	2.101399458	0.035605916
Brevundimonas diminuta	−5.607547967	2.699838683	−2.076993712	0.03780214

Four species decreased, with six increasing abundance (Figure 1. Volcano plot of A versus non-A, after xylitol and before xylitol). *Rhodococcus qingshengii* (soil organism with anti-fungal properties) significantly increased, but *Akkermansia muciniphilia* substantially decreased. *Akkermansia muciniphilia* is considered a probiotic in the gut and the oral cavity. *Brevundimonas diminuta* was also decreased and is ubiquitous in humans but has been associated with infections in immunocompromised and cancer patients. *Acinetobacter ursingii* increased in predominance and has been associated with immunocompromised patient infections (see Table 1). It is essential to note that almost all microbes may cause an illness if in the wrong place at the wrong time. We lack sufficient knowledge of the nasal gateway microbiome to properly judge which bacteria are acceptable commensals and which are pathobionts. For instance, *Streptomyces armeniacus* is a spore-forming soil organism, a probiotic producing Streptopyrrole, and the abundance is increased by xylitol therapy. The presence of so many soil-based microorganisms is to be expected as they may become airborne via dust particles. One can only imagine the bacterial load a runner takes

in nasally as they sprint on a dirt path. The Streptopyrrole probiotic bacteria *Streptomyces armeniacus* may be an essential link in nasal health.

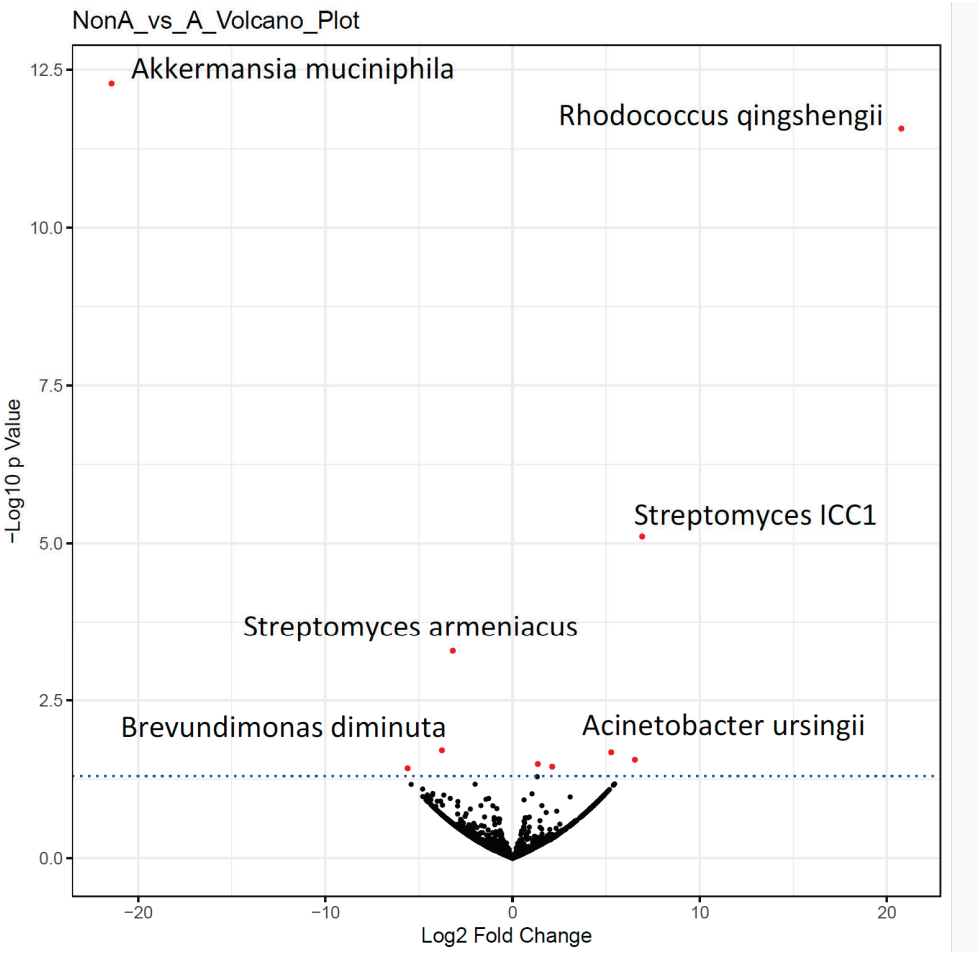


Figure 1. Non-A_vs._A Volcano Plot demonstrating microbiome shift from control to xylitol-treated nasal microbiome. This is a labeled volcano plot that plots the *p*-values against the log fold change. Red dots are considered significantly different. The further a dot is away from the origin, the more significant the difference. Any taxa above the blue dotted line displayed significant changes after nasal spray use.

The greater majority of taxa were bacterial, although these viruses were more predominant, Pahexavirus ATCC29399BT, Propionibacterium phage BruceLethal (infects *Propionibacterium acnes*), Streptococcus phage IPP9, and Streptococcus phage phiARI0455b. The fungus *Puccinia graminis*, which causes significant disease in cereal crops, was also found in the samplings. Crop species affected by the disease include bread wheat, durum wheat, barley, and triticale. Another fungus was *Candida tropicalis*, a pathogen similar to *Candida albicans* and capable of forming a pathogenic biofilm. *Blastocytis hominis*, a single-cell parasite responsible for gastrointestinal infections, but with 22 strains, of which only nine can cause infections, was also isolated. *Methanosarcina mazei* an anaerobic archaeon that can metabolize various substrates and produce methane was present. The 50 most abundant taxa are listed in Table 2.

Table 2. The list is of the top 50 most common taxa, with bacterial species being predominant. Two key oral probiotics, *Streptococcus oralis* and *Streptococcus sanguinis*, are present in the nasal microbiome.

Staphylococcus aureus
Vibrio sp. *Scap24*
Corynebacterium segmentosum
Aeromona scaviae
Cutibacterium acnes
Salmonella enterica
Streptomyces sp. *ICC1*
Dolosigranulum pigrum
Klebsiella pneumoniae
Delftia lacustris
Corynebacteriumm kefirresidentii
Corynebacterium propinquum
Staphylococcus epidermidis
Finegoldia magna
Corynebacterium tuberculostearicum
Lawsonella clevelandensis
Corynebacterium macginleyi
Bartonella krasnovii
Peptoniphilus harei
Cutibacterium granulosum
Pseudomonas aeruginosa
Enterococcus faecium
Xanthobacter autotrophicus
Aeromonas hydrophila
Haemophilus ducreyi
Streptococcus mitis
Burkholderia dolosa
Xanthomonas citri
Staphylococcus capitis
Klebsiella grimontii
Escherichia coli
Cutibacterium avidum
Corynebacterium striatum
Klebsiella variicola
Streptomyces armeniacus
Enterococcus faecalis
Delftia acidovorans
Citrobacter koseri
Alistipes communis
Streptococcus oralis
Corynebacterium diphtheriae

Table 2. *Cont.*

Actinomyces oris
Streptococcus pneumoniae
Mycolicibacter sp. *MYC340*
Streptococcus sanguinis
Micrococcus luteus
Ralstonia pickettii
Mycolicibacterium austroafricanum
Bacillus sp. *BD59S*

Pearson correlation demonstrated statistical differences between the A and Non-A groups. A group refers to the subjects using a xylitol spray for 21 days, non-A was the group before using the spray. A and non-A nomenclature was used to blind the laboratory personnel to the nature of the research (see Figure 2). Similar to a correlation matrix, a heatmap demonstrated the differences between the A and non-A groups (see Figure 3).

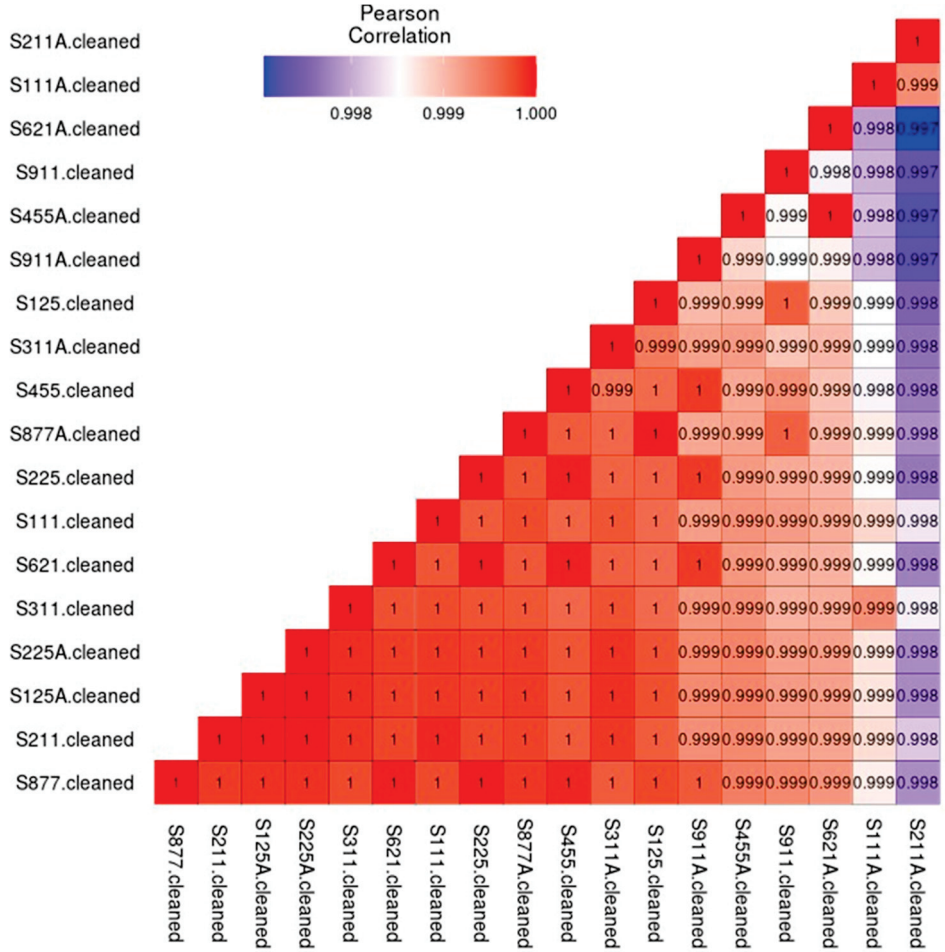


Figure 2. Pearson Correlation demonstrates the statistical differences between the A and non-A groups. In this figure, A refers to subjects after using a xylitol spray for 21 days. This is a correlation matrix displaying the R-squared value of the correlation between each pair of samples. A perfect score is 1.0; the higher the score, the more similar the samples. Negative scores mean negative correlations; that is, the microbe abundances are high in one sample and low in the other. Cleaned refers to the samples having all contaminants and human DNA removed.

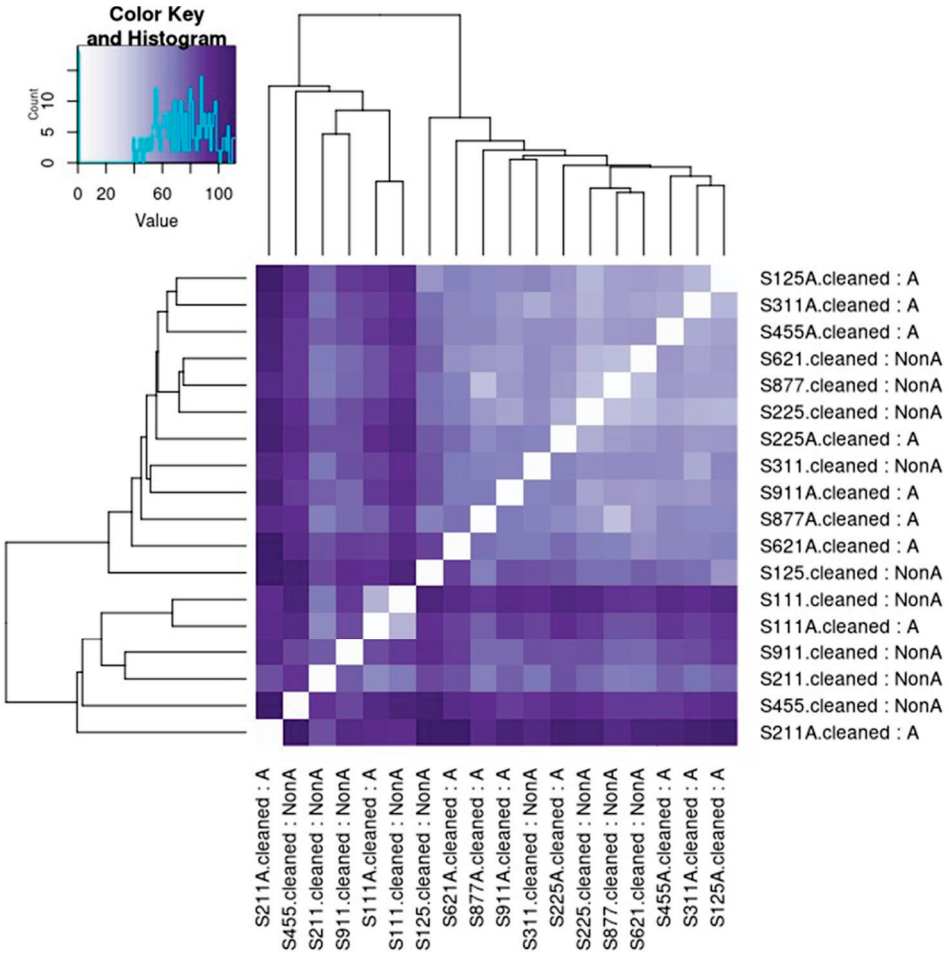


Figure 3. Heatmap demonstrating the differences between the A and non-A groups. Non-A is before the use of the xylitol spray. Cleaned means that all containment has been removed, especially human DNA. This is a heatmap showing the similarity between samples, similar to a correlation matrix. The Euclidean distance is computed between each pair of samples, and the distribution of distances is shown in the legend of the upper left. The lighter the shade of purple, the smaller the distance. The smaller the distance, the more similar the samples are. This graph shows an extra feature of clustering the samples based on similarity. Even though both are shown, the x and y axes are redundant.

4. Discussion

The microbial analysis included all bacteria, archaea, viruses, molds, and yeasts obtained via deep sequencing for species analysis. Artificial intelligence analyses by the Northwestern University Sequencing Center revealed shifts in species and strains due to the use of the OTC nasal spray, resulting in changes in diversity or richness. Deep sequencing and artificial intelligence analysis may be used to determine changes in Richness and Shannon diversity. Genus, species, and strain shifts determined the precise response of the nasal microbiome to the most common OTC nasal spray. Moreover, the complete nasal microbiome by deep sequencing was yet to be discovered. In addition, more information was needed on the microbiome shifts that occur in response to xylitol nasal spray. Before the widespread use of COVID-19 precautions, what may be considered the normal nasal microbiome needed to be determined. The nasal microbiome of healthy adults before and after xylitol exposure has yet to be analyzed using whole-genome deep sequencing methods.

The large number of taxa found may not represent the total number likely present in the nasal cavity. The study ended at the number reported mainly owing to its economic restraints and the use of artificial intelligence with machine language development, which demanded technology resources. In addition, there are often constraints owing to the

availability of DNA libraries for comparison. Regardless of these constraints, we found that routine xylitol nasal spray use significantly affects the prevalence of at least ten taxa.

Several probiotic bacteria have been identified in the nasal cavity; oral or gut probiotics are often abundant. The nasal microbiome also contains several skin commensals, such as *Staphylococcus epidermidis* and *Cutibacterium acnes*. This indicates that the nasal microbiome combines the features of both the oral and skin microbiomes. Additionally, bacteria are often specific to the nasal cavity. Previous studies have shown that the most abundant bacteria are *Staphylococcus aureus*, *Staphylococcus epidermidis*, and *Propionibacterium acnes* [38,39].

Of evolutionary importance, *Bartonella kraznovii* is associated with sub-Saharan black African rats [40]. During the great "throttling", *Homo sapiens* moved into caves along the seacoast because of significant climate changes (Marine Isotope Stage 6) [41]. The surviving H. sapiens' diet included tubers, often roasted in caves [42]. The roasted tubers were consumed as a survivor's food, but any leftovers would have been available for the cave rodents. The association of this nasal bacterium with rodents that share *Homo sapiens* caves should not be surprising and is an interesting result of this study.

Burkholderia dolosa is a species of bacteria member of the *Burkholderia cepacia* complex. This strain is highly drug-resistant and primarily found in immunocompromised patients [43]. *B. dolosa* chronic infection in cystic fibrosis is associated with an accelerated loss of lung function and decreased survival [44]. However, this was observed in several participants. Another rare pathogen found in patients with cystic fibrosis, *Achromobacter xylosoxidans*, was also found in a few participants [45]. However, as in all diseases, it is not just the presence of pathogens but the absence of compensatory commensals and probiotic microorganisms that influence the expression of pathology [46]. Xylitol solutions may inhibit pathogens without negatively reducing the levels of probiotic bacteria, as has often been reported in the literature [47]. Xylitol nebulizers have proven effective in patients with cystic fibrosis [48].

Nasal microbiome changes and sleep-disturbed breathing may be correlated, which warrants significant research [49]. This research project utilized young adults, not children, and examined the possible correlation between the nasal microbiome and nasal obstruction due to chronic inflammation, which leads to nasal airway obstruction [50]. The connection between the airway and microbiome has been well established. Although many researchers believe that the airway affects the microbiome, they do not consider that it affects respiration, albeit indirectly. For instance, the oral microbiome produces nitrites from nitrates, eventually processed by stomach acid into nitric oxide [51]. Salivary nitric oxide inhibits decay, reduces periodontal disease, lowers CRP, and potentially affects airway resistance [52]. Systemic serum levels are associated with normosystolic blood pressure during pregnancy [53]. Nitrate-reducing oral bacterial levels are also linked to a normal pregnancy, and oral dysbiosis due to *Porphyromonas gingivalis* causes pre-eclampsia [54–56].

Ideally, future studies should use larger population samples to examine the association between diseases and microbiome shifts. The first step in our study was to analyze the healthy microbiome before and after a 21-day course of xylitol (see Figure 4). All patients were asymptomatic; therefore, no clinical correlations could be established. Animal studies have demonstrated an association between the gut and feline upper respiratory tract disease, specifically in felines. FURTD, often caused by infectious etiologies, is a multifactorial syndrome that affects feline populations. The eco-phylogenetic method identified 136 and 89 microbial features, respectively, within the gut and nasal microbiomes, significantly associated with active FURTD clinical signs [57]. Nasal and gut microbial community members are associated with a chronic clinical course [58]. Studies have shown that endogenous microbiome dysbiosis can affect mucosal health and disease severity. Some bacterial species exhibit protective properties, whereas others are pathogenic [59]. Antimicrobial agents can create a similar disruption, affect the nasal microbiome balance, and provoke allergic responses [60]. Probiotics offer a promising avenue for developing systemic and topical therapies to strategically manipulate the biological host load, thereby augmenting immune homeostasis [61]. Combining probiotics indigenous to oral and nasal

cavities with prebiotics, such as xylitol, may inhibit pathogens and restore nasal health. This combination therapy may benefit the developing fetus and infant from maternal usage and prevent pathogenic microbiome development [62–64]. An example of a potential nasal probiotic therapy is the discovery that *Lactobacilli* taxa can be present on the facial skin, in the nasal cavity, and in the vagina. Hypothetically, multiple microniches provide commensal protection to multiple microniches [65]. Further discovery that *Lactobacilli casei* (AMBR2) is present in healthy individuals but not in those with chronic rhinosinusitis led to the new designation of "keystone" probiotic commensal [66].

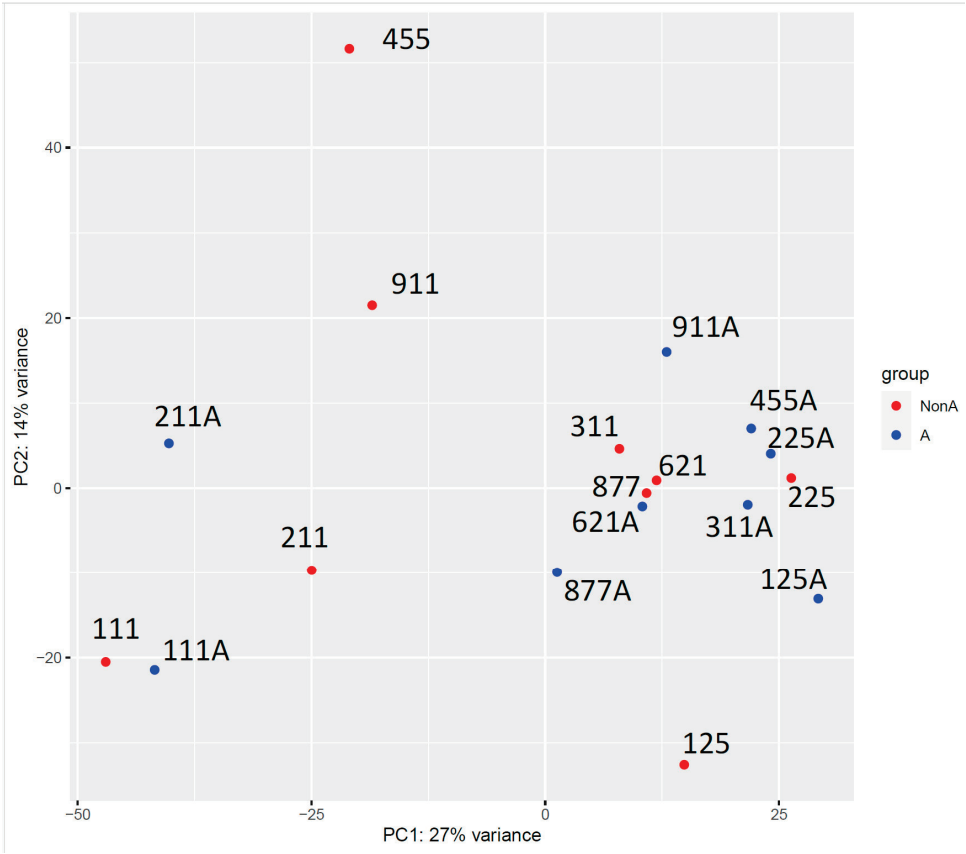


Figure 4. This is a PCA (principal components analysis) plot that shows the similarities between samples, the nearer samples are on the plot, the more similar. Each sample is labeled.

Because so many other respiratory illnesses are associated with a dysbiotic nasal microbiome, discovering a eubiotic nasal microbiome should be an important objective [3–6]. A study by Chin et al. showed that the upper airway microbiome is associated with airway inflammation disorders and the level of asthma control [67]. This would indicate that the goal of healthcare professionals should always be not just to treat symptoms but to re-establish an eubiotic nasal microbiome [68]. This treatment could be started in infancy and then successfully prevent later respiratory disease [69]. The knowledge of what constitutes a healthy nasal microbiome in symptomless young adults and how to maintain the microbiome is, therefore, very significant. The present study helps provide the necessary data on what is bacterially present in asymptomatic young adults.

Another important aspect of the nasal microbiome is its importance in overall health. Nasal vaccines may prevent many respiratory diseases by increasing the IgA antibody response and improving mucosal defenses [70]. Nasal vaccines induce protective immunity that prevents the invasion of pathogens from mucosal surfaces and inhibits severe disease. Nasal vaccines are most appropriate for respiratory infectious diseases because they induce effective immunity in the upper and lower respiratory tracts [71]. Intranasal vaccination

induces the nasal IgA antibody that protects against respiratory viruses like influenza and SARS-CoV-2 [72]. Using commensal bacteria with hemagglutinin in nasal vaccines increased the antibody response [73,74]. Future nasal vaccines may utilize newly discovered nasal commensal bacteria as adjutant carriers of the antigen that stimulates the protective immune response.

5. Conclusions

The nasal microbiome is rich and diverse, containing taxa often associated with oral and skin microbiomes. Microbial interventions, even with OTC nasal rinses, may significantly affect the nasal microbiome, and associated microbiomes.

Author Contributions: Conceptualization, M.C.; methodology, M.S.; software, M.S.; validation, M.C., M.T. and G.F.; formal analysis, M.S.; investigation, M.C.; resources, M.C.; data curation, M.S.; writing—original draft preparation, M.C.; writing—review and editing, G.F.; visualization, M.T.; supervision, M.C.; project administration, M.C.; funding acquisition, M.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by The Swanson Foundation, a Northwestern University private research fund.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board, Pearl IRB Indianapolis IN, identified as 18-Grov-101 for studies involving humans.

Informed Consent Statement: Informed consent and/or assent was obtained from all subjects involved in the study.

Data Availability Statement: Due to the need to maintain strict privacy of the subjects (children), data were protected, and consent contained specific language assuring complete safeguarding measures.

Acknowledgments: The donors of the Swanson Fund, specifically R William Cornell, Alonzo Jones, and Michael Milligan.

Conflicts of Interest: The authors declare no conflicts of interest. The Swanson Fund had no role in the design of the study, in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

References

1. Salo, P.M.; Arbes, S.J., Jr.; Jaramillo, R.; Calatroni, A.; Weir, C.H.; Sever, M.L.; Hoppin, J.A.; Rose, K.M.; Liu, A.H.; Gergen, P.J.; et al. Prevalence of allergic sensitization in the United States: Results from the National Health and Nutrition Examination Survey (NHANES) 2005–2006. *J. Allergy Clin. Immunol.* **2014**, *134*, 350–359. [CrossRef] [PubMed]
2. Wheatley, L.M.; Togias, A. Clinical practice. *Allergic rhinitis. N. Engl. J. Med.* **2015**, *372*, 456–463. [CrossRef] [PubMed]
3. Settipane, R.A. Demographics and epidemiology of allergic and nonallergic rhinitis. *Allergy Asthma Proc.* **2001**, *22*, 185–189. [PubMed]
4. Bousquet, J.; Khaltaev, N.; Cruz, A.A.; Denburg, J.; Fokkens, W.J.; Togias, A.; Zuberbier, T.; Baena-Cagnani, C.E.; Canonica, G.W.; van Weel, C.; et al. (Allergic Rhinitis and its Impact on Asthma (ARIA) 2008 update (in collaboration with the World Health Organization, GA(2)LEN and AllerGen). *Allergy* **2008**, *63* (Suppl. S86), 8–160. [CrossRef] [PubMed]
5. Cruz, A.A.; Popov, T.; Pawankar, R.; Annesi-Maesano, I.; Fokkens, W.; Kemp, J.; Ohta, K.; Price, D.; Bousquet, J.; ARIA Initiative Scientific Committee. Common characteristics of upper lower airways in rhinitis asthma: ARIA update in collaboration with, GA(2)LEN. *Allergy* **2007**, *62* (Suppl. S84), 1–41. [CrossRef]
6. Shaaban, R.; Zureik, M.; Soussan, D.; Neukirch, C.; Heinrich, J.; Sunyer, J.; Wjst, M.; Cerveri, I.; Pin, I.; Bousquet, J.; et al. Rhinitis and onset of asthma: A longitudinal population-based study. *Lancet* **2008**, *372*, 1049–1057. [CrossRef] [PubMed]
7. Chirakalwasan, N.; Ruxrungtham, K. The linkage of allergic rhinitis and obstructive sleep apnea. *Asian Pac. J. Allergy Immunol.* **2014**, *32*, 276–286. [PubMed]
8. Zheng, M.; Wang, X.; Ge, S.; Gu, Y.; Ding, X.; Zhang, Y.; Ye, J.; Zhang, L. Allergic and non-allergic rhinitis are common in obstructive sleep apnea but not associated with disease severity. *J. Clin. Sleep Med.* **2017**, *13*, 959–966. [CrossRef] [PubMed]
9. Zheng, M.; Wang, X.; Zhang, L. Association between allergic and nonallergic rhinitis and obstructive sleep apnea. *Curr. Opin. Allergy Clin. Immunol.* **2018**, *18*, 16–25. [CrossRef]
10. Cao, Y.; Wu, S.; Zhang, L.; Yang, Y.; Cao, S.; Li, Q. Association of allergic rhinitis with obstructive sleep apnea: A meta-analysis. *Medicine* **2018**, *97*, e13783. [CrossRef]

11. Lavigne, F.; Petrof, B.J.; Johnson, J.R.; Lavigne, P.; Binothman, N.; Kassissia, G.O.; Al Samri, M.; Giordano, C.; Dubé, N.; Hercz, D.; et al. Effect of topical corticosteroids on allergic airway inflammation and disease severity in obstructive sleep apnoea. *Clin. Exp. Allergy* **2013**, *43*, 1124–1133. [CrossRef] [PubMed]
12. Zhao, K.; Jiang, J. What is normal nasal airflow? A computational study of 22 healthy adults. *Int. Forum Allergy Rhinol.* **2014**, *4*, 435–446. [CrossRef] [PubMed]
13. Seppänen, T.M.; Alho, O.P.; Seppänen, T. Concomitant dynamic changes in autonomic nervous system function and nasal airflow resistance during allergen provocation. In Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 25–29 August 2015; pp. 3339–3342. [CrossRef]
14. Hanif, J.; Jawad, S.S.; Eccles, R. The nasal cycle in health and disease. *Clin. Otolaryngol. Allied Sci.* **2000**, *25*, 461–467. [CrossRef] [PubMed]
15. Letzel, J.; Darbinjan, A.; Hummel, T. The nasal cycle before and after nasal septoplasty. *Eur. Arch. Otorhinolaryngol.* **2022**, *279*, 4961–4968. [CrossRef] [PubMed]
16. Soler, E.; de Mendoza, A.; Cuello, V.I.; Silva-Vetri, M.G.; Núñez, Z.H.; Ortega, R.G.; Rizvi, S.A.; Sanchez-Gonzalez, M.; Ferrer, G. Intranasal xylitol for the treatment of COVID-19 in the outpatient setting: A pilot study. *Cureus* **2022**, *14*, e27182. [CrossRef] [PubMed]
17. Cannon, M.L.; Westover, J.B.; Bleher, R.; Sanchez-Gonzalez, M.A.; Ferrer, G. In vitro Analysis of the antiviral Potential of nasal spray constituents against SARS-CoV-2. *bioRxiv* **2020**. [CrossRef]
18. Winchester, S.; John, S.; Jabbar, K.; John, I. Clinical efficacy of nitric oxide nasal spray (NONS) for the treatment of mild COVID-19 infection. *J. Infect.* **2021**, *83*, 237–279. [CrossRef] [PubMed]
19. Tomonobu, N.; Komalasari, N.L.G.Y.; Sumardika, I.W.; Jiang, F.; Chen, Y.; Yamamoto, K.I.; Kinoshita, R.; Murata, H.; Inoue, Y.; Sakaguchi, M. Xylitol acts as an anticancer monosaccharide to induce selective cancer death via regulation of the glutathione level. *Chem. Biol. Interact.* **2020**, *324*, 109085. [CrossRef] [PubMed]
20. Shi, C.Y.; Yu, C.H.; Yu, W.Y.; Ying, H.Z. Gut-lung microbiota in chronic pulmonary diseases: Evolution, pathogenesis, and therapeutics. *Can. J. Infect. Dis. Med. Microbiol.* **2021**, *2021*, 9278441. [CrossRef]
21. Li, R.; Dee, D.; Li, C.M.; Hoffman, H.J.; Grummer-Strawn, L.M. Breastfeeding and risk of infections at 6 years. *Pediatrics* **2014**, *134* (Suppl. S1), S13–S20. [CrossRef]
22. Rashidi, K.; Darand, M.; Garousi, N.; Dehghani, A.; Alizadeh, S. Effect of infant formula supplemented with prebiotics and probiotics on incidence of respiratory tract infections: A systematic review and meta-analysis of randomized clinical trials. *Complement. Ther. Med.* **2021**, *63*, 102795. [CrossRef] [PubMed]
23. Sjödin, K.S.; Sjödin, A.; Ruszczyński, M.; Kristensen, M.B.; Hernell, O.; Szajewska, H.; West, C.E. Targeting the gut-lung axis by synbiotic feeding to infants in a randomized controlled trial. *BMC Biol.* **2023**, *21*, 38. [CrossRef] [PubMed]
24. Larsen, N.; Cahú, T.B.; Isay Saad, S.M.; Blennow, A.; Jespersen, L. The effect of pectins on survival of probiotic *Lactobacillus* spp. in gastrointestinal juices is related to their structure and physical properties. *Food Microbiol.* **2018**, *74*, 11–20. [CrossRef] [PubMed]
25. Konstantinidis, K.T.; Ramette, A.; Tiedje, J.M. The bacterial species definition in the genomic era. *Philos. Trans. R. Soc. B* **2006**, *361*, 1929–1940. [CrossRef] [PubMed]
26. Johnson, J.S.; Spakowicz, D.J.; Hong, B.Y.; Petersen, L.M.; Demkowicz, P.; Chen, L.; Leopold, S.R.; Hanson, B.M.; Agresta, H.O.; Gerstein, M.; et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat. Commun.* **2019**, *10*, 5029. [CrossRef] [PubMed]
27. Silva, C.F.F.S.D.; Silva, F.E.R.D.; Pauna, H.F.; Hurtado, J.G.G.M.; Dos Santos, M.C.J. Symptom assessment after nasal irrigation with xylitol in the postoperative period of endonasal endoscopic surgery. *Braz. J. Otorhinolaryngol.* **2022**, *88*, 243–250. [CrossRef] [PubMed]
28. Salli, K.; Lehtinen, M.J.; Tiihonen, K.; Ouwehand, A.C. Xylitol's health benefits beyond Dental Health: A comprehensive review. *Nutrients* **2019**, *11*, 1813. [CrossRef] [PubMed]
29. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **2011**, *17*, 10–12. [CrossRef]
30. Wood, D.E.; Lu, J.; Langmead, B. Improve Metagenomic Anal Kraken. *Genome Biol.* **2019**, *20*, 257. [CrossRef]
31. Lu, J.; Breitwieser, F.P.; Thielen, P.; Salzberg, S.L. Bracken: Estimating species abundance in metagenomics data. *PeerJ Comput. Sci.* **2017**, *3*, e104. [CrossRef]
32. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [CrossRef] [PubMed]
33. Azarpazhooh, A.; Lawrence, H.P.; Shah, P.S. Xylitol for preventing acute otitis media in children up to 12 years of age. *Cochrane Database Syst. Rev.* **2016**, *2016*, CD007095. [CrossRef]
34. Uhari, M.; Tapiainen, T.; Kontiokari, T. Xylitol in preventing acute otitis media. *Vaccine* **2000**, *19* (Suppl. S1), S144–S147. [CrossRef] [PubMed]
35. Hernández, P.; Sánchez, M.C.; Llama-Palacios, A.; Ciudad, M.J.; Collado, L. Strategies to combat caries by maintaining the integrity of biofilm and homeostasis during the rapid phase of supragingival plaque formation. *Antibiotics* **2022**, *11*, 880. [CrossRef] [PubMed]
36. Gasmi Benahmed, A.; Gasmi, A.; Doşa, A.; Chirumbolo, S.; Mujawdiya, P.K.; Aaseth, J.; Dadar, M.; Bjørklund, G. Association between the gut and oral microbiome with obesity. *Anaerobe* **2021**, *70*, 102248. [CrossRef] [PubMed]

37. Olsen, I.; Yamazaki, K. Can oral bacteria affect the microbiome of the gut? *J. Oral Microbiol.* **2019**, *11*, 1586422. [CrossRef] [PubMed]
38. Bassis, C.M.; Tang, A.L.; Young, V.B.; Pynnonen, M.A. The nasal cavity microbiota of healthy adults. *Microbiome* **2014**, *2*, 27. [CrossRef] [PubMed]
39. Rawls, M.; Ellis, A.K. The microbiome of the nose. *Ann. Allergy Asthma Immunol.* **2019**, *122*, 17–24. [CrossRef] [PubMed]
40. Gutiérrez, R.; Shalit, T.; Markus, B.; Yuan, C.; Nachum-Biala, Y.; Elad, D.; Harrus, S. *Bartonella kosoyi* sp. nov. and *Bartonella krasnovii* sp. nov., two novel species closely related to the zoonotic *Bartonella elizabethae*, isolated from black rats and wild desert rodent-fleas. *Int. J. Syst. Evol. Microbiol.* **2020**, *70*, 1656–1665. [CrossRef]
41. Marean, C.W. When the sea saved humanity. *Sci. Am.* **2010**, *303*, 54–61. [CrossRef]
42. Marlowe, F.W.; Berbesque, J.C. Tubers as fallback foods and their impact on Hadza hunter-gatherers. *Am. J. Phys. Anthropol.* **2009**, *140*, 751–758. [CrossRef]
43. Jones, A.M.; Dodd, M.E.; Webb, A.K. *Burkholderia cepacia*: Current clinical issues, environmental controversies and ethical dilemmas. *Eur. Respir. J.* **2001**, *17*, 295–301. [CrossRef]
44. Kalish, L.A.; Waltz, D.A.; Dovey, M.; Potter-Bynoe, G.; McAdam, A.J.; LiPuma, J.J.; Gerard, C.; Goldmann, D. Impact of *Burkholderia dolosa* on lung function and survival in cystic fibrosis. *Am. J. Respir. Crit. Care Med.* **2006**, *173*, 421–425. [CrossRef]
45. Isler, B.; Kidd, T.J.; Stewart, A.G.; Harris, P.; Paterson, D.L. Achromobacter infections and treatment options. *Antimicrob. Agents Chemother.* **2020**, *64*, e01025-20. [CrossRef] [PubMed]
46. Lyons, K.E.; Ryan, C.A.; Dempsey, E.M.; Ross, R.P.; Stanton, C. Breast milk, a source of beneficial microbes and associated benefits for infant health. *Nutrients* **2020**, *12*, 1039. [CrossRef]
47. Sakallioğlu, Ö.; Güvenç, I.A.; Cingi, C. Xylitol and its usage in ENT practice. *J. Laryngol. Otol.* **2014**, *128*, 580–585. [CrossRef] [PubMed]
48. Singh, S.; Hornick, D.; Fedler, J.; Launspach, J.L.; Teresi, M.E.; Santacroce, T.R.; Cavanaugh, J.E.; Horan, R.; Nelson, G.; Starner, T.D.; et al. Randomized controlled study of aerosolized hypertonic xylitol versus hypertonic saline in hospitalized patients with pulmonary exacerbation of cystic fibrosis. *J. Cyst. Fibros.* **2020**, *19*, 108–113. [CrossRef]
49. Cai, Y.; Goldberg, A.N.; Chang, J.L. The nose and nasal breathing in sleep apnea. *Otolaryngol. Clin. N. Am.* **2020**, *53*, 385–395. [CrossRef]
50. Nosetti, L.; Piacentini, G.; Macchi, A.; De Bernardi, F.; Simoncini, D.; Nicoloso, M.; Agosti, M.; Zaffanello, M. Nasal cytology in children with primary snoring and obstructive sleep apnoea syndrome. *Int. J. Pediatr. Otorhinolaryngol.* **2019**, *122*, 133–137. [CrossRef] [PubMed]
51. Rosier, B.T.; Takahashi, N.; Zaura, E.; Krom, B.P.; Martínez-Espinosa, R.M.; van Breda, S.G.J.; Marsh, P.D.; Mira, A. The importance of nitrate reduction for Oral Health. *J. Dent. Res.* **2022**, *101*, 887–897. [CrossRef]
52. de Farias, J.O.; de Freitas Lima, S.M.; Rezende, T.M.B. Physiopathology of nitric oxide in the oral environment and its biotechnological potential for new oral treatments: A literature review. *Clin. Oral Investig.* **2020**, *24*, 4197–4212. [CrossRef] [PubMed]
53. Altemani, F.; Barrett, H.L.; Callaway, L.K.; McIntyre, H.D.; Dekker Nitert, M. Reduced abundance of nitrate-reducing bacteria in the oral microbiota of women with future preeclampsia. *Nutrients* **2022**, *14*, 1139. [CrossRef] [PubMed]
54. Owusu Darkwa, E.; Djagbletey, R.; Sottie, D.; Owoo, C.; Vanderpuye, N.M.; Essuman, R.; Aryee, G. Serum nitric oxide levels in healthy pregnant women: A case- control study in a tertiary facility in Ghana. *Matern. Health Neonatol. Perinatol.* **2018**, *4*, 3. [CrossRef] [PubMed]
55. Vanterpool, S.F.; Been, J.V.; Houben, M.L.; Nikkels, P.G.; De Krijger, R.R.; Zimmermann, L.J.; Kramer, B.W.; Progulske-Fox, A.; Reyes, L. *Porphyromonas gingivalis* within Placental Villous Mesenchyme and Umbilical Cord stroma Is Associated with Adverse Pregnancy Outcome. *PLoS ONE* **2016**, *11*, e0146157. [CrossRef] [PubMed]
56. León, R.; Silva, N.; Ovalle, A.; Chaparro, A.; Ahumada, A.; Gajardo, M.; Martinez, M.; Gamonal, J. Detection of *Porphyromonas gingivalis* in the amniotic fluid in pregnant women with a diagnosis of threatened premature labor. *J. Periodontol.* **2007**, *78*, 1249–1255. [CrossRef] [PubMed]
57. Arnold, H.K.; Hanselmann, R.; Duke, S.M.; Sharpton, T.J.; Beechler, B.R. Chronic clinical signs of upper respiratory tract disease associate with gut and respiratory microbiomes in a cohort of domestic felines. *PLoS ONE* **2022**, *17*, e0268730. [CrossRef] [PubMed]
58. Miraglia Del Giudice, M.; Parisi, G.F.; Indolfi, C.; Manti, S.; Leonardi, S.; Decimo, F.; Ciprandi, G. Nasal microbiome in chronic rhinosinusitis. *Minerva Pediatr.* **2022**, *74*, 586–592. [CrossRef] [PubMed]
59. Cho, D.Y.; Hunter, R.C.; Ramakrishnan, V.R. The microbiome and chronic rhinosinusitis. *Immunol. Allergy Clin. N. Am.* **2020**, *40*, 251–263. [CrossRef]
60. Tramper-Stranders, G.; Ambrożej, D.; Arcolaci, A.; Atanaskovic-Markovic, M.; Boccabella, C.; Bonini, M.; Karavelia, A.; Mingomataj, E.; O' Mahony, L.; Sokolowska, M.; et al. Dangerous liaisons: Bacteria, antimicrobial therapies, and allergic diseases. *Allergy* **2021**, *76*, 3276–3291. [CrossRef] [PubMed]
61. Luo, C.; Peng, S.; Li, M.; Ao, X.; Liu, Z. The efficacy and safety of probiotics for allergic rhinitis: A systematic review and meta-analysis. *Front. Immunol.* **2022**, *13*, 848279. [CrossRef]
62. Ren, Z.; Jeckel, H.; Simon-Soro, A.; Xiang, Z.; Liu, Y.; Cavalcanti, I.M.; Xiao, J.; Tin, N.N.; Hara, A.; Drescher, K.; et al. Interkingdom assemblages in human saliva display group-level surface mobility and disease-promoting emergent functions. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2209699119. [CrossRef]

63. Alamoudi, N.M.; Hanno, A.G.; Sabbagh, H.J.; Masoud, M.I.; Almushayt, A.S.; El Derwi, D.A. Impact of maternal xylitol consumption on mutans streptococci, plaque and caries levels in children. *J. Clin. Pediatr. Dent.* **2012**, *37*, 163–166. [CrossRef] [PubMed]
64. Talattof, Z.; Azad, A.; Zahed, M.; Shahradnia, N. Antifungal activity of xylitol against Candida albicans: An in vitro study. *J. Contemp. Dent. Pract.* **2018**, *19*, 125–129. [CrossRef] [PubMed]
65. Sarah, L.; Oerlemans, E.F.M.; Claes, I.; Wuyts, S.; Henkens, T.; Spacova, I.; van den Broek, M.F.L.; Tuyaerts, I.; Wittouck, S.; De Boeck, I.; et al. Topical cream with live lactobacilli modulates the skin microbiome and reduce acne symptoms. *bioRxiv* **2018**. [CrossRef]
66. De Boeck, I.; van den Broek, M.F.L.; Allonsius, C.N.; Spacova, I.; Wittouck, S.; Martens, K.; Wuyts, S.; Cauwenberghs, E.; Jokicevic, K.; Vandenheuvel, D.; et al. Lactobacilli have a niche in the human nose. *Cell Rep.* **2020**, *31*, 107674. [CrossRef] [PubMed]
67. Chen, M.; He, S.; Miles, P.; Li, C.; Ge, Y.; Yu, X.; Wang, L.; Huang, W.; Kong, X.; Ma, S.; et al. Nasal Bacterial Microbiome Differs between Healthy Controls and Those with Asthma and Allergic Rhinitis. *Front. Cell. Infect. Microbiol.* **2022**, *12*, 841995. [CrossRef]
68. Yang, G.; Liu, Z.Q.; Yang, P.C. Treatment of allergic rhinitis with probiotics: An alternative approach. *N. Am. J. Med. Sci.* **2013**, *5*, 465–468. [CrossRef] [PubMed]
69. Morin, A.; McKennan, C.G.; Pedersen, C.T.; Stokholm, J.; Chawes, B.L.; Malby Schoos, A.M.; Naughton, K.A.; Thorsen, J.; Mortensen, M.S.; Vercelli, D.; et al. Epigenetic landscape links upper airway microbiota in infancy with allergic rhinitis at 6 years of age. *J. Allergy Clin. Immunol.* **2020**, *146*, 1358–1366. [CrossRef] [PubMed]
70. Nakahashi-Ouchida, R.; Fujihashi, K.; Kurashima, Y.; Yuki, Y.; Kiyono, H. Nasal vaccines: Solutions for respiratory infectious diseases. *Trends Mol. Med.* **2023**, *29*, 124–140. [CrossRef]
71. Chavda, V.P.; Vora, L.K.; Pandya, A.K.; Patravale, V.B. Intranasal vaccines for SARS-CoV-2: From challenges to potential in COVID-19 management. *Drug Discov. Today* **2021**, *26*, 2619–2636. [CrossRef]
72. Yahalom-Ronen, Y.; Melamed, S.; Politi, B.; Erez, N.; Tamir, H.; Bar-On, L.; Ryvkin, J.; Leshkowitz, D.; Israeli, O.; Weiss, S.; et al. Induction of Superior Systemic and Mucosal Protective Immunity to SARS-CoV-2 by Nasal Administration of a VSV-ΔG-Spike Vaccine. *Vaccines* **2024**, *12*, 491. [CrossRef] [PubMed]
73. van der Ley, P.A.; Zariri, A.; van Riet, E.; Oosterhoff, D.; Kruiswijk, C.P. An Intranasal OMV-Based Vaccine Induces High Mucosal and Systemic Protecting Immunity against a SARS-CoV-2 Infection. *Front. Immunol.* **2021**, *12*, 781280. [CrossRef] [PubMed]
74. Nian, X.; Zhang, J.; Huang, S.; Duan, K.; Li, X.; Yang, X. Development of Nasal Vaccines and the Associated Challenges. *Pharmaceutics* **2022**, *14*, 1983. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

microorganisms

MDPI

Article

Salmonella enterica serovar Schwarzengrund: Distribution, Virulence, and Antimicrobial Resistance

Monique A. Felix [1,2], Jing Han [1], Bijay K. Khajanchi [1,3], Yasser M. Sanad [1,2], Shaohua Zhao [4] and Steven L. Foley [1,*]

[1] Division of Microbiology, National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, AR 72079, USA; moniquefelix148@gmail.com (M.A.F.); jing.han1@fda.hhs.gov (J.H.); bijay.khajanchi@fda.hhs.gov (B.K.K.); sanady@uapb.edu (Y.M.S.)
[2] Department of Agriculture, University of Arkansas at Pine Bluff, Pine Bluff, AR 71601, USA
[3] Office of Applied Research and Safety Assessment, Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration, Laurel, MD 20708, USA
[4] Office of Applied Research, Center for Veterinary Medicine, U.S. Food and Drug Administration, Laurel, MD 20708, USA; shaohua.zhao@fda.hhs.gov
* Correspondence: steven.foley@fda.hhs.gov; Tel.: +1-870-543-7547

Abstract: The global incidence of *Salmonella enterica* serovar Schwarzengrund has risen in recent years. This serotype has been isolated from poultry, retail meat, and other food products, leading to multiple outbreaks. Alongside the increase in infections, there are growing concerns about the increasing levels of antimicrobial resistance (AMR) among *S.* Schwarzengrund strains. This study aims to better understand the genetic factors possibly contributing to the rising prevalence of *S.* Schwarzengrund by analyzing the sequences of 2058 isolates from both human patients (N = 313) and food- and animal-associated sources, including chicken (N = 1145), turkey (N = 300), pork (N = 132), and other sources (N = 168). Data were obtained from GenBank and analyzed for AMR genes using AMRFinder. Additionally, putative virulence genes and plasmid transfer genes were assessed using the Virulence and AMR Plasmid Transfer Factor Database. AMR genes were found in 1269 (61.7%) of the isolates, with a total of 2478 AMR genes among the isolates, the most common being *aph(3″)-Ib* (N = 969, 47.1%), *tet(A)* (N = 190, 9.2%), and *sul2* (N = 150, 7.3%), which are responsible for resistance to aminoglycoside, tetracycline, and sulfonamide, respectively. Additionally, 1060 (51.5%) isolates carried multiple plasmid transfer genes associated with IncFIB-FIC(FII) plasmids. Other plasmid types found in at least 1% of the strains included IncI1 (N = 101, 4.9%), IncHI2 (N = 62, 3.0%), or IncHI1 (N = 24, 1.2%). The virulence gene profiles of human isolates showed diversity but largely overlapped with those from different food sources. Notably, the aerobactin iron acquisition genes, associated with *Salmonella*'s virulence and colonization, were highly prevalent among chicken isolates (N = 1019, 89.0%) but less frequent in isolates from other sources (N = 65, 7.2%). IncFIB-FIC(FII) plasmids, commonly harboring the aerobactin operon, were highly prevalent among chicken-related isolates and present in about 10% of human isolates. The diverse plasmid, AMR, and virulence gene profiles in human-associated isolates suggest that multiple factors may contribute to the increased virulence in *S.* Schwarzengrund.

Keywords: IncFIB-IncFIC(FII) plasmids; genotypic analyses; virulence database; plasmid transfer genes; phylogenetics

1. Introduction

In the United States (U.S.), there are over one million cases of *Salmonella* infection annually, leading to approximately 20,000 hospitalizations and nearly 400 deaths, according to estimates by the Centers for Disease Control and Prevention (CDC) [1]. *Salmonella* infects and invades

the intestinal tract and is disseminated through the excretion of feces. The disease spectrum of salmonellosis in humans ranges from relatively mild food poisoning and gastroenteritis to more severe illnesses that can include enteric fever and septicemia [2]. Salmonellosis is often associated with fecal contamination of water or food sources such as meat, poultry, and eggs, as well as through the contact with cross-contaminated surfaces and utensils in poor hygienic conditions. Common signs and symptoms of salmonellosis include stomach cramps, bloody stools, chills, diarrhea, fever, headaches, muscle pain, nausea, vomiting, and dizziness [3,4]. *Salmonella* infections are categorized as either typhoidal (human-specific) or non-typhoidal (which affect a broad range of hosts) [2,5]. In the U.S., most infections are caused by non-typhoidal *Salmonella,* which infects both humans and animals. One serovar that causes non-typhoidal infections is *S*. Schwarzengrund [6], the primary focus of this research.

S. Schwarzengrund has been an emerging serovar over the last few decades, rising to be one of the more common serovars in countries such as Thailand, Slovakia, New Zealand, and Venezuela [6]. The first report of *S.* Schwarzengrund in the U.S. was in the early 1950s. In this case, nine-day old turkeys from a farm in Indiana were sent to a diagnostic lab. At the farm, 754 of 13,000 turkeys (5.8%) died in the first 4 days following infection. *S.* Schwarzengrund strains were subsequently isolated from the liver of the turkey. This isolation was the first in the U.S. and the second recorded worldwide [7]. Since that time, the number of *S*. Schwarzengrund infections in food animals and humans has increased. Data from the CDC showed that the isolation rates of *S*. Schwarzengrund had increased over the years, especially in the upper eastern region of the U.S. [8]. From 1968 to 2011, 77.7% of reported human isolates, as documented by the CDC, originated from stool samples; however, 9.8% were from more invasive sources such as blood, cerebral spinal fluids, joints, or wounds/abscesses, and 3.6% were from urine [8]. These high rates of extraintestinal infections are particularly alarming given that they often require antimicrobial therapy, making AMR in this serovar especially concerning.

Globally, the prevalence of *S*. Schwarzengrund has also increased. In Asian countries, including Japan and Thailand, its prevalence in chickens increased from the mid-1990s to the mid-2000s, and it has maintained its status as a top serovar from chickens in Japan [9–12]. *S*. Schwarzengrund has also been among the top four most prevalent serovars isolated from human patients in Japan in recent years [13]. Pulsed-field gel electrophoresis (PFGE) analyses of *S*. Schwarzengrund isolates revealed that antibiotic-resistant strains were spreading among broiler chickens and retail chicken meat with cross contamination occurring during processing, likely through contaminated transport crates and utensils [12]. In the U.S., *S*. Schwarzengrund ranks among the top five *Salmonella* serovars isolated from pig farms and is one of the top fifteen serovars found in food, animals, and humans in Thailand, Slovakia, New Zealand, and Venezuela [6].

A study conducted by Chen et al. (2010) found that *S*. Schwarzengrund were more prevalent during the summer compared to winter and on certain parts of the chicken, such as the buttock and neck [14]. A study by Duc et al. (2020) showed that 17.8% (546/3069) of broiler chicken cecal samples collected between 2003 and 2016 were positive for *Salmonella* spp., and 21.3% (116/546) were identified as *S*. Schwarzengrund, a significant increase compared to 2.1% (5/243) in a previous study conducted from 2009 to 2012 [15]. In the U.S., *S*. Schwarzengrund outbreaks among humans have been linked to contact with dry pet food and consumption of contaminated ground turkey [16,17].

AMR in *S*. Schwarzengrund is of particular concern, as it complicates the treatment of invasive infections. Internationally, there has been a rise in multidrug-resistant (MDR) *S.* Schwarzengrund strains which can resist antibiotics such as tetracyclines, sulfonamides, aminoglycosides, β-lactams, and fluoroquinolones [18–20]. In Brazil, 11% of isolates from poultry carried the extended-spectrum β-lactamase (ESBL) resistance gene $bla_{\text{CTX-M-2}}$,

which encodes resistance to critical antimicrobials [21]. In Taiwan, 30% of S. Schwarzengrund strains from retail chicken meat were resistant to antibiotics, with 15.8% resistant to quinolones and 28.1% resistant to ciprofloxacin [14]. Notably, *S.* Schwarzengrund was the first *Salmonella* serotype in the U.S. found to be resistant to fluoroquinolones, following a case involving a patient who had traveled to the Philippines [21].

The pathogenicity of *S.* Schwarzengrund is driven by a wide range of virulence factors, including type 3 secretion systems (T3SSs), encoded by *Salmonella* Pathogenicity Islands (SPIs) 1 and 2, which facilitate invasion and survival within host cells [2]. Additional virulence genes are found on both chromosomally and plasmid-encoded gene clusters, further enhancing the bacterium's ability to infect hosts. Plasmids also play a critical role in *S.* Schwarzengrund's AMR and virulence. Fusion plasmids, such as IncFIB-IncFIC(FII), carry genes for iron acquisition (*iucABCD* and *iutA* of the aerobactin operon), virulence factors, and AMR. These plasmids can enhance the bacterium's survival under nutrient-limited conditions and facilitate horizontal gene transfer, spreading AMR traits across bacterial populations [22]. In addition to their fusion plasmids, distinct IncFII and IncFIC plasmids are also found in *S.* Schwarzengrund; these can facilitate horizontal transfer of various AMR genes such as the extended-spectrum β-lactamase genes [22–25]. In addition to these IncF-related plasmids, other AMR-associated plasmid types such as IncI1, IncHI1, and IncHI2 plasmids have also been detected among *S.* Schwarzengrund isolates.

Given their rising prevalence, and the AMR and virulence profiles of many *S.* Schwarzengrund strains, especially in food animal hosts, there is a need for more research to better understand their genetic diversity. The goal of this project was to provide deeper insights into the virulence factors, AMR genes, and plasmids present in *S.* Schwarzengrund by conducting an extensive analysis of over 2000 publicly available whole genome sequences from the GenBank database.

2. Materials and Methods

2.1. Bacterial Strains Analyzed

To investigate the genetic characteristics of *S.* Schwarzengrund, sequenced isolates were identified using the National Center for Biotechnology Information (NCBI) Pathogen Detection Isolate Browser. Relevant metadata, including isolate source/type, geographical location, year of isolation, and SNP cluster information, were extracted from both GenBank and the NCBI Pathogen Detection Isolate Browser (https://www.ncbi.nlm.nih.gov/pathogens/ (accessed on 21 May 2024)). The whole genome sequence (WGS) FASTA files of 2058 *S.* Schwarzengrund isolates were extracted from GenBank in November of 2022. These isolates included samples from human patients (N = 313) and food- and animal-associated sources, including chicken (N = 1145), turkey (N = 300), and pork (N = 132), representing all the available *S.* Schwarzengrund WGSs at the time of data collection.

2.2. Virulence, Plasmid, and AMR Gene Detection

Putative virulence genes and plasmid transfer genes were identified using the Virulence and AMR Plasmid Transfer Factor Database (https://virulence.fda.gov) [26,27]. The WGS FASTA files for each strain were uploaded into the database and analyzed using the *Salmonella* Virulence Factor Comparison and Plasmid Transfer Gene Comparison tools, respectively. For AMR gene detection, data were either extracted directly from the NCBI Pathogen Detection Isolate Browser or, where unavailable, the WGS FASTA files were analyzed with AMRFinder (ver. 26.1) within the GalaxyTrakr platform [28,29]. This allowed for the identification of AMR genes across all of the *S.* Schwarzengrund isolates.

2.3. Data Analysis

The sample metadata, along with presence/absence data for virulence, and AMR and plasmid transfer genes were extracted from the respective analysis tools and formatted as delimited files. These were then imported into Excel (Microsoft, Redlands, WA, USA), for initial descriptive statistics of prevalence of AMR, putative virulence, and plasmid transfer-associated genes. For more detailed analyses, the data were uploaded into BioNumerics (ver. 8.1; Applied Maths, Austin, TX, USA). BioNumerics was used for phylogenetic analyses based on the presence/absence of genes across different metadata categories. These analyses included the identification of polymorphic characters (i.e., those genes detected in 1–99% of isolates), performing principal components analysis (PCA), and constructing minimal spanning trees using default parameters for the character data sets.

3. Results and Discussion

3.1. S. Schwarzengrund Demographics

S. Schwarzengrund infections have been increasing globally over the last several years; however, most strains that have undergone whole genome sequencing are from the United States [6,15]. In this dataset, the vast majority (85.6%, N = 1762) of the strains originated in the U.S., followed by the United Kingdom (8.5%, N = 174) (Table 1). The majority of the U.S. isolates were collected through surveillance efforts by the Food and Drug Administration (FDA) and the U.S. Department of Agriculture's Food Safety and Inspection Service (FSIS). When looking at the isolation locations within the U.S., the southeastern states, particularly Georgia (GA; 21.7% of U.S. strains, N = 383), Alabama (AL; 9.9%, N = 175), Tennessee (TN; 8.2%, N = 144), and North Carolina (NC; 4.9%, N = 87), reported the highest number of isolates (Figure 1). In terms of isolate sources, chicken-related samples dominated (55.6%, N = 1145), followed by human patients (15.2%, N = 313), turkey (14.6%, N = 300), and porcine-related samples (6.4%, N = 132). Among the U.S. isolates, food animal sources were even more prominent, with chicken accounting for 62.9% (N = 1102), turkey, 17.0% (N = 299), and porcine, 7.0% (N = 124). Conversely, the percentage of human isolates from the U.S. was lower (6.0%, N = 106) than the overall percentage. Interestingly, the largest proportion of isolates from human patients was from the U.K., at 52.7% (N = 165). In terms of state-level comparisons within the U.S., isolates from GA, AL, and TN were predominantly chicken-related, while NC isolates were more often turkey- and pork-related. Unfortunately, the absence of detailed state-level data for many human isolates, likely due to privacy concerns, limits further comparisons. This lack of metadata complicates attempts to correlate human isolates with food source origins, emphasizing the need for better documentation of human cases to improve understanding of transmission pathways.

Table 1. Isolate demographics, including the source, country of origin, and year of isolation.

Isolate Source	Number	Percent	Country Isolated	Number	Percent	Year Isolated	Number	Percent
Animal	8	0.4	**Brazil**	23	1.1	**2022**	282	13.7
Animal feed	17	0.8	**Canada**	27	1.3	**2021**	217	10.5
Avian	13	0.6	**France**	4	0.2	**2020**	194	9.4
Bovine	62	3.0	**Germany**	5	0.2	**2019**	162	7.9
Canine	6	0.3	**India**	6	0.3	**2018**	325	15.8
Chicken	1145	55.6	**Taiwan**	32	1.6	**2017**	297	14.4
Environmental	21	1.0	**Thailand**	6	0.3	**2016**	151	7.3
Fish	10	0.5	**United Kingdom**	174	8.5	**2015**	86	4.2
Food	23	1.1	**United States**	1762	85.6	**2014**	78	3.8
Human	313	15.2	**Vietnam**	4	0.2	**2013**	32	1.6
Porcine	132	6.4	**Other**	9	0.4	**2012**	36	1.7
Turkey	300	14.6	**Not Reported**	6	0.3	**2011**	21	1.0
Not Reported	8	0.4	**Total:**	**2058**	**100**	**2001–2010**	157	7.6
Total:	**2058**	**100**				**<2000**	6	0.3
						Not Reported	14	0.7
						Total	**2058**	**100**

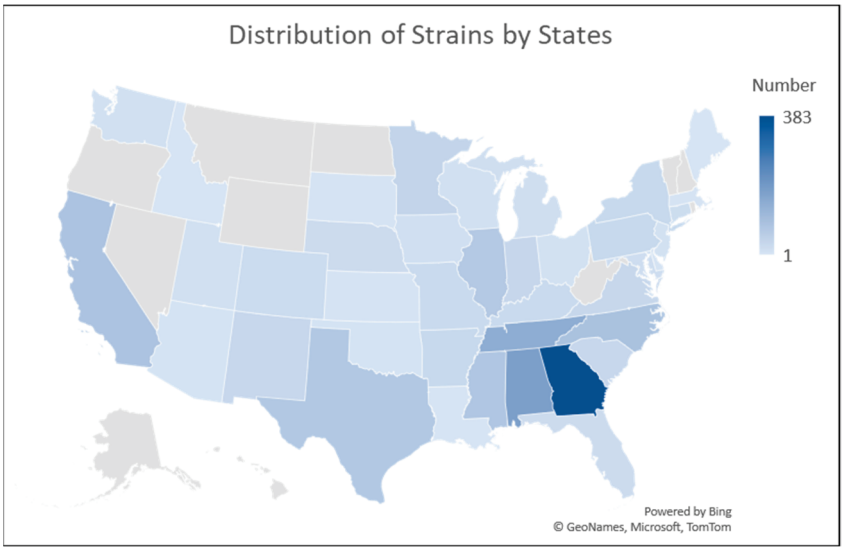


Figure 1. Distribution of *S.* Schwarzengrund strains isolated within the United States. The heatmap to the right of the figure indicates the numbers of isolates from each state that were present in the GenBank dataset.

3.2. Antimicrobial Resistance

AMR genes were relatively common among the isolates, with 1269 (61.7%) carrying at least one AMR gene. Overall, there were 2478 AMR genes among the isolates, the most commonly observed being *aph(3″)-Ib* (aminoglycoside; N = 969, 47.1%), followed by *tet(A)* (tetracycline; N = 190, 9.2%) and *sul2* (sulfonamide; N = 150, 7.3%) (Figure 2). Other notable AMR genes included $bla_{\text{TEM-1}}$ (β-lactamase), *sul1* (sulfonamide), and aminoglycoside resistance genes *aadA1* and *aph(6)-Id*, each present in at least 3% of the strains. Detailed results of the AMR profiles are presented in Table S2. There were two major AMR gene profiles: those without any detectable AMR genes (upper ball in Figures 3A and S1A) and those carrying only *aph(3″)-Ib*. Strains carrying *aph(3″)-Ib* were predominantly from chickens and frequently associated with IncFIB-FIC(FII) plasmids, which is consistent with our earlier findings [29]. This plasmid has been shown to be transmissible in vitro, raising concerns about the potential spread of resistance within poultry populations and, subsequently, to humans.

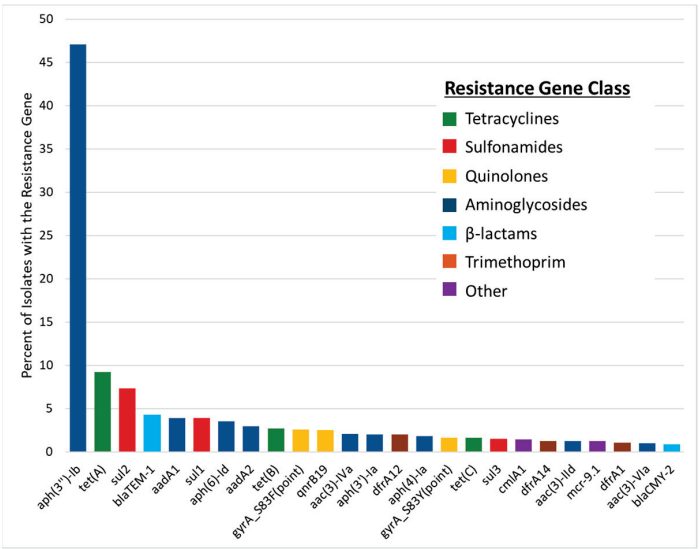


Figure 2. Percent of *S.* Schwarzengrund isolated in this study that carried the respective AMR genes. The bars are color-coded based on antimicrobial classes where resistance genes were detected.

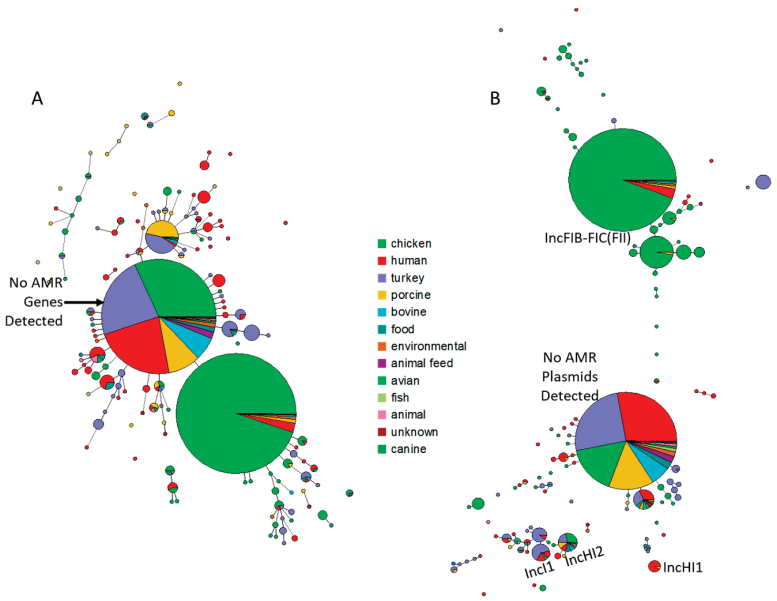


Figure 3. Minimum spanning tree analyses based on the AMR (panel (**A**)) and plasmid transfer gene (**B**) profiles of the isolates included in this study. The trees are color-coded based on isolation source. The relative size of the circles is proportional to the group size. In panel (**A**), the larger circle in the middle represents the strains without an identified AMR gene. In panel (**B**), the larger ball near the bottom is the group without a detected AMR plasmid. The groups representing the major plasmid types are annotated. The colors are described by the legend included in the figure.

Although the majority of isolates fell into two main AMR profiles, smaller profiles exhibited more diverse AMR gene combinations, particularly among non-chicken sources. The high prevalence of *aph(3″)-Ib* among chicken-related strains, and its association with transmissible plasmids, highlights the potential risk of widespread of antimicrobial resistance within the poultry industry and its broader impact on human health, especially in the context of treatment efficacy for related infections [30].

When compared to AMR profiles in other *Salmonella* serovars, the predominance of *aph(3″)-Ib* in *S.* Schwarzengrund isolates highlights a distinctive pattern of resistance, suggesting that plasmid-mediated transmission may play a more significant role in this serovar's resistance profile. Although *aph(3″)-Ib* dominates the AMR profiles, the presence of β-lactamase ($bla_{\text{TEM-1}}$) and sulfonamide (*sul1*) resistance genes in over 3% of isolates further underscores the need for continued surveillance, particularly for emerging multidrug-resistant strains in non-poultry sources.

3.3. Plasmid Characterization

Among the identified plasmid types, approximately 51.5% (N = 1060) carried multiple plasmid transfer genes associated with IncFIB-FIC(FII) plasmids, while other plasmid types, such as IncI1 (N = 101, 4.9%), IncHI2 (N = 62, 3.0%), or IncHI1 (N = 24, 1.2%), were detected in at least 1% of the strains (Figures 3B and S1B). The high prevalence of IncFIB-FIC(FII) plasmids is of particular concern given their carriage of a range of virulence and antimicrobial resistance genes, potentially contributing to increased pathogenicity and resistance spread. The detailed results of the plasmid transfer gene analysis are presented in Table S3 in the supplemental data. The approach of using the transfer genes associated with the different plasmid types has value in both predicting the plasmid types present and identifying potential limitations on conjugal ability if genes are absent. The current study did not evaluate conjugation ability; however, our recent study has demonstrated that some of the IncFIB-FIC(FII) fusion plasmids from *S.* Schwarzengrund can transfer under in vitro conditions, although not in all cases [30], highlighting the risk of horizontal gene

transfer in the agriculture environment [31]. The presence of transmissible plasmids in poultry isolates may have important implications for both food safety and public health [32]. Given the predominance of these plasmids in isolates from chicken-related sources, further investigation into their mobility and contribution to the spread of AMR genes across different environments is warranted.

3.4. Virulence Genetics Characterization

The virulence gene profiles were determined using the *Salmonella* Virulence Factor Comparison tool. Of the 590 genes examined, 43.9% (N = 259) were present in all isolates (Table S4). These genes included those associated with SPIs 1–5 and 9, as well as certain fimbrial genes (e.g., *csg* and *std* operons). Additionally, 23.1% (N = 136) were present in more than 99% of the strains. In contrast, 15.6% (N = 92) of the genes were detected in less than 1% of the strains, including 44 genes that were absent in all strains. These low prevalence and absent genes include those associated with different SPIs (e.g., SPI-7 associated with *S.* Typhi) and many of the *Salmonella* genomic island (SGI)-1 genes. The remaining 17.5% (N = 103) of genes, which represented those present in between 1 and 99% of the isolates, were considered polymorphic characters which provided useful markers for comparing the virulence gene profiles across different isolates from different sources. As shown in Figure S2, phylogenetic analyses of virulence profiles revealed that most chicken-related isolates clustered into four large groups, each containing over 70 strains. In addition, there were two large groups that contained over 300 strains each from multiple different sources, including many isolates from human patients. These results highlight that the human isolates exhibited greater diversity of virulence gene profiles yet frequently overlapped with those from different food/animal sources. One of the key differences between the four predominantly chicken clusters and those with the more diverse sources is that isolates from chickens carried a higher prevalence of IncFIB-FIC(FII) plasmid-associated virulence factors, such as the aerobactin iron acquisition genes (*iucABCD* and *iutA*) and *traT*. These genes were significantly more common in chicken-related isolates (88.5%, N = 1019) than in those from other sources (7.1%, N = 65) (Table 2). Certain genes, such as *allD, envF, sseK1*, and *sspH2*, as well as several fimbrial operons (including *stc, ste, stf,* and *stk*), were more commonly found in strains from human patients compared to those from chicken sources. These fimbrial genes were particularly interesting because strains exhibited different combinations of operons (*sta, stb, stc, ste, stf, stk,* and *stj*), while *std, sth,* and *sti* were present in nearly all (>99%) strains, and the *stg* operon was absent in all strains. When looking at the VF profiles based on geographic location, the vast majority of the strains carrying the IncFIB-FIC(FII) plasmids and associated VFs originated from the U.S., while those lacking the plasmids were from a wider range of locations.

As nearly 86% of the isolates originated from the U.S., we conducted a more detailed comparison of the VF profiles of these isolates. As shown in Figure 4, the strains were separated based on the presence of the IncFIB-FIC(FII) plasmid-associated genes and then, within these groups, based on whether they carried *int, xis,* and *sopE* genes. In the 1217 strains that carry at least one of these genes, 1204 (98.9%) carry all three of the genes. The *int* and *xis* genes encode integrase and excisionase functions, respectively, which are associated with the integration and excision of bacteriophages from the bacterial chromosome [29]. The *int* and *xis* genes are also part of *Salmonella* Genomic Island 1 (SGI-1) which typically carries AMR and genes predicted to contribute to increased virulence [33,34]. Among the analyzed *S.* Schwarzengrund strains, none carried the SGE-1-associated genes (S004, S013–S026), while 61.3% and 61.1% carried *int* and *xis*, respectively. The highest percent of isolates carrying these two genes were those from chickens, with over 77% carriage (Table 2). Approximately a third of the isolates from human patients carried these genes.

The presence of a high number of *int*/*xis* elements in the isolates from chickens and the lack of other SGE-1 genes may be evidence for the integration of chicken-associated phages into these strains [35]. The presence of phage genes was evident when examining the WGS assemblies of some of the strains, in that the genes flanking *int*/*xis* included those associated with phages (e.g., FSIS1609433, GenBank accession CP119491.1; FSIS1608447, accession CP119495.1). *sopE* also has a phage origin and encodes the guanine nucleotide exchange factor SopE, an effector for the SPI-1 type 3 secretion system (T3SS) [36,37]. SopE is translocated into the host cell via the T3SS where it interacts with the cytoskeleton to facilitate entry of the bacterium into the host [37]. In the two example strains note above, all three genes are located next to each other in their chromosomes. These features highlight the need for further examination of the role of integrated phages within the *S.* Schwarzengrund genome, as the strains isolated from blood and urine had an apparent higher prevalence of these phage-associated genes compared to those isolated from stool. Further research is needed to understand the role of integrated phages and the corresponding mobility of virulence genes in clinical and agricultural contexts.

Table 2. Percentage of *S.* Schwarzengrund strains with the respective virulence gene by source type.

Source	Animal	Animal feed	Avian	Bovine	Canine	Chicken	Environmental	Fish	Food	Human	Porcine	Turkey	Not Reported
allD	12.5	5.9	0	1.6	0	0.8	14.3	20.0	13.0	8.3	0	0.3	0
cdtB	87.5	94.1	100	98.4	100	99.0	85.7	80.0	87.0	92.0	100	99.7	100
envF	12.5	5.9	0	1.6	0	0.8	14.3	20.0	13.0	8.6	0	0.3	0
hlyE	87.5	88.2	100	98.4	100	99.2	85.7	80.0	87.0	92.0	100	99.7	100
icmF	100	100	100	100	100	99.6	95.2	90.0	73.9	96.2	100	99.7	100
int	37.5	41.2	30.8	17.7	50.0	77.5	66.7	0	17.4	33.9	48.5	52.0	25.0
iucA	0	0	23.1	8.1	16.7	88.5	28.6	0	0	8.6	6.8	0.7	25.0
iutA	0	0	23.1	8.1	16.7	88.5	28.6	0	0	8.6	6.8	0.7	25.0
lpfA	12.5	0	0	0	0	0.8	14.3	20.0	13.0	1.9	0	0.3	0
pagO	100	100	100	100	100	99.8	90.5	90.0	87.0	93.6	100	100	100
pltA	87.5	94.1	100	98.4	100	99.0	85.7	80.0	87.0	92.0	100	99.7	100
safA	100	94.1	100	98.4	100	98.9	90.5	90.0	87.0	99.0	100	98.3	100
sciL	100	100	100	100	100	99.5	95.2	90.0	73.9	95.8	100	99.7	100
sopE	25.0	35.3	53.8	17.7	50.0	77.1	66.7	20.0	21.7	31.6	53.0	53.0	62.5
sseK1	0	0	0	0	0	0.3	9.5	10.0	13.0	8.0	0	0	0
sspH2	12.5	5.9	0	1.6	0	0.8	14.3	10.0	13.0	8.3	0.8	0.3	0
staC	87.5	94.1	100	98.4	100	99.2	85.7	80.0	87.0	92.0	100	99.7	100
stbC	100	100	100	100	100	97.4	100	100	100	100	100	100	100
stcC	12.5	0	0	0	0	0.6	14.3	20.0	13.0	7.7	0	0.3	0
steC	12.5	0	0	0	0	0.7	14.3	20.0	13.0	7.3	0	0.3	0
stfC	12.5	5.9	0	1.6	0	0.8	14.3	20.0	13.0	8.0	0	0.3	0
stjC	0	5.9	0	1.6	0	0.8	14.3	10.0	13.0	2.2	0	0	0
stkC	0	0	0	0	0	0.5	9.5	10.0	4.3	6.1	0	0	0
tcfC	100	94.1	100	98.4	100	99	95.2	100	82.6	88.5	100	100	100
traT	0	5.9	23.1	8.1	16.7	85.9	28.6	0	0	11.5	6.8	7.7	25.0
xis	37.5	41.2	30.8	17.7	50.0	77.4	66.7	0	17.4	33.2	48.5	52.0	0

Note: the color schema is based on the percentage of isolates with the respective gene, ranging from dark green (0%) to yellow (~25%) to orange (~50%) to dark red (100%).

As noted above, the IncFIB-FIC(FII) fusion plasmids carrying the aerobactin, and likely Sit, iron acquisition operons were common among isolates from chicken-related sources. It is important to note that all of the strains carried a chromosomally encoded Sit operon, which made it somewhat challenging to tease out when they are also located on plasmids. Related plasmids encoding iron acquisition operons (aerobactin and Sit) are common among avian pathogenic *Escherichia coli* (APEC), where they likely play a role in extraintestinal infections, and in *S.* Kentucky, where they may enhance the ability to colonize chickens [30,38]. These iron acquisition genes are important to overcome the innate immune response to APEC and *Salmonella* infections, which involves sequestration of iron from the pathogens [39,40]. Among the isolates from human patients in the current study, only about 10% carried one of the plasmids. The distribution of the plasmids is highlighted

by Figure 3B, with the plasmid-containing group being predominately from chickens, while the groups lacking the plasmids had much more diverse sources, including most of those from human patients. The lack of diversity in the plasmid-carrying isolates was also borne out with the SNP typing, where most isolates carrying the fusion plasmids were of SNP type PDS000002218.839, two were of type PDS000112154.2, and four lacked SNP-type information (Figure S2C). Conversely, the strains lacking the plasmids represented a wide range of SNP types, the most common being PDS000027514.392; these plasmid-lacking strains were isolated from a number of different source types (including from blood, urine, and stool of human patients), states, and years of isolation (Table S1).

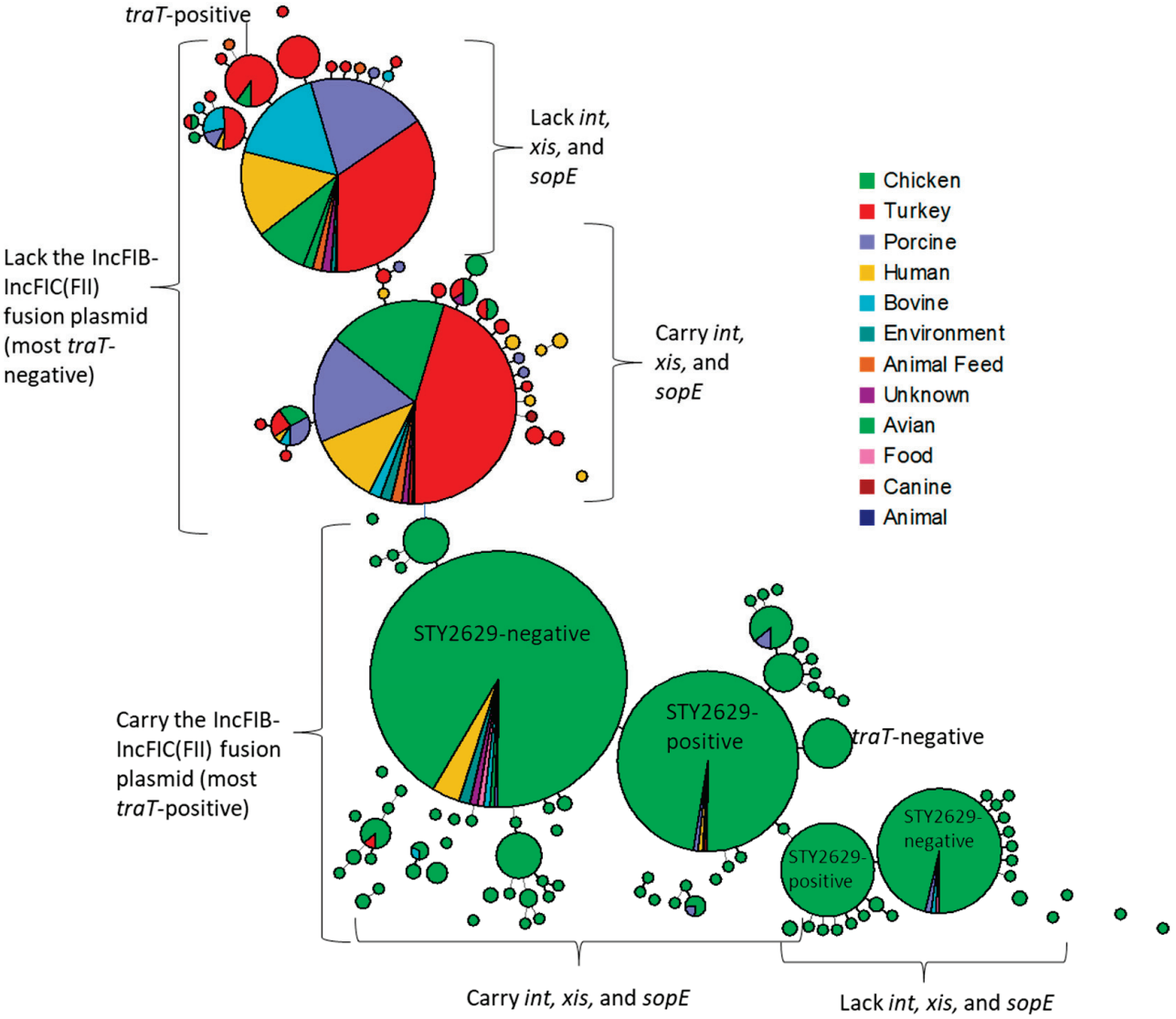


Figure 4. Minimum spanning tree analyses of strains originating within the United States based on their VF profiles. The trees are color-coded based on isolation source and the relative size of the circles is proportional to the group size. Key VF traits of the different clusters are annotated within the figure, including the presence or absence of key genetic factors. The colors are described by the legend included in the figure.

Among the strains collected from human patients, diverse virulence gene profiles were observed, suggesting that multiple factors may contribute to increased virulence in *S.* Schwarzengrund. For example, in the strains that were of SNP type PDS000002218.839 and carried the IncFIB-FIC(FII) plasmid, the aerobactin genes along with phage-associated

genes were detected. A relatively high proportion (>20%) of these strains were isolated from urine, which may be related to the carriage of iron acquisition genes on the plasmids, which are also common among *E. coli* that cause urinary tract infections [40]. Among the fourteen isolates identified from blood (over half of all isolates from human patients lacked specimen types), 12 (86%) carried the *int*, *xis*, and *sopE* genes, compared to 53% (N = 50/96) of isolates from stool. There was also diversity among the fimbrial gene operons among many strains [41], with a group of SNP type PDS000001366.21 isolates from the U.K. carrying genes for the *stc*, *stf*, *stg*, and *stk* operons, while lacking the *sta* and *sti* operons carried by almost all (~98%) of the other strains.

It is important to note that this study is not without potential limitations, including the fact that it was reliant upon the data that is present in the GenBank database, which is reliant upon submissions from different investigators. This dependence could introduce bias, as submissions might be weighted towards strains of particular interest to the submitting investigator. Another challenge was the quality of the metadata, especially for isolates from human patients. In many cases, the type of sample (stool, blood, urine, etc.) or detailed geography (e.g., state) was not included in the available metadata, which limited the ability for more detailed examinations of virulence and clinical disease. Improving the quality of metadata in public databases is crucial for enabling more accurate epidemiological studies.

4. Conclusions

Overall, the results of this study provided a broad picture of the genetic diversity, especially as it relates to virulence factors, AMR genes, and plasmids, through the detailed analyses of WGS data from *S.* Schwarzengrund and lays the foundation for more extensive examination of the serotype. The serotype's widespread presence in both human and food animal populations, combined with its ability to carry transmissible plasmids, poses significant public health challenges. Furthermore, the overlap in virulence profiles between human and food animal isolates highlights the potential zoonotic risk posed by *S.* Schwarzengrund and underscores its significance as a health concern. These findings emphasize the need for improved surveillance, particularly in the poultry industry, to mitigate the risk of AMR and virulence gene spread. Additionally, this study highlights the importance of enhancing metadata quality in public databases to facilitate more robust epidemiological investigations. Future studies should focus on characterizing the role of integrated phages and the potential for horizontal gene transfer in both clinical and agricultural settings.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/microorganisms13010092/s1: Table S1: Sample Metadata for the strains examined in this study; Table S2: AMR Gene Analyses (1=present, 2=absent) provides the AMR gene profiles for all of the strains in this study; Table S3: Plasmid Transfer Gene Analyses (1=present, 2=absent) provides the plasmid transfer gene profiles for all of the strains in this study; Table S4: Virulence Gene Analyses (1=present, 2=absent) provides the virulence gene profiles for all of the strains in this study; Figure S1: Minimum spanning tree analyses based on the AMR and plasmid transfer gene profiles; Figure S2: Minimum spanning tree analyses based on the virulence genes of the *S.* Schwarzengrund strains included in this study. These trees are delineated based on isolation source, country of origin, and SNP cluster.

Author Contributions: Conceptualization, M.A.F., B.K.K., S.Z., Y.M.S. and S.L.F.; methodology, J.H., S.Z., M.A.F. and S.L.F.; software, J.H. and S.L.F.; validation, M.A.F., B.K.K., Y.M.S. and S.L.F.; formal analysis, M.A.F., J.H. and S.L.F.; resources, Y.M.S. and S.L.F.; data curation, J.H. and S.L.F.; writing—original draft preparation, M.A.F., B.K.K. and S.L.F.; writing—review and editing, M.A.F., J.H., B.K.K., Y.M.S., S.Z. and S.L.F.; visualization, J.H.; supervision, B.K.K., Y.M.S. and S.L.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Center for Toxicological Research and the U.S. Food and Drug Administration.

Data Availability Statement: The data analyzed in this study is publicly available through GenBank and the corresponding accession numbers are identified in Table S1.

Acknowledgments: All the authors thank Ahmed Hikal and Ashraf Khan for their critical reviews of the manuscript prior to submission. The views expressed in this article are those of the authors and do not necessarily reflect the official policy of the Department of Health and Human Services, the U.S. Food and Drug Administration, or the U.S. Government. Reference to any commercial materials, equipment, or process does not in any way constitute approval, endorsement, or recommendation by the Food and Drug Administration.

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

References

1. Scallan, E.; Hoekstra, R.M.; Angulo, F.J.; Tauxe, R.V.; Widdowson, M.A.; Roy, S.L.; Jones, J.L.; Griffin, P.M. Foodborne illness acquired in the United States—Major pathogens. *Emerg. Infect. Dis.* **2011**, *17*, 7–15. [CrossRef] [PubMed]
2. Han, J.; Aljahdali, N.; Zhao, S.; Tang, H.; Harbottle, H.; Hoffman, M.; Frye, J.G.; Foley, S.L. Infection biology of *Salmonella enterica*. *EcoSal Plus* **2024**, *12*, eesp-0001. [CrossRef] [PubMed]
3. Hoffmann, S.; Batz, M.B.; Morris, J.G. Annual cost of illness and quality-adjusted life year losses in the United States due to 14 foodborne pathogens. *J. Food Prot.* **2012**, *75*, 1292–1302. [CrossRef] [PubMed]
4. Andino, A.; Hanning, I. *Salmonella enterica*: Survival, colonization, and virulence differences among serovars. *Sci. World J.* **2015**, *2015*, 520179. [CrossRef]
5. Jajere, S.M. A review of *Salmonella enterica* with particular focus on the pathogenicity and virulence factors, host specificity and antimicrobial resistance including multidrug resistance. *Vet. World* **2019**, *12*, 504–521. [CrossRef]
6. Aarestrup, F.M.; Hendriksen, R.S.; Lockett, J.; Gay, K.; Teates, K.; McDermott, P.F.; White, D.G.; Hasman, H.; Sørensen, G.; Bangtrakulnonth, A.; et al. International spread of multidrug-resistant *Salmonella* Schwarzengrund in food products. *Emerg. Infect. Dis.* **2007**, *13*, 726–731. [CrossRef]
7. Henderson, W. Isolation of *Salmonella* Schwarzengrund from turkey poults. *Science* **1952**, *116*, 669. [CrossRef]
8. Department of Health and Human Services (Ed.) *Centers for Disease Control and Prevention, An Atlas of Salmonella in the United States, 1968–2011: Laboratory-based Enteric Disease Surveillance*; Department of Health and Human Services: Washington, DC, USA; CDC: Atlanta, GA, USA, 2013; pp. 1–248.
9. Bangtrakulnonth, A.; Pornreongwong, S.; Pulsrikarn, C.; Sawanpanyalert, P.; Hendriksen, R.S.; Lo Fo Wong, D.M.; Aarestrup, F.M. *Salmonella* serovars from humans and other sources in Thailand, 1993–2002. Emerg. *Infect. Dis.* **2004**, *10*, 131–136. [CrossRef]
10. Noenchat, P.; Direksin, K.; Sornplang, P. The phenotypic and genotypic antimicrobial resistance patterns of *Salmonella* isolated from chickens and meat at poultry slaughterhouses in Japan and Thailand. *Vet. World* **2023**, *16*, 1527–1533. [CrossRef]
11. Sasaki, Y.; Ikeda, T.; Momose, Y.; Yonemitsu, K.; Uema, M.; Asai, T. Geographical Variation of Antimicrobial Resistance of *Salmonella* in Japanese Chicken. *Food Safety* **2024**, *12*, 59–66. [CrossRef]
12. Asai, T.; Murakami, K.; Ozawa, M.; Koike, R.; Ishikawa, H. Relationships between multidrug-resistant *Salmonella enterica* Serovar Schwarzengrund and both broiler chickens and retail chicken meats in Japan. *Jpn. J. Infect. Dis.* **2009**, *62*, 198–200. [CrossRef] [PubMed]
13. Sasaki, Y.; Ikeda, T.; Yonemitsu, K.; Kuroda, M.; Ogawa, M.; Sakata, R.; Uema, M.; Momose, Y.; Ohya, K.; Watanabe, M.; et al. Antimicrobial resistance profiles of *Campylobacter jejuni* and *Salmonella* spp. isolated from enteritis patients in Japan. *J. Vet. Med. Sci.* **2023**, *85*, 463–470. [CrossRef] [PubMed]
14. Chen, M.H.; Wang, S.W.; Hwang, W.Z.; Tsai, S.J.; Hsih, Y.C.; Chiou, C.S.; Tsen, H.Y. Contamination of *Salmonella* Schwarzengrund cells in chicken meat from traditional marketplaces in Taiwan and comparison of their antibiograms with those of the human isolates. *Poult. Sci.* **2010**, *89*, 359–365. [CrossRef] [PubMed]
15. Duc, V.M.; Shin, J.; Nagamatsu, Y.; Fuhiwara, A.; Toyofuku, H.; Obi, T.; Chuma, T. Increased Salmonella Schwarzengrund prevalence and antimicrobial susceptibility of Salmonella enterica isolated from broiler chickens in Kagoshima Prefecture in Japan between 2013 and 2016. *J. Vet. Med. Sci.* **2020**, *82*, 585–589. [CrossRef]
16. Centers for Disease Control and Prevention, Multistate outbreak of human *Salmonella* infections caused by contaminated dry dog food—United States, 2006–2007. MMWR Morb. *Mortal. Wkly. Rep.* **2008**, *57*, 521–524.
17. Centers for Disease Control and Prevention. *Outbreak of Salmonella Infections Linked to Butterball Brand Ground Turkey*; CDC: Atlanta, GA, USA, 2019.

18. Monte, D.F.; Lincopan, N.; Berman, H.; Cerdeira, L.; Keelara, S.; Thakur, S.; Fedorka-Cray, P.J.; Landgraf, M. Genomic features of high-priority *Salmonella enterica* serovars circulating in the food production chain, Brazil, 2000–2016. *Sci. Rep.* **2019**, *9*, 11058. [CrossRef]
19. Olsen, S.J.; DeBess, E.E.; McGivern, T.E.; Marano, N.; Eby, T.; Mauvais, S.; Balan, V.K.; Zirnstein, G.; Cieslak, P.R.; Angulo, F.J. A nosocomial outbreak of fluoroquinolone-resistant *Salmonella* infection. *N. Engl. J. Med.* **2001**, *344*, 1572–1579. [CrossRef]
20. Penha Filho, R.A.C.; Ferreira, J.C.; Galetti, R.; Kanashiro, A.M.I.; Berchieri, A., Jr.; da Costa Darini, A.L. The rise of multidrug resistant *Salmonella* isolates in healthy chickens in Brazil by successful establishment of plasmid IncHI2A carrying several antibiotic resistance genes. *Braz. J. Microbiol.* **2023**, *54*, 469–474. [CrossRef]
21. Silva, K.C.; Fontes, L.C.; Moreno, A.M.; Astolfi-Ferreira, C.S.; Ferreira, A.J.; Lincopan, N. Emergence of extended-spectrum-beta-lactamase CTX-M-2-producing *Salmonella enterica* serovars Schwarzengrund and Agona in poultry farms. *Antimicrob. Agents Chemother.* **2013**, *57*, 3458–3459. [CrossRef]
22. Akiyama, T.; Khan, A.A. Molecular characterization of strains of fluoroquinolone-resistant *Salmonella enterica* serovar Schwarzengrund carrying multidrug resistance isolated from imported foods. *J. Antimicrob. Chemother.* **2012**, *67*, 101–110. [CrossRef]
23. Khajanchi, B.K.; Xu, J.; Grim, C.J.; Ottesen, A.R.; Ramachandran, P.; Foley, S.L. Global transcriptomic analyses of *Salmonella enterica* in Iron-depleted and Iron-rich growth conditions. *BMC Genom.* **2019**, *20*, 490. [CrossRef] [PubMed]
24. de Jesus Bertani, A.M.; Vieira, T.; Reis, A.D.; Dos Santos, C.A.; de Almeida, E.A.; Camargo, C.H.; Casas, M.R.T. Whole genome sequence analysis of the first reported isolate of *Salmonella* Agona carrying blaCTX-M-55 gene in Brazil. *Sci. Rep.* **2023**, *13*, 2299. [CrossRef] [PubMed]
25. Yoon, S.; Lee, Y.J. Molecular characteristics of ESBL-producing Escherichia coli isolated from chickens with colibacillosis. *J. Vet. Sci.* **2022**, *23*, e37. [CrossRef] [PubMed]
26. Han, J.; Tang, H.; Zhao, S.; Foley, S.L. *Salmonella enterica* virulence databases and analysis tools development. *Sci. Rep.* **2024**, *14*, 25228. [CrossRef] [PubMed]
27. Algarni, S.; Foley, S.L.; Tang, H.; Zhao, S.; Gudeta, D.D.; Khajanchi, B.K.; Ricke, S.C.; Han, J. Development of an antimicrobial resistance plasmid transfer gene database for enteric bacteria. *Front. Bioinform.* **2023**, *3*, 1279359. [CrossRef]
28. Feldgarden, M.; Brover, V.; Gonzalez-Escalona, N.; Frye, J.G.; Haendiges, J.; Haft, D.H.; Hoffmann, M.; Pettengill, J.B.; Prasad, A.B.; Tillman, G.E.; et al. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance. *; stress response.; and virulence. Sci. Rep.* **2021**, *11*, 12728.
29. Gangiredla, J.; Rand, H.; Benisatto, D.; Payne, J.; Strittmatter, C.; Sanders, J.; Wolfgang, W.J.; Libuit, K.; Herrick, J.B.; Prarat, M.; et al. GalaxyTrakr: A distributed analysis tool for public health whole genome sequence data accessible to non-bioinformaticians. *BMC Genom.* **2021**, *22*, 114. [CrossRef]
30. Felix, M.A.; Sopovski, D.; Commichaux, S.; Yoskowitz, N.; Aljahdali, N.H.; Grim, C.J.; Abbott, C.N.; Carlton, A.; Han, J.; Sanad, Y.M.; et al. Genetic relatedness and virulence potential of *Salmonella* Schwarzengrund strains with or without an IncFIB-IncFIC(FII) fusion plasmid isolated from food and clinical sources. *Front. Microbiol.* **2024**, *15*, 1397068. [CrossRef]
31. Oxendine, A.; Walsh, A.A.; Young, T.; Dixon, B.; Hoke, A.; Rogers, E.E.; Lee, M.D.; Maurer, J.J. Conditions necessary for the transfer of antimicrobial resistance in poultry litter. *Antibiotics* **2023**, *12*, 1006. [CrossRef]
32. Abreu, R.; Semedo-Lemsaddek, T.; Cunha, E.; Tavares, L.; Oliveira, M. Antimicrobial drug resistance in poultry production: Current status and innovative strategies for bacterial control. *Microorganisms* **2023**, *11*, 953. [CrossRef]
33. Braetz, S.; Schwerk, P.; Figueroa-Bossi, N.; Tedin, K.; Fulde, M. Prophage Gifsy-1 Induction in *Salmonella enterica* Serovar Typhimurium Reduces Persister Cell Formation after Ciprofloxacin Exposure. *Microbiol. Spectr.* **2023**, *11*, e0187423. [CrossRef] [PubMed]
34. Siebor, E.; Neuwirth, C. Overview of *Salmonella* Genomic Island 1-related elements among Gamma-Proteobacteria reveals their wide distribution among environmental species. *Front. Microbiol.* **2022**, *13*, 857492. [CrossRef] [PubMed]
35. Doublet, B.; Golding, G.R.; Mulvey, M.R.; Cloeckaert, A. Potential integration sites of the *Salmonella* Genomic Island 1 in Proteus mirabilis and other bacteria. *J. Antimicrob. Chemother.* **2007**, *59*, 801–803. [CrossRef] [PubMed]
36. Zaworski, J.; McClung, C.; Ruse, C.; Weigele, P.R.; Hendrix, R.W.; Ko, C.C.; Edgar, R.; Hatfull, G.F.; Casjens, S.R.; Raleigh, E.A. Genome analysis of *Salmonella enterica* serovar Typhimurium bacteriophage L.; indicator for StySA (StyLT2III) restriction-modification system action. *G3*, 2021; 11, kaa037.
37. Mirold, S.; Rabsch, W.; Rohde, M.; Stender, S.; Tschape, H.; Russmann, H.; Igwe, E.; Hardt, W.D. Isolation of a temperate bacteriophage encoding the type III effector protein SopE from an epidemic *Salmonella typhimurium* strain. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9845–9850. [CrossRef]
38. Lopez, C.A.; Winter, S.E.; Rivera-Chavez, F.; Xavier, M.N.; Poon, V.; Nuccio, S.P.; Tsolis, R.M.; Baumler, A.J. Phage-mediated acquisition of a type III secreted effector protein boosts growth of *Salmonella* by nitrate respiration. *mBio* **2021**, *3*, e00143-12. [CrossRef]

39. Gao, Q.; Wang, X.; Xu, H.; Xu, Y.; Ling, J.; Zhang, D.; Gao, S.; Liu, X. Roles of iron acquisition systems in virulence of extraintestinal pathogenic *Escherichia coli*: Salmochelin and aerobactin contribute more to virulence than heme in a chicken infection model. *BMC Microbiol.* **2012**, *12*, 143. [CrossRef]
40. Johnson, J.R. Microbial virulence determinants and the pathogenesis of urinary tract infection. *Infect. Dis. Clin. N. Am.* **2003**, *17*, 261–278. [CrossRef]
41. Khajanchi, B.K.; Hasan, N.A.; Choi, S.Y.; Han, J.; Zhao, S.; Colwell, R.R.; Cerniglia, C.E.; Foley, S.L. Comparative genomic analysis and characterization of incompatibility group FIB plasmid encoded virulence factors of *Salmonella enterica* isolated from food sources. *BMC Genom.* **2017**, *18*, 570. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

microorganisms

MDPI

Article

Staphylococcus aureus Proteins Implicated in the Reduced Virulence of *sarA* and *sarA/agr* Mutants in Osteomyelitis

Karen E. Beenken [1,2,*], Mara J. Campbell [1], Stephanie D. Byrum [2], Rick D. Edmondson [2], Samuel G. Mackintosh [2], Alan J. Tackett [2] and Mark S. Smeltzer [1]

[1] Department of Microbiology and Immunology, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA; mjcampbell@uams.edu (M.J.C.); smeltzermarks@uams.edu (M.S.S.)
[2] Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA; rdedmondson@uams.edu (R.D.E.); mackintoshsamuelg@uams.edu (S.G.M.); ajtackett@uams.edu (A.J.T.)
* Correspondence: beenkenkarene@uams.edu; Tel.: +1-501-686-5408

Abstract: Using a murine osteomyelitis model, we recently demonstrated that *Staphylococcus aureus sarA* and *sarA/agr* mutants generated in the USA300 strain LAC are attenuated to a greater extent than an isogenic *agr* mutant and that this can be attributed to a significant extent to the increased production of extracellular proteases in both mutants. Based on this, we used a mass-based proteomics approach to compare the proteomes of LAC, its isogenic *agr*, *sarA*, and *sarA/agr* mutants, and isogenic derivatives of all four of these strains unable to produce the extracellular proteases aureolysin, SspA, SspB, ScpA, or SplA-F. This allowed us to identify proteins that were present in reduced amounts in *sarA*, and *sarA/agr* mutants owing to the increased production of extracellular proteases. A total of 1039 proteins were detected in conditioned media (CM) from overnight cultures of LAC, and protease-mediated degradation was shown to contribute to the reduced abundance of 224 of these (21.6%) in CM from the *sarA* and *sarA/agr* mutants. Among these were specific proteins previously implicated in the pathogenesis and therapeutic recalcitrance of *S. aureus* osteomyelitis. This demonstrates that the ability of *sarA* to limit protease production plays a key role in post-translational remodeling of the *S. aureus* proteome to a degree that can be correlated with reduced virulence in our osteomyelitis model, and that it does so irrespective of the functional status of *agr*. This also suggests that at least some of these 224 proteins may be viable targets for prophylactic or therapeutic intervention.

Keywords: *Staphylococcus*; osteomyelitis; *sarA*; *agr*; protease; proteomics

1. Importance

Staphylococcus aureus produces a diverse array of virulence factors. This complicates attempts to identify virulence factors that are critical to the pathogenesis of *S. aureus* infections. Mutation of *sarA* limits the virulent *S. aureus* infection including osteomyelitis irrespective of the functional status of *agr*, and an important factor contributing to this attenuation is the increased production of extracellular proteases and the impact of these proteases on its virulence factor repertoire. Identifying proteins whose abundance is limited in *sarA* and *sarA/agr* mutants can facilitate efforts to identify virulence factors of particular importance in osteomyelitis and perhaps other forms of *S. aureus* infection. Indeed, the critical role of *sarA* in limiting protease production suggests a mass-based proteomics approach capable of distinguishing between truncated and full-length proteins is the only way to accomplish this task.

2. Introduction

Staphylococcus aureus is a leading cause of all forms of orthopedic infection including osteomyelitis, septic arthritis, and infections associated with indwelling hardware [1–3]. It is becoming increasingly difficult to treat these infections owing to the persistent emergence of strains resistant to key antibiotics. Moreover, these infections are characterized by formation of a biofilm, the presence of which confers a therapeutically relevant level of intrinsic resistance to both conventional antibiotics and host defenses [4,5]. Thus, the treatment of *S. aureus* orthopedic infections requires long-term intensive antibiotic therapy accompanied by surgical debridement to remove infected tissues and indwelling hardware, and even after such intensive medical and surgical intervention, the recurrence rate is unacceptably high [1–3].

These observations have rekindled interest in virulence-associated therapies that could be used to overcome this therapeutic recalcitrance [6]. The staphylococcal accessory regulator (*sarA*) and the accessory gene regulator (*agr*) are the two *S. aureus* regulatory loci that have been studied extensively in this regard [6–10]. The *sarA* locus encodes a 15 kDa DNA-binding protein (SarA) that impacts the transcription of many *S. aureus* genes including *agr* [11]. Specifically, SarA is required for maximum expression of *agr*, activation of which modulates the production of multiple *S. aureus* virulence factors in a coordinated fashion [12]. In this regulatory scenario, *agr* is downstream of *sarA*, suggesting that inhibition of *agr* could disrupt the coordinated production of these virulence factors irrespective of the functional status of *sarA*. However, *sarA* also serves regulatory functions that are independent of *agr,* one critical observation in this regard being that mutation of *sarA* results in increased protease production and decreased biofilm formation, while mutation of *agr* has the opposite effect on both phenotypes.

Extracellular proteases produced by *S. aureus* include aureolysin, SspA (V8), SspB, ScpA, and as many as six *spl*-encoded proteases (SplA-F) [13,14]. These proteases collectively serve vital roles on behalf of the bacterium by promoting tissue invasion, nutrient acquisition, and avoidance of host defenses [13–19]. This suggests that the ability of *S. aureus* to survive in vivo and cause disease would be limited in strains unable to produce proteases. Conversely, strains that produce elevated levels of these proteases would be predicted to exhibit increased virulence. However, a derivative of the methicillin-resistant USA300 strain LAC (Los Angeles County clone) that does not produce any of these 10 extracellular proteases is hypervirulent in murine models of bacteremia and osteomyelitis [13,20], while strains that produce significantly increased amounts of these proteases are attenuated in the same models [20–22].

S. aureus sarA mutants are the prototype example of the latter in that they produce elevated levels these extracellular proteases, and this has been shown to play a significant role in defining the attenuation of *sarA* mutants in diverse animal models of infection including osteomyelitis [20,22–25]. Based on this, we used a mass-based proteomics approach to identify proteins that are present in reduced amounts in *sarA* mutants by comparison to their isogenic parent strains owing to protease-mediated degradation [21,26]. Mutation of other *S. aureus* regulatory loci also results in increased protease production, but we confirmed that mutation of *sarA* results in a greater increase in protease production, and a greater decrease in virulence, than mutation of any other *S. aureus* regulatory locus examined to date [23,27,28]. However, these studies were limited to regulatory mutants that exhibit increased protease production, and as noted above this does not include *agr*.

To address this, and to define the impact of *sarA* on protease production in an *agr* mutant, we used our murine osteomyelitis model to directly evaluate the relative virulence of LAC, isogenic *sarA*, *agr*, and *sarA/agr* mutants, and isogenic derivatives of all four strains unable to produce aureolysin, ScpA, SspA, SspB, or SplA-F [20]. The results confirmed that

sarA and *sarA/agr* mutants are attenuated to a greater extent than the isogenic *agr* mutant, and that this is due to, a statistically significant extent, the increased production of these extracellular proteases. This suggests that *S. aureus* proteins of interest in the specific context of osteomyelitis could be identified and prioritized based on their reduced abundance in both *sarA* and *sarA/agr* mutants by comparison to LAC. Indeed, the reduced virulence of these mutants was correlated with reduced biofilm formation and reduced cytotoxicity for osteoblasts and osteoclasts, and these phenotypes were correlated in turn with the reduced abundance of full-length and presumably functional forms of specific virulence factors implicated in these phenotypes including α-toxin, both components of the Panton–Valentine leukocidin (LukF and LukS), the primary *S. aureus* extracellular nuclease (Nuc1), and protein A (Spa) [20]. However, our studies investigating these correlations were limited by the availability of antibodies appropriate for targeted Western blot analysis. To overcome this limitation, we used a mass-based proteomics approach capable of quantitatively distinguishing between full-length and truncated *S. aureus* proteins [21,26] to assess the global impact of increased protease production on the proteomes of *sarA* and *sarA/agr* mutants and to identify *S. aureus* proteins present in reduced amounts in both attenuated mutants owing to protease-mediated degradation.

3. Materials and Methods

Bacterial strains and growth conditions. Mutants were generated as previously described [20,21,24] and stored at −80 °C in tryptic soy broth (TSB) supplemented with 25% (*v*/*v*) glycerol. Each strain was removed from cold storage and plated on tryptic soy agar (TSA) with antibiotic selection. To prepare conditioned medium (CM), multiple colonies of each strain were used to inoculate TSB without antibiotics. Cultures were grown 16 h and standardized to an optical density (OD_{560}) of 10.0. An aliquot was removed and used to verify viable cell density by plating on TSA with and without antibiotics. CM was then prepared for proteome analysis by centrifugation and filter sterilization using 0.22-micron filters.

Proteome analysis. Proteins in CM were resolved by SDS-PAGE using 4–12% Bolt Bis-Tris Plus gradient gels (Thermo Fisher Scientific, Waltham, MA, USA). To focus on the abundance of full-length proteins, gels were cut into 12 equal segments and each segment subjected to proteomic analysis as previously described [26]. Briefly, gel slices were destained in 50% methanol (Fisher, Hampton, NH, USA), 50 mM ammonium bicarbonate (Sigma-Aldrich, St. Louis, MO, USA), followed by reduction in 10 mM Tris [2-carboxyethyl]phosphine (Pierce, Woodland Hills, CA, USA) and alkylation in 50 mM iodoacetamide (Sigma-Aldrich). They were then dehydrated in acetonitrile (Fisher), followed by addition of 100 ng porcine sequencing grade modified trypsin (Promega, Madison, WI, USA) in 50 mM ammonium bicarbonate (Sigma-Aldrich) and incubation at 37 °C for 12–16 h. Peptide products were then acidified in 0.1% formic acid (Pierce). Tryptic peptides were separated by reverse phase XSelect CSH C18 2.5 µm resin (Waters, Milford, MA, USA) on an in-line 150 × 0.075 mm column using an UltiMate 3000 RSLCnano system (Thermo). Peptides were eluted using a 60 min gradient from 98:2 to 65:35 buffer A:B ratio (Buffer A = 0.1% formic acid, 0.5% acetonitrile, Buffer B = 0.1% formic acid, 99.9% acetonitrile). Eluted peptides were ionized by electrospray (2.2 kV) followed by MS/MS analysis using higher-energy collisional dissociation (HCD) on an Orbitrap Fusion Tribrid mass spectrometer (Thermo) in the top-speed data-dependent mode. MS data were acquired using the FTMS analyzer in the profile mode at a resolution of 240,000 over a range of 375–1500 m/z. Following HCD activation, MS/MS data were acquired using the ion trap analyzer in the centroid mode and normal mass range with precursor mass-dependent normalized collision energy between 28.0 and 31.0. Proteins were identified by database search using

Mascot (Matrix Science, version 2.6.2, Columbus, OH, USA) against the *USA300 S. aureus* database (2607 entries, Genebank accession JTJK01000002) as previously described [21,26]. Protein probabilities were assigned by the Protein Prophet algorithm [29]. Total spectral counts for each of the 5 biological replicates were exported from Scaffold into Microsoft Excel for further analysis. CM from LAC was used to determine the gel band with the maximum MS2 spectral count for each identified protein. Spectral counts for full-length proteins in other CM were then determined by adding the spectral counts in that gel band with those in the gel bands immediately above and below, thus accounting for minor variations in full-length protein migration through the gel. The spectral counts observed were then averaged for the same gel slices across all biological replicates from all other strains.

Statistical analysis. The spectral counts from the gel slice for which the abundance of each protein was highest in CM from LAC was added to the spectral counts for the same protein in the two slices above and below this maximum were used to calculate the fold change and a Student's *t*-test for significance [26]. The missing value heatmap was generated using R 4.4.1 and the visdat package.

Western blot analysis. Western blots were performed using commercially available antibodies (Sigma (St. Louis, MO, USA) and Toxin Technologies (Sarasota, FL, USA), abCAM (Cambridge, UK), and United States Biological (Salem, MA, USA)) as previously described [24].

4. Results

Characterization of mutants. We first evaluated our results in the context of the mutations that define each of the strains included in these studies. For instance, although SarA is an intracellular protein it was detected at comparable levels in conditioned medium (CM) from overnight (16 h) stationary phase cultures of LAC and its protease-deficient mutant, it was not detected in CM from *sarA* or *sarA*/*agr* mutants or their protease-deficient derivatives (Table 1). AgrA is also an intracellular protein, and it was not detected in CM from LAC or any of its isogenic regulatory mutants irrespective of their ability to produce extracellular proteases. However, delta toxin (δ-toxin), which is encoded within the regulatory RNA associated with *agr* (RNAIII) that is produced under the positive regulatory control of AgrA was detected at comparable levels in LAC and its protease-deficient derivative, but not in CM from the *agr* or *sarA*/*agr* mutants (Table 1). The abundance of δ-toxin was also reduced in CM from the LAC *sarA* mutant by comparison to CM from LAC. This is consistent with studies demonstrating that *sarA* is required for maximum expression of *agr* [11,12]. However, the abundance of δ-toxin was restored in the protease-deficient *sarA* mutant, demonstrating that the increased production of extracellular proteases further limits the abundance of δ-toxin at a post-translational level.

Similarly, the only *S. aureus* proteins encoded outside the *agr* operon and known to be produced under the regulatory control of AgrA are the genes encoding phenol-soluble modulins (PSMs) [30], and all alpha class PSMs (PSMα1, PSMα2, PSMα3, and PSMα4) were absent in CM from the LAC *agr* and *sarA*/*agr* mutants and their protease-deficient derivatives (Table 1). PSMs were present at comparable levels in LAC and its protease-deficient mutant, and absent or only present at very low levels ($\leq$0.6 average spectral counts) in the isogenic *sarA* mutant. However, the abundance of all four alpha class PSMs was increased in CM from the protease-deficient *sarA* mutant. This confirms the importance of *agr* in PSM production, and the importance of *sarA* in limiting proteases to facilitate their accumulation.

Table 1. Spectral count of full-length proteins associated with each mutant. Spectral counts shown are the average of five biological replicates of CM from LAC, isogenic *sarA* (S), *agr* (A), sarA/agr (SA) mutants, and protease deficient derivatives of each strain (P, SP, AP and SAP respectively).

Protein	LAC	P	S	SP	A	AP	SA	SAP
SarA	27.6	21.8	0.0	0.0	2.6	2.2	0.0	0.0
Delta Toxin	26.6	19.0	8.2	21.4	0.2	0.0	0.0	0.0
PSMα1	18.2	19.8	0.4	5.8	0.4	0.4	0.0	0.0
PSMα2	14.4	12.6	0.0	3.0	0.0	0.0	0.0	0.0
PSMα3	10.2	9.0	0.0	2.6	0.0	0.0	0.0	0.0
PSMα4	24.4	18.6	0.6	8.0	0.0	0.0	0.0	0.0
Aurolysin	6.4	0.0	14.4	0.0	2.0	0.0	36.2	0.0
SspA	19.4	0.2	105.0	1.2	4.6	0.0	19.4	0.0
SspB	43.2	0.2	2.0	0.4	5.6	0.0	6.6	0.0
ScpA	21.2	0.0	73	0.2	6.2	0.0	29.4	0.0
SplA	5.6	0.0	34.6	0.0	0.6	0.0	4.8	0.0
SplB	22.4	0.0	107.6	1.0	3.8	0.0	18.4	0.0
SplC	6.4	0.0	23.0	0.0	1.2	0.0	3.4	0.0
SplD	12.2	0.0	30.8	0.2	3.0	0.0	5.2	0.0
SplE	10.2	0.0	30.6	0.0	0.0	0.0	0.6	0.0
SplF	10.4	0.0	30.8	0.2	3.0	0.0	5.2	0.0

The protease-deficient mutants have null mutations in all of the genes encoding aureolysin, ScpA, SspA, SspB, and SplA-F [26], so it would be anticipated that none of these proteases would be detected. We did detect some of these proteases in CM from the protease-deficient *sarA* mutant and, to a lesser extent, the protease-deficient derivative of LAC (Table 1). However, the abundance of these proteases even in CM from the protease-deficient *sarA* mutant was very low ($\leq$1.2 average spectral counts across five biological replicates), and none were detected in the protease-deficient *agr* or *sarA*/*agr* mutants. Our previous phenotypic characterization of these mutants also confirmed that protease activity was limited to the same degree in all four protease mutants [20].

The primary strain in which proteases were detected in a protease mutant was the protease-deficient *sarA* mutant. This is relevant in that every extracellular protease other than SspB was present in an increased amount in CM from the *sarA* mutant by comparison to CM from LAC (Table 1). This suggests that the low levels detected in CM from the protease-deficient *sarA* mutant are background noise. However, the observation that SspB, the abundance of which was decreased in CM from *sar, agr* and *sarA*/*agr* mutants, was detected in CM from the protease-deficient *sarA* mutant contradicts this conclusion. At the same time, we previously demonstrated that spectral counts for full-length SspB are decreased in CM from a *sarA* mutant, but total spectral counts as assessed independently of mass are increased [26]. This is consistent with our previous demonstration that transcription from all four protease-encoding genes and operons is increased in *sarA* mutants [23], and it confirms that mutation of *sarA* results in the increased production of all 10 of the primary extracellular proteases produced by LAC. By comparison to LAC, the abundance of all proteases was reduced in the LAC *agr* mutant (Table 1).

Most proteases were present in CM from the *sarA*/*agr* mutant in amounts comparable to CM from LAC and higher than those observed in the isogenic *agr* mutant (Table 1). How-

ever, the abundance of most was reduced in CM from the *sarA/agr* mutant by comparison to the isogenic *sarA* mutant. The exception was aureolysin, which was most abundant in CM from the *sarA/agr* mutant. This likely accounts for the increase in overall protease activity in CM from the *sarA/agr* mutant relative to CM from LAC, while the fact that none of the other proteases were increased likely explains why overall protease activity was decreased in the *sarA/agr* mutant by comparison to the *sarA* [20]. Importantly, aureolysin is one of two proteases shown to contribute to the greatest extent to the attenuation of *sarA* mutants in osteomyelitis [24]. The other is staphopain A (ScpA), the abundance of which was increased to a greater extent in the *sarA* mutant than the isogenic *sarA/agr* mutant (Table 1). In fact, the abundance of ScpA was not significantly increased in CM from the *sarA/agr* mutant by comparison to CM from LAC. Mutation of the genes encoding these same proteases (*aur* and *scpA*) was also shown to replicate the hypervirulent phenotype of a LAC protease null mutant in a murine sepsis model [13].

The fact that the abundance of aureolysin was increased in both *sarA* and *sarA/agr* mutants while the abundance of ScpA was only increased in the *sarA* mutant suggests that the increased production of aureolysin may be particularly important in the pathogenesis of osteomyelitis. Indeed, aureolysin-mediated degradation of alpha PSMs was shown to be a primary factor defining the attenuation of a LAC *saePQRS* (*sae*) mutant in osteomyelitis [31]. The reduced abundance of PSMs was also confirmed to contribute to the attenuation of a LAC *sarA* mutant, although whether this could be attributed specifically to degradation by aureolysin was not examined [32].

Our previous Western blots demonstrated that the increased production of extracellular proteases in a LAC *sarA* mutant limits the abundance of LukF, LukS, and Nuc1 in their full-length forms with a corresponding increase in the abundance of a truncated but immunoreactive protein [20]. Using our mass-based proteomics approach, we confirmed the presence of truncated versions of LukF, LukS, and Nuc1 in CM from the *sarA* mutant but not the protease-deficient *sarA* mutant (Figure 1), thus confirming the ability to quantitatively distinguish between full-length and truncated proteins. We also confirmed that these proteins, whether in truncated or full-length form, are present in increased amounts in CM from the *sarA* mutant and its protease-deficient derivative by comparison to CM from LAC (Figure 1), confirming that mutation of *sarA* results in increased production of LukF, LukS, and Nuc1, but that the increased production of proteases limits their abundance in full-length and presumably functional forms.

Additionally,, we previously concluded that LukF is only present in a truncated form in CM from the *sarA* mutant and that this might be a limiting factor in the activity of the bi-component Panton–Valentine leukocidin (PVL) [20]. Here, we confirmed the presence of this truncated product and the fact that it was absent in CM from the protease-deficient *sarA* mutant, but also detected a larger version of LukF that was not apparent by Western blot (Figure 1). This size of this protein was similar to full-length LukF, but it was nevertheless truncated as evidenced by the loss of an antibody-binding site [20]. Thus, this does not contradict the suggestion that the absence of full-length LukF may be a limiting factor in the phenotypic impact of PVL, but it does further illustrate the utility of our mass-based approach. This did not appear to be a concern with LukS and Nuc1 in that both full-length and truncated versions of these proteins were detected by both Western blot and in our proteomics analysis (Figure 1).

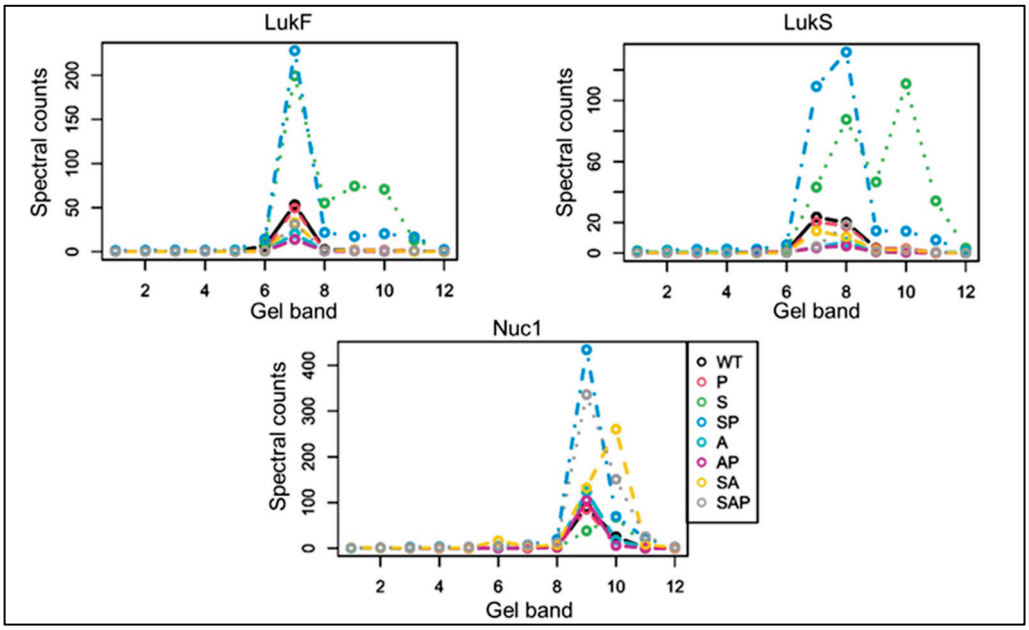


Figure 1. Spectral counts of specific proteins as a function of molecular weight (gel-slice). CM from overnight cultures of LAC, isogenic derivatives with mutations in *sarA* (S), *agr* (A), *sarA*/*agr* (SA), and protease deficient derivatives of all these strains (P, SP, AP, and SAP, respectively) were resolved by SDS-PAGE and each slice subjected to proteomic analysis. Graphs indicate the average spectral count from 5 biological replicates as determined by tandem mass spectrometry plotted for each gel slice.

Impact of protease production on the proteome of *S. aureus sarA*, *agr*, and *sarA*/*agr* mutants. The results discussed above are consistent with the expected phenotypes of all mutants included in our comparisons. They also validate our mass-based approach including its ability to distinguish between truncated and full-length proteins. Based on this, we extended our analysis in an unbiased, comprehensive, and antibody-independent fashion. We detected 1039 full-length proteins in conditioned media (CM) from stationary phase (16 h) cultures of LAC (Supplementary Tables S1 and S2). Only 139 (13.4%) were detected in CM from the *sarA* mutant, while 870 (83.7%) were detected in the protease-deficient *sarA* mutant (Figure 2). A total of461 full-length proteins (44.4%) were detected in CM from the *sarA*/*agr* mutant, with 563 (54.2%) being detected in CM from the protease-deficient *sarA*/*agr* mutant. The lower number of proteins detected in CM from the *sarA*/*agr* mutant by comparison to the *sarA* mutant is consistent with the observation that protease production is higher in CM from the *sarA* mutant than in CM from the *sarA*/*agr* mutant [20]. Conversely, the fewer proteins detected in CM from the protease-deficient *sarA*/*agr* mutant by comparison to the protease-deficient *sarA* mutant is consistent with the impact of mutating *agr* on protein production. This is further reflected in the reduced number of proteins detected in CM from the *agr* mutant (496), which was not increased in CM from the isogenic protease-deficient *agr* mutant. Indeed, the *agr* mutant was the only regulatory mutant examined in which the number of proteins detected was not increased in the protease-deficient derivative (Figure 2).

Correlations between protein abundance and virulence. LAC *sarA* and *sarA*/*agr* mutants are more attenuated in our osteomyelitis model than an isogenic *agr* mutant, and virulence is restored to a significant extent in a protease-deficient derivatives of both mutants [20]. This suggests that proteins that are present in decreased amounts in CM from *sarA* and *sarA*/*agr* mutants owing to protease-mediated degradation would be of interest in the pathogenesis of osteomyelitis. We used a $\log_2$ fold change >2.0 and p value <0.05 to identify these proteins. Because statistical analysis is not possible when considering a protein that was not detected in one of the strains included in the comparison, we also iden-

tified proteins that were not detected in CM *sarA* and *sarA/agr* mutants but were detected in CM from both of their protease-deficient derivatives (Supplementary Tables S1 and S2).

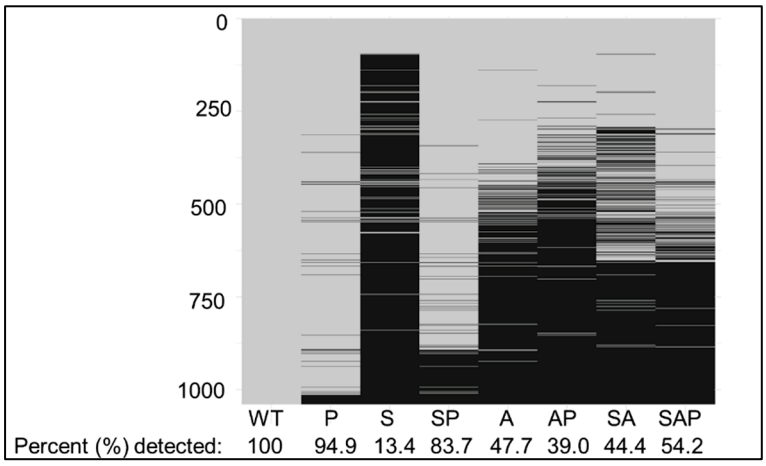


Figure 2. Heatmap reflecting the detection of proteins in LAC and its regulatory mutants as a function of protease production. CM from overnight cultures of LAC, isogenic derivatives with mutations in *sarA* (S), *agr* (A), *sarA/agr* (SA), and protease deficient derivatives of all these strains (P, SP, AP, and SAP, respectively). 1039 proteins were detected in CM from LAC. Horizontal black lines in each column indicate proteins that were absent in CM from each mutant, with the percentage of the 1039 proteins detected in CM from LAC that were detected in CM from each mutant indicated.

The ability to identify some proteins of potential interest in CM from the *sarA/agr* mutant was limited by the influence of *agr* on protein production. For instance, PSMs contribute to the pathogenesis of osteomyelitis and are present in reduced amounts in CM from *sarA* mutants owing to protease-mediated degradation [32], but the impact of proteases on the abundance of PSMs is not apparent in CM from *sarA/agr* mutant because PSMs are not made owing to the *agr* mutation (Table 1). In fact, 543 of the 1039 proteins detected in CM from LAC (52.3%) were not detected in CM from the isogenic *agr* mutant, and virtually all of these were detected at very low levels if at all in CM from the protease-deficient *agr* mutant (Supplementary Table S3). The abundance of an additional 426 proteins was decreased to a significant extent in CM from the *agr* mutant by comparison to CM from LAC (Supplementary Table S4), with the remaining 70 proteins being detected in CM from LAC and its *agr* mutant in comparable amounts (Supplementary Table S5). None of the 426 proteins that were present in decreased amounts in CM from the *agr* mutant were significantly increased in CM from the protease-deficient *agr* mutant. This demonstrates that the reduced abundance of proteins in the *agr* mutant is a function of limited production rather than protease-mediated degradation. Nevertheless, because *sarA* and *sarA/agr* mutants are attenuated in our osteomyelitis model owing at least in part to protease-mediated degradation [20], identifying proteins with reduced abundance in *sarA* and *sarA/agr* mutants remains important.

A total of 45 proteins were present in a significantly increased amount in CM from the protease-deficient *sarA/agr* mutant by comparison to the *sarA/agr* mutant itself, with an additional 182 being detected in CM from the protease-deficient *sarA/agr* mutant but not in CM from the *sarA/agr* mutant (Supplementary Table S6). Thus, the abundance of 227 proteins was limited to some degree by increased protease production in the *sarA/agr* mutant (Figure 3). The number of proteins limited by increased protease production in the *sarA* mutant was much greater, but this would be expected given the influence of *agr* on protein production and the observation that protease activity in the *sarA* mutant is higher than in the isogenic *sarA/agr* mutant [20].

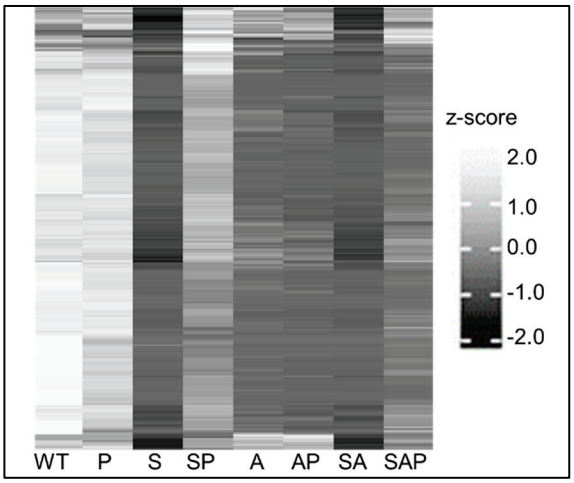


Figure 3. Heatmap reflecting strain dependent variation of prioritized *S. aureus* proteins. Average spectral counts from 5 biological replicates were determined for LAC (WT), isogenic derivatives with mutations in *sarA* (S), *agr* (A), *sarA/agr* (SA), and protease deficient derivatives of all these strains (P, SP, AP, and SAP, respectively). The relative abundance of the 224 proteins that were found to be present in limited amounts in *sarA* and *sarA/agr* mutants owing to protease mediated degradation are shown after transforming data to the Z-scale to reflect the relative differences of each protein to the same protein in all other samples.

Of the 227 proteins found to be limited by extracellular proteases in CM from the *sarA/agr* mutant, 3 were undetectable in CM from both the *sarA* mutant and its protease-deficient derivative (Supplementary Table S6), thus excluding them from the common pool of proteins present in reduced amounts in both *sarA* and *sarA/agr* mutants owing to protease-mediated degradation. Ten of the remaining 224 were present in a significantly increased amount in CM from the protease-deficient *sarA* mutant by comparison to CM from the *sarA* mutant, while 214 were detected in CM from the protease-deficient *sarA* mutant but not in CM from the *sarA* mutant itself (Supplementary Table S6). These results are consistent with the hypothesis that the increased production of extracellular proteases in *sarA* mutants limits the abundance of proteins that are potentially important in the pathogenesis of osteomyelitis, and that it does so irrespective of the functional status of *agr*, suggesting that these 224 proteins may be of particular interest in this regard.

Impact of eliminating protease production on the proteome of LAC. While our primary focus was on identifying proteins that are present in reduced amounts owing to increased protease production in *sarA* and *sarA/agr* mutants, it was previously reported that eliminating the ability produce extracellular proteases increases the virulence of LAC itself in a murine sepsis model [13]. Since it might be anticipated that this could be attributed to an increase in the abundance of some of the same virulence factors that were limited by protease production in *sarA* and *sarA/agr* mutants, we also identified proteins that were present in increased amounts in CM the protease-deficient derivative of LAC by comparison to CM from LAC itself. However, the only two proteins we identified were ArcB and MreC. MreC was among the proteins present in reduced amounts in CM from both the *sarA* and *sarA/agr* mutants owing to protease-mediated degradation, but ArcB was not (Supplementary Table S6).

Additionally, neither ArcB nor MerC were more abundant in CM from the protease-deficient derivative of LAC that was found to be hypervirulent in a murine sepsis model [13]. Moreover, none of the six proteins that were identified in this study (LukA, PSMα4, Sbi, SEK, SPIN, and a putative chitinase encoded by SAUSA300_0964) were identified in our proteomic comparisons of LAC and its protease-deficient derivative (Supplementary Tables S1 and S2). However, the abundance of all six was reduced in CM from

the *sarA* mutant, and other than Sbi the abundance of these proteins was increased to a statistically significant extent by eliminating the production of extracellular proteases (Figure 4). The abundance of Sbi, LukA, SAUSA300_0964, and PSMα4 was also reduced in CM from the *sarA/agr* mutant, although the reduced abundance of LukA and PSMα4 is likely due to the *agr* mutation (Table 1).

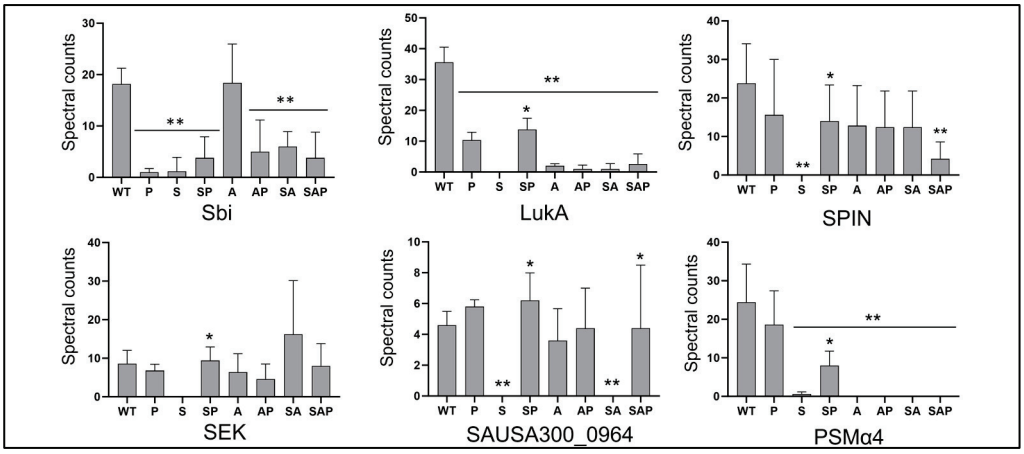


Figure 4. Targeted proteomic comparisons. The abundance of full-length versions of each of the six proteins identified by Gimza et al. [13] was determined using CM from LAC (WT), isogenic derivatives with mutations in *sarA* (S), *agr* (A), *sarA/agr* (SA), and protease deficient derivatives of all of these strains (P, SP, AP, and SAP, respectively). Double asterisks indicate a statistically significant reduction in spectral counts by comparison to CM from LAC as determined by one-way ANOVA. Single asterisks indicate a statistically significant increase in CM from protease deficient derivative of each strain by comparison to its isogenic regulatory mutant as determined by *t*-test.

Impact of *sarA* and *agr* on oxacillin resistance. We also found that the penicillin-binding proteins PBP1, PBP2, PBP2A, and PBP3 were detected in CM from LAC but not in CM from its isogenic *sarA* mutant, and that the abundance of all four PBPs was increased in CM from the protease-deficient *sarA* mutant (Figure 5). Mutation of *sarA* increased susceptibility to oxacillin and was increased even more in *sarA/agr* and *agr* mutants (Figure 6). However, eliminating the ability to produce extracellular proteases did not impact the susceptibility of any of these regulatory mutants. These results are consistent with a previous report concluding that the impact of *sarA* on oxacillin resistance is a function of decreased transcription of the gene (*mecA*) encoding PBP2A [33]. Irrespective of the mechanism(s) involved, this suggests that targeting *sarA* could have the added benefit of enhancing the ability to overcome the therapeutic recalcitrance of osteomyelitis, particularly when caused by methicillin-resistant strains.

Impact on factors involved in invasion of the osteocyte lacuna-canalicular network (OLCN). The reduced abundance of PBP3 may also be relevant in the pathogenesis of osteomyelitis because the absence of PBP3 has been correlated with a decreased ability of *S. aureus* to invade the OLCN [34]. Mutation of the gene encoding the primary *S. aureus* autolysin (*atl*) also limits OLCN invasion, and like PBP3 the abundance of Atl was reduced in CM from *sarA* and *sarA/agr* mutants. However, unlike PBP3, the abundance of Atl was restored to a statistically significant extent in CM from the protease-deficient *sarA* and *sarA/agr* mutants (Figure 7). Since invasion of the OLCN has been proposed to provide *S. aureus* with a protective niche in bone, this suggests another mechanism by which targeting *sarA* could enhance the ability to overcome the therapeutic recalcitrance of osteomyelitis to conventional antibiotic therapy.

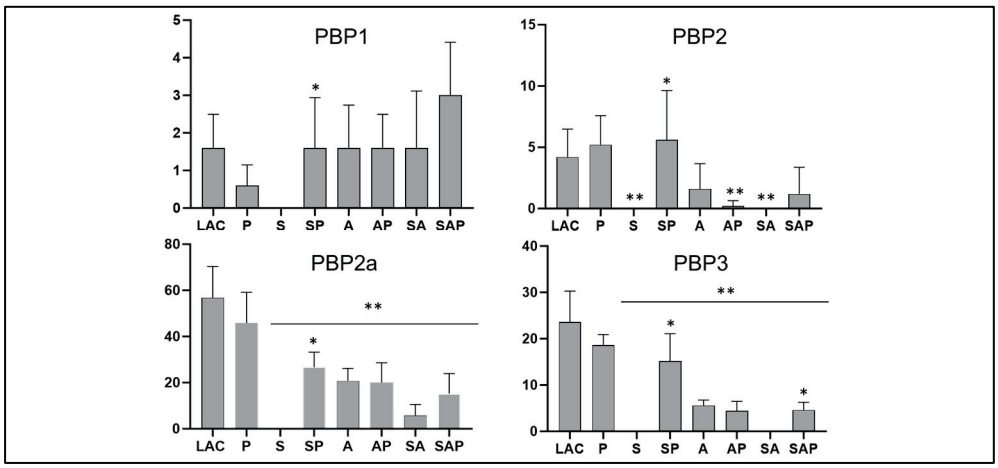


Figure 5. Targeted proteomic analysis of penicillin binding proteins (PBPs). The abundance of full-length PBPs was determined using CM from LAC (WT), its *sarA* (S), *agr* (A), and *sarA/agr* (SA) mutants, and protease deficient derivatives of all 4 strains. Double asterisks indicate a statistically significant reduction in spectral counts as determined by one-way ANOVA comparisons to CM from LAC. Single asterisk indicates a significant increase in the protease-deficient derivative by comparison to its isogenic regulatory mutant as determined by *t*-test.

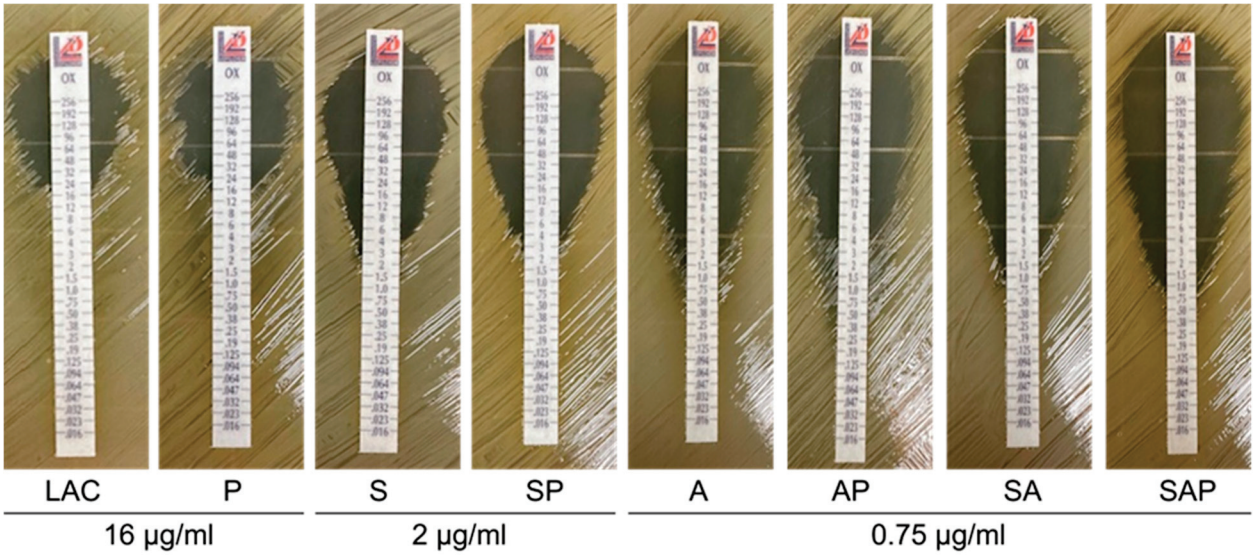


Figure 6. Impact of *sarA*, agr, and extracellular proteases on oxacillin susceptibility. Susceptibility was assessed by E-strip for LAC, isogenic *sarA* (S), *agr* (A), and *sarA/agr* (SA) mutants, and protease-deficient derivatives of all strains (P, SP, AP, and SAP, respectively). The approximate minimum inhibitory concentration shown below the figure was not affected by protease production in any strain.

Impact on Atl on exoprotein abundance. It was recently demonstrated that mutation of *atl* results in the increased abundance of extracellular LukAB owing to changes in peptidoglycan cleavage [35]. Atl was essentially undetectable in CM from the *sarA* mutant, so it might be anticipated that the phenotype of the *sarA* mutant would mimic the phenotype of an *atl* mutant. Since eliminating protease production restored the abundance of Atl in the *sarA* mutant, this leaves open the possibility changes in Atl rather than changes in protease production account for the increased abundance of at least some proteins in CM from the protease-deficient *sarA* mutant by comparison to CM from the *sarA* mutant itself. To test this, we used Western blots to examine the abundance of α-toxin, LukF, and LukS in CM from LAC, its *sarA* and protease-deficient *sarA* mutants, and an isogenic *atl* mutant. The

results confirmed our previous findings [20] with respect to the decreased abundance of these proteins in their full-length form in CM from the *sarA* mutant and their increased abundance relative to LAC in CM from the protease-deficient *sarA* mutant (Figure 8). They also demonstrated that the abundance of α-toxin and LukS, but not LukF, was increased relative to LAC in CM from the *atl* mutant. However, the abundance of all three exoproteins was still lower in CM from the *atl* mutant than CM from the protease-deficient *sarA* mutant (Figure 8).

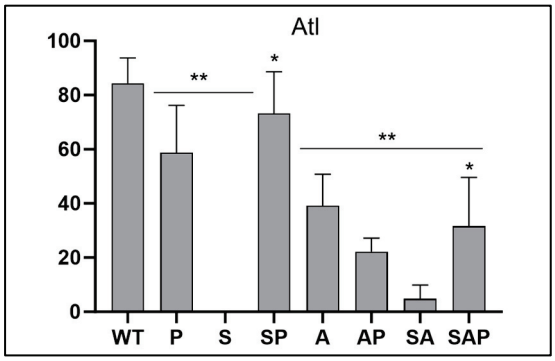


Figure 7. Proteomic comparison for autolysin (Atl). The abundance of full-length Atl was determined using CM from LAC (WT), its sarA (S), agr (A), and *sarA*/*agr* (SA) mutants, and protease-deficient derivatives of all four strains. Double asterisks indicate a statistically significant reduction in spectral counts as determined by one-way ANOVA comparisons to CM from LAC. Single asterisk indicates a significant increase in each protease-deficient strain by comparison to its isogenic regulatory mutant as determined by *t*-test.

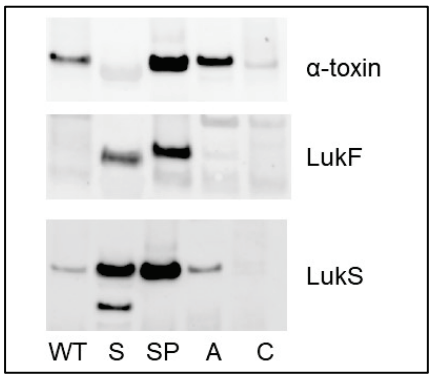


Figure 8. Impact of *atl* on the abundance of *S. aureus* exoproteins. Western blots were done with CM from overnight cultures of LAC and isogenic *sarA* (S), protease deficient *sarA* mutant (SP), an *atl* mutant (A). CM from an *hla* and *lukFS* mutant was included as a control (C).

5. Discussion

Staphylococcus aureus produces a diverse array of virulence factors and causes an equally diverse array of infections. This has complicated attempts to identify specific virulence factors that can be studied to gain a better understanding of pathogenesis mechanisms or perhaps even be targeted to prophylactic and/or therapeutic advantage. One alternative is to focus on regulatory elements that modulate the production of multiple *S. aureus* virulence factors, and the two regulatory targets that have been investigated most extensively in this regard are *sarA* and *agr* [6–10]. Although *sarA* is known to function through an agr-dependent pathway [11,12,36,37], the impact of mutating *sarA* and *agr* is often different. One example of specific relevance in this report is their opposite impact on the production of extracellular proteases [38]. Specifically, mutation of *agr* results in decreased protease production, while mutation of *sarA* has the opposite effect.

We have extensively investigated the impact of increased protease production in *S. aureus sarA* mutants and demonstrated that mutation of *sarA* results in a greater increase in protease production, and a greater decrease in biofilm formation, than mutation of any other *S. aureus* regulatory locus identified to date, and that this is true in diverse clinical isolates including both methicillin-sensitive and methicillin-resistant strains [23,27]. These collective studies led us to propose that the ability of *sarA* to repress the production of extracellular proteases is a key mechanism of post-translational regulation in *S. aureus* that prevents these proteases from compromising the *S. aureus* proteome [21–26]. However, none of these earlier studies considered the functional status of *agr*, which is an important consideration given that spontaneous *agr* mutants are known to arise in vivo and may in fact promote the transition between acute and chronic forms of *S. aureus* infection including osteomyelitis [39–42].

To address this, we used our osteomyelitis model to assess the relative virulence of sarA, agr, and sarA/agr mutants generated in the USA300 strain LAC, and the results confirmed that *sarA* and *sarA/agr* mutants are more attenuated in our osteomyelitis model than the isogenic *agr* mutant [20]. This suggests that *sarA* is not only a viable target for therapeutic intervention but is perhaps the preferred target by comparison to other regulatory loci including *agr*. The observation that the reduced virulence of both *sarA* and *sarA/agr* mutants can be attributed to a significant extent to the increased production of extracellular proteases also suggests that a proteomics approach could be used to go beyond the regulatory loci themselves to identify and prioritize specific virulence factors of interest based on a correlation between reduced virulence and the reduced abundance of these virulence factors. In fact, to the extent that our collective studies have also demonstrated that mutation of *sarA* results in the increased production of specific *S. aureus* virulence factors, but limited accumulation of these virulence factors in their full-length and presumably functional forms owing to protease-mediated degradation [20], it suggests that a proteomics approach capable of quantitatively distinguishing between full-length and truncated proteins is the only way to accomplish this important task.

Using this approach, we identified 224 full-length *S. aureus* proteins that are present in reduced amounts in CM from LAC *sarA* and *sarA/agr* mutants by comparison to CM from their isogenic protease-deficient derivatives. Many of these proteins were present at very low levels, and in fact most (214) were only identified because they were not detected in CM from *sarA* and *sarA/agr* mutants but were detected in CM from their protease-deficient derivatives. The absence of individual proteins in some samples precludes statistical analysis, but we included these among the 224 proteins of potential interest because it is not possible to know how much of a given protein is required to have a phenotypic effect.

The studies we report are an extension of previous studies that used the same mass-based proteomics approach to examine the correlation between protein abundance and virulence in our osteomyelitis model using six isogenic LAC mutants that differed in the functional status of *saeRS* and *sarA* relative to each other [21]. This was based on the demonstration that *saeRS* and *sarA* act synergistically to limit protease production [43] and that mutation of *saeRS* also limits virulence in an osteomyelitis model [21]. In fact, we demonstrated that LAC and its constitutively active *saeRS* variant (*sae*^{C}) are more virulent that isogenic *saeRS*, *sarA*, *saeRS/sarA*, and *sae*^{C}*/sarA* mutants, all of which are attenuated to a comparable degree. Proteomic analysis led to the identification of 114 proteins that were significantly more abundant in CM from both virulent strains by comparison to all 4 attenuated strains. Of these 114 proteins, 25 were among the 224 proteins identified in this study (Supplementary Table S6).

Since mutation of *saeRS* also results in increased protease production, albeit to a lesser degree than mutation of *sarA* [43], it might be anticipated that the overlap between these

two datasets would be larger. However, most of the 224 proteins we identified in this study were identified only because they were undetectable in CM from *sarA* and *sarA/agr* mutants and detectable in CM from their protease-deficient derivatives, and in our previous study we excluded proteins from further consideration if they were not detected in CM from all 6 strains [21]. Moreover, *sae* also functions in a coordinated fashion with *agr* to regulate the production of *S. aureus* virulence factors in response to different environmental signals [44]. This further illustrates the complex and highly interactive nature of *S. aureus* regulatory circuits and the critical need to maintain an appropriate balance between protein production and protein degradation. This also provides additional evidence supporting the hypothesis that limiting protease production is a key factor in the post-translational remodeling of the *S. aureus* virulence factor repertoire and suggests that the 25 proteins (Supplementary Table S6) that were identified in our collective studies based on correlations with virulence may warrant prioritization even among the 224 identified in this report.

In addition to *sarA* and *saeRS*, mutation of several other *S. aureus* regulatory proteins has been implicated in protease production, and this suggests that additional proteomic comparisons could further enhance the ability to prioritize specific targets for further study. To date, we have examined the impact of 21 *S. aureus* regulatory loci implicated in protease production and biofilm formation. The results demonstrate that mutation of *sarA* results in a greater increase in protease production than mutation of any of these other regulatory loci, but also suggest that *codY*, *sigB*, and *rot* warrant further consideration in this regard [22,25,27]. It is also important to extend such comparisons to other clinically relevant strains. Indeed, the methicillin-sensitive, USA200 strain UAMS-1 is virulent in our osteomyelitis model, and mutation of *sarA* attenuates this virulence owing to increased protease production, but LAC and UAMS-1 differ in significant ways beyond their methicillin-resistance status including their ability to produce important cytolytic including α-toxin and the Panton–Valentine leucocidin [23–25,32]. This suggests alternative pathogenesis pathways to the same end, thus making it important to prioritize based not only common proteins impacted by mutation of different regulatory loci but also on common proteins impacted in diverse clinical isolates of *S. aureus*.

Additionally, comparison of datasets obtained independently of each other can be informative, but direct comparisons using the same experimental and statistical methods are clearly preferred, particularly given the sensitivity of current proteomics technology. Indeed, while our primary focus was on identifying proteins that were present in reduced amounts in both *sarA* and *sarA/agr* mutants owing to protease-mediated degradation, we included CM samples from a protease-deficient derivative of LAC itself, and we did not identify any of the six proteins that Gimza et al. [13] found were more abundant in a protease-deficient derivative of LAC by comparison to LAC itself. This could potentially be explained by differences in the specific experimental approaches employed in these two studies, and this accounts for our efforts to fully validate our proteomics approach including its ability to distinguish between truncated and full-length proteins. We also employed a different model to assess virulence, and we did not observe a statistically significant increase in virulence with the protease-deficient derivative of LAC in the specific context of osteomyeltis, but we did demonstrate what we believe are clear trends suggesting increased virulence in the protease-deficient derivative of LAC [20]. Moreover, whether virulence was significantly increased in osteomyelitis vs. sepsis is irrelevant in the context of this discussion since both proteomic comparisons were performed using CM from stationary phase cultures grown in vitro in tryptic soy broth.

There were, however, other differences with respect to sample preparation and analysis that could impact the results. For instance, Gimza et al. [13] standardized cultures based on optical density before precipitating proteins and standardizing samples for com-

parative analysis based on overall protein content. In contrast, we examined CM after standardization based on optical density without further manipulation based on the logic that standardizing based on protein content might mask important differences owing to the global impact of extracellular proteases. Both studies also used a mass-based approach after resolution by SDS-PAGE, but we used gradient gels and focused on a narrower size range to distinguish between the abundance full-length and truncated proteins [21,24,26]. It is reasonable to suggest that such differences could impact the results, but it is difficult to envision how they can fully explain this discrepancy since we also did not identify any of these six proteins even after removing the size restriction and examining total abundance in any form. Thus, we do not have a definitive explanation for this discrepancy, but this nevertheless further illustrates the need to make direct comparisons between multiple strains and multiple isogenic mutants proven to exhibit significant differences in virulence, and to make these comparisons using consistent methods that have been validated to the greatest extent possible. Given that our proteomics analysis confirmed the identity of all mutants included in our comparisons, and also validated recognized phenotypes beyond the regulatory mutations themselves, we believe the experimental approach we describe is appropriate to address these issues moving forward.

Indeed, while we did not find that any of the six proteins identified by Gimza et al. [13] were more abundant in CM from the protease-deficient derivative of LAC by comparison to CM from LAC itself, we did find that the abundance of four of these six proteins were reduced in CM from the attenuated LAC *sarA* and *sarA*/*agr* mutants. In some cases, the reduced abundance of these proteins in CM from the *sarA/agr* mutant could not be attributed to increased protease production, but this does not detract from the possibility that the reduced abundance of these proteins in CM from *sarA* and *sarA/agr* mutants contributes to their reduced virulence. Indeed, this further illustrates the importance of *agr* in enhancing the production of *S. aureus* virulence factors, and the importance of *sarA* in concomitantly repressing the production of extracellular proteases to maintain a critical balance in the abundance of these virulence factors that is to the best advantage of *S. aureus* in specific in vivo environments including bone.

Identifying critical virulence factors is important because these factors contribute to in vivo survival and ultimately to the damage to the host associated with ongoing infection. To the extent that it compromises the local vasculature, this damage contributes to the therapeutic recalcitrance of *S. aureus* orthopedic infections to conventional antibiotic therapy. However, the limited abundance of these virulence factors is not the only relevant observation in defining this therapeutic recalcitrance. Indeed, penicillin-binding proteins including PBP2A were also identified in this report. This is consistent with previous reports demonstrating that mutation of *sarA* increases β-lactam susceptibility [33]. These reports concluded that the increased β-lactam susceptibility of *sarA* mutants was correlated with reduced *mecA* transcription, and our results demonstrating that the increased susceptibility to oxacillin observed in a *sarA* mutant is not impacted by increased protease production support this conclusion. There is also a report demonstrating that SarA binds the *mecA* promoter [45], further suggesting a protease-independent transcriptional mechanism. However, irrespective of the mechanism involved, it is important to note that mutation of *sarA* increased oxacillin susceptibility to a level comparable to the breakpoint minimum inhibitory concentration (MIC) for a methicillin-resistant strain ($\leq$2.0 μg/mL).

Additionally, PBP3 and the primary *S. aureus* autolysin Atl were also present in low levels in CM from *sarA* and *sarA*/*agr* mutants, and the abundance of both was significantly restored by eliminating protease production. The impact of protease production on the abundance of Atl and PBP3 may be particularly relevant to both the pathogenesis and therapeutic recalcitrance of *S. aureus* osteomyelitis. Specifically, mutation of *atl* and *pbp3* has

been shown to limit septic implant loosening and abscess formation in the medullary cavity, while mutation of *pbp3* has been correlated with reduced peri-implant osteolysis, reduced osteoclast activity, reduced invasion of the OLCN, and reduced production of the activator of nuclear factor kappa-B ligand (RANKL), a central modulator of the balance between osteoblast and osteoclast activity in bone remodeling [34,46]. Thus, these observations may be particularly relevant in the specific clinical context of *S. aureus* orthopedic infections.

Finally, one unexpected result we observed is the reduced abundance of SarA in CM from the *agr* and protease-deficient *agr* mutants (Table 1). It has been demonstrated that *sarA* is required for maximum expression of *agr*, but we are unaware of any reports suggesting that *agr* enhances expression of *sarA*. Our studies were performed using a null mutant that does not produce SarA, but different levels of SarA production in different strains has been correlated with relative levels of protease production [47]. Thus, if this were a true regulatory effect with a phenotypic impact, it would be anticipated that the limited production of SarA in the *agr* mutant would result in increased protease production, and this was not the case for any of the 10 extracellular proteases (Table 1). Based on the observation that the abundance of Atl was reduced in CM from the *agr* and protease-deficient *agr* mutants, one plausible explanation is that the reduced abundance of SarA in CM from *agr* mutants reflects changes in the abundance of Atl leading to changes in cell wall turnover and release of intracellular proteins, particularly since our comparisons were performed using CM from stationary phase cultures.

The abundance of SarA was comparable in CM from the *agr* and protease-deficient *agr* mutants, as was the abundance of Atl and SarA (Table 1). Atl abundance was also lower in *sarA* and *sarA/agr* mutants by comparison to the isogenic *agr* mutants, and in both mutants it was significantly increased in the isogenic protease-deficient strains. This leaves open the possibility that differences in the abundance of SarA and presumably other proteins in CM from *sarA* and *sarA/agr* mutants by comparison to their protease-deficient derivatives is a function of the abundance of Atl. Indeed, it was recently demonstrated that mutation of *atl* results in the increased abundance of extracellular LukAB owing to changes in peptidoglycan cleavage [34]. Additionally, Atl was essentially undetectable in CM from the *sarA* mutant, so it might be anticipated that the phenotype of the *sarA* mutant would mimic the phenotype of an *atl* mutant. LukA and LukB (SAUSA300_1975 and SAUSA300_1974, respectively) were undetectable in CM from *sarA* and *sarA/agr* mutants but readily detectable in CM from their protease-deficient derivatives (Supplementary Table S2). This is consistent with the hypothesis that the abundance of these proteins in CM from *sarA* mutants is limited directly by protease-mediated degradation, but it could also be explained at least to some degree by the limited abundance of Atl in *sarA* and *sarA/agr* mutants.

However, despite the reduced abundance of Atl in CM from the *sarA* and *sarA/agr* mutants by comparison to their protease-deficient derivatives, the abundance of LukF, LukS, and Nuc1 was increased to a comparable degree in both regulatory mutants irrespective of their ability to produce proteases, with the only protease-dependent difference being whether these proteins were present in truncated or full-length and presumably functional forms. Additionally, we confirmed by Western blot that, while mutation of *atl* does result in an increase in the abundance of some exoproteins (e.g., α-toxin and LukS), the abundance of these proteins, as well as LukF, remains below that observed in an isogenic protease-deficient *sarA* mutant. Thus, while we cannot exclude the possibility that changes in the abundance of Atl contribute to differences in the abundance of some of the 224 proteins defined as protease dependent by our comparisons, we do not believe they can account for our cumulative results to an extent that detracts from the conclusion that increased protease production plays a phenotype-defining role in *sarA* and *sarA/agr* mutants that

includes reduced virulence in the clinical context of osteomyelitis even if differences in the abundance of Atl, which are themselves protease dependent, do in fact contribute to the proteome phenotype of these mutants.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/microorganisms13010181/s1.

Author Contributions: Conceptualization, K.E.B.; Methodology, K.E.B., M.J.C., S.D.B. and M.S.S.; Formal analysis, K.E.B., S.D.B., S.G.M., R.D.E., A.J.T. and M.S.S.; Resources, S.D.B.; Data curation, K.E.B. and M.J.C.; Writing—original draft, M.S.S.; Writing—review & editing, K.E.B. and S.D.B.; Supervision, M.S.S.; Funding acquisition, M.S.S. All authors have read and agreed to the published version of the manuscript.

Funding: These experiments were supported by the National Institute of Allergy and Infectious Disease (NIAID), R01AI119380-06 (MSS), the National Institute of General Medical Sciences, NIGMS P30-GM145393 (MSS) and NIGMS R24 GM137786 (AJT), and a generous gift from the Texas Hip and Knee Center (MSS).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The authors confirm that the data supporting the findings of this study are available within this article and its Supplementary Materials.

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

References

1. Urish, K.L.; Cassat, J.E. Staphylococcus aureus osteomyelitis: Bone, bugs, and surgery. *Infect. Immun.* **2020**, *88*, e00932-19. [CrossRef]
2. Graan, D.; Balogh, Z.J. Microbiology of fracture related infections. *J. Orthop. Surg.* **2022**, *30*, 10225536221118512. [CrossRef]
3. Mirzaei, R.; Campoccia, D.; Ravaioli, S.; Arciola, C.R. Emerging issues and initial insights into bacterial biofilms: From orthopedic infection to metabolomics. *Antibiotics* **2024**, *13*, 184. [CrossRef] [PubMed]
4. Chen, Y.T.; Lohia, G.K.; Chen, S.; Riquelme, S.A. Immunometabolic regulation of bacterial infection, biofilms, and antibiotic susceptibility. *J. Innate Immun.* **2024**, *16*, 143–158. [CrossRef] [PubMed]
5. Bhattacharya, M.; Horswill, A.R. The role of human extracellular matrix proteins in defining *Staphylococcus aureus* biofilm infections. *FEMS Microbiol. Rev.* **2024**, *48*, fuae002. [CrossRef]
6. Ford, C.A.; Hurford, I.M.; Cassat, J.E. Antivirulence strategies for the treatment of *Staphylococcus aureus* infections: A mini review. *Front. Microbiol.* **2021**, *11*, 632706. [CrossRef]
7. Arya, R.; Princy, S.A. An insight into pleiotropic regulators Agr and Sar: Molecular probes paving the new way for antivirulent therapy. *Future Microbiol.* **2013**, *8*, 1339–1353. [CrossRef]
8. Cai, X.; Zheng, W.; Li, Z. High-throughput screening strategies for the development of anti-virulence inhibitors against Staphylococcus aureus. *Curr. Med. Chem.* **2019**, *26*, 2297–2312. [CrossRef]
9. Horswill, A.R.; Gordon, C.P. Structure-activity relationship studies of small molecule modulators of the staphylococcal accessory gene regulator. *J. Med. Chem.* **2020**, *63*, 2705–2730. [CrossRef] [PubMed]
10. Salam, A.M.; Quave, C.L. Targeting virulence in *Staphylococcus aureus* by chemical inhibition of the accessory gene regulator system in vivo. *mSphere* **2018**, *3*, e00500-17. [CrossRef]
11. Cheung, A.L.; Projan, S.J. Cloning and sequencing of *sarA* of *Staphylococcus aureus*, a gene required for the expression of *agr*. *J. Bacteriol.* **1994**, *176*, 4168–4172. [CrossRef] [PubMed]
12. Chien, Y.; Manna, A.C.; Cheung, A.L. SarA level is a determinant of *agr* activation in *Staphylococcus aureus*. *Mol. Microbiol.* **1998**, *30*, 991–1001. [CrossRef] [PubMed]
13. Gimza, B.D.; Jackson, J.K.; Frey, A.M.; Budny, B.G.; Chaput, D.; Rizzo, D.N.; Shaw, L.N. Unraveling the impact of secreted proteases on hypervirulence in *Staphylococcus aureus*. *mBio* **2021**, *12*, e03288-20. [CrossRef] [PubMed]
14. Kolar, S.L.; Ibarra, J.A.; Rivera, F.E.; Mootz, J.M.; Davenport, J.E.; Stevens, S.M.; Horswill, A.R.; Shaw, L.N. Extracellular proteases are key mediators of Staphylococcus aureus virulence via the global modulation of virulence-determinant stability. *MicrobiologyOpen* **2013**, *2*, 18–34. [CrossRef] [PubMed]
15. Kantyka, T.; Shaw, L.N.; Potempa, J. Papain-like proteases of *Staphylococcus aureus*. *Adv. Exp. Med. Biol.* **2011**, *712*, 1–14.

16. Krysko, O.; Teufelberger, A.; Van Nevel, S.; Krysko, D.V.; Bachert, C. Protease/antiprotease network in allergy: The role of *Staphylococcus aureus* protease-like proteins. *Allergy* **2019**, *74*, 2077–2086. [CrossRef]
17. Lehman, M.K.; Nuxoll, A.S.; Yamada, K.J.; Kielian, T.; Carson, S.D.; Fey, P.D. Protease-mediated growth of *Staphylococcus aureus* on host proteins is opp3 dependent. *mBio* **2019**, *10*, e02553-18. [CrossRef] [PubMed]
18. Pietrocola, G.; Nobile, G.; Rindi, S.; Speziale, P. *Staphylococcus aureus* manipulates innate immunity through own and host-expressed proteases. *Front. Cell Infect. Microbiol.* **2017**, *7*, 166. [CrossRef] [PubMed]
19. Tam, K.; Torres, V.J. *Staphylococcus aureus* secreted toxins and extracellular enzymes. *Microbiol. Spectr.* **2019**, *7*. [CrossRef]
20. Beenken, K.E.; Campbell, M.J.; Smeltzer, M.S. The ability of *sarA* to limit protease production plays a key role in the pathogenesis of *Staphylococcus aureus* osteomyelitis irrespective of the functional status of *agr*. *Infect. Immun.* **2024**, e0047324. [CrossRef]
21. Ramirez, A.M.; Byrum, S.D.; Beenken, K.E.; Washam, C.; Edmondson, R.D.; Mackintosh, S.G.; Spencer, H.J.; Tackett, A.J.; Smeltzer, M.S. Exploiting correlations between protein abundance and the functional status of *saeRS* and *sarA* to identify virulence factors of potential importance in the pathogenesis of *Staphylococcus aureus* Osteomyelitis. *ACS Infect. Dis.* **2020**, *6*, 237–249. [CrossRef] [PubMed]
22. Rom, J.S.; Atwood, D.N.; Beenken, K.E.; Meeker, D.G.; Loughran, A.J.; Spencer, H.J.; Lantz, T.L.; Smeltzer, M.S. Impact of Staphylococcus aureus regulatory mutations that modulate biofilm formation in the USA300 strain LAC on virulence in a murine bacteremia model. *Virulence* **2017**, *8*, 1776–1790. [CrossRef] [PubMed]
23. Ramirez, A.M.; Beenken, K.E.; Byrum, S.D.; Tackett, A.J.; Shaw, L.N.; Gimza, B.D.; Smeltzer, M.S. SarA plays a predominant role in controlling the production of extracellular proteases in the diverse clinical isolates of *Staphylococcus aureus* LAC and UAMS-1. *Virulence* **2020**, *11*, 1738–1762. [CrossRef] [PubMed]
24. Campbell, M.J.; Beenken, K.E.; Ramirez, A.M.; Smeltzer, M.S. Increased production of aureolysin and staphopain A is a primary determinant of the reduced virulence of *Staphylococcus aureus sarA* mutants in osteomyelitis. *mBio* **2024**, *15*, e0338323. [CrossRef]
25. Rom, J.S.; Beenken, K.E.; Ramirez, A.M.; Walker, C.M.; Echols, E.J.; Smeltzer, M.S. Limiting protease production plays a key role in the pathogenesis of the divergent clinical isolates of *Staphylococcus aureus* LAC and UAMS-1. *Virulence* **2021**, *12*, 584–600. [CrossRef]
26. Byrum, S.D.; Loughran, A.J.; Beenken, K.E.; Orr, L.M.; Storey, A.J.; Mackintosh, S.G.; Edmondson, R.D.; Tackett, A.J.; Smeltzer, M.S. Label-free proteomic approach to characterize protease-dependent and-independent effects of *sarA* inactivation on the *Staphylococcus aureus* exoproteome. *J. Proteome Res.* **2018**, *17*, 3384–3395. [CrossRef]
27. Atwood, D.N.; Loughran, A.J.; Courtney, A.P.; Anthony, A.C.; Meeker, D.G.; Spencer, H.J.; Gupta, R.K.; Lee, C.Y.; Beenken, K.E.; Smeltzer, M.S. Comparative impact of diverse regulatory loci on *Staphylococcus aureus* biofilm formation. *Microbiologyopen* **2015**, *4*, 436–451. [CrossRef]
28. Fey, P.D.; Endres, J.L.; Yajjala, V.K.; Widhelm, T.J.; Boissy, R.J.; Bose, J.L.; Bayles, K.W. A genetic resource for rapid and comprehensive phenotype screening of nonessential Staphylococcus aureus genes. *mBio* **2013**, *4*, e00537-12. [CrossRef]
29. Nesvizhskii, A.I.; Keller, A.; Kolker, E.; Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **2003**, *75*, 4646–4658. [CrossRef]
30. Chatterjee, S.S.; Chen, L.; Hwang-Soo, J.; Cheung, G.Y.C.; Kreiswirth, B.N.; Otto, M. Distribution and regulation of the mobile genetic element-encoded phenol-soluble modulin PSM-mec in methicillin-resistant *Staphylcoccus aureus*. *PLoS ONE* **2011**, *6*, e28781. [CrossRef]
31. Cassat, J.E.; Hammer, N.D.; Campbell, J.P.; Benson, M.A.; Perrien, D.S.; Mrak, L.N.; Smeltzer, M.S.; Torres, V.J.; Skaar, E.P. A secreted bacterial protease tailors the *Staphylococcus aureus* virulence repertoire to modulate bone remodeling during osteomyelitis. *Cell Host Microbe* **2013**, *13*, 759–772. [CrossRef] [PubMed]
32. Loughran, A.J.; Gaddy, D.; Beenken, K.E.; Meeker, D.G.; Morello, R.; Zhao, H.; Byrum, S.D.; Tackett, A.J.; Cassat, J.E.; Smeltzer, M.S. Impact of *sarA* and phenol-soluble modulins on the pathogenesis of osteomyelitis in diverse clinical isolates of *Staphylococcus aureus*. *Infect. Immun.* **2016**, *84*, 2586–2594. [CrossRef] [PubMed]
33. Liang, L.; Cheung, A.; Bayer, A.S.; Chen, L.; Abdelhady, W.; Kreiswirth, B.N.; Yeaman, M.R.; Xiong, Y.Q. The global regulon *sarA* regulates β-lactam antibiotic resistance in methicillin-resistant *Staphylococcus aureus* in vitro and in endovascular infections. *J. Infect. Dis.* **2016**, *214*, 1421–1429.
34. Masters, E.A.; Muthukrishnan, G.; Ho, L.; Gill, A.L.; de Mesy Bentley, K.L.; Galloway, C.A.; McGrath, J.L.; Awad, H.A.; Gill, S.R.; Schwarz, E.M. *Staphylococcus aureus* cell wall biosynthesis modulates bone invasion and osteomyelitis pathogenesis. *Front. Microbiol.* **2021**, *12*, 723498. [CrossRef]
35. Zheng, X.; Ma, S.X.; St John, A.; Torres, V.J. The major autolysin Atl regulates the virulence of *Staphylococcus aureus* by controlling the sorting of LukAB. *Infect. Immun.* **2022**, *90*, e0005622. [CrossRef]
36. Jenul, C.; Horswill, A.R. Regulation of *Staphylococcus aureus* virulence. *Microbiol. Spectr.* **2019**, *7*. [CrossRef]
37. Priest, N.K.; Rudkin, J.K.; Feil, E.J.; van den Elsen, J.M.; Cheung, A.; Peacock, S.J.; Laabei, M.; Lucks, D.A.; Recker, M.; Massey, R.C. From genotype to phenotype: Can systems biology be used to predict *Staphylococcus aureus* virulence? *Nat. Rev. Microbiol.* **2012**, *10*, 791–797. [CrossRef] [PubMed]

38. Beenken, K.E.; Mrak, L.N.; Griffin, L.M.; Zielinska, A.K.; Shaw, L.N.; Rice, K.C.; Horswill, A.R.; Bayles, K.W.; Smeltzer, M.S. Epistatic relationships between *sarA* and *agr* in *Staphylococcus aureus* biofilm formation. *PLoS ONE* **2010**, *5*, e10790. [CrossRef]
39. Gor, V.; Ohniwa, R.L.; Morikawa, K. No Change, No Life? What we know about phase variation in *Staphylococcus aureus*. *Microorganisms* **2021**, *9*, 244. [CrossRef] [PubMed]
40. Suligoy, C.M.; Díaz, R.E.; Gehrke, A.K.; Ring, N.; Yebra, G.; Alves, J.; Gómez, M.I.; Wendler, S.; Fitzgerald, J.R.; Tuchscherr, L.; et al. Acapsular *Staphylococcus aureus* with a non-functional *agr* regains capsule expression after passage through the bloodstream in a bacteremia mouse model. *Sci. Rep.* **2020**, *10*, 14108. [CrossRef]
41. Traber, K.E.; Lee, E.; Benson, S.; Corrigan, R.; Cantera, M.; Shopsin, B.; Novick, R.P. *agr* function in clinical *Staphylococcus aureus* isolates. *Microbiology* **2008**, *154*, 2265–2274. [CrossRef] [PubMed]
42. Yarwood, J.M.; Paquette, K.M.; Tikh, I.B.; Volper, E.M.; Greenberg, E.P. Generation of virulence factor variants in Staphylococcus aureus biofilms. *J. Bacteriol.* **2007**, *189*, 7961–7967. [CrossRef] [PubMed]
43. Mrak, L.N.; Zielinska, A.K.; Beenken, K.E.; Mrak, I.N.; Atwood, D.N.; Griffin, L.M.; Lee, C.Y.; Smeltzer, M.S. *saeRS* and *sarA* act synergistically to repress protease production and promote biofilm formation in *Staphylococcus aureus*. *PLoS ONE* **2012**, *7*, e38453. [CrossRef]
44. Novick, R.P.; Jiang, D. The staphylococcal *saeRS* system coordinates environmental signals with *agr* quorum sensing. *Microbiology* **2003**, *149*, 2709–2717. [CrossRef] [PubMed]
45. Kim, B.; Lee, H.J.; Jo, S.H.; Kim, M.G.; Lee, Y.; Lee, W.; Kim, W.; Joo, H.S.; Kim, Y.G.; Kim, J.S.; et al. Study of SarA by DNA Affinity Capture Assay (DACA) employing three promoters of key virulence and resistance genes in methicillin-resistant *Staphylococcus aureus*. *Antibiotics* **2022**, *11*, 1714. [CrossRef] [PubMed]
46. Campbell, M.J.; Bustamante-Gomez, C.; Fu, Q.; Beenken, K.E.; Reyes-Pardo, H.; Smeltzer, M.S.; O'Brien, C.A. RANKL-mediated osteoclast formation is required for bone loss in a murine model of Staphylococcus aureus osteomyelitis. *Bone* **2024**, *187*, 117181. [CrossRef]
47. Karlsson, A.; Arvidson, S. Variation in extracellular protease production among clinical isolates of *Staphylococcus aureus* due to different levels of expression of the protease repressor *sarA*. *Infect. Immun.* **2002**, *70*, 4239–4246. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

microorganisms

MDPI

Article

Functional Characterization, Genome Assembly, and Annotation of *Geobacillus* sp. G4 Isolated from a Geothermal Field in Tacna, Peru

Alonso R. Poma Ticona [1,*,†], Karita C. R. Santos [2,†], Heber E. Ramirez-Arua [3], Roberto Castellanos [1], Jéssica Pinheiro Silva [4], Pedro R. Vieira Hamann [4], Eliane F. Noronha [4] and Fabyano A. C. Lopes [2,*]

[1] Enzyme Biotechnology Research Laboratory, Science Faculty, Universidad Nacional Jorge Basadre Grohmann, Tacna 23003, Peru; rcastellanosc@unjbg.edu.pe
[2] Laboratório de Microbiologia, Universidade Federal Do Tocantins, Porto Nacional 77500-000, TO, Brazil; karita.cristine@uft.edu.br
[3] Laboratorio de Biología Molecular, Facultad de Farmacia y Bioquímica, Universidad Nacional Mayor de San Marcos, Lima 15081, Peru; heber.ramirez@unmsm.edu.pe
[4] Department of Cell Biology, Institute of Biological Sciences, University of Brasilia, Brasilia 70910-900, DF, Brazil; wpinheiro.jessica@gmail.com (J.P.S.); pedrorvhamann@gmail.com (P.R.V.H.); enoronha@unb.br (E.F.N.)
* Correspondence: arpomat@unjbg.edu.pe (A.R.P.T.); flopes@uft.edu.br (F.A.C.L.)
† These authors contributed equally to this work.

Abstract: The genome of *Geobacillus* sp. G4, a thermophilic bacterium isolated from a geothermal field in Peru, was sequenced and analyzed to evaluate its taxonomic and biotechnological potential. This strain exhibits optimal growth at temperatures between 50 and 70 °C and at a pH range of 6.0–7.5. Phenotypic assays demonstrated extracellular enzymatic activities, including amylases, cellulases, pectinases, and xylanases, highlighting its potential for efficient polysaccharide degradation. The assembled genome comprises approximately 3.4 Mb with a G+C content of 52.59%, containing 3,490 genes, including coding sequences, rRNAs, and tRNAs. Functional annotation revealed genes associated with key metabolic pathways such as glycogen and trehalose biosynthesis, indicating adaptation to carbohydrate-rich environments. Phylogenetic analyses based on ANI and dDDH values identified *Geobacillus thermoleovorans* KCTC3570 as its closest relative, suggesting a strong evolutionary relationship. Additionally, the genome harbors gene clusters for secondary metabolites such as betalactone and fengycin, suggesting potential industrial and pharmaceutical applications, including bioremediation. The identification of antibiotic resistance genes, specifically those conferring glycopeptide resistance, underscores their relevance for antimicrobial resistance studies. The presence of enzymes like amylases and pullulanase further emphasizes its biotechnological potential, particularly in starch hydrolysis and biofuel production. Overall, this research highlights the significant potential of *Geobacillus* species as valuable sources of thermostable enzymes and biosynthetic pathways for industrial applications.

Keywords: bacterial thermophily; biotechnological potential; industrial enzymes; phylogenomic analysis

1. Introduction

The genus *Geobacillus* comprises thermophilic, Gram-positive, spore-forming, rod-shaped bacteria that are either aerobic or facultatively anaerobic, constituting a phylogenetically coherent clade within the family Bacillaceae [1]. These microorganisms inhabit a

wide range of environments, including hot springs, cold regions, and ocean sediments [2]. The genus *Geobacillus* was proposed by Nazina et al. (2001) [3] to differentiate it from the genus *Bacillus* based on distinct physiological traits, variations in growth temperatures, and phenotypic characteristics among strains. *Geobacillus* species exhibit broad catabolic versatility, facilitating their widespread occurrence across diverse environments. These bacteria are recognized for secreting commercially important enzymes, including lipases, xylanases, and amylases, widely utilized in various industrial and biotechnological processes.

According to the List of Prokaryotic Names with Standing in Nomenclature (LPSN, 2025), the genus *Geobacillus* currently includes 30 valid species, including *Geobacillus thermodenitrificans*, *Geobacillus uzenensis*, *Geobacillus thermocatenulatus*, *Geobacillus kaustophillus*, *Geobacillus thermoglucosidasius*, *Geobacillus thermoleovorans*, *Geobacillus subterraneus*, *Geobacillus stearothermophilus*, *Geobacillus toebii*, *Geobacillus jurassicus*, *Geobacillus thermoglucosidans*, *Geobacillus stearothermophilus*, and *Geobacillus thermodenitrificans.* *Geobacillus stearothermophilus* serves as the type species, with DSM 22 designated as its type strain [1,4].

Analysis based on the 16S rRNA gene revealed a cohesive grouping of 56 *Geobacillus* strains, exhibiting intra-species similarities between 99.3% and 100%, and inter-species similarities ranging from 96.0% to 99.4%, facilitating reliable distinctions [5]. However, limitations in taxonomic resolution using the 16S rRNA gene led researchers to adopt additional genetic markers, such as recN, infB, rpoB, and spo0A, providing enhanced phylogenetic clarity [6,7].

Distinctions between *Geobacillus* species in previous studies have been based on small chemotaxonomic differences and phenotypic characteristics [5]. Growth temperatures range from 35 °C to 80 °C, with most isolated strains growing between 45 °C and 70 °C [3]. The core genome of *Geobacillus* comprises approximately 1665 genes, 85% of which are shared with the genome of *Bacillus subtilis* 168, suggesting a common survival strategy centered around rapid growth and sporulation under nutrient-limited conditions [8].

Traditionally, bacterial classification has relied on methods such as 16S rRNA sequencing, DNA-DNA hybridization (DDH), and phenotypic characterization. Nonetheless, high sequence similarity within the 16S rRNA gene among *Geobacillus* species complicates precise identification and has resulted in misidentifications [9]. The growing number of studies classifying new species using only single-pathway markers has led to the rise of unreliable and confusing classifications within the genus [3,10–12]. To overcome these challenges, integrated approaches combining genomic and phenotypic analyses, including whole-genome sequencing, have been proposed to achieve higher phylogenetic resolution.

The Calientes geothermal field in Tacna, Peru, represents an underexplored source of microbial biodiversity with significant biotechnological potential [13,14]. Characterized by elevated temperatures and unique mineral compositions, these geothermal springs provide optimal conditions for thermophilic bacteria like *Geobacillus* [15,16]. These microorganisms have attracted considerable interest due to their ability to produce thermostable enzymes applicable in biotechnology, particularly in industrial processes involving the bioconversion and biodegradation of complex organic polymers such as lignocellulosic biomass (e.g., cellulose, hemicellulose, and lignin) [17].

Geobacillus species are highly valuable biotechnological resources due to their capability of producing stable and active enzymes under extreme environmental conditions, such as high temperatures and variable pH levels. These enzymes have found extensive industrial applications, notably in biofuel production and the bioconversion of industrial waste [18]. Moreover, certain *Geobacillus* strains efficiently degrade complex organic polymers like lignin and cellulose, offering promising opportunities for bioethanol production from renewable biomass [19]. Ongoing genomic and metabolic engineering research efforts

further enhance the biotechnological potential of these bacteria, optimizing their utilization in industrial applications [20].

Geobacillus species have attracted considerable attention not only due to their capacity to produce thermostable enzymes but also because of their potential to biosynthesize secondary metabolites with diverse industrial and pharmaceutical applications. These metabolites include compounds like fengycin, known for its antimicrobial and antifungal properties, valuable in bioremediation and biological control applications, and lanthipeptides or betalactones, which exhibit potential therapeutic properties, including antibiotic and anticancer activities, relevant to pharmaceutical development [21]. The identification and characterization of secondary metabolite biosynthetic gene clusters (BGCs) in thermophilic bacteria can reveal novel compounds beneficial for biotechnological and biomedical use [17,18].

The study of genomic diversity and characterization of new *Geobacillus* strains is essential from both a taxonomic and biotechnological perspective. Genomic analysis of new strains allows for the identification of genes involved in specific metabolic processes, such as the degradation of starch and hemicellulose, which are crucial for biomass conversion into biofuels and other value-added products [22,23].

In this study, we aimed to characterize *Geobacillus* sp. G4, a thermophilic strain isolated from a geothermal environment in southern Peru, through phenotypic and enzymatic analyses. Furthermore, we sought to sequence, assemble, and annotate its genome; determine its taxonomic position; and identify genes and metabolic pathways with potential biotechnological relevance, particularly those involved in the production of thermostable enzymes and the degradation of complex carbohydrates for industrial applications.

2. Materials and Methods

2.1. Bacterium Isolation and Characterization of Geobacillus sp. *G4*

Sediment samples were obtained from a geothermal spring in the Calientes geothermal field, Candarave, Tacna, Peru. The spring, designated as G8, was selected based on the presence of sediment, elevated temperatures (>60 °C), and accessibility. The G8 spring, with a water temperature of 65.3 °C and a pH of 7.35, is located at an altitude of 4162 m above sea level (UTM coordinates: 0.379908 E, 81.09185 N). The area surrounding the spring was characterized by sparse vegetation and the accumulation of brownish sediments (Figure 1). Approximately 250 g of sediment was collected from the upper 3 cm layer of the sediment bed, at a depth of up to 15 cm from the water surface, at the edge of the geothermal spring using a sterile plastic bottle. A suspension of sediment and spring water was transferred into a sterile thermos bottle to maintain a temperature close to that of the sample during both sampling and transport to preserve microbial viability.

A sediment sample (10 g) was initially enriched in Luria–Bertani (LB) liquid medium (consisting of 10 g/L peptone, 5 g/L yeast extract, and 10 g/L sodium chloride) at a 10% (v/v) concentration. The enrichment culture was incubated at 60 °C for 48 h. After incubation, 100 µL of the enriched culture (OD 600 nm = 0.2) was spread onto LB agar plates supplemented with 1% (w/v) starch as the sole carbon source to selectively isolate amylolytic bacteria. The plates were incubated at 60 °C until bacterial colonies appeared. After two days, bacterial colonies were individually collected and streaked three times on the same LB agar to ensure purity. Five amylolytic colonies (G1, G2, G4, G16, and G23) were isolated, and the strain G4, which exhibited the highest amylase activity levels, was chosen for further characterization and genomic analysis. Amylase activity was quantified using the Miller method [24]: 100 µL of starch solution (0.5% w/v) and 100 µL of G4 culture supernatant were incubated at 60 °C for 15 min. Then, 300 µL of DNS reagent was added, and absorbance was measured at 540 nm.

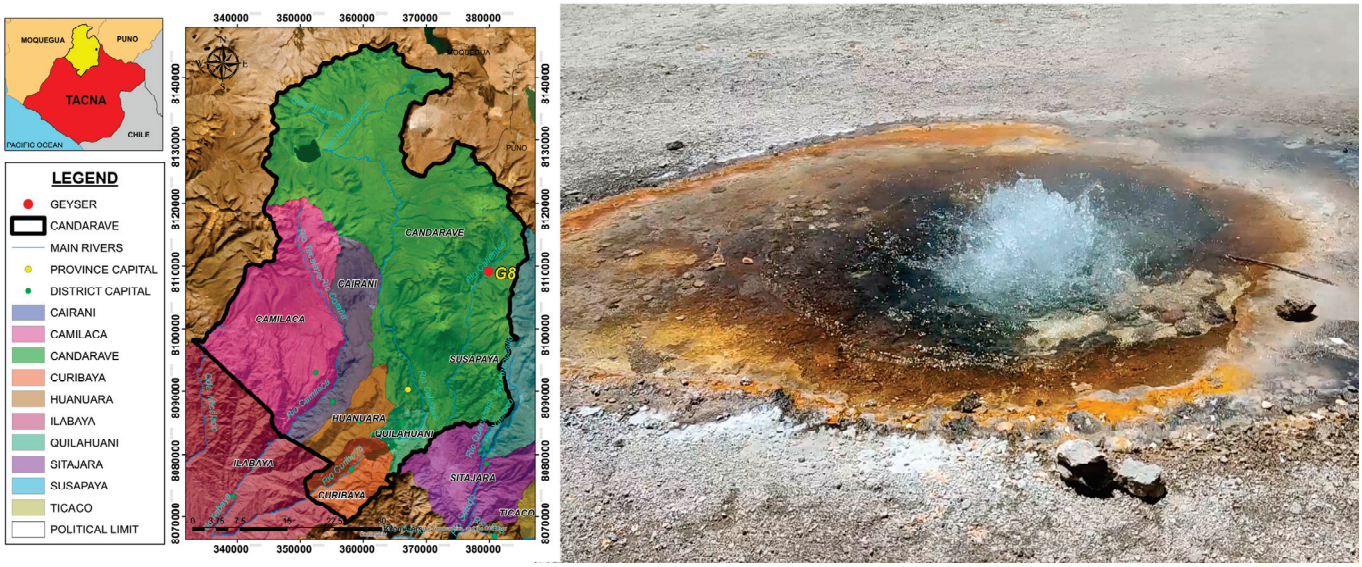


Figure 1. Map of the geothermal sampling area in Calientes, Candarave, using ArcGIS version 10.5. Image of the geothermal spring sampled in this study, located in the Calientes geothermal field, Candarave, Tacna, Peru.

The morphology of the G4 strain was evaluated by analyzing colony shape, pigmentation, and microscopic characteristics using Gram staining, endospore formation assays, and motility tests. The effect of temperature on the growth of strain G4 was assessed by culturing it in LB liquid medium (pH 7) at a range of temperatures (37, 45, 50, 55, 60, 65, 70, 75, and 80 °C) for 48 h. Similarly, the effect of pH on growth was evaluated by incubating G4 in LB medium buffered at pH values ranging from 4 to 8 (100 mM sodium citrate for pH 4 and 5, 100 mM sodium phosphate for pH 6 and 7, and 100 mM Tris-HCl for pH 8) at a fixed temperature of 60 °C for 48 h. For acid production assays, substrates such as lactose, cellobiose, glycerol, sucrose, and xylose were used, following the methodology described by Mac Faddin (2000) [25].

Carbon utilization of strain G4 was determined by culturing it in LB liquid medium (pH 7) supplemented with various substrates, including carboxymethylcellulose (CMC), starch, pectin, microcrystalline cellulose, casein, lactose, xylose, glucose, sucrose, cellobiose, and maltose, each at 1% (w/v) as a carbon source (Sigma Aldrich, St. Louis, MO, USA). Cultures were incubated at 60 °C for 48 h. Bacterial growth was monitored by measuring optical density at 600 nm.

The extracellular hydrolytic enzyme activities, including cellulase, pectinase, xylanase, amylase, protease, and lipase, were assessed. The G4 isolate was cultivated on LB agar plates supplemented individually with specific substrates: carboxymethylcellulose (CMC) 1% (w/v), pectin 0.5% (w/v), oat spelt xylan 1% (w/v), potato starch 1% (w/v), casein 1% (w/v), and tributyrin 0.5% (v/v), following the protocol described by Guta et al. (2024) with minor modifications [26]. Plates were incubated at 60 °C for 24–48 h. The presence of clear zones surrounding colonies indicated positive enzymatic activity.

To analyze enzyme production by isolate G4, time-course experiments were performed. G4 was inoculated into 250 mL Erlenmeyer flasks containing 50 mL of LB liquid medium at pH 7.0, individually supplemented with the enzyme-specific substrates mentioned above. Cultures were incubated at 60 °C, and 2 mL samples were withdrawn at intervals of 3, 6, 12, 24, 30, and 48 h. Samples were centrifuged at 10,000× g for 10 min at 4 °C, and the resulting supernatants were used for enzyme assays.

The enzymatic activities of cellulase, pectinase, xylanase, and amylase were determined using their respective substrates buffered with 100 mM sodium phosphate (pH 7.0),

quantifying the reducing sugars released according to the Ticona method [27]. One unit of enzymatic activity was defined as the amount of enzyme required to release 1 µmol of reducing sugar per minute.

Protease activity was assessed by mixing 25 µL of enzyme solution with 130 µL of 0.5% (*w*/*v*) casein buffered in 100 mM Tris-acetate (pH 7.5), following the methodology described by Oluoch et al. (2018) [28]. One unit of protease activity was defined as the amount of enzyme that released 1 µmol of tyrosine per minute. Lipase activity was measured titrimetrically using an olive oil emulsion as the substrate, as described by Sharma et al. (2017) [29]. One unit of lipase activity was defined as the enzyme amount releasing 1 µmol of fatty acid per minute under assay conditions.

2.2. DNA Extraction and Sequencing

Genomic DNA was extracted from bacterial strain G4 for whole-genome sequencing. Strain G4 was first cultured in 10 mL of Luria–Bertani (LB) broth and incubated at 60 °C for 48 h. The culture was centrifuged at 8000× *g* for 15 min at 4 °C, and the supernatant was discarded. The resulting cell pellet was resuspended in 5 mL of Tris-HCl buffer (pH 8.0). Genomic DNA was isolated using the Bacterial Genomic DNA Isolation Kit (Norgen Biotek Corp., Thorold, ON, Canada) following the manufacturer's protocol. The concentration and purity of the DNA were measured using a BioTek Epoch 2T spectrophotometer (Agilent Technologies, Inc., Santa Clara, CA, USA).

DNA libraries were prepared using the Illumina TruSeq Nano DNA Library Prep Kit (350), and sequencing was performed using the Illumina NovaSeq platform with paired-end 150 bp reads at Macrogen Inc. (Seoul, Republic of Korea).

2.3. Quality Analysis, Genome Assembly, and Annotation of Geobacillus sp. *G4*

The raw fastq files (paired-end) comprising 13,778,078 sequences with an average length of 151 bp were imported into the FastQC v0.12.1 tool for quality assessment. The quality filtering was performed using Trimmomatic v0.38 to remove sequences with low-quality scores, applying the following parameters: SLIDINGWINDOW (4:30) and PHRED 30. The remaining sequences were then imported into the SPAdes v3.15.4 tool for genome assembly using the parameters k-mer (21, 33, 55, and 77) and a Phred score of 64 (Illumina). Short contigs (<200 bp) were removed from the assembly using seqtk v.1.4 (https://github.com/lh3/seqtk, accessed on 13 November 2024). The completeness and contamination of the genome were analyzed using the CheckM v1.2.2 tool and CoverM v0.7.0 (https://github.com/wwood/CoverM, accessed on 13 November 2024), while the size and quality of the assembled genome were assessed using Quast v5.2.0 and Busco v5.5.0 [30–32].

Functional genome annotation was performed in the Galaxy (https://usegalaxy.org/) environment using Prokka v1.14.6 with default parameters. Prokka predicted CDSs, rRNAs, and tRNAs, assigning functions initially through BLAST+ searches against core databases (ISfinder, NCBI AMR, and UniProtKB/SwissProt), followed by refinement using HMMER3 [33–37]. After annotation, the genome was mapped according to its functional categories and enzyme classes using the online mapper eggNOG-mapper v2 (eggNOG-mapper) with the eggNOG 5.0 database [38]. Type strains from the most similar species, based on the results from the Digital DNA-DNA hybridization (dDDH) analysis, were used for comparison with the literature.

To analyze and highlight the biotechnological potential of the *Geobacillus* sp. G4 strain, three metabolic pathways were highlighted in the KEGG annotation: glycogen biosynthesis (M00854), glycogen degradation (M00855), and trehalose biosynthesis (M00565) (Supplementary Figure S1). The visualization of the pathways of interest was performed

using the KEGG mapper (www.genome.jp/kegg/mapper, accessed on 20 November 2024) via the KEGG database.

To identify gene clusters related to secondary metabolism (BCGs) in microorganisms, AntiSMASH v7.1.0 software was used [39]. The Antibiotic Resistance Genes Database (ARDB) was used to select and annotate genes potentially responsible for antibiotic resistance [40].

The genome of *Geobacillus* sp. G4, along with three reference genomes (*G. kaustophilus* HTA426, *G. kaustophilus* NBRC12445, and *G. thermoleovorans* KCTC3570), was functionally annotated using the eggNOG-mapper online tool (http://eggnog-mapper.embl.de/, accessed on 6 January 2025). Genes related to four functional categories of interest—carbon starvation, DNA repair and supercoiling, heat-shock response, and oxidative stress—were selected based on their presence in the genomes. Gene presence/absence analyses and graphical visualization were performed using R statistical software (https://www.r-project.org/).

2.4. Taxonomic Classification of Geobacillus sp. *G4*

2.4.1. Digital DNA Hybridization (dDDH) and Average Nucleotide Identity (ANI)

Digital DNA hybridization (dDDH) was performed on the TYGS platform—DSMZ (tygs.dsmz.de) for the taxonomic classification of the strain and identity confirmation [41]. For phylogenomic inference to compare complete genomes in dDDH, the GBDP method (Genome BLAST Distance Phylogeny) was used, employing the trimming algorithm, formula d5, and bootstrap 100 [42]. To calculate genomic similarity using DDQ (Digital DNA-DNA Hybridization), GGDC v4.0 software integrated into the TYGS platform was used. The average nucleotide identity (ANI) analysis was calculated using the JSpeciesWS platform (https://jspecies.ribohost.com/), which provides ANIb, representing the average nucleotide identity based on BLAST calculations [43], and ANIm, representing the average nucleotide identity based on MUMmer calculations [43].

2.4.2. 16S rRNA Gene-Based Phylogenetic Tree Reconstruction

16S rRNA gene sequences from the genomes were extracted using barnnap v0.9 (https://github.com/tseemann/barrnap, accessed on 11 January 2025). All information regarding the sequences used can be found in Supplementary Table S1. Sequence alignment was performed using MAFFT v7.526 in default mode [44]. After alignment, all sequences were trimmed using TrimAl v1.4.1 [45]. The selection of the most appropriate model for tree reconstruction was performed using ModelTest-NG v0.1.7 [46], suggesting the TIM1+I model. The phylogenetic tree was constructed using the Maximum Likelihood method [47], utilizing RAxML-NG X v1.2.2 [48] with the following parameters: 10 random initial parsimony trees, nucleotide substitution matrix model (TIM1+I), and 1000 bootstrap replications. The *Bacillus subtilis* DSM10 strain was used as an outgroup.

2.4.3. Phylogenetic Tree Reconstruction

The genomes used were obtained from complete genomes deposited in GenBank. The selection of type strains for valid *Geobacillus* and *Bacillus* species used was based on the LPSN (List of Prokaryotic names with Standing in Nomenclature—lpsn.dsmz.de) [41,49]. All information regarding the complete genomes used can be found in Supplementary Table S2.

Two phylogenetic trees were constructed using the studied genome. The first tree was reconstructed using molecular markers extracted from the genomes using AMPHORA2 [50], and data on the extracted markers are available in Supplementary Table S3. Sequence alignment was performed using the MAFFT v7.526 aligner in default mode [44]. After alignment, all sequences were trimmed using TrimAl v1.4.1 [45]. The selection of the most appropriate model for tree reconstruction was performed using ModelTest-NG v0.1 [46], with the LG+I+G4+F model suggested. The phylogenetic tree was constructed using the

Maximum Likelihood method [47], utilizing RAxML-NG X v1.2.2 [48] with the following parameters: 10 random initial parsimony trees, nucleotide substitution matrix model (LG+I+G4+F), and 1000 bootstrap replications. The *Bacillus subtilis* DSM10 strain was used as an outgroup.

The second tree was reconstructed using the FastME v2.1.6.1 tool [51] with the GBDP method (Genome BLAST Distance Phylogeny) integrated into the TYGS platform (tygs.dsmz.de). The clade lengths are scaled in terms of GBDP distance formula d5, and 100 pseudo-bootstrap replicates were performed. Only pseudo-bootstrap support values above 60% were accepted. The tree was rooted at the midpoint [52]. The *Bacillus subtilis* DSM10 strain was used as an outgroup.

For genome comparison, the online platform Proksee (proksee.ca) was used [53]. Comparisons were made using the genome of the type strains *G. thermoleovorans* KCTC3570 and *G. kaustophilus* NBRC102445 in comparison with the *Geobacillus* sp. G4 strain. BLAST v1.3.1 and BLAST Formatter v1.0.3 tools were used to check identity and similarity profiles between the genomes [42].

3. Results

3.1. Characterization of Geobacillus sp. *G4*

The *Geobacillus* sp. G4 strain is a Gram-positive, rod-shaped bacterium forming round, white-yellow colonies with an average diameter of 1.4 mm (Supplementary Table S4). Characterization revealed that *Geobacillus* sp. G4 is aerobic, non-motile, and forms endospores. Optimal growth conditions (50 °C to 70 °C and pH 6.0 to 7.5) were determined in LB liquid medium based on maximal bacterial growth, as measured by optical density at 600 nm. Additionally, acid production from lactose, cellobiose, glycerol, sucrose, and xylose was observed (Table 1).

Table 1. Phenotypic characteristics for differentiation among *G. kaustophilus*, *G. thermoleovorans*, and *Geobacillus* sp. G4.

Phenotypic Characteristics	***Geobacillus* sp. G4 ***	***G. kaustophilus* ATCC 8005 ****	***G. thermoleovorans* ATCC 43513 ****
Gram Staining	Positive	Positive	Positive
Motility	Non-motile	Non-motile	Non-motile
Morphology	Rod shape	Rod shape	Rod shape
Oxygen Requirement	Aerobic	Aerobic	Aerobic
G+C (%)	52	51–58	52–58
Endospore Formation	+	+	+
Voges-Proskauer	-	-	-
Methyl red	+	+	no data
Catalase	+	+	+
Oxidase	+	+	+
Hydrolysis of:			
Casein	+	+	Variable
Starch	+	+	+
Acid from:			
Lactose	+	-	+
Cellobiose	+	+	Variable
Glycerol	+	+	+
Sucrose	+	+	Variable
Xylose	+	+	Variable
Growth at/in:			
pH range	6.0–7.5	6.2–7.5	5.0–9.0

Table 1. *Cont.*

Phenotypic Characteristics	*Geobacillus* sp. G4 *	*G. kaustophilus* ATCC 8005 **	*G. thermoleovorans* ATCC 43513 **
NaCl (%, *w*/*v*)	0–2.5	0–2	0–4
Temperature range (°C)	50–70	40–75	37–70
Acetate	+	+	+
Citrate	-	+	-

* Unpublished data; ** from Najar and Thakur (2020) [1].

The ability of isolate G4 to produce extracellular hydrolytic enzymes was evaluated on agar plates. After incubation at 60 °C for 24–48 h, isolate G4 demonstrated positive enzymatic activities for cellulase, pectinase, xylanase, amylase, protease, and lipase, indicated by distinct clear zones around colonies. Among these, *Geobacillus* sp. G4 showed the largest halo formation for amylase (starch), followed by pectinase, xylanase, and cellulase (Figure 2A).

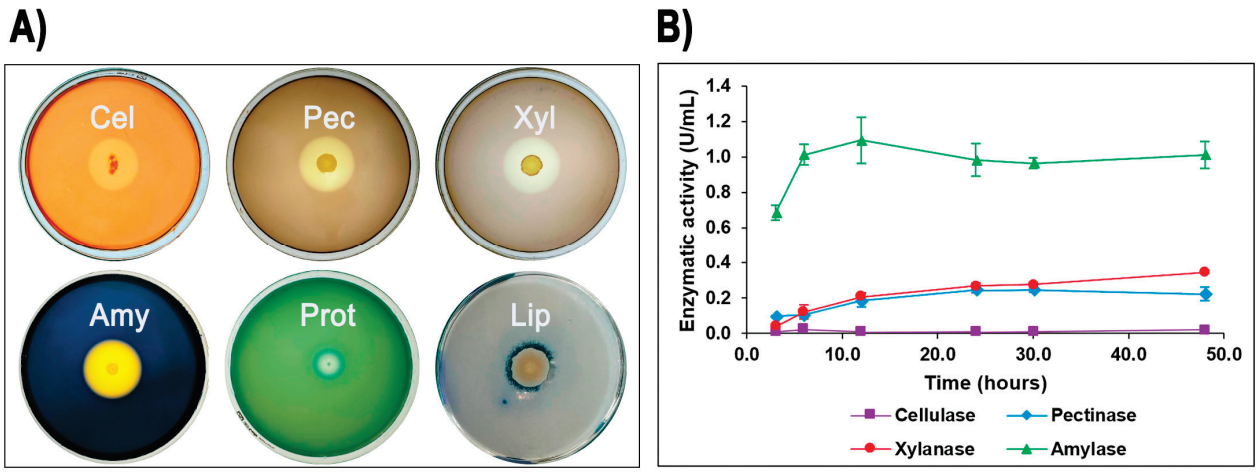


Figure 2. Quantitative and qualitative enzymatic activity of *Geobacillus* sp. G4. Extracellular enzymatic activities of *Geobacillus* sp. G4. (**A**) Formation of clear halos on agar plates indicating positive extracellular enzyme production: Amy (amylase), Pec (pectinase), Xyl (xylanase), Cel (cellulase), Prot (protease), and Lip (lipase). (**B**) Enzyme production profiles measured in LB liquid media over time.

To further quantify these enzymatic activities, isolate G4 was cultured in liquid media containing potato starch, oat spelt xylan, pectin, carboxymethylcellulose (CMC), olive oil, and casein as carbon sources. Enzyme assays revealed measurable production of amylase (maximum activity: 1.09 U/mL at 12 h), xylanase (0.34 U/mL at 48 h), and pectinase (0.24 U/mL at 30 h), as well as low but detectable cellulase activity (0.02 U/mL at 48 h) (Figure 2B). In contrast, protease and lipase activities, although clearly evident in agar plate assays, were below detection limits in liquid media under the tested conditions.

3.2. Genome Description and Taxonomic Classification of Geobacillus sp. *G4*

The results regarding the preliminary genome draft of *Geobacillus* sp. G4 is presented in Table 2. The strain genome has a size of approximately 3.4 Mb with a coverage of ~1225x. The assembly comprised 164 contigs, with a total G+C content of 52.59%. The genome completeness and contamination were 99.45% and 0.95%, respectively. The N50 and L50 values for the assembly were 236,119 bp and 5 contigs, respectively. The annotation identified a total of 3490 genes: 2047 coding sequences (CDS), 5 rRNAs, 88 tRNAs, 1 transfer-messenger RNA, and 1349 putative proteins. Regarding the gene clusters related

to secondary metabolism (BCGs), seven clusters were identified: betalactone, T3PKS, Lanthipeptide class-i, NI-siderophore, terpene, fengycin, and RiPP-like.

Table 2. Genome assembly data for *Geobacillus* sp. G4 was performed using SPAdes v.3.15.4. The annotation was conducted using Prokka v.1.14.6, and quality parameters were assessed using Busco v.5.5.0 and Quast v.5.2.0. All processes were carried out on the Galaxy v24.1.2 server.

General Data	Assembly De Novo *Geobacillus* sp. G4	Reference Genome *G. kaustophilus* HTA426	Reference Genome *G. thermoleovorans* KCTC 3570	Reference Genome *G. kaustophilus* NBRC102445
Contigs	164	2	2	1
Total length	3,362,411	3,544,776	3,450,609	3,670,957
N50 (KB)	236,119	3,716,831	1,192,538	750,728
L50	5	1	2	1
GC (%)	52.59	52.35	52.50	51.73
Coverage	1224.95	-	224.2	150.5
Completeness (%)	99.45	99.31	99.44	93.81
Contamination (%)	0.95	0.22	30.89	0.24
CDS	2047	5668	4714	5548
rRNA	5	37	32	35
tRNA	88	129	100	111
tmRNA	1	2	1	3
Genes	2047	3655	3622	3734
Putative protein	1349	3386	3369	3401

The nucleotide average identity (ANI) values between the *Geobacillus* sp. G4 genome and other species of the *Geobacillus* genus show similarity with *Geobacillus kaustophilus* and *Geobacillus thermoleovorans*, with ANI values above 97% (Table 3). Thus, the obtained results suggest that *Geobacillus thermoleovorans* KCTC3570 and *Geobacillus* sp. G4 are sufficiently similar to be considered as belonging to the same species.

Table 3. Average nucleotide identity (ANI) calculated for *Geobacillus* sp. G4 using type strains of each species of the *Geobacillus* genus, calculated on the JSpeciesWS platform (ANIb and ANIm). Information is available at the following link: (https://jspecies.ribohost.com/jspeciesws/, accessed on 11 January 2025).

Isolated Strain	Type Reference Strains	ANIb (%)	ANIm (%)
Geobacillus sp. G4	*G. thermoleovorans* KCTC3570	97.65	98.30
	G. kaustophilus HTA426	97.19	97.91
	G. kaustophilus NBRC102445	97.11	97.85
	G. proteiniphilus 1017	94.66	95.51
	G. thermocatenulatus BGSC93A1	93.15	94.06
	G. jurassicus NBRC107829	90.57	91.52
	Geobacillus vulcani PSS1	90.41	91.54
	G. stearothermophilus ATCC12980	88.93	90.65
	G. uzenensis BGSC92A1	85.05	87.16
	G. subterraneus KCTC3922	85.04	87.12
	G. icigianus G1w11	84.64	87.28
	G. thermodenitrificans DSM465	82.64	85.04

In the dDDH analysis, the *Geobacillus* sp. G4 strain showed dDDH values above 70% with *Geobacillus kaustophilus* HTA426 (79.1%) and *Geobacillus kaustophilus* NBRC102445 (78.5%) (Table 4). However, the *Geobacillus thermoleovorans* KCTC 3570 strain exhibited the highest similarity with *Geobacillus* sp. G4, with a dDDH value of 83.2% and a GC content difference of only 0.31%, shows a high genomic proximity. This suggests that both organisms are closely related, meaning *Geobacillus* sp. G4 could be a variant of *G. thermoleovorans*. Other species showed dDDH values below 70% and a larger GC content difference. These data indicate that, among the species compared, *Geobacillus thermoleovorans* is the closest to *Geobacillus* sp. G4 in terms of genomic similarity.

Table 4. Digital DNA hybridization (dDDH) generated by the TYGS-DSMZ platform between the *Geobacillus* sp. G4 genome and the species described for the genus up to August 2024. All type strains of each species according to the List of Prokaryotic Names with Standing in Nomenclature (https://lpsn.dsmz.de/, accessed on 9 January 2025).

***Geobacillus* sp. G4**	***dDDH Formula 4-d_4 (Similarity Based on Sequence Identity)***		
Reference Genomes	**dDDH (d_4%)**	**Model C.I. (%)**	**G+C Difference (%)**
G. thermoleovorans KCTC3570	83.2	[80.4–85.7]	0.31
G. kaustophilus HTA426	79.1	[76.2–81.8]	0.61
G. kaustophilus NBRC102445	78.5	[75.6–81.2]	0.74
G. proteiniphilus 1017	61.5	[58.6–64.3]	0.84
G. thermocatenulatus BGSC93A1	53.4	[50.7–56.1]	0.82
G. vulcani PSS1	43.3	[40.8–45.9]	0.19
G. jurassicus NBRC107829	42.8	[40.3–45.4]	0.38
G. stearothermophilus ATCC12980	38.7	[36.2–41.2]	0.18
G. uzenensis BGSC92A1	30.4	[28.1–32.9]	0.37
G. subterraneus KCTC3922	30.6	[28.2–33.1]	0.4
G. icigianus G1w11	30.4	[28.0–32.9]	1.13
G. thermodenitrificans DSM465	26.3	[24.0–28.8]	3.55

The 16S rRNA gene sequence of the *Geobacillus* sp. G4 strain showed 99.87% identity with *Geobacillus kaustophilus* strain BGSC 90A1 and 99.80% identity with *Geobacillus thermoleovorans* strain LEH-1 (Supplementary Table S5). The *Geobacillus thermoleovorans* BGSC 96A1 lineage also had 100% query coverage and 99.61% identity. Other *Geobacillus* and *Bacillus* strains showed identity percentages ranging from 99.08% to 99.87%, with query coverages ranging from 92% to 100%.

In the phylogenetic tree constructed based on the 16S rRNA gene, the *Geobacillus* strains are distributed across several clades without statistical support (Figure 3), indicating significant diversification within this genus. However, *Geobacillus* sp. G4 is positioned within a clade supported by a bootstrap value of 71%, containing the strains *Geobacillus thermoleovorans* KCTC3570, *Geobacillus kaustophilus* HTA426, *Geobacillus kaustophilus* NBRC102445, *Geobacillus proteiniphilus* 1017, and *Geobacillus vulcani* PSS1.

In the phylogenomic tree reconstructed using molecular markers extracted with AMPHORA2, the strain *Geobacillus* sp. G4 is grouped with *Geobacillus kaustophilus* HTA426 and *Geobacillus thermoleovorans* KCTC3570, forming a clade supported by a bootstrap value of 95% (Supplementary Figure S1). In the phylogenomic tree reconstructed on the TYGS platform (Figure 4), the strain *Geobacillus* sp. G4 is in the clade with the strain *Geobacillus thermoleovorans* KCTC3570, supported by a bootstrap of 72%.

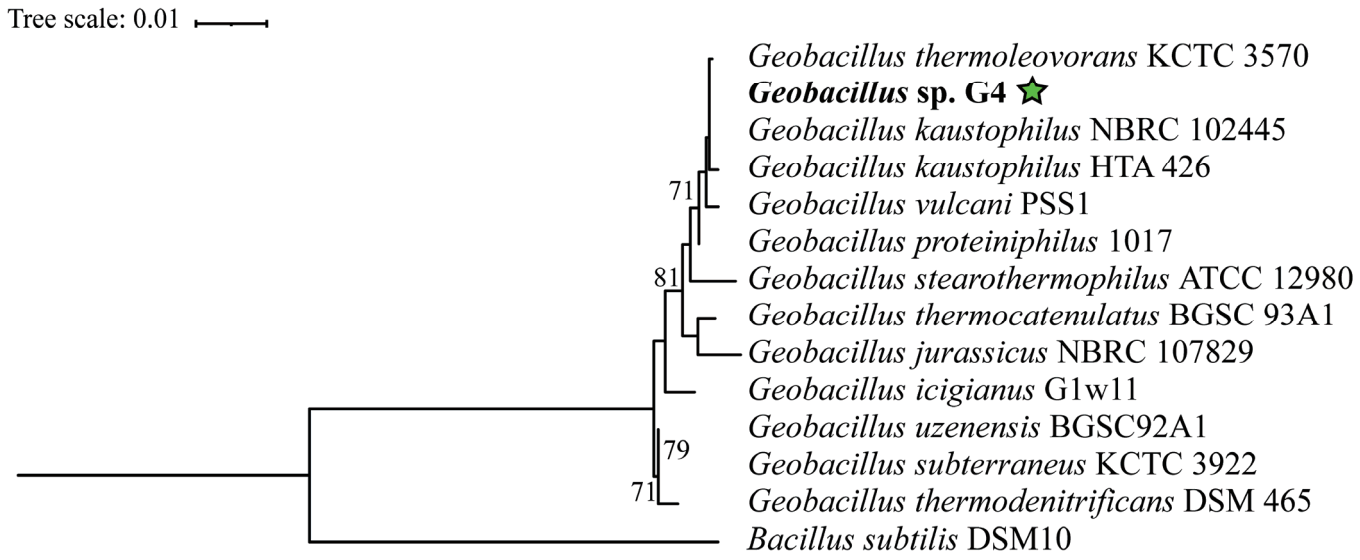


Figure 3. Phylogenetic tree based on the 16S rRNA gene of 13 type strains of *Geobacillus* and *Bacillus*, extracted from the complete genome using the (barrnap) tool. Accession numbers and references are available in Supplementary Table S1. Evolutionary analyses were conducted using the RAxML tool. The strain *Geobacillus* sp. G4, in bold, corresponds to the 16S gene from the genome assembled in this study. The green star indicates the position of the strain isolated in this study.

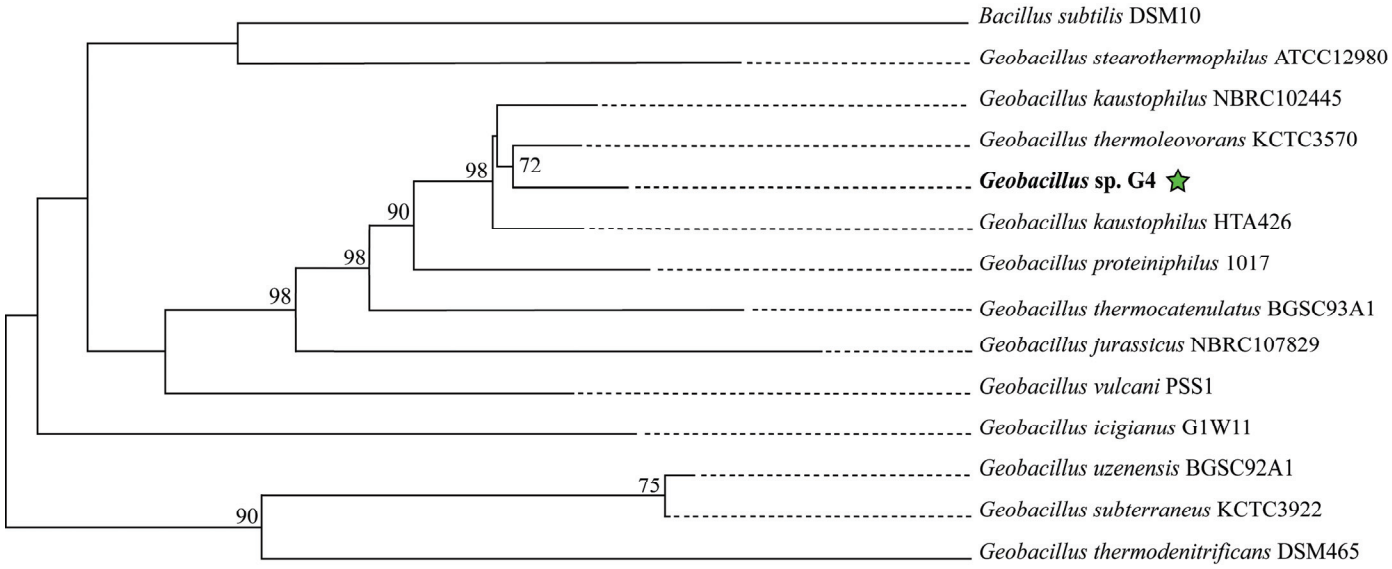


Figure 4. Phylogenetic tree of 13 type strains of *Geobacillus* and *Bacillus*, constructed on the online platform TYGS available at (https://tygs.dsmz.de). The green star indicates the position of the strain isolated in this study.

Comparing the genomes of *Geobacillus* sp. G4 with the genomes of *G. kaustophilus* NBRC102445 and *G. thermoleovorans* KCTC3570 reveals several gaps between them, with a higher frequency observed against *G. kaustophilus* NBRC102445 (Figure 5A). It was also noted that *Geobacillus* sp. G4 showed greater similarity across several regions of its genome compared to *G. thermoleovorans* KCTC3570 (Figure 5B).

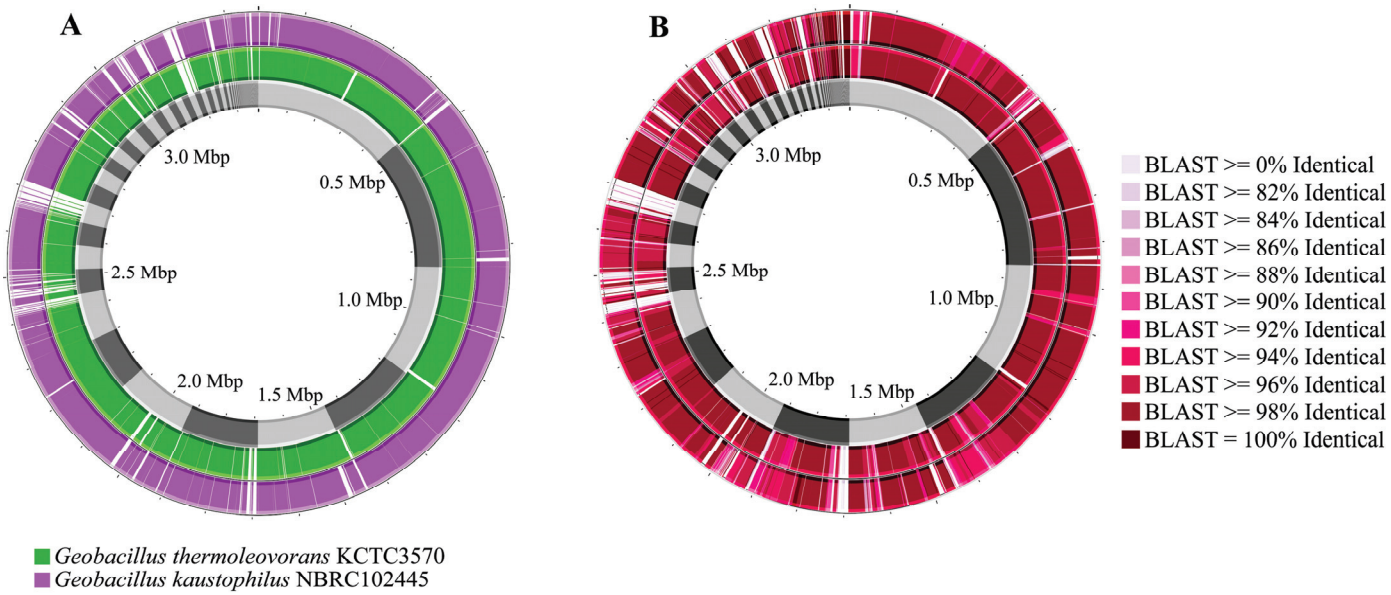


Figure 5. Comparison of the genomes of *G. thermoleovorans* KCTC3570 and *G. kaustophilus* NBRC102445 with the genome of the isolated strain *Geobacillus* sp. G4. (**A**) *Geobacillus* sp. G4 genome (gray), *G. thermoleovorans* KCTC3570 genome (green), and *G. kaustophilus* NBRC102445 genome (purple). (**B**) Similarity comparison between the genomes of *Geobacillus* sp. G4 (gray), *G. thermoleovorans* KCTC3570 (inner circle), and *G. kaustophilus* NBRC102445 (outer circle).

3.3. Functional Analysis of Geobacillus thermoleovorans G4 Strain

The functional analysis of the *G. thermoleovorans* G4 genome revealed a predominance of genes involved in different various metabolic categories, including metabolism, transport, transcription, replication, and repair, among others (Figure 6). Additionally, a large portion of genes were not classified into any known COG category, highlighting the presence of genes with functions that remain unclear within the context of the *G. thermoleovorans* G4 genome.

The genome of *G. thermoleovorans* G4 shows a predominance of enzymes belonging to the various classes, with hydrolases, transferases, and oxidoreductases being the most abundant (Figure 7). Additionally, many sequences lacked known annotations for any of the classes.

In the set of metabolic pathways involved in the degradation of starch and sucrose present in the genome of *Geobacillus thermoleovorans* G4, enzymes involved in the degradation, conversion, and biosynthesis of sugars such as glucose, fructose, and their phosphorylated derivatives were observed (Figure 8). In the glycogen biosynthesis pathway (Supplementary Figure S2A), the enzymes annotated were 1,4-alpha-glucan branching enzyme (GBE1, glgB), glucose-1-phosphate adenylyltransferase (glgC), UTP-glucose-1-phosphate uridylyltransferase (UGP2, galU, galF), and starch synthase (glgA), all classified under transferases, related to starch synthesis. In the glycogen degradation pathway (Supplementary Figure S2B), the annotated enzymes were glycogen phosphorylase (PYG, glgP), pullulanase (pulA), phosphoglucomutase (pgm), and phosphomannomutase/phosphoglucomutase (pmm-pgm), classified under transferases, isomerases, and hydrolases, related to the mobilization and breakdown of glycogen and starch. In the trehalose biosynthesis pathway (Supplementary Figure S2C), the annotated enzymes were glucose-1-phosphate adenylyltransferase (glgC), starch synthase (glgC), and 1,4-alpha-glucan branching enzyme (GBE1, glgB), classified under transferases, related to starch synthesis.

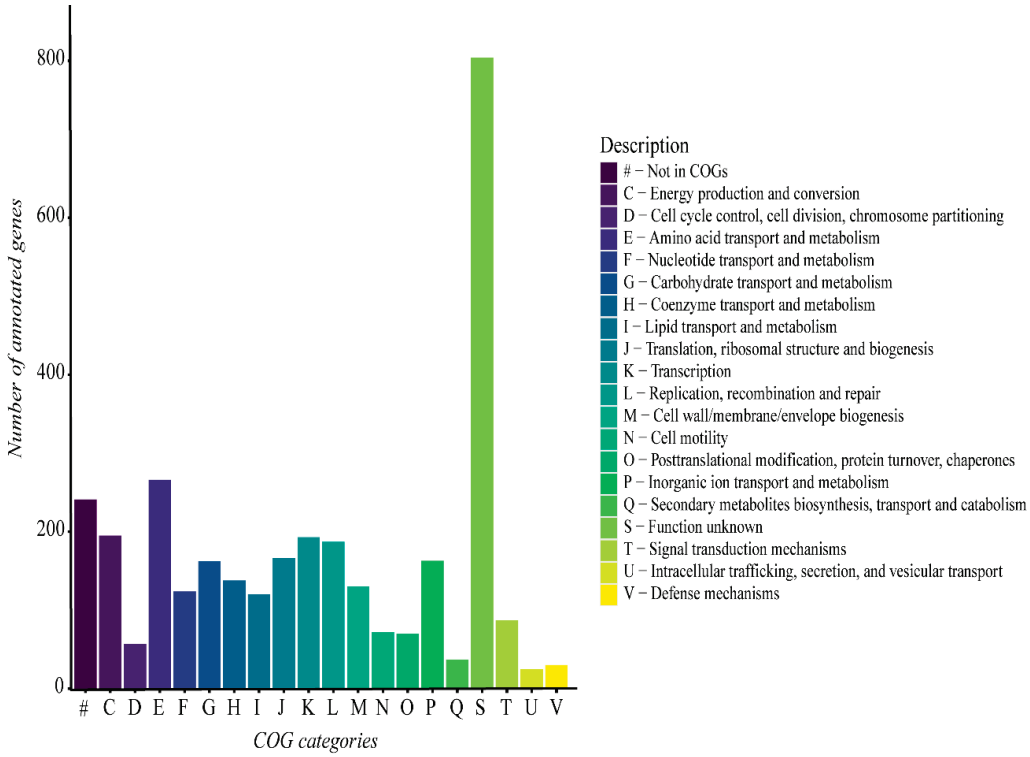


Figure 6. Functional gene categories derived from the functional annotation of the *G. thermoleovorans* G4 genome using Prokka software and the UniProt database on the Galaxy server. Categories were classified using the eggNOG-mapper online tool with the eggNOG5 database.

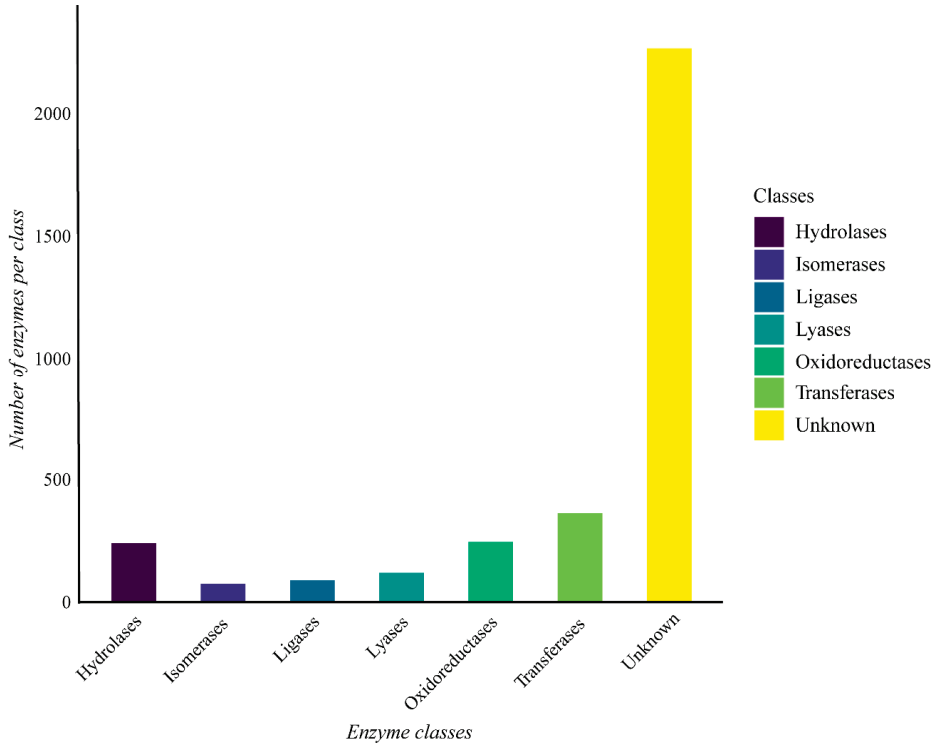


Figure 7. Enzyme classes were obtained through functional annotation of the assembled genome of *Geobacillus thermoleovorans* G4 using Prokka software on the Galaxy server (https://usegalaxy.org/) and classified using the eggNOG-mapper online tool with the eggNOG 5 database (http://eggnog-mapper.embl.de/). The "Unknown" category refers to proteins encoded in the genome that have been annotated as possible enzymes, but do not have an assigned EC number.

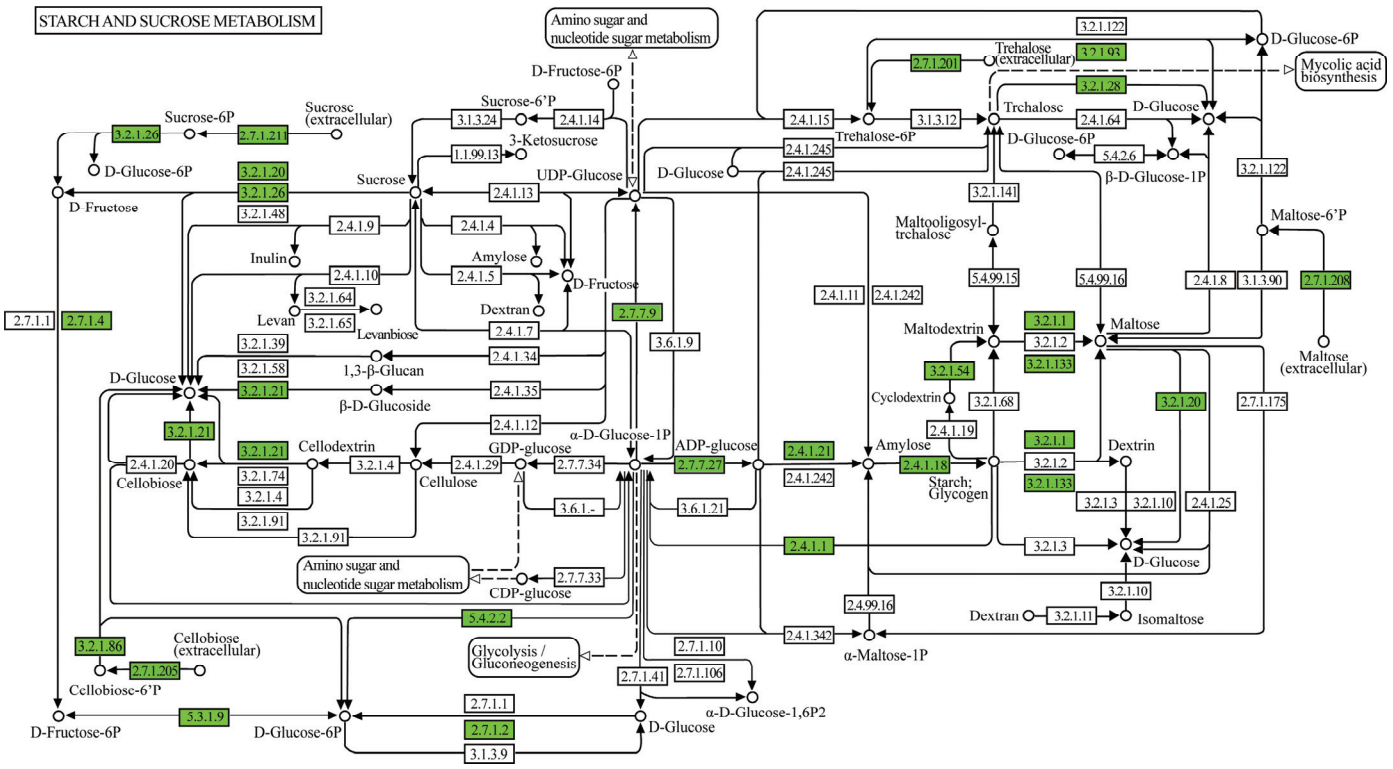


Figure 8. Metabolic pathway map of starch and sucrose metabolism displaying genes and enzymes of interest annotated in the genome of *G. thermoleovorans* G4. Annotations were performed using the KEGG Mapper tool (www.genome.jp/kegg/mapper, accessed on 20 November 2024), identifying key enzymes involved in the metabolism of sugars and complex carbohydrates.

The analysis of antibiotic resistance genes in the genome of *G. thermoleovorans* G4 identified two clusters related to glycopeptide resistance: vanY/vanB_cl and vanT/vanG_cl (Supplementary Table S6). The similarity of these genes between *Geobacillus* sp. G4 and other species ranged from 34.15% to 36.41% for vanY/vanB_cl and from 35.9% to 34.15% for vanT/vanG_cl, with both species compared showing high similarity.

For the *G. thermoleovorans* G4 strain, clusters for seven different types of secondary metabolites were identified: betalactone, T3PKS, Lanthipeptide class-i, NI-siderophore, terpene, fengycin, and RiPP-like (Supplementary Table S7). Comparatively, the strains *Geobacillus kaustophilus* HTA425 (G.K HTA425), *Geobacillus kaustophilus* NBRC102445 (G.K NBRC102445), and *Geobacillus thermoleovorans* KCTC3570 (G.T KCTC3570) also presented clusters for some of the analyzed secondary metabolites, except for NI-siderophore, fengycin, and RiPP-like, where no clusters were found. These observations suggest a shared biosynthetic capability among the strains, with minimal variations in the types of secondary metabolites detected.

The *Geobacillus thermoleovorans* G4 strain presents a complete repertoire of genes (Figure 9) associated with carbon starvation, including *relA*, *cstA*, *ackA*, and *accA*. Complementarily, in the DNA repair and supercoiling category, a diverse set of genes was observed, such as *uvrABC*, *recA*, *recO*, *recR*, *mutS*, *mutL*, and *gyrAB*. In addition, classical heat-shock response genes were detected, including *hrcA*, *dnaK*, *groEL*, *clpB*, and *ibpB*. Finally, the presence of oxidative stress-related genes, such as *sodA*, *katA*, *ahpC*, *trxB*, and *msrA*, was also confirmed. The complete list of selected genes, along with their corresponding COG reference numbers, can be found in Supplementary Table S8.



Figure 9. Distribution of presence/absence of stress-related genes in the genome of *G. thermoleovorans* G4 and in reference genomes of other *Geobacillus* strains. Each row represents an annotated gene, while each column corresponds to a strain. Black circles indicate the presence of the gene, and white circles indicate its absence. Genes are grouped into four main functional categories: carbon starvation, DNA repair and supercoiling, heat-shock response, and oxidative stress.

4. Discussion

The genomic characterization and functional analysis of *Geobacillus* sp. G4, isolated from the geothermal field of Calientes in Tacna, Peru, provided valuable insights into its biotechnological potential. *Geobacillus* sp. G4 exhibited morphological and physiological characteristics typical of the *Geobacillus* genus. Initially, a direct comparison with *Geobacillus kaustophilus* and *Geobacillus thermoleovorans* was established, given that preliminary phylogenetic analyses and previous studies identified these species as the closest phylogenetic relatives to strain G4, sharing similar physiological profiles and temperature ranges [1,46].

Morphologically, strain G4 is characterized as a Gram-positive, endospore-forming rod, producing yellowish-white colonies approximately 1.4 mm in diameter, features that are commonly observed among thermophilic *Geobacillus* strains [3,5]. Its growth within a temperature range of 50–70 °C and an optimal pH of 6.0–7.5 further confirms its adaptation to thermophilic environments.

Table 1 specifically compares the properties of G4 with *G. kaustophilus* and *G. thermoleovorans*, highlighting relevant similarities and differences. All three strains are aerobic, Gram-positive, non-motile, and endospore-forming; however, notable differences were observed regarding acid production from various substrates. For instance, *G. kaustophilus* does not produce acid from lactose, whereas G4 does, along with cellobiose, glycerol, sucrose, and xylose, indicating broader or more efficient metabolic pathways for sugar degradation. The shared ability of G4 and *G. thermoleovorans* to metabolize lactose and glycerol supports their close relationship [1].

Regarding the optimal pH range, *Geobacillus* sp. G4 exhibits growth between pH 6.0 and 7.5, slightly narrower compared to some *G. thermoleovorans* strains that can tolerate conditions up to approximately pH 9.0. Additionally, the tolerance of strain G4 to NaCl concentrations up to 2.5% (*w*/*v*) indicates adaptations to moderately saline environments, a trait shared by many thermophilic *Geobacillus* species inhabiting geothermal springs [19,54,55].

Geobacillus sp. G4 exhibited extracellular enzymatic activities for cellulase, pectinase, xylanase, amylase, protease, and lipase on agar plates, with notably prominent hydrolysis halos observed particularly on starch-containing media. Similar enzyme profiles were previously reported for other thermophilic strains, such as *G. stearothermophilus*, *Geobacillus* sp. GS33, *Geobacillus* sp. WSUCF1, and *Geobacillus* sp. DUSELR13, particularly regarding the stability of amylase and xylanase at elevated temperatures [56–59]. Although protease and lipase activities were detectable in solid media assays, they were not quantifiable in liquid cultures, potentially due to differences in gene regulation or specific culture conditions [60,61].

Moreover, the diverse enzymatic profile identified in *Geobacillus* sp. G4 highlights a promising opportunity for developing enzyme cocktails tailored for efficient lignocellulosic biomass conversion into bioethanol and other valuable bioproducts. Despite their recognized biotechnological potential, relatively few studies have focused on formulating enzyme cocktails based on multiple thermostable enzymes derived from *Geobacillus* species [2,18].

Genome sequencing revealed a 3.4 Mb genome with a G+C content of 52.59%, consistent with genomic attributes previously reported for the *Geobacillus* genus, including strains such as *Geobacillus stearothermophilus* EF60045 and SJEF4-2 and *Geobacillus* sp. E55-1, and *Geobacillus* sp. 56T2 and 56T3 [22,62,63]. Studies demonstrate that the pan-genome of these bacteria reveals great genetic diversity, facilitated by horizontal gene transfer (HGT), which enables adaptation to various ecological niches [64]. However, classification based solely on 16S rRNA sequences has proven to be limited due to the high similarity between

species [9]. To address this issue, ANI and dDDH analyses, combined with complete genome sequences, provide more precise phylogenetic resolution [65].

In our study, ANI and dDDH analyses revealed that *Geobacillus* sp. G4 is closely related to *G. thermoleovorans* KCTC3570, with ANI values above 97% and dDDH values exceeding 83%. These results are consistent with previous studies indicating that species such as *G. kaustophilus* and *G. thermoleovorans* could be conspecific, suggesting ongoing taxonomic revisions [66]. Phylogenetic analysis corroborates this relationship, showing that *Geobacillus* sp. G4 aligns within the *G. thermoleovorans* clade, indicating an evolutionary proximity.

In the genome of *G. thermoleovorans* G4, key enzymes for the degradation of complex carbohydrates, such as amylases and pullulanase, were identified. These enzymes play essential roles in breaking down starch and pullulan into simpler sugars. These enzymes are widely reported in other *Geobacillus* strains [57], reflecting the metabolic versatility of this bacterium in starch and sucrose degradation.

The enzymatic classes of hydrolases, transferases, and oxidoreductases are essential for the metabolism and adaptation of *Geobacillus* to extreme environments. Hydrolases enable the breakdown of biomolecules and the acquisition of nutrients under limited conditions. Transferases regulate biosynthesis and metabolic pathways, facilitating environmental adaptation. Oxidoreductases are crucial for energy production and defense against oxidative stress. The abundance of these enzymes in the *G. thermoleovorans* sp. G4 genome highlights its ability to adapt and survive in harsh environments.

Concerning the biotechnological potential of *G. thermoleovorans*, extensive research has focused on identifying enzymes capable of withstanding extreme environmental conditions, such as elevated temperatures and the presence of inhibitory substances [67,68]. Several enzymes from extremophiles have been identified as interesting tools for industrial processes, including biomass deconstruction in biorefinery [69,70].

Here, we describe a new strain of *G. thermoleovorans* with the potential to produce several industrially relevant enzymes, including more specific endoglucanases (evaluated against carboxymethylcellulose) and pectinases. Cellulases derived from organisms capable of surviving extreme temperatures are extensively documented targets for biotechnological applications because of their ability to withstand prolonged exposure to high temperatures while maintaining catalytic activity [71,72]. The capacity of *G. thermoleovorans* to grow in the presence of pectin highlights its potential to degrade other complex carbohydrates present in plant cell walls.

Despite an increasing number of reports describing thermophiles that can grow in the presence of cellulose-based polymers, there are few studies describing such organisms as capable of growing in the presence of other complex carbohydrates, such as pectin. The ability of *G. thermoleovorans* to grow in the presence of pectin may indicate its diverse potential to degrade such complex polymers, which require enzymes with different modes of action, including esterases, lyases, and glycosyl hydrolases [73,74]. To date, the potential of this bacterium to degrade complex carbohydrates has not been explored; however, closely related species have been described as potential producers of pectin lyases [75].

The strain explored in the present study could grow in the presence of starch, which may be a direct consequence of the diversity of starch-degrading enzymes produced by this organism. *G. thermoleovorans* has been reported to be a good producer of hyperthermostable amylases, including maltogenic α-amylase [76]. The results shown in the present study and previous literature call attention to further heterologous expression and characterization of such enzymes with untapped biotechnological potential.

Genes involved in metabolic categories such as nutrient metabolism, molecule transport, transcription, DNA replication, and repair are fundamental to ensuring the survival

and adaptation of *Geobacillus* sp. G4 in extreme environments [18]. Nutrient metabolism and molecule transport ensure that the bacterium can absorb and process a wide variety of carbon and energy sources, which is essential in habitats with limited resources or adverse conditions. Genes related to transcription and replication ensure the regulation and efficient reproduction of genetic material, while DNA repair genes are crucial for maintaining genomic integrity, especially in environments that expose the organism to stress factors such as heat, radiation, or oxidative compounds [77].

The presence of pathways for glucose-6-phosphate and fructose-6-phosphate metabolism highlights the adaptation of *G. thermoleovorans* G4 to carbohydrate-rich environments. Furthermore, enzymes associated with cellulose and cellobiose processing indicate a robust capacity to degrade lignocellulosic materials, an important resource for biofuel production, as well as efficient adaptation for energy production via glycolysis, reflecting the metabolic versatility and adaptation of *Geobacillus* to various ecological niches, including extreme environments [78].

The *G. thermoleovorans* G4 strain also harbors genes related to antibiotic resistance, such as the vanY/vanB_cl and vanT/vanG_cl clusters, associated with glycopeptide resistance [79], with similarity ranging from 34% to 36% with other *Geobacillus* species. These findings suggest potential for studies on antimicrobial resistance, which could have important implications for understanding the evolution of bacterial resistance.

This study highlights the identification of seven biosynthetic gene clusters associated with secondary metabolites, including betalactones, terpenes, and, notably, fengycin. Fengycin is a well-known antimicrobial lipopeptide typically produced by mesophilic *Bacillus* species; however, its occurrence in thermophilic *Geobacillus* species has not been previously reported [80,81]. Studying fengycin from *G. thermoleovorans* G4 could uncover novel structural features, enhanced stability, or unique biological activities, highlighting its potential for innovative industrial and pharmaceutical applications.

During carbon starvation, bacteria activate specific mechanisms, including the expression of genes such as *cstA*, *relA*, and *spoT*, to maintain cellular viability. The *cstA* gene encodes a peptide transporter that is upregulated during nutrient limitation and regulated by *CsrA*, while *relA* and *spoT* mediate the stringent response through the synthesis of (p)ppGpp, which modulates global gene expression and protein synthesis to adapt to stress conditions [82–84].

To preserve genomic integrity, *G. thermoleovorans* G4 possesses conserved DNA repair systems such as nucleotide excision repair (NER) and mismatch repair (MMR). NER, involving *uvrA*, *uvrB*, and *uvrC*, is critical for correcting UV-induced lesions and bulky adducts [85]. Likewise, MMR, mediated by *mutS* and *mutL*, ensures replication fidelity by removing base mismatches [86]. Extremophilic organisms have adapted these systems with functional versatility to cope with high-stress environments.

Oxidative stress responses in *G. thermoleovorans* G4 include enzymes such as superoxide dismutase (*sodA*), catalase (*katA*), and alkyl hydroperoxide reductase (*ahpC*), which detoxify reactive oxygen species (ROS). The overexpression of *ahpC*, for instance, enhances tolerance to oxidative damage, as shown in other bacterial models [87].

Under heat stress, genes such as *dnaK*, *groEL*, and *clpXP* are upregulated, functioning as molecular chaperones and proteases that maintain protein folding and quality control. These genes are regulated by the transcriptional repressor *hrcA*, which functions as a protein thermosensor in thermophilic organisms [88].

Finally, these genomic adaptations spanning nutrient limitation, DNA repair, and oxidative and thermal stress responses emphasize the ecological resilience of *G. thermoleovorans* G4 and its promising value for industrial applications requiring stability under extreme environmental conditions.

5. Conclusions

The genomic and functional data of *Geobacillus thermoleovorans* G4 demonstrate its value as a highly promising strain for biotechnological applications. Its thermostable enzymes, along with biosynthetic diversity and antibiotic resistance genes, make this strain an important target for biofuel production, industrial biotechnology, and antimicrobial resistance studies.

The detailed analysis of the *G. thermoleovorans* G4 genome reinforces the importance of combining modern molecular tools, such as ANI, dDDH, and phylogenomics, for accurate bacterial taxonomy and understanding of phylogenetic relationships within the *Geobacillus* genus. This not only improves the classification of new strains but also facilitates the development of innovative biotechnological solutions based on the adaptive and metabolic capabilities of these extremophilic microorganisms.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/microorganisms13061374/s1, Table S1: The type strains of the genera *Bacillus* and *Geobacillus* used to build the phylogenetic tree. Table S2: The complete genomes of type strains from the genera *Bacillus* and *Geobacillus* used to construct the phylogenetic tree. Table S3: A list of genes and markers extracted from complete genomes using the AMPHORA2 tool for phylogenetic tree reconstruction. Table S4: The phenotypic characteristics of *Geobacillus* sp. G4. Table S5: A comparison of the 16S rRNA gene of *Geobacillus* sp. G4 with the NCBI database. Table S6: An analysis of the annotation of possible antibiotic resistance genes. Table S7: An analysis of biosynthetic gene clusters for secondary metabolites. Table S8: Genes associated with the functional categories: heat-shock response, DNA repair and supercoiling, oxidative stress, and Carbon starvation identified in the genomes of *Geobacillus* sp. G4, *G. kaustophilus* HTA426, *G. kaustophilus* NBRC12445, and *G. thermoleovorans* KCTC3570. Figure S1: Phylogenomic tree constructed from type strains of Geobacillus and Bacillus species, retrieved from NCBI (GenBank). Figure S2: Metabolic pathways displaying genes and enzymes of interest annotated in the genome of *G. thermoleovorans* G4.

Author Contributions: Conceptualization, A.R.P.T., F.A.C.L. and R.C.; methodology, A.R.P.T., H.E.R.-A. and K.C.R.S.; software, validation, formal analysis, K.C.R.S., A.R.P.T., F.A.C.L. and R.C.; investigation, H.E.R.-A., A.R.P.T., J.P.S. and P.R.V.H.; resources, A.R.P.T., R.C., P.R.V.H. and E.F.N.; data curation, A.R.P.T. and K.C.R.S.; writing—original draft preparation, K.C.R.S. and A.R.P.T.; writing—review and editing, A.R.P.T., E.F.N., R.C., P.R.V.H. and F.A.C.L.; visualization, H.E.R.-A. and J.P.S.; supervision, E.F.N. and F.A.C.L.; project administration, R.C. and A.R.P.T.; funding acquisition, A.R.P.T., E.F.N. and F.A.C.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research and APC was funded by the Jorge Basadre Grohmann National University through "Fondos del canon sobrecanon y regalias mineras", approved by Rectoral Resolution N°11174-2023-UNJBG, under the project "Aplicación de tecnología del ADN recombinante en la generación de cocteles enzimáticos para la biodegradación de residuo orgánico vegetal".

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding authors.

Acknowledgments: Alonso R. Poma Ticona acknowledges the financial support of CONCYTEC through the PROCIENCIA Program, within the framework of Contest E067-2022-042 (contract No. PE501080334-2022-PROCIENCIA). Kárita C. R. Santos was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Fabyano A. C. Lopes was supported by Fundação de Amparo à Pesquisa do Tocantins (FAPT). E.F Noronha is a recipient of CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) research fellowship. Pedro R.V Hamann is a recipient of the CAPES-PIPD postdoctoral fellowship process number 88887.084629/2024-00. Jéssica P. Silva is a recipient of CNPq postdoctoral fellowship process number 381641/2025-02.

Conflicts of Interest: The authors mutually agree on this submission as it is in the present format; there are no conflicts of interest, and neither human nor animal samples were used during the study. Credit is given to the authors who contributed, and the funding agencies are acknowledged.

References

1. Najar, I.N.; Thakur, N. A Systematic Review of the Genera *Geobacillus* and *Parageobacillus*: Their Evolution, Current Taxonomic Status and Major Applications. *Microbiology* **2020**, *166*, 800–816. [CrossRef] [PubMed]
2. Dumorné, K.; Córdova, D.C.; Astorga-Eló, M.; Renganathan, P. Extremozymes: A Potential Source for Industrial Applications. *J. Microbiol. Biotechnol.* **2017**, *27*, 649–659. [CrossRef] [PubMed]
3. Nazina, T.N.; Tourova, T.P.; Poltaraus, A.B.; Novikova, E.V.; Grigoryan, A.A.; Ivanova, A.E.; Lysenko, A.M.; Petrunyaka, V.V.; Osipov, G.A.; Belyaev, S.S.; et al. Taxonomic Study of Aerobic Thermophilic Bacilli: Descriptions of *Geobacillus subterraneus* Gen. Nov., Sp. Nov. and *Geobacillus uzenensis* Sp. Nov. from Petroleum Reservoirs and Transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, G. th. *Int. J. Syst. Evol. Microbiol.* **2001**, *51*, 433–446. [CrossRef]
4. LPSN—List of Prokaryotic Names with Standing in Nomenclature. Available online: https://lpsn.dsmz.de/ (accessed on 30 March 2025).
5. Coorevits, A.; Dinsdale, A.E.; Halket, G.; Lebbe, L.; de Vos, P.; van Landschoot, A.; Logan, N.A. Taxonomic Revision of the Genus *Geobacillus*: Emendation of *Geobacillus*, *G. stearothermophilus*, *G. jurassicus*, *G. toebii*, *G. thermodenitrificans* and *G. Thermoglucosidans* (Nom. Corrig., Formerly 'Thermoglucosidasius'); Transfer of *Bacillus thermantarcticus* to the genus as *G. thermantarcticus* comb. nov.; proposal of *Caldibacillus debilis* gen. nov., comb. nov.; transfer of *G. tepidamans* to *Anoxybacillus* as A. tepidamans comb. nov.; and proposal of *Anoxybacillus caldiproteolyticus* sp. nov. *Int. J. Syst. Evol. Microbiol.* **2012**, *62*, 1470–1485. [CrossRef]
6. Kuisiene, N.; Raugalas, J.; Chitavichius, D. *Geobacillus lituanicus* Sp. Nov. *Int. J. Syst. Evol. Microbiol.* **2004**, *54*, 1991–1995. [CrossRef]
7. Meintanis, C.; Chalkou, K.I.; Kormas, K.A.; Lymperopoulou, D.S.; Katsifas, E.A.; Hatzinikolaou, D.G.; Karagouni, A.D. Application of RpoB Sequence Similarity Analysis, REP-PCR and BOX-PCR for the Differentiation of Species within the Genus *Geobacillus*. *Lett. Appl. Microbiol.* **2008**, *46*, 395–401. [CrossRef]
8. Zeigler, D.R. The *Geobacillus* Paradox: Why Is a Thermophilic Bacterial Genus so Prevalent on a Mesophilic Planet? *Microbiology* **2014**, *160*, 1–11. [CrossRef] [PubMed]
9. Studholme, D.J. Some (Bacilli) like It Hot: Genomics of *Geobacillus* Species. *Microb. Biotechnol.* **2015**, *8*, 40–48. [CrossRef]
10. Nazina, T.N.; Lebedeva, E.V.; Poltaraus, A.B.; Tourova, T.P.; Grigoryan, A.A.; Sokolova, D.S.; Lysenko, A.M.; Osipov, G.A. *Geobacillus gargensis* sp. Nov., a Novel Thermophile from a Hot Spring, and the Reclassification of *Bacillus vulcani* as *Geobacillus vulcani* Comb. Nov. *Int. J. Syst. Evol. Microbiol.* **2004**, *54*, 2019–2024. [CrossRef]
11. Nazina, T.N.; Sokolova, D.S.; Grigoryan, A.A.; Shestakova, N.M.; Mikhailova, E.M.; Poltaraus, A.B.; Tourova, T.P.; Lysenko, A.M.; Osipov, G.A.; Belyaev, S.S. *Geobacillus jurassicus* sp. Nov., a New *Thermophilic bacterium* Isolated from a High-Temperature Petroleum Reservoir, and the Validation of the *Geobacillus* Species. *Syst. Appl. Microbiol.* **2005**, *28*, 43–53. [CrossRef]
12. Miñana-Galbis, D.; Pinzón, D.L.; Lorén, J.G.; Manresa, À.; Oliart-Ros, R.M. Reclassification of *Geobacillus pallidus* (Scholz et al. 1988) Banat et al. 2004 as *Aeribacillus pallidus* Gen. Nov., Comb. Nov. *Int. J. Syst. Evol. Microbiol.* **2010**, *60*, 1600–1604. [CrossRef] [PubMed]
13. Valdez, S.; de la Vega, F.V.; Pairazaman, O.; Castellanos, R.; Esparza, M. Hyperthermophile Diversity Microbes in the Calientes Geothermal Field, Tacna, Peru. *Braz. J. Microbiol.* **2023**, *54*, 2927–2937. [CrossRef] [PubMed]
14. Castellanos, R.; Santos, K.C.R.; Ticona, A.R.P.; Ramirez-arua, H.E.; Bornás-acosta, S.A.; Silva, P.; Hamann, P.R.V.; Lopes, F.A.C. Draft Genome Sequence of *Geobacillus* sp. Strain G4 Isolated from the Calientes Geothermal Field in Tacna, Peru. *Microbiol. Resour. Announc.* **2024**, *13*, e0071724. [CrossRef] [PubMed]
15. Cruz, V.; Vargas, V.; Matsuda, K. Geochemical Characterization of Thermal Waters in the Calientes Geothermal Field, Tacna, South of Peru. *Proc. World Geotherm. Congr.* **2010**, *1*, 1–7.
16. Vargas, V.; Cruz Pauccara, V.; Antayhua Vera, Y.; Rivera Porras, M.; Cacya Dueñas, L. *Estudio Geotérmico Del Campo Calientes*; Instituto Geológico, Minero y Metalúrgico–INGEMMET: Lima, Peru, 2012; Volume 48.
17. Cortez, Y.; Paul, S.; Villena, G.K.; Gutiérrez-Correa, M.; Rodríguez-Herrera, R.; Kharat, A.S. Isolation and Characterization of Cellulase Producing Bacterial Strains from an Amazonian Geothermal Spring in Peru. *Microbiol. Res. J. Int.* **2016**, *15*, 1–8. [CrossRef]
18. Khaswal, A.; Chaturvedi, N.; Mishra, S.K.; Kumar, P.R.; Paul, P.K. Current Status and Applications of Genus *Geobacillus* in the Production of Industrially Important Products—A Review. *Folia Microbiol.* **2022**, *67*, 389–404. [CrossRef]
19. Suzuki, H. Peculiarities and Biotechnological Potential of Environmental Adaptation by *Geobacillus* Species. *Appl. Microbiol. Biotechnol.* **2018**, *102*, 10425–10437. [CrossRef]

20. Mol, M.; de Maayer, P. Elucidating the Biotechnological Potential of the Genera *Parageobacillus* and *Saccharococcus* through Comparative Genomic and Pan-Genome Analysis. *BMC Genom.* **2024**, *25*, 723. [CrossRef]
21. van Staden, A.D.P.; van Zyl, W.F.; Trindade, M.; Dicks, L.M.T.; Smith, C. Therapeutic Application of Lantibiotics and Other Lanthipeptides: Old and New Findings. *Appl. Environ. Microbiol.* **2021**, *87*, e00186-21. [CrossRef]
22. Brumm, P.J.; De Maayer, P.; Mead, D.A.; Cowan, D.A. Genomic Analysis of Six New *Geobacillus* Strains Reveals Highly Conserved Carbohydrate Degradation Architectures and Strategies. *Front. Microbiol.* **2015**, *6*, 137987. [CrossRef]
23. De Maayer, P.; Brumm, P.J.; Mead, D.A.; Cowan, D.A. Comparative Analysis of the *Geobacillus* Hemicellulose Utilization Locus Reveals a Highly Variable Target for Improved Hemicellulolysis. *BMC Genom.* **2014**, *15*, 836. [CrossRef] [PubMed]
24. Miller, G.L. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Anal. Chem.* **1959**, *31*, 426–428. [CrossRef]
25. Mac Faddin, J.F. *Biochemical Tests for Identification of Medical Bacteria*, 3rd ed.; Williams and Wilkins: Baltimore, MD, USA, 2000; ISBN 0683053183.
26. Guta, M.; Abebe, G.; Bacha, K.; Cools, P. Screening and Characterization of Thermostable Enzyme-Producing Bacteria from Selected Hot Springs of Ethiopia. *Microbiol. Spectr.* **2024**, *12*, e0371023. [CrossRef]
27. Ticona, A.R.P.; Ullah, S.F.; Hamann, P.R.V.; Lopes, F.A.C.; Noronha, E.F. *Paenibacillus barengoltzii* A1_50L2 as a Source of Plant Cell Wall Degrading Enzymes and Its Use on Lignocellulosic Biomass Hydrolysis. *Waste Biomass Valorization* **2021**, *12*, 393–405. [CrossRef]
28. Oluoch, K.R.; Okanta, P.W.; Hatti-Kaul, R.; Mattiasson, B.; Mulaa, F.J. Protease-, Pectinase- and Amylase- Producing Bacteria from a Kenyan Soda Lake. *Open Biotechnol. J.* **2018**, *12*, 33–45. [CrossRef]
29. Sharma, P.; Sharma, N.; Pathania, S.; Handa, S. Purification and Characterization of Lipase by *Bacillus methylotrophicus* PS3 under Submerged Fermentation and Its Application in Detergent Industry. *J. Genet. Eng. Biotechnol.* **2017**, *15*, 369–377. [CrossRef] [PubMed]
30. 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]
31. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality Assessment Tool for Genome Assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [CrossRef]
32. Seppey, M.; Manni, M.; Zdobnov, E.M. BUSCO: Assessing Genome Assembly and Annotation Completeness. *Methods Mol. Biol.* **2019**, *1962*, 227–245. [CrossRef]
33. 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]
34. Eddy, S.R. Accelerated Profile HMM Searches. *PLoS Comput. Biol.* **2011**, *7*, e1002195. [CrossRef] [PubMed]
35. Feldgarden, M.; Brover, V.; Gonzalez-Escalona, N.; Frye, J.G.; Haendiges, J.; Haft, D.H.; Hoffmann, M.; Pettengill, J.B.; Prasad, A.B.; Tillman, G.E.; et al. AMRFinderPlus and the Reference Gene Catalog Facilitate Examination of the Genomic Links among Antimicrobial Resistance, Stress Response, and Virulence. *Sci. Rep.* **2021**, *11*, 12728. [CrossRef] [PubMed]
36. Siguier, P.; Perochon, J.; Lestrade, L.; Mahillon, J.; Chandler, M. ISfinder: The Reference Centre for Bacterial Insertion Sequences. *Nucleic Acids Res.* **2006**, *34*, 32–36. [CrossRef]
37. Bateman, A.; Martin, M.J.; Orchard, S.; Magrane, M.; Ahmad, S.; Alpi, E.; Bowler-Barnett, E.H.; Britto, R.; Bye-A-Jee, H.; Cukura, A.; et al. The UniProt Consortium. UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **2023**, *51*, D523–D531. [CrossRef]
38. Huerta-Cepas, J.; Szklarczyk, D.; Heller, D.; Hernández-Plaza, A.; Forslund, S.K.; Cook, H.; 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.* **2019**, *47*, D309–D314. [CrossRef]
39. Blin, K.; Shaw, S.; Augustijn, H.E.; Reitz, Z.L.; Biermann, F.; Alanjary, M.; Fetter, A.; Terlouw, B.R.; Metcalf, W.W.; Helfrich, E.J.N.; et al. AntiSMASH 7.0: New and Improved Predictions for Detection, Regulation, Chemical Structures and Visualisation. *Nucleic Acids Res.* **2023**, *51*, W46–W50. [CrossRef]
40. Liu, B.; Pop, M. ARDB—Antibiotic Resistance Genes Database. *Nucleic Acids Res.* **2009**, *37*, D443–D447. [CrossRef] [PubMed]
41. Meier-Kolthoff, J.P.; Carbasse, J.S.; Peinado-Olarte, R.L.; Göker, M. TYGS and LPSN: A Database Tandem for Fast and Reliable Genome-Based Classification and Nomenclature of Prokaryotes. *Nucleic Acids Res.* **2022**, *50*, D801–D807. [CrossRef]
42. 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]
43. Richter, M.; Rosselló-Móra, R.; Oliver Glöckner, F.; Peplies, J. JSpeciesWS: A Web Server for Prokaryotic Species Circumscription Based on Pairwise Genome Comparison. *Bioinformatics* **2016**, *32*, 929–931. [CrossRef]
44. Katoh, K.; Misawa, K.; Kuma, K.I.; Miyata, T. MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066. [CrossRef] [PubMed]
45. Capella-Gutiérrez, S.; Silla-Martínez, J.M.; Gabaldón, T. TrimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses. *Bioinformatics* **2009**, *25*, 1972–1973. [CrossRef] [PubMed]

46. Darriba, D.; Posada, D.; Kozlov, A.M.; Stamatakis, A.; Morel, B.; Flouri, T. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Mol. Biol. Evol.* **2020**, *37*, 291–294. [CrossRef]
47. Hasegawa, M.; Kishino, H.; Saitou, N. On the Maximum Likelihood Method in Molecular Phylogenetics. *J. Mol. Evol.* **1991**, *32*, 443–445. [CrossRef]
48. Kozlov, A.M.; Darriba, D.; Flouri, T.; Morel, B.; Stamatakis, A. RAxML-NG: A Fast, Scalable and User-Friendly Tool for Maximum Likelihood Phylogenetic Inference. *Bioinformatics* **2019**, *35*, 4453–4455. [CrossRef] [PubMed]
49. Parte, A.C.; Carbasse, J.S.; Meier-Kolthoff, J.P.; Reimer, L.C.; Göker, M. List of Prokaryotic Names with Standing in Nomenclature (LPSN) Moves to the DSMZ. *Int. J. Syst. Evol. Microbiol.* **2020**, *70*, 5607–5612. [CrossRef]
50. Wu, M.; Scott, A.J. Phylogenomic Analysis of Bacterial and Archaeal Sequences with AMPHORA2. *Bioinformatics* **2012**, *28*, 1033–1034. [CrossRef]
51. Lefort, V.; Desper, R.; Gascuel, O. FastME 2.0: A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program. *Mol. Biol. Evol.* **2015**, *32*, 2798–2800. [CrossRef]
52. Farris, J.S. Estimating Phylogenetic Trees from Distance Matrices. *Am. Nat.* **1972**, *106*, 645–668. [CrossRef]
53. Grant, J.R.; Enns, E.; Marinier, E.; Mandal, A.; Herman, E.K.; Chen, C.Y.; Graham, M.; Van Domselaar, G.; Stothard, P. Proksee: In-Depth Characterization and Visualization of Bacterial Genomes. *Nucleic Acids Res.* **2023**, *51*, W484–W492. [CrossRef]
54. Hussein, A.H.; Lisowska, B.K.; Leak, D.J. The Genus *Geobacillus* and Their Biotechnological Potential. In *Advances in Applied Microbiology*; Elsevier: Amsterdam, The Netherlands, 2015; Volume 92, pp. 1–48. [CrossRef]
55. Meslé, M.M.; Mueller, R.C.; Peach, J.; Eilers, B.; Tripet, B.P. Isolation and Characterization of Lignocellulose-Degrading *Geobacillus thermoleovorans* from Yellowstone National Park. *Appl. Environ. Microbiol.* **2022**, *88*, e00958-21. [CrossRef] [PubMed]
56. Gandhi, S.; Salleh, A.B.; Rahman, R.N.Z.R.A.; Chor Leow, T.; Oslan, S.N. Expression and Characterization of *Geobacillus stearothermophilus* Sr74 Recombinant α-Amylase in Pichia Pastoris. *Biomed. Res. Int.* **2015**, *2015*, 529059. [CrossRef] [PubMed]
57. Burhanoğlu, T.; Sürmeli, Y.; Şanlı-Mohamed, G. Identification and Characterization of Novel Thermostable α-Amylase from *Geobacillus* sp. GS33. *Int. J. Biol. Macromol.* **2020**, *164*, 578–585. [CrossRef] [PubMed]
58. Bhalla, A.; Bischoff, K.M.; Sani, R.K. Highly Thermostable Xylanase Production from A Thermophilic *Geobacillus* sp. Strain WSUCF1 Utilizing Lignocellulosic Biomass. *Front. Bioeng. Biotechnol.* **2015**, *3*, 84. [CrossRef]
59. Bibra, M.; Kunreddy, V.R.; Sani, R.K. Thermostable Xylanase Production by *Geobacillus* sp. Strain Duselr13, and Its Application in Ethanol Production with Lignocellulosic Biomass. *Microorganisms* **2018**, *6*, 93. [CrossRef]
60. Shaheen, M.; Shah, A.A.; Hameed, A.; Hasan, F. Influence of Culture Conditions on Production and Activity of Protease from *Bacillus subtilis* Bs1. *Pak. J. Bot.* **2008**, *40*, 2161–2169.
61. Morabandza, C.J.; Dibangou, V.; Mabika, F.A.S.; Gatse, E.V.; Ngoulou, T.B.; Itsouhou, N.; Nguimbi, E. Optimization of Culture Conditions for Protease Production Using Three Strains of *Bacillus*. *J. Pure Appl. Microbiol.* **2021**, *15*, 621–629. [CrossRef]
62. Sung, J.; Ganbat, D.; Kim, S.B.; Lee, S.; Lee, D. Complete Genome Sequences of *Geobacillus stearothermophilus* Strains EF60045 and SJEF4-2 from Korean Hot Springs. *Microbiol. Resour. Announc.* **2024**, *13*, e0057324. [CrossRef]
63. Miyazaki, K.; Hase, E.; Tokito, N. Complete Genome Sequence of *Geobacillus* sp. Strain E55-1, Isolated from Mine Geyser in Japan. *Microbiol. Resour. Announc.* **2020**, *9*, 10–1128. [CrossRef]
64. Esin, A.; Ellis, T.; Warnecke, T. Horizontal Gene Flow into *Geobacillus* Is Constrained by the Chromosomal Organization of Growth and Sporulation. *bioRxiv* **2018**, 381442. [CrossRef]
65. Aliyu, H.; Lebre, P.; Blom, J.; Cowan, D.; De Maayer, P. Phylogenomic Re-Assessment of the Thermophilic Genus *Geobacillus*. *Syst. Appl. Microbiol.* **2016**, *39*, 527–533. [CrossRef] [PubMed]
66. Sunna, A.; Tokajian, S.; Burghardt, J.; Rainey, F.; Antranikian, G.; Hashwa, F. Identification of *Bacillus kaustophilus*, *Bacillus thermocatenulatus* and *Bacillus* Strain HSR as Members of *Bacillus thermoleovorans*. *Syst. Appl. Microbiol.* **1997**, *20*, 232–237. [CrossRef]
67. GArcía-Aparicio, M.P.; Ballesteros, I.; González, A.; Oliva, J.; Ballesteros, M.; Negro, M.J. Effect of Inhibitors Released During Steam-Explosion Pretreatment of Barley Straw on Enzymatic Hydrolysis. *Appl. Biochem. Biotechnol.* **2006**, *129*, 278–288. [CrossRef]
68. Duarte, G.C.; Moreira, L.R.S.; Jaramillo, P.M.D.; Filho, E.X.F. Biomass-Derived Inhibitors of Holocellulases. *Bioenergy Res.* **2012**, *5*, 768–777. [CrossRef]
69. Haki, G.D.; Rakshit, S.K. Developments in Industrially Important Thermostable Enzymes: A Review. *Bioresour. Technol.* **2003**, *89*, 17–34. [CrossRef]
70. Dutta, T.; Sahoo, R.; Sengupta, R.; Ray, S.S.; Bhattacharjee, A.; Ghosh, S. Novel Cellulases from an Extremophilic Filamentous Fungi *Penicillium citrinum*: Production and Characterization. *J. Ind. Microbiol. Biotechnol.* **2008**, *35*, 275–282. [CrossRef] [PubMed]
71. Liu, D.; Zhang, R.; Yang, X.; Wu, H.; Xu, D.; Tang, Z.; Shen, Q. Thermostable Cellulase Production of *Aspergillus fumigatus* Z5 under Solid-State Fermentation and Its Application in Degradation of Agricultural Wastes. *Int. Biodeterior. Biodegrad.* **2011**, *65*, 717–725. [CrossRef]
72. Waheeb, M.S.; Elkhatib, W.F.; Yassien, M.A.; Hassouna, N.A. Optimized Production and Characterization of a Thermostable Cellulase from *Streptomyces thermodiastaticus* Strain. *AMB Express* **2024**, *14*, 129. [CrossRef]

73. Jayani, R.S.; Saxena, S.; Gupta, R. Microbial Pectinolytic Enzymes: A Review. *Process Biochem.* **2005**, *40*, 2931–2944. [CrossRef]
74. Gummadi, S.N.; Panda, T. Purification and Biochemical Properties of Microbial Pectinases—A Review. *Process Biochem.* **2003**, *38*, 987–996. [CrossRef]
75. Demir, N.; Nadaroglu, H.; Tasgin, E.; Adiguzel, A.; Gulluce, M. Purification and Characterization of a Pectin Lyase Produced by *Geobacillus stearothermophilus* Ah22 and Its Application in Fruit Juice Production. *Ann. Microbiol.* **2011**, *61*, 939–946. [CrossRef]
76. Uma, J.L.; Rao, M.; Satyanarayana, T. Purification and Characterization of a Hyperthermostable and High Maltogenic α-Amylase of an Extreme Thermophile *Geobacillus thermoleovorans*. *Appl. Biochem. Biotechnol.* **2007**, *142*, 179–193. [CrossRef]
77. Wang, M.; Zhu, H.; Kong, Z.; Li, T.; Ma, L.; Liu, D.; Shen, Q. Pan-Genome Analyses of *Geobacillus* spp. Reveal Genetic Characteristics and Composting Potential. *Int. J. Mol. Sci.* **2020**, *21*, 3393. [CrossRef]
78. Najar, I.N.; Sharma, P.; Das, R.; Mondal, K.; Singh, A.K.; Radha, A.; Sharma, V.; Sharma, S.; Thakur, N.; Gandhi, S.G.; et al. Exploring Probiotic Potential: A Comparative Genomics and In Silico Assessment of Genes within the Genus *Geobacillus*. *bioRxiv* **2024**, *1*, 2024.05.15.594408. [CrossRef]
79. Evers, S.; Courvalin, P. Regulation of VanB-Type Vancomycin Resistance Gene Expression by the VanS B-VanR B Two-Component Regulatory System in Enterococcus Faecalis V583. *J. Bacteriol.* **1996**, *178*, 1302–1309. [CrossRef] [PubMed]
80. Lin, L.Z.; Zheng, Q.W.; Wei, T.; Zhang, Z.Q.; Zhao, C.F.; Zhong, H.; Xu, Q.Y.; Lin, J.F.; Guo, L.Q. Isolation and Characterization of Fengycins Produced by *Bacillus amyloliquefaciens* JFL21 and Its Broad-Spectrum Antimicrobial Potential Against Multidrug-Resistant Foodborne Pathogens. *Front. Microbiol.* **2020**, *11*, 579621. [CrossRef] [PubMed]
81. Markelova, N.; Chumak, A. Antimicrobial Activity of *Bacillus* Cyclic Lipopeptides and Their Role in the Host Adaptive Response to Changes in Environmental Conditions. *Int. J. Mol. Sci.* **2025**, *26*, 336. [CrossRef]
82. Matin, A. The Molecular Basis of Carbon-Starvation-Induced General Resistance in *Escherichia coli*. *Mol. Microbiol.* **1991**, *5*, 5–10. [CrossRef]
83. Dubey, A.K.; Baker, C.S.; Suzuki, K.; Jones, A.D.; Pandit, P.; Romeo, T.; Babitzke, P. CsrA Regulates Translation of the *Escherichia coli* Carbon Starvation Gene, CstA, by Blocking Ribosome Access to the CstA Transcript. *J. Bacteriol.* **2003**, *185*, 4450–4460. [CrossRef]
84. Flardh, K.; Axberg, T.; Albertson, N.H.; Kjelleberg, S. Stringent Control during Carbon Starvation of Marine *Vibrio* sp. Strain S14: Molecular Cloning, Nucleotide Sequence, and Deletion of the RelA Gene. *J. Bacteriol.* **1994**, *176*, 5949–5957. [CrossRef]
85. Yeung, A.T.; Mattes, W.B.; Oh, E.Y.; Grossman, L. Enzymatic Properties of Purified *Escherichia coli* UvrABC Proteins. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 6157–6161. [CrossRef] [PubMed]
86. Zhou, H.; Zhang, L.; Xu, Q.; Zhang, L.; Yu, Y.; Hua, X. The Mismatch Repair System (*MutS* and *MutL*) in *Acinetobacter baylyi* ADP1. *BMC Microbiol.* **2020**, *20*, 40. [CrossRef] [PubMed]
87. Zuo, F.L.; Yu, R.; Khaskheli, G.B.; Ma, H.Q.; Chen, L.L.; Zeng, Z.; Mao, A.J.; Chen, S.W. Homologous Overexpression of Alkyl Hydroperoxide Reductase Subunit C (AhpC) Protects Bifidobacterium Longum Strain NCC2705 from Oxidative Stress. *Res. Microbiol.* **2014**, *165*, 581–589. [CrossRef]
88. Versace, G.; Palombo, M.; Menon, A.; Scarlato, V.; Roncarati, D. Feeling the Heat: The *Campylobacter jejuni* HrcA Transcriptional Repressor Is an Intrinsic Protein Thermosensor. *Biomolecules* **2021**, *11*, 1413. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

microorganisms

MDPI

Article

Comparative Genomics of Sigma Factors in *Acidithiobacillia* Sheds Light into the Transcriptional Regulatory Networks Involved in Biogeochemical Dynamics in Extreme Acidic Environments

Pedro Sepúlveda-Rebolledo [1,2], Carolina González-Rosales [3], Mark Dopson [3], Ernesto Pérez-Rueda [4], David S. Holmes [5,6,*] and Jorge H. Valdés [2,*]

[1] Microbial Ecophysiology Laboratory, CCTE Ciencia y Vida, Fundación Ciencia y Vida, Huechuraba 8580704, Chile; pedro.sepulveda.reb@gmail.com
[2] Centro de Bioinformática y Biología Integrativa, Facultad de Ciencias de la Vida, Universidad Andrés Bello, Santiago 8370146, Chile
[3] Centre for Ecology and Evolution in Microbial Model Systems, Linnaeus University, SE-391 82 Kalmar, Sweden; carola.mgr@gmail.com (C.G.-R.); mark.dopson@lnu.se (M.D.)
[4] Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Unidad Académica del Estado de Yucatán, Universidad Nacional Autónoma de México, Mérida 97302, Yucatán, Mexico; ernesto.perez@iimas.unam.mx
[5] Center for Bioinformatics and Genome Biology, CCTE Ciencia y Vida, Santiago 7780272, Chile
[6] Facultad de Medicina y Ciencia, Universidad San Sebastián, Santiago 7510156, Chile
* Correspondence: dsholmes2000@yahoo.com (D.S.H.); jorge.valdes@gmail.com (J.H.V.)

Abstract: Extreme acidophiles from the *Acidithiobacillia* class thrive in highly acidic environments where they rely on diverse regulatory mechanisms for adaptation. These mechanisms include sigma factors, transcription factors (TFs), and transcription factor binding sites (TFBS), which control essential pathways. Comparative genomics and bioinformatics analyses identified sigma factors and TFs in *Acidithiobacillia*, showing similarities but key differences from reference neutrophiles. This study highlights sigma54-dependent one- and two-component systems that are crucial for survival in energy acquisition from sulfur compounds and hydrogen as well as nutrient assimilation. Furthermore, the data suggested evolutionary divergence in regulatory elements distinguishes S-oxidizing from Fe-S-oxidizing members of *Acidithiobacillia*. Conservation of gene clusters, synteny, and phylogenetic analyses supported the expected phenotypes in each species. Notable examples include HupR's role in hydrogenase-2 oxidation in Fe-S-oxidizers, TspR/TspS regulation of the sulfur oxidation complex, and FleR/FleS control of flagellar motility in S-oxidizers. These regulatory mechanisms act as master controllers of bacterial activity, reflecting adaptation to distinct metabolic needs within *Acidithiobacillia*.

Keywords: sigma-factors; RpoN; motility; nitrogen metabolism; hydrogen; sulfur metabolism; enhancer-binding proteins; *Acidithiobacillia*

1. Introduction

Microorganisms that thrive in extremely acidic environments, characterized by low pH levels (below 3.0) and high metal ion concentrations [1–3], are vital catalysts for mineral oxidation. This process generates acidic solutions loaded with metals, commonly known as "acid rock drainage" and "acid mine drainage". Their exceptional capabilities have led to various applications, including the biomining of metal sulfides, desulfurization

of coal and natural gas, and numerous industrial processes [4–6]. Extreme acidophiles exhibit remarkable adaptability, adjusting their metabolism to tackle a wide range of challenges and environmental fluctuations, such as temperature variations, pH levels, electron acceptor/donor availability, CO_2 concentrations, and nutrient deficiencies [4].

Over time, diverse acidophilic microorganisms have been isolated and classified, providing a wealth of genetic and phenotypic information from extremely acidic environments [7]. Within this spectrum, the chemolithotrophic extreme acidophilic bacteria from the *Acidithiobacillia* class [8] stand out, encompassing two primary families: the acidophilic *Acidithiobacillaceae* and the neutral-pH-adapted *Thermithiobacillaceae*. A recent reclassification of species within the genera distributed in *Acidithiobacillia* was introduced [9], and this reclassification has been further elaborated upon in subsequent studies [10,11]. These microorganisms exhibit a shared metabolic profile, relying on the oxidation of inorganic sulfur compounds for energy generation, known as thiotrophy [12]. Additionally, they autonomously convert CO_2 into biomass by fixation processes [13]. Some species further display the capacity to oxidize ferrous iron for energy acquisition [14], while others are capable of performing hydrogen oxidation [15]. These phenotypes can be triggered as a response to environmental fluctuations such as energy or nutrient availability. Appropriate responses and metabolic switching strongly relies on transcriptional regulation, with a key role played by transcriptional regulatory elements including sigma factors (SFs), transcription factors (TFs), and their associated transcription factor binding sites (TFBS) [16].

Recently, a group of TFs was identified within the *Acidithiobacillia* class [11], contributing to extending our understanding of the transcriptional regulatory mechanisms in these extremophile bacteria and providing novel insights on the regulation of critical pathways necessary for their growth and survival in extremely acidic environments. However, there is little information regarding the distribution and function of SFs as master regulators of key physiological roles in response to environmental fluctuations.

SFs are multi-domain protein subunits of bacterial RNA polymerase that play critical roles in transcription initiation, including the recognition and opening of promoters for the initial steps in RNA synthesis [17]. These proteins play an essential role in the regulation of specific genes that participate in several metabolic processes. SFs can be classified first by the major sigma70 family that includes the housekeeping sigma factor genes *rpoD* (sigma70) from group 1, and *rpoS* (sigma38), *rpoH* (sigma32), *fliA* (sigma28), and *rpoE* (sigma24) from groups 2–4 [17]. On the other hand, the sigma54 family includes the *rpoN* (sigma 54) [18,19]. Sigma54 proteins contain an activator interaction domain (Sigma54_AID, PF00309), a core binding region (Sigma54_CBD, PF04963), and a DNA-binding component (Sigma54_DBD, PF04552), defining a single differentiated protein family. In contrast, Sigma70 representatives in *Acidithiobacillia* are distributed into several groups [20,21] with Sigma70_r1_2 (PF00140) and Sigma70_r3 (PF04539) characteristic domains supporting the classification of groups 1–2 when both domains are present; these have been identified as RpoD, RpoS, and RpoH. Group 3 Sigma70 has a single Sigma70_r3 with high similarity to the classic FliA protein involved in flagellar expression. For a schematic representation of sigma factors domains architectures see Figure 1A. Group 4 (RpoE) shows only the presence of Sigma70_r2 and Sigma70_r4_2 domains. Domain architecture analysis strengthens the initial similarity-based identification and helps to define sigma factor diversity across *Acidithiobacillia* representatives. Unlike sigma70-mediated transcription initiation, sigma54-dependent regulation absolutely requires an activator to couple ATP hydrolysis energy to RNA polymerase–sigma54 complex isomerization for activation [22]. Sigma54 exhibits versatile characteristics, participating in diverse regulatory activities, which can respond to varied environmental signals [23].

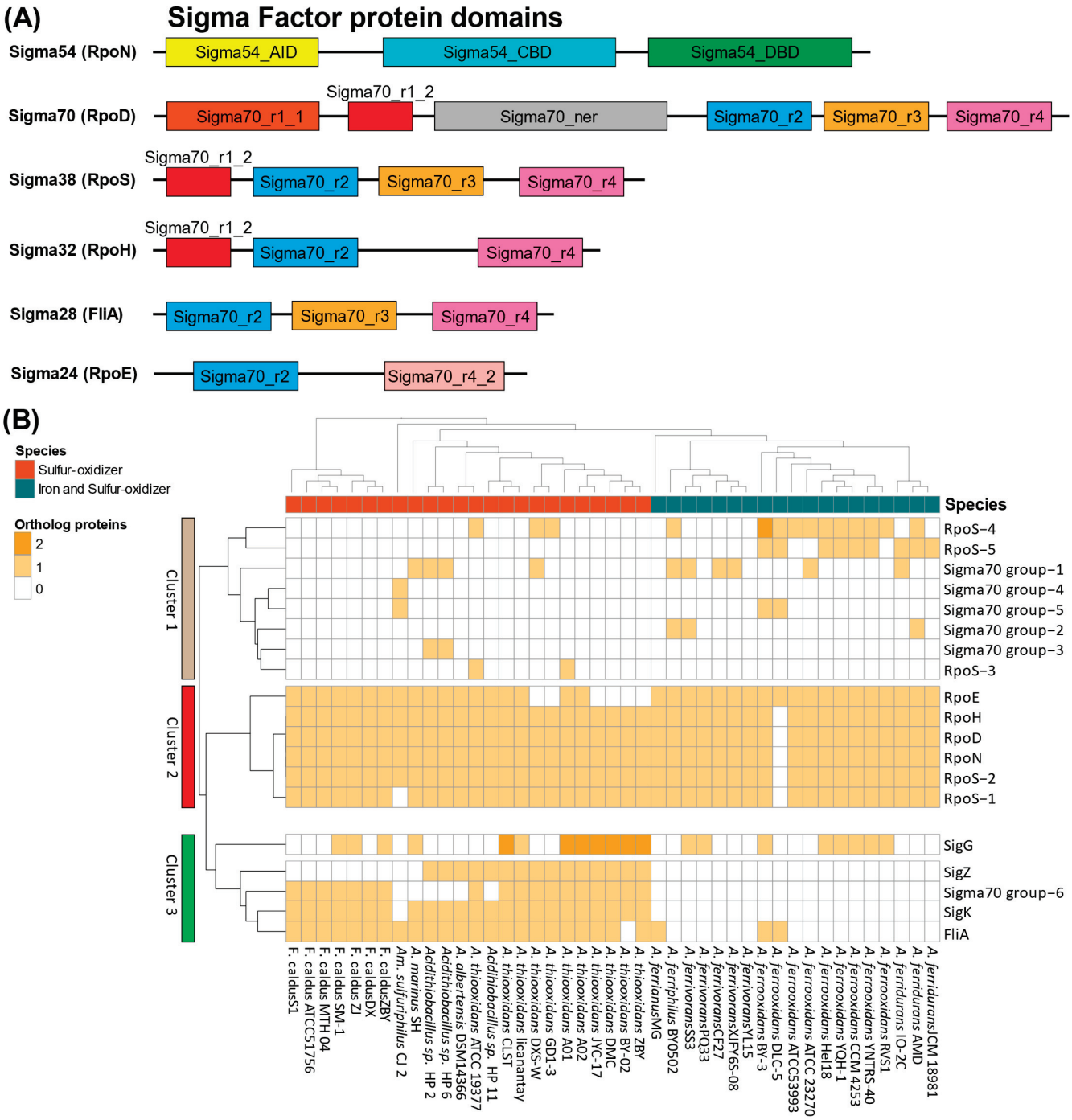


Figure 1. Phylogenetic distribution of *Acidithiobacillia* sigma factors. (**A**) Protein domain architecture of sigma54 and sigma70 proteins and (**B**) heatmap indicating presence/absence of known and potential SFs, based on protein orthology analysis. *Acidithiobacillia* representatives are phylogenetically arranged as represented by the dendrogram. Orthologous SFs counts are colored by the number (0–2) and rearranged in clusters according to Euclidean distance (left side dendrogram).

SFs have been characterized in several model bacteria, including *Escherichia coli* [24], *Bacillus subtilis* [25], and *Clostridium* [26]. Their role has been linked to the regulation of diverse functions, including essential life processes and responses activated by specific stimuli, such as stress and morphological development [17,27]. Sigma factors also play a role in adaptive responses, such as sporulation, toxin production, and cell surface modifications [28].

SFs families can be recognized by their characteristic conserved binding site in the bacterial promoter. The sigma70 family contains the binding motif "TTGACA" from −35 and "TATAAT" from the −10 upstream promoter sequence [27]. In the case of the sigma54, conservation of "GC" from −24, and "TGC" from −12 were observed; additionally, this sigma factor strictly requires the ATP hydrolysis activity of an activating TF for the formation of open promoter complex [29]. These TFs are considered bacterial enhancer binding proteins (bEBPs) [22] and could work as one-component regulator, or two-component regulatory systems (TCSs). However, there is little information regarding the distribution and function of SFs as master regulators of key physiological roles in response to environmental fluctuations.

In *Acidithiobacillia*, SFs reports are restricted to gene expression analysis in the S-oxidizing '*Fervidacidithiobacillus caldus*' (Formerly *Acidithiobacillus caldus*). These studies include investigations into the response to acid stress through RpoS regulatory activity [30] and functions related to flagellar biosynthesis and motility functions by FliA and RpoN [31].

TCSs trigger specific transcriptional responses of many important pathways, where a phosphotransfer reaction takes place between two conserved components: a histidine kinase sensor (HKS) that senses the stimulus, and a response regulator TF [32]. TFs, whose structure contains a DNA-binding domain (DBD), can recognize a transcription factor binding site motif (TFBS) in the promoter and/or regulatory regions. In addition, some TFs contain a sigma54 activating domain, which are part of the bEBPs families [22,33] and are able to work in a sigma54-dependent fashion, promoting coordinated responses of several pathways based one specific or several stimuli.

The transcriptional regulators NtrC, DctD, HydG, PilR, AlgB, HoxA, Hup1, RocR, and FlbD have been described as bEBPs [23]. bEBPs have been characterized in diverse bacterial models, including *Desulfovibrio vulgaris* [31], *E. coli* [34], *Salmonella typhimurium* [35], and *Sinorhizobium meliloti* [36]. Examples of these TCSs in *Acidithiobacillia* include the specific sigma54-dependent the TspS/TspR system, involved in the regulation of energy metabolism in '*F. caldus*' MTH-04 [33], and the *Acidithiobacillus ferrooxidans* NtrB/NtrC pair, involved in nitrogen metabolism [37]. These TCSs have been shown to be essential for energy gathering and nutrient assimilation pathways in extreme acidophiles.

In this study, we present a comparative genomic exploration of 43 complete and draft genomes of *Acidithiobacillia* to detect SFs family proteins, their potentially associated TFs, and their regulatory targets. This comparative analysis was conducted to assess the function, distribution, and phylogeny of SFs and TFs to shed light on the roles of these systems in the *Acidithiobacillia* representatives, their association to speciation and the regulation of specific metabolic traits among clades. Furthermore, specific sigma54-dependent TCSs, their bEBPs, and their potential TFBSs have been also computationally identified. This provides an integrated view of the regulatory networks that might be responsible for specific physiological processes such as energy acquisition and nutrient assimilation key for survival and proliferation in highly acidic environments.

2. Materials and Methods

2.1. Genome Accessions

Complete and draft *Acidithiobacillia* genomes (n = 43) were obtained from NCBI for analysis (Table S1). CheckM (v1.2.2) [38] assessed the genome quality and completeness, requiring $\leq$10% contamination, $\geq$90% completeness, and >50% normalized compatible split length [39] (Table S2). Unannotated genomes underwent Prokka (v1.14.6) [40] prediction of coding regions and proteins with default parameters. EggNOG (v2.1.9) [41] and InterProScan (v5.55-88.0) [42] then assigned functional annotation and classification using default parameters.

2.2. Phylogenetic Analysis

A phylogenomic tree was then constructed to corroborate the *Acidithiobacillia* taxonomy. Concatenated alignments were generated from core conserved proteins extracted from the 43 genomes using MAFFT (v7.407) with L-INS-I iterative refinement [43]. GBLOCKS (v0.91b) [44] trimmed unreliable alignment regions before concatenation. Maximum likelihood trees were then built by IQtree (v 1.6.12) [45] using the concatenated refined alignments, with 1000 bootstraps for robustness, and a substitution model from IQtree [46]. The final tree was visualized in Figtree (http://tree.bio.ed.ac.uk/software/figtree/; accessed on 1 May 2023).

2.3. Orthology and Protein Family Analysis

Comparative genomics focused on protein orthologous groups (OGs) across *Acidithiobacillia* representatives. OGs were defined using ProteinOrtho (V6.0.22) [47] with bidirectional best BLASTp hits between genomes. The criteria were: BLASTp E-value under 10^{-10}, over 60% identity, 60% query coverage, and other default parameters. Each OG then underwent HMMER3 (v3.4) [48] hidden Markov model searches against Pfam- database v33.1 [49] for protein family classification.

2.4. Identification of Genes and Pathways Across Acidithiobacillia Genomes

The metabolic reconstruction analysis involved selecting curated gene- and protein-coding sequences from species with available experimental data. Literature mining and manual curation identified genes encoding proteins crucial for processes such as sulfur oxidation, nitrogen assimilation/reduction, and hydrogen uptake/evolution (Table S3).

2.5. Recovery of Transcriptional Units and Upstream Regions

Gene clusters and potential transcriptional units (TUs) were identified using available experimental data from sources including [7,50–54] and the AciDB database (v1.0) [55]. Additionally, putative gene clusters and TUs were inferred based on intergenic distances between neighboring genes, following established bacterial protocols. In this context, an "upstream region" was defined as having more than 40 bases between coding sequences.

2.6. Identification and Phylogeny of Sigma Factors in Acidithiobacillia Representatives

Sigma factors were identified by comparative genomics using experimentally validated SFs from the model species *E. coli* K-12 (from RegulonDB; https://regulondb.ccg.unam.mx/; accessed on 1 May 2022) and *B. subtilis* 168 (from DBTBS; https://dbtbs.hgc.jp/; accessed on 1 May 2022). Additionally, gene annotations, descriptions, and Pfam domains were data-mined to find further Sigma70-like proteins. Phylogenetic trees for sigma70 and sigma54 families were built using previous protocols, with validated SF sequences from *E. coli* K-12 and *B. subtilis* 168 as outgroup references. Sigma54 and sigma70 protein family orthologs were assessed via earlier orthology analysis. Heatmaps showing SF ortholog groups diversity, quantity, and distribution were generated in Rstudio using the pheatmap v1.0.12 package. *Acidithiobacillia* representatives were clustered according to their phylogeny, and ortholog SF groups were clustered by Euclidean distance.

2.7. Identification of Sigma54-Dependent Two-Component Systems

Identification of bEBPs that could be participating in TCSs was implemented by searching Pfam domains to select proteins with the sigma54 activator (PF00158.27) and response regulator (PF00072.25) domains. The HKS proteins were identified by a curated search examining the genomic architecture of previously recovered transcriptional units and locating those containing both HKS and bEBP genes. Additionally, the signal transduction histidine kinase domain (PF00512) Pfam annotation was identified in HKS proteins.

2.8. Identification and Analysis of Putative DNA Binding Sites

A motif footprint approach was used in MEME (Multiple Em for Motif Elicitation, which detects novel conserved motifs in input sequences) (v 5.4.1) to identify potential DNA binding sites [56] while focusing on upstream regions of studied transcription units containing known genes. The MEME parameters were: up to 10 motifs, the ZOOPS algorithm to allow zero or one site per sequence, a 4th-order Markov background model to normalize biases, and motif lengths from 5 to 20 nucleotides. Sequence logos were also generated using WebLogo (v 2.8.2) [57] to visualize sequence conservation and relative nucleic acid frequencies at each position for the predicted motifs from transcription factors and sigma factors by applying the Schneider and Stephens method [58].

2.9. Matrix Scanning of Binding Sites

Potential DNA binding sites were identified using Position Weight Matrices (PWMs) in 15 *Acidithiobacillia* reference genomes along with experimentally validated *E. coli* K-12 binding site sequences for bacterial transcription factors and sigma factors from RegulonDB. "sites2meme" (from MEME suite) was then used to construct PWMs from these known binding sites. Fourth-order Markov background models were built for *E. coli* K-12 and the 15 reference species using the "create-background-model" tool from the RSAT suite (v2018) [59]. The FIMO tool then scanned (with default parameters) the PWMs against upstream regions of transcription units, using matched background models. Retrieved sites were filtered based on *p*-value thresholds from RegulonDB or a default *p*-value under 1×10^{-4}. Overlapping sites were removed, keeping those over 50% sequence coverage with the best score and the remaining sites were restricted according to the selected reference genes. Sequence logos summarize these putative binding sites.

3. Results

3.1. Acidithiobacillia Display a Diverse Sigma Factor Repertoire Providing Clues on Extreme Acidophile Adaptability

Acidithiobacillia representatives were organized into the five major clades proposed by Moya-Beltrán et al. (2021) [9], where three phylogroups were identified as mesophilic sulfur-oxidizers (S-oxidizers), thermotolerant S-oxidizers, and mesophile/eurypsychrophilic iron–sulfur oxidizers (Fe-S-oxidizers) (Figure S1).

Identification of *Acidithiobacillia* SFs using the complete repertoire from the *E. coli* protein models (RpoN, RpoD, RpoS, RpoH, FliA, RpoE), a partial group from *B. subtilis* (SigK, SigZ, SigG), and proteins containing sigma70 conserved domains from the Pfam database is presented in Figure 1. A schematic representation of these domain architectures is shown in Figure 1A, using '*F. caldus*' ATCC 51756 plus *Acidithiobacillus thiooxidans* ATCC 19377 for S-oxidizers and *A. ferrooxidans* ATCC 23270 for Fe-S-oxidizers as representative examples.

An orthology-based SFs distribution analysis identified three main clusters in the *Acidithiobacillia* (Figure 1B). Cluster 1 included five sigma70-like proteins and three orthologous groups of RpoS (RpoS_3, RpoS_4, and RpoS_5). Notably, RpoS_4 and RpoS_5 exhibited a predominant distribution among Fe-S-oxidizers in comparison to other clades. In cluster 2, a core of conserved groups of sigma factors was identified, encompassing five out of the six sigma factor groups present in *E. coli* (RpoD, RpoN, RpoE, RpoH, and two RpoS groups). This cluster was present in nearly all analyzed genomes, except for *A. ferrooxidans* DLC-5, notably characterized by its fragmented assembly. Within cluster 3, the FliA sigma factor was identified alongside with SigK, SigZ, and sigma70-like group 6, showing a predominant presence among S-oxidizers. Notably, SigG was widely distributed across almost all taxonomic groups considered in the analysis, showing differences in gene dosage in *A. thiooxidans* and the complete absence of SigZ in '*F. caldus*' strains.

3.2. Sigma Factors Phylogeny Reflects Divergence Signatures of Acidithiobacillia Phenotypes

To assess the phylogenetic relationships between *Acidithiobacillia* Sigma70 candidates, a maximum-likelihood phylogeny was generated using all potential SFs proteins for the two primary families, sigma70 and sigma54. The generated sigma70 phylogeny (Figure 2) revealed several clades containing individual groups of SFs, including RpoD, RpoE, RpoH, RpoS_2, FliA, SigG, SigK, and SigZ, along with sigma70-like groups (groups 1, 2, and 6). However, there was an exception of one sub-clade containing multiple protein groups including RpoS (RpoS_1, RpoS_3, RpoS_4, and RpoS_5) and sigma70-like groups (groups 3, 4, and 5). RpoS_2 was found to be distributed in an independent branch distantly separated from the other RpoS groups. Selected outgroups in this analysis included SFs from *E. coli* and *B. subtilis* (Figure 2, depicted in black) that show a distinctive position typically outside the *Acidithiobacilli* branches. As expected from their taxonomical origin, *B. subtilis* SFs were more distantly related than *E. coli* representatives.

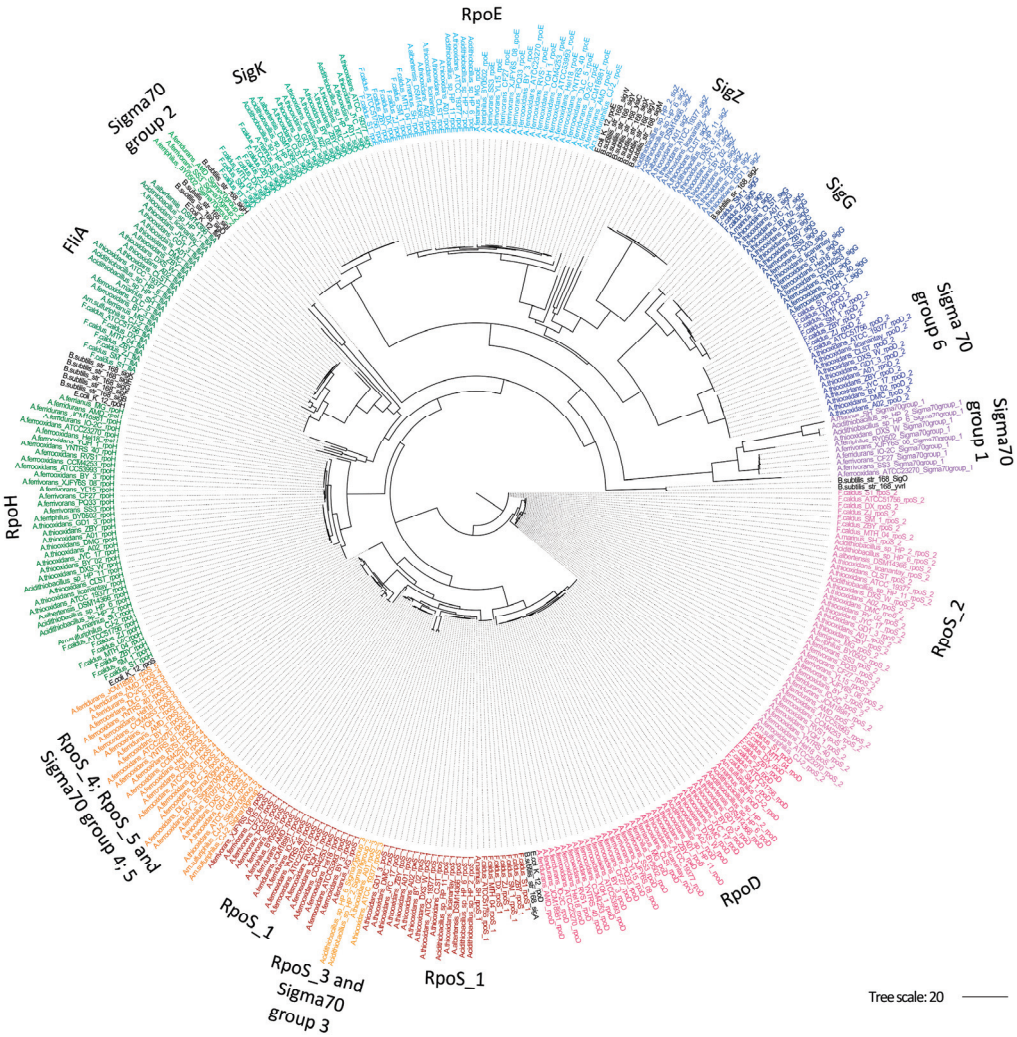


Figure 2. Phylogenetic relationships between sigma70 family proteins in *Acidithiobacillia*. Maximum-likelihood phylogenetic tree based on the sigma70 proteins identified in *Acidithiobacillia*. Groups of SFs were colored by species. Model *E. coli*, and *B. subtilis* representative SFs are shown in black.

As expected, the sigma54 phylogeny (Figure 3), characterized by a unique RpoN protein family, showed high concordance with the species' phylogenetic distribution supporting the vertical inheritance of this sigma factor.

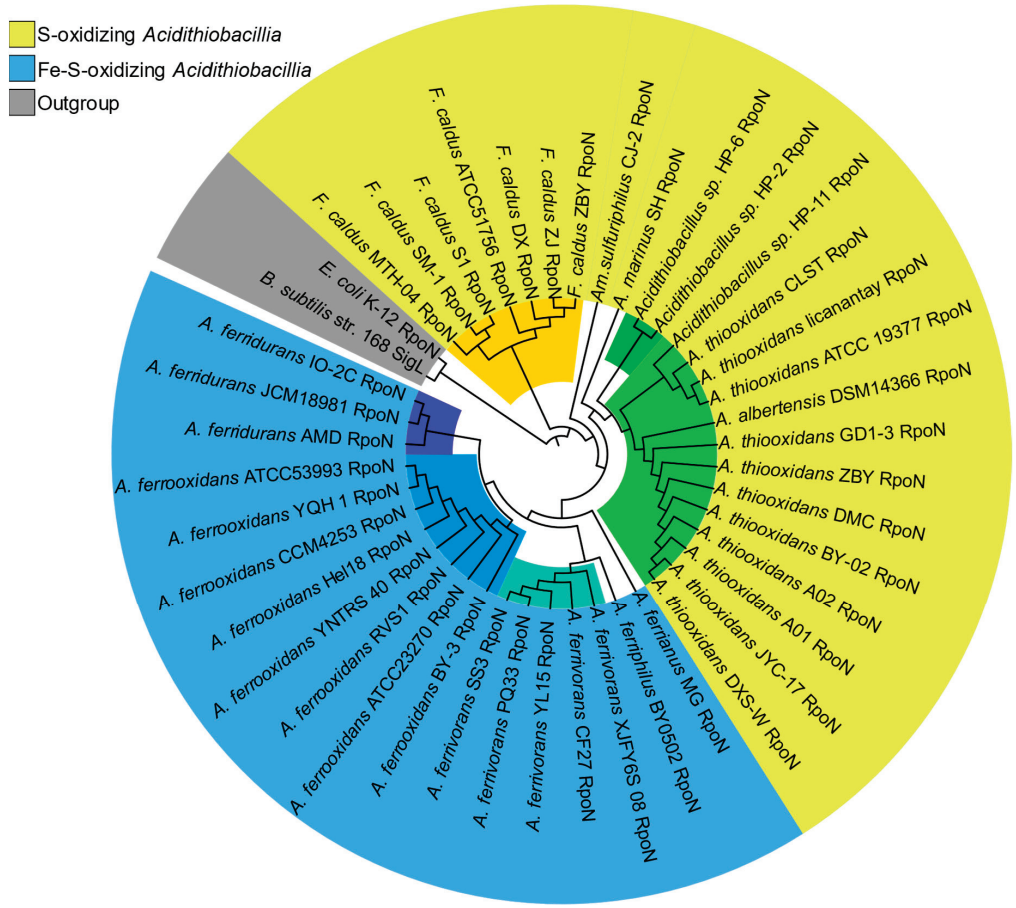


Figure 3. Phylogeny of *Acidithiobacillia* sigma54 orthologs. Maximum-likelihood phylogenetic tree based on *Acidithiobacillia* sigma54 (RpoN) representatives. Sigma54s are colored by phenotype, represented by outer layout color, and species lineages are represented by inner sub-clade colors (colored tree branches). Model species outgroups (*E. coli* and *B. subtilis*) are highlighted in gray.

3.3. Sigma54: A Master Regulator for Energy and Nutrient Assimilation in Acidithiobacillia?

To elucidate Sigma54's influence across key metabolic pathways, a protein domain prediction approach was implemented based on Sigma54 activator (PF00158.27) and response regulator (PF00072.25) domain searches across the *Acidithiobacillia* representatives. A total of 546 proteins were identified containing a Sigma54 activator domain and 617 with a response regulator domain (Table S4, Supplementary Data). Of these, 318 proteins contained both domains, constituting the identified bEBPs from one-component and TCS (bEBP and HKS). These proteins were distributed across nine OGs, including HupR, NtrC/NtrB, NtrX/NtrY, TspR/TspS (or TcsR/TcsS), FleR/FleS, PilR/PilS, GlrR/GlrK, GlrR_2/QseE, and AtoC/AtoS (Figure 4). The TCSs comparison using referential representatives consistently revealed three-domains for the bEBP response regulator component, including the response regulator (also called as REC_2), sigma54 activator (AAA^+), and the DNA-binding (HTH_8) domains [22]. However, some variations were observed in HKS proteins containing cytoplasmic domains involved in signal input/transmission (PAS domain) or signal transduction domains (HAMP domain). In addition, TspS showed an additional domain, a single cache domain 3 (sCache_3_2), described within the extracellular sensors in bacteria [60]. Nonetheless, all the HKS found in *Acidithiobacillia* contain the conserved domains phosphoacceptor Histidine Kinase A (HisKA) and the catalytic domain HATPase_c (Figure 4A) that are essential for TCS functions. These conserved orthologous genes highlighted conserved systems in all *Acidithiobacillia* (Figure 4B), except for AtoS/AtoC and FleS/FleR exclusive to most S-oxidizers. However, the HKS was absent for the HupR/HupT

hydrogen oxidation TCS in all *Acidithiobacillia*, with only the HupR bEBP one-component regulator conserved.

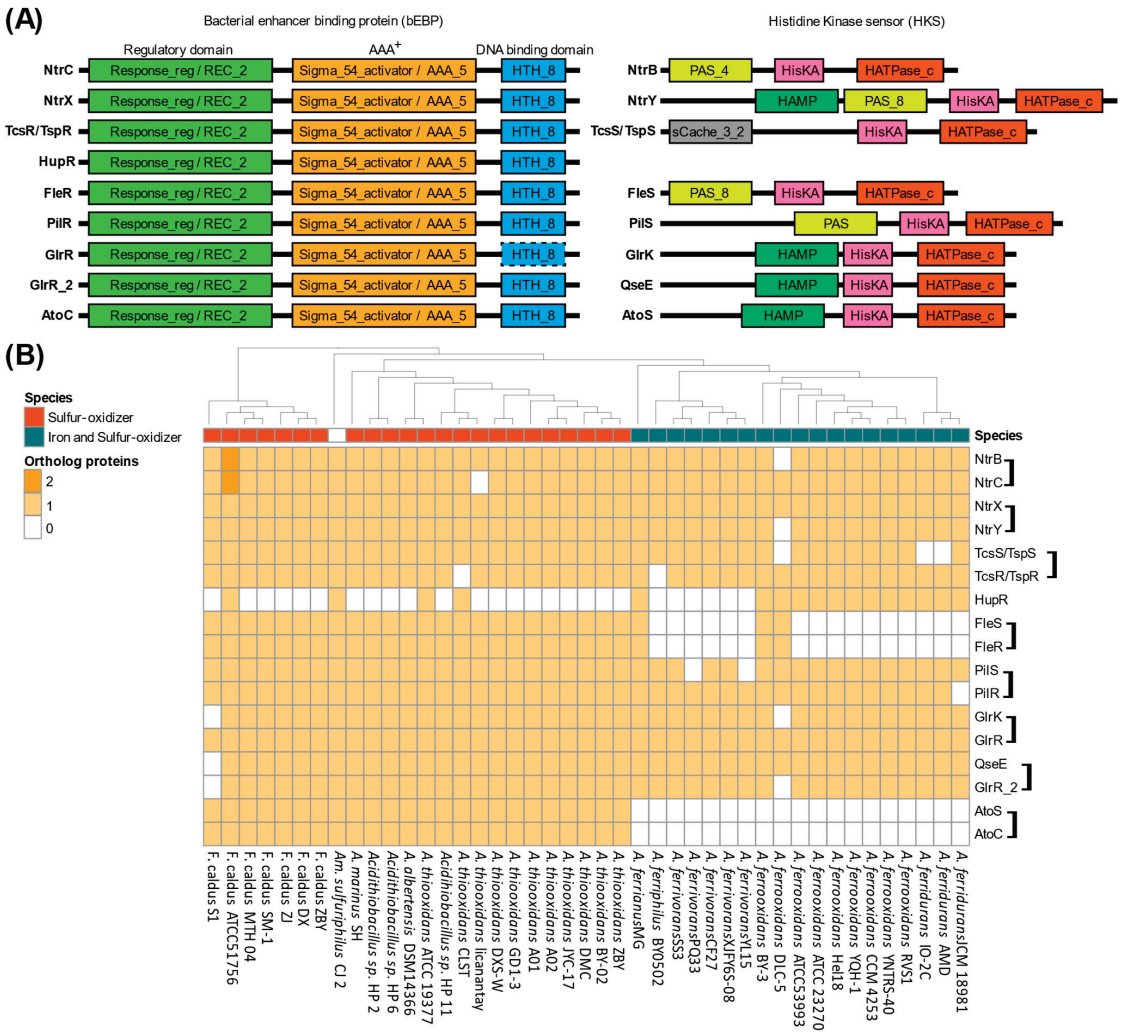


Figure 4. Sigma54-dependent two-component systems repertoire in *Acidithiobacillia*. (**A**) Protein domain architecture of sigma54-dependent TCSs separated by response regulator bEBP component and histidine kinase sensor component. AAA$^+$, ATPase associated with cellular activities; REC, phosphoacceptor receiver domain; HTH, helix–turn–helix domain; PAS, Per–Arnt–Sim domain; HisKa, Signal transduction histidine kinase; HATPase_c, histidine kinase/HSP90-like ATPase. A dashed frame denotes that the HTH_8 domain was not detected in GlrR with the applied parameters, but its presence is inferred from its identification in the GlrR_2 group. (**B**) Protein ortholog groups in *Acidithiobacillia* representatives with TCSs grouped in a bracket with the histidine kinase sensor first, followed by the transcription factor.

The identification of putative DNA binding sites, the genes and operon potentially under control, cognate transcription, and sigma factors enabled the reconstruction of transcriptional regulatory circuit (Figure 5). This reconstruction was centered on curated and conserved orthologous genes across the *Acidithiobacillia* (Figure S2) with reported bEBP-controlled operons or suggested pathway-linked regulation. Information from binding site prediction from sigma54 (RpoN) (Figure 5A) and key bEBPs (Figure 5B) were generated allowing the identification of regulators for hydrogen (HupR), nitrogen (NtrC and NtrX), and sulfur (TspS) pathways. Experimental information from validated sites from other bacteria enabled recognition of likely regulated targets via PWM scanning, such as RpoN and NtrC. On the other hand, for non-reported binding sites (without known PWM) from bEBP HupR, TspR, and NtrX, the utilization of a motif footprint approach pinpointed con-

served upstream binding motifs in suggested pathway-linked regulation genes. Referential *Acidtihobacillia* species were utilized to represent operon organization (Figure 5B, right-side). HupR and TspR co-localized with the suggested genes they regulate (*hupS-hupL*), whereas the remaining studied bEBPs were located in separate genomic regions, distinct from their target genes. The NtrC binding site scanning also enabled proposals for regulatory interactions in diverse transcriptional units. Bringing these together allowed the reconstruction of an NtrC regulon centered on nitrogen pathways in the *Acidithiobacillia*. Since NtrX shares related regulatory activities, participation in similar regulatory interactions as NtrC can be inferred. A summarized regulatory network captures sigma factor master regulator connectivity as well as bEBP regulatory roles across examined metabolic pathways (Figure 5C).

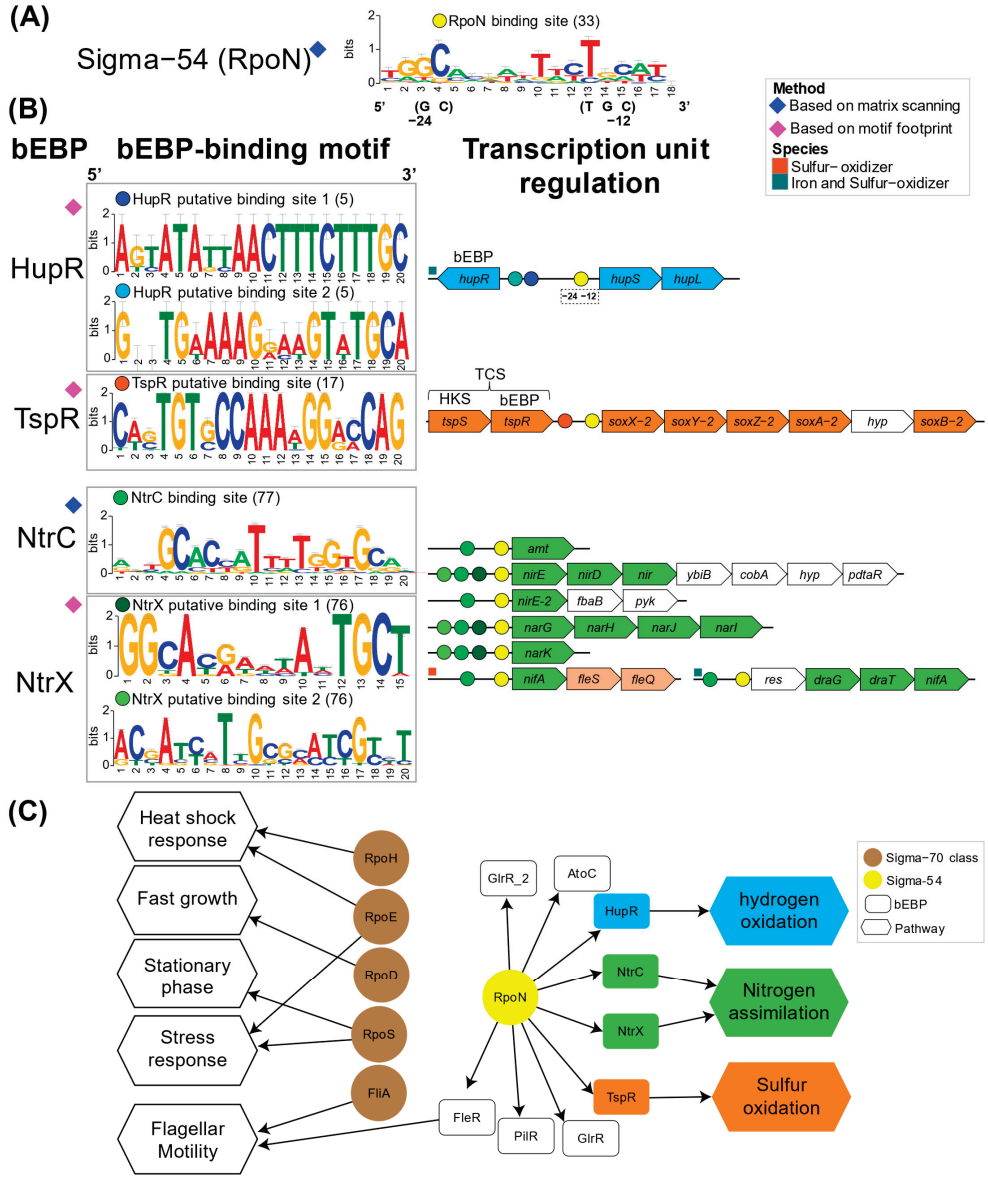


Figure 5. Transcriptional regulatory circuits of bacterial enhancer binding proteins in *Acidithiobacillia*. (**A**) RpoN binding site and (**B**) bEBP binding sites and regulatory activity retrieved for known and de novo inferences. The number of binding sites that build the logos graph are given in parentheses. Transcriptional units are represented by referential species, where bEBP binding sites are depicted in color circles in the promoter regions. Genes encoding metabolic pathways of hydrogen in cyan, nitrogen in green, sulfur in orange, and additional genes in white; *res*, response regulator protein. (**C**) Transcriptional regulatory network representation of sigma factors and bEBPs.

4. Discussion

4.1. An Acidophilic Lifestyle Demands Robust Gene Regulation Circuitry

While sigma70 family regulatory proteins participate globally in all essential bacterial processes, sigma54 has been demonstrated as an indispensable master regulator specifically controlling vital pathways [18]. The *Acidithiobacillia* SFs (Figure 1A) revealed the presence of identical domains as observed in the sigma factor family classification [20,61]. This suggested adaptations to regulate essential processes enabling acidophile survival and growth in extremely acidic settings. However, additional proteins with sigma54-associated domains were identified that lacked a defined annotation and without comparison in other bacterial models. Almost the entire set of *E. coli* SFs was conserved in the *Acidithiobacillia* class except for FliA, which was conserved mainly in S-oxidizers (Figure 1B). FliA controls the expression of flagella-related genes [62]. Interestingly, the S-oxidizing '*F. caldus*' and *A. thiooxidans* share the gene potential and capacity to develop flagellum and chemotaxis [30,63–65]. Therefore, it was suggested that FliA may also participate in regulating flagellar genes. In addition, FliA was identified in some Fe-S-oxidizers, such as *A. ferrooxidans* strains where genes involved in chemotaxis have been previously identified [65,66].

4.2. Sigma Factor Master Regulators Tend to Mirror Acidithiobacillia Evolution

The sigma70 phylogeny showed most groups forming individual sub-clades near reference outgroups. Sigma70-like group 6 displayed just one domain shared by RpoE (Table S5, Supplementary Data) and had the shortest protein sequence, suggesting a new group of sigma factors with intriguing regulatory activity in acidophiles yet to be determined. Additionally, we find RpoS instances. One sub-clade contained four RpoS and three sigma70-like groups, indicating these latter may associate with RpoS. Furthermore, the deep-branching RpoS_2 cluster differed from the reference outgroups and lacked one conserved RpoS domain (Table S5, supplementary material), distinguishing it from other RpoS proteins. RpoS_2 may have emerged within the class *Acidithiobacillia* or been acquired through horizontal gene transfer from another member of the acidophilic community, highlighting the dynamic genetic exchanges that shape adaptation to extreme acidic environments and its differentiation from neutrophilic species described.

The sigma54 (RpoN) phylogeny (Figure 3) mirrored the taxonomic structure of the broader *Acidithiobacillia* phylogenomic tree. This agreement provides evidence cementing *rpoN* as a fundamental housekeeping gene perpetuating acidophilic viability across the taxon.

In the case of SigZ, originally identified in *B. subtillis*, it constituted the main difference in the repertoires of '*F. caldus*' strains in comparison to mesophilic sulfur-oxidizers like *A. thiooxidans*, which clustered separately despite their phenotypic similarities. While '*F. caldus*' shared a similar sulfur oxidation profile and many sigma factor orthologs with mesophilic sulfur-oxidizers such as *A. thiooxidans*, the separation into a distinct cluster was likely explained by lineage-specific gene differences due to their gene dosage adaptations, potentially driven by moderately thermophilic conditions.

4.3. Sigma54-Dependent Two-Component Systems in Acidithiobacillia Were Potentially Linked to Key Metabolic Pathways

The discovery of *Acidithiobacillia* sigma54-dependent two-component systems that were involved in key metabolic pathways advances understanding of their dominance in extremely acidic environments [67]. Sulfur oxidation TCSs regulate protein complexes that enable the utilization of diverse reduced inorganic sulfur compounds for energy acquisition [52]. In '*F. caldus*', TspS/TspR is co-located with the *sox* sulfur oxidation complex gene cluster [33,68], mainly conserved in S-oxidizing representatives and notably absent in

other Fe-S-oxidizers except for *A. ferrivorans*, which shows a partial conservation of this gene cluster [9]. TspS/TspR's DNA-binding site and Sigma54 promoter interaction suggest transcriptional regulation of *sox* genes in S-oxidizers [11]. Furthermore, the TcsS/TcsR system described in Fe-S-oxidizers and localized adjacent to the *tetH-doxDA* gene cluster [69] was orthologous to TspS/TspR, (Figure 4B) and conserved identical domains Their specialized regulatory roles were likely adapted to control a phenotype-specific metabolism in S-oxidizers and Fe-S-oxidizers.

Hydrogen oxidation also supplies energy, a process carried out by different protein complexes categorized into hydrogenase groups [70]. The one-component HupR system was situated in a divergent orientation relative to hydrogenase group 2 (*hupS-hupL*) and primarily conserved in Fe-S-oxidizers. While HupR has been suggested to function as part of a two-component regulatory system, it may also exert its regulatory activity independently, without requiring phosphorylation by a histidine kinase [71], suggesting a versatile role in hydrogenase regulation.

Nitrogen scarcity in extremely acidic environments is offset by the nitrogen fixation capability of some *Acidithiobacillia* [72] to produce bioavailable nitrogen compounds, supporting the broader microbial communities in these environments. The NtrB/NtrC and NtrX/NtrY TCSs regulate some of these nitrogen assimilation pathways genes [73] while NtrB/NtrC is described in other microorganisms [74] such as *Rhodobacter capsulatus* [75], *A. ferrooxidans* DSM 16786, and *A. thiooxidans* DSM 17318 [76]. NtrB/NtrC senses nitrogen sources like molecular nitrogen, ammonia, and nitrate, and it was suggested to control nitrogen assimilation in most *Acidithiobacillia* due to their complete conservation (Figure 4). Assimilation mechanisms for nitrate (*nar*) and nitrite (*nir*) are present in some *Acidithiobacillia* [76] (Figure S2). Therefore, NtrB/NtrC may regulate these compounds' utilization by modulating related gene transcription. Though *ntrB-ntrC* clusters individually within genomes, there were no nearby genes associated with nitrogen metabolism. Additionally, the TCS NtrX/NtrY detects ammonia levels, demonstrated in the diazotrophs such as *Azospirillum brasilense* [77], and identified in the acidophilic *Leptospirillum ferriphilum* DSM 17947 [76]. NtrX/NtrY appeared widely in *Acidithiobacillia* (Figure 4), suggesting NtrY-mediated ammonia sensing and NtrX regulation. The *ntrX-ntrY* genes fell within a five or six gene operon across *Acidithiobacillia*, also lacking nearby nitrogen metabolism genes. Notably in *A. ferrooxidans* ATCC 23270, NtrX/NtrY was nearby to TcsS/TcsR. NtrC/NtrB and NtrX/NtrY TCSs share protein domains with those in *Herbaspirillum seropedicae* [73]. Thus, this supported the nitrate and nitrite regulatory activity via these TCS components found in *Acidithiobacillia*.

Flagellar motility enables chemotaxis and environmental signaling response in bacteria [78]. Flagellar regulation involves transcriptional cascades controlled by the bacterial enhancer-binding protein FleQ, demonstrated in organisms like *Pseudomonas aeruginosa* [79] and *Legionella pneumophila* [80]. Flagellar motility uniquely occurs in S-oxidizing *Acidithiobacillia*, which conserves the Class II flagellar genes [30]. These likely fall under regulation by RpoN, FleQ, FleS/FleR, and FliA [30]. Notably, the sigma54-dependent *fleS/fleR* genes were collocated with *fleQ*, suggesting a key conserved regulatory gene cluster. The bEBP FleR was proposed to control the transcription of the Class III flagellar rod encoding genes (*flgBCDEFGHIJKL*), which encode the flagellar basal body rod [30].

5. Conclusions

This work unveils previously uncharacterized regulatory circuits that may govern the dominance of extreme acidophiles in natural and anthropogenic extreme acidic environments. Sigma factors (SFs) and bacterial enhancer-binding proteins (bEBPs) orchestrate the regulation of key pathways involved in hydrogen, nitrogen, and sulfur metabolism in

the *Acidithiobacillia* class. These regulatory tools not only empower acidophiles to exploit metals and nutrients in industrial settings, but also likely played a role in early metabolic evolution and continue to shape contemporary acidophilic microbial communities via biogeochemical cycles. Through the comprehensive identification of SF family proteins and their associated TFs across 43 *Acidithiobacillia* genomes, this study offers insights into their distribution, evolutionary patterns, and potential involvement in speciation and clade-specific metabolic traits. The computational detection of sigma54-dependent TCSs, their bEBPs, and potential TFBSs suggested a coordinated regulatory framework governing essential metabolic functions. These elements are likely critical for energy acquisition and nutrient assimilation, underpinning survival and proliferation in extremely acidic environments. By revealing complex yet adaptable transcriptional regulatory networks, this research not only highlights a sophisticated acid life strategy but also lays the groundwork for future experimental validation of these computational predictions, aiming to confirm their physiological significance in *Acidithiobacillia*.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/microorganisms13061199/s1, Table S1: Properties of 43 *Acidithiobacillia* assemblies; Table S2: CheckM assessment results in *Acidithiobacillus* genus; Table S3: List of selected genes coding sequences involved in metabolic pathways; Figure S1: Phylogeny and reported phenotypes of 43 strains of *Acidithiobacillia* representatives; Figure S2: Phylogenetic distribution of selected proteins ortholog groups involved in assessed metabolic pathways in *Acidithiobacillia*. Supplementary material (Excel file), Table S4: Identification of Sigma54 activators and Response regulators across all strains. Each regulator type is presented in a separate section, side by side. Table S5: Domain and Family identification of Sigma Factor proteins

Author Contributions: J.H.V., D.S.H. and P.S.-R. conceived the study and drafted the manuscript. All authors contributed to data collection, analysis, and manuscript preparation. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Fondecyt 1181717 (DSH), ANID (Chilean National Research and Development Agency) ANILLO ATE220007 (JHV), ANID FONDECYT 3190792 (CGR), ANID 21180808 (PSR), and Centro Ciencia & Vida, FB210008, Financiamiento Basal para Centros Científicos y Tecnológicos de Excelencia de ANID, Dirección General de Asuntos del Personal Académico-Universidad Nacional Autónoma de México (IN-220523).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used in this study were obtained from public databases hosted by NCBI and are available in online repositories.

Acknowledgments: We want to thank the support given by grant ANID/FONDECYT 1221035.

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

Abbreviations

The following abbreviations are used in this manuscript:

TF	transcription factor
TFBS	transcription factor binding site
TCS	two-component system
SF	sigma factor
bEBP	bacterial enhancer binding proteins
HKS	histidine kinase sensor
DBD	DNA-binding domain
OG	orthologous group

TU	transcriptional unit
PWM	Position Weight Matrix

References

1. Goltsman, D.S.A.; Comolli, L.R.; Thomas, B.C.; Banfield, J.F. Community Transcriptomics Reveals Unexpected High Microbial Diversity in Acidophilic Biofilm Communities. *ISME J.* **2015**, *9*, 1014–1023. [CrossRef] [PubMed]
2. Quatrini, R.; Johnson, D.B. Microbiomes in Extremely Acidic Environments: Functionalities and Interactions That Allow Survival and Growth of Prokaryotes at Low pH. *Curr. Opin. Microbiol.* **2018**, *43*, 139–147. [CrossRef] [PubMed]
3. Huang, Y.; Li, X.T.; Jiang, Z.; Liang, Z.L.; Wang, P.; Liu, Z.H.; Li, L.Z.; Yin, H.Q.; Jia, Y.; Huang, Z.S.; et al. Key Factors Governing Microbial Community in Extremely Acidic Mine Drainage (pH < 3). *Front. Microbiol.* **2021**, *12*, 761579. [CrossRef]
4. Méndez-García, C.; Peláez, A.I.; Mesa, V.; Sánchez, J.; Golyshina, O.V.; Ferrer, M. Microbial Diversity and Metabolic Networks in Acid Mine Drainage Habitats. *Front. Microbiol.* **2015**, *6*, 475. [CrossRef]
5. Jerez, C.A. Biomining of Metals: How to Access and Exploit Natural Resource Sustainably. *Microb. Biotechnol.* **2017**, *10*, 1191–1193. [CrossRef]
6. Nguyen, T.H.; Won, S.; Ha, M.G.; Nguyen, D.D.; Kang, H.Y. Bioleaching for Environmental Remediation of Toxic Metals and Metalloids: A Review on Soils, Sediments, and Mine Tailings. *Chemosphere* **2021**, *282*, 131108. [CrossRef]
7. Nuñez, H.; Moya-Beltrán, A.; Covarrubias, P.C.; Issotta, F.; Cárdenas, J.P.; González, M.; Atavales, J.; Acuña, L.G.; Johnson, D.B.; Quatrini, R. Molecular Systematics of the Genus *Acidithiobacillus*: Insights into the Phylogenetic Structure and Diversification of the Taxon. *Front. Microbiol.* **2017**, *8*, 30. [CrossRef]
8. Williams, K.P.; Kelly, D.P. Proposal for a New Class within the Phylum Proteobacteria, *Acidithiobacillia* Classis Nov., with the Type Order Acidithiobacillales, and Emended Description of the Class Gammaproteobacteria. *Int. J. Syst. Evol. Microbiol.* **2013**, *63*, 2901–2906. [CrossRef] [PubMed]
9. Moya-Beltrán, A.; Beard, S.; Rojas-Villalobos, C.; Issotta, F.; Gallardo, Y.; Ulloa, R.; Giaveno, A.; Degli Esposti, M.; Johnson, D.B.; Quatrini, R. Genomic Evolution of the Class *Acidithiobacillia*: Deep-Branching Proteobacteria Living in Extreme Acidic Conditions. *ISME J.* **2021**, *15*, 3221–3238. [CrossRef]
10. González-Rosales, C.; Vergara, E.; Dopson, M.; Valdés, J.H.; Holmes, D.S. Integrative Genomics Sheds Light on Evolutionary Forces Shaping the *Acidithiobacillia* Class Acidophilic Lifestyle. *Front. Microbiol.* **2022**, *12*, 822229. [CrossRef]
11. Sepúlveda-Rebolledo, P.; González-Rosales, C.; Dopson, M.; Pérez-Rueda, E.; Holmes, D.S.; Valdés, J.H. Comparative Genomics Sheds Light on Transcription Factor-Mediated Regulation in the Extreme Acidophilic *Acidithiobacillia* Representatives. *Res. Microbiol.* **2024**, *175*, 104135. [CrossRef] [PubMed]
12. Dopson, M.; Johnson, D.B. Biodiversity, Metabolism and Applications of Acidophilic Sulfur-Metabolizing Microorganisms. *Environ. Microbiol.* **2012**, *14*, 2620–2631. [CrossRef]
13. Esparza, M.; Jedlicki, E.; Dopson, M.; Holmes, D.S. Expression and Activity of the Calvin-Benson-Bassham Cycle Transcriptional Regulator CbbR from *Acidithiobacillus ferrooxidans* in *Ralstonia eutropha*. *FEMS Microbiol. Lett.* **2015**, *362*, 3–8. [CrossRef] [PubMed]
14. Kupka, D.; Liljeqvist, M.; Nurmi, P.; Puhakka, J.A.; Tuovinen, O.H.; Dopson, M. Oxidation of Elemental Sulfur, Tetrathionate and Ferrous Iron by the Psychrotolerant *Acidithiobacillus* strain SS3. *Res. Microbiol.* **2009**, *160*, 767–774. [CrossRef] [PubMed]
15. Hedrich, S.; Johnson, D.B. *Acidithiobacillus ferridurans* Sp. Nov., an Acidophilic Iron-, Sulfur- and Hydrogen-Metabolizing Chemolithotrophic Gammaproteobacterium. *Int. J. Syst. Evol. Microbiol.* **2013**, *63*, 4018–4025. [CrossRef]
16. Leyn, S.A.; Novichkov, P.S.; Stepanova, V.V.; Ravcheev, D.A.; Kazakov, A.E.; Suvorova, I.A.; Rodionov, D.A. Comparative Genomics and Evolution of Transcriptional Regulons in Proteobacteria. *Microb. Genom.* **2016**, *2*, e000061. [CrossRef]
17. Paget, M.S. Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution. *Biomolecules* **2015**, *5*, 1245–1265. [CrossRef]
18. Zhang, N.; Buck, M. A Perspective on the Enhancer Dependent Bacterial RNA Polymerase. *Biomolecules* **2015**, *5*, 1012–1019. [CrossRef]
19. Doucleff, M.; Malak, L.T.; Pelton, J.G.; Wemmer, D.E. The C-Terminal RpoN Domain of Sigma54 Forms an Unpredicted Helix-Turn-Helix Motif Similar to Domains of sigma70. *J. Biol. Chem.* **2005**, *280*, 41530–41536. [CrossRef]
20. Paget, M.S.B.; Helmann, J.D. The Σ70 Family of Sigma Factors. *Genome Biol.* **2003**, *4*, 203. [CrossRef]
21. Otani, H.; Udwary, D.W.; Mouncey, N.J. Comparative and Pangenomic Analysis of the Genus *Streptomyces*. *Sci. Rep.* **2022**, *12*, 18909. [CrossRef]
22. Bush, M.; Dixon, R. The Role of Bacterial Enhancer Binding Proteins as Specialized Activators of σ 54-Dependent Transcription. *Microbiol. Mol. Biol. Rev.* **2012**, *76*, 497–529. [CrossRef] [PubMed]
23. Shingler, V. Signal Sensing by Sigma-54-Dependent Regulators Derepression as a Control Mechanism. *Mol. Microbiol.* **1996**, *19*, 409–416. [CrossRef] [PubMed]
24. Malhotra, A.; Severinova, E.; Darst, S.A. Crystal Structure of a Σ^{70} Subunit Fragment from *E. coli* RNA Polymerase. *Cell* **1996**, *87*, 127–136. [CrossRef]

25. Haldenwang, W.G. The Sigma Factors of *Bacillus subtilis*. *Microbiol. Mol. Biol. Rev.* **1995**, *59*, 1–30. [CrossRef]
26. Sauer, U.; Santangelo, J.D.; Treuner, A.; Buchholz, M.; Dürre, P. Sigma Factor and Sporulation Genes in *Clostridium*. *FEMS Microbiol. Rev.* **1995**, *17*, 331–340. [CrossRef]
27. Feklístov, A.; Sharon, B.D.; Darst, S.A.; Gross, C.A. Bacterial Sigma Factors: A Historical, Structural, and Genomic Perspective. *Annu. Rev. Microbiol.* **2014**, *68*, 357–376. [CrossRef]
28. Clark, A.E.; Adamson, C.C.; Carothers, K.E.; Roxas, B.A.P.; Viswanathan, V.K.; Vedantam, G. The Alternative Sigma Factor SigL Influences *Clostridioides difficile* Toxin Production, Sporulation, and Cell Surface Properties. *Front. Microbiol.* **2022**, *13*, 871152. [CrossRef]
29. Kazakov, A.E.; Rajeev, L.; Chen, A.; Luning, E.G.; Dubchak, I.; Mukhopadhyay, A.; Novichkov, P.S. Σ^{54}-Dependent Regulome in *Desulfovibrio vulgaris* Hildenborough. *BMC Genom.* **2015**, *16*, 919. [CrossRef]
30. Mangold, S.; Rao Jonna, V.; Dopson, M. Response of *Acidithiobacillus caldus* toward Suboptimal pH Conditions. *Extremophiles* **2013**, *17*, 689–696. [CrossRef]
31. Yang, C.L.; Chen, X.K.; Wang, R.; Lin, J.Q.; Liu, X.M.; Pang, X.; Zhang, C.J.; Lin, J.Q.; Chen, L.X. Essential Role of σ Factor Rpof in Flagellar Biosynthesis and Flagella-Mediated Motility of *Acidithiobacillus caldus*. *Front. Microbiol.* **2019**, *10*, 1130. [CrossRef]
32. Gao, R.; Stock, A.M. Biological Insights from Structures of Two-Component Proteins. *Annu. Rev. Microbiol.* **2009**, *63*, 133–154. [CrossRef] [PubMed]
33. Li, L.F.; Fu, L.J.; Lin, J.Q.; Pang, X.; Liu, X.M.; Wang, R.; Wang, Z.B.; Lin, J.Q.; Chen, L.X. The Σ^{54}-Dependent Two-Component System Regulating Sulfur Oxidization (Sox) System in *Acidithiobacillus caldus* and Some Chemolithotrophic Bacteria. *Appl. Microbiol. Biotechnol.* **2017**, *101*, 2079–2092. [CrossRef] [PubMed]
34. Kern, D.; Volkman, B.F.; Luginbuhl, P.; Nohaile, M.J.; Kustu, S.; Wemmer, D.E. Structure of a Transiently Phosphorylated Switch in Bacterial Signal Transduction. *Nature* **1999**, *402*, 894–897. [CrossRef]
35. Sallai, L.; Tucker, P.A. Crystal Structure of the Central and C-Terminal Domain of the Σ^{54}-Activator ZraR. *J. Struct. Biol.* **2005**, *151*, 160–170. [CrossRef]
36. Park, S.; Meyer, M.; Jones, A.D.; Yennawar, H.P.; Yennawar, N.H.; Nixon, B.T. Two-Component Signaling in the AAA+ ATPase DctD: Binding Mg^{2+} and BeF^{3-} Selects between Alternate Dimeric States of the Receiver Domain. *FASEB J.* **2002**, *16*, 1964–1966. [CrossRef]
37. Kilkenny, C.A.; Berger, D.K.; Rawlings, D.E. Isolation of the *Thiobacillus ferrooxidans NtrBC* Genes Using a *T. ferrooxidans NifH-LacZ* Fusion. *Microbiology* **1994**, *140*, 2543–2553. [CrossRef]
38. Parks, D.H.; Imelfort, M.; Skennerton, C.T.; Parks, D.H.; Tyson, G.W.; Hugenholtz, P. CheckM: Assessing the Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes. *Genome Res.* **2015**, *25*, 1043–1055. [CrossRef]
39. Parks, D.H.; Rinke, C.; Chuvochina, M.; Chaumeil, P.A.; Woodcroft, B.J.; Evans, P.N.; Hugenholtz, P.; Tyson, G.W. Recovery of Nearly 8000 Metagenome-Assembled Genomes Substantially Expands the Tree of Life. *Nat. Microbiol.* **2017**, *2*, 1533–1542. [CrossRef]
40. Seemann, T. Prokka: Rapid Prokaryotic Genome Annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [CrossRef]
41. Huerta-Cepas, J.; Forslund, K.; Coelho, L.P.; Szklarczyk, D.; Jensen, L.J.; Von Mering, C.; Bork, P. Fast Genome-Wide Functional Annotation through Orthology Assignment by EggNOG-Mapper. *Mol. Biol. Evol.* **2017**, *34*, 2115–2122. [CrossRef] [PubMed]
42. 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]
43. González, C.; Lazcano, M.; Valdés, J.; Holmes, D.S. Bioinformatic Analyses of Unique (Orphan) Core Genes of the Genus *Acidithiobacillus*: Functional Inferences and Use as Molecular Probes for Genomic and Metagenomic/Transcriptomic Interrogation. *Front. Microbiol.* **2016**, *7*, 2035. [CrossRef]
44. Castresana, J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Mol. Biol. Evol.* **2000**, *17*, 540–552. [CrossRef]
45. 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]
46. 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]
47. 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]
48. Mistry, J.; Finn, R.D.; Eddy, S.R.; Bateman, A.; Punta, M. Challenges in Homology Search: HMMER3 and Convergent Evolution of Coiled-Coil Regions. *Nucleic Acids Res.* **2013**, *41*, e121. [CrossRef]
49. Mistry, J.; Chuguransky, S.; Williams, L.; Qureshi, M.; Salazar, G.A.; Sonnhammer, E.L.L.; Tosatto, S.C.E.; Paladin, L.; Raj, S.; Richardson, L.J.; et al. Pfam: The Protein Families Database in 2021. *Nucleic Acids Res.* **2021**, *49*, D412–D419. [CrossRef]
50. Valdés, J.; Veloso, F.; Jedlicki, E.; Holmes, D. Metabolic Reconstruction of Sulfur Assimilation in the Extremophile *Acidithiobacillus ferrooxidans* Based on Genome Analysis. *BMC Genom.* **2003**, *4*, 51. [CrossRef]

51. Quatrini, R.; Valdés, J.; Jedlicki, E.; Holmes, D.S. The Use Of Bioinformatics And Genome Biology To Advance Our Understanding Of Bioleaching Microorganisms. In *Microbial Processing of Metal Sulfides*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 1–314. [CrossRef]
52. Valdés, J.; Pedroso, I.; Quatrini, R.; Dodson, R.J.; Tettelin, H.; Blake, R.; Eisen, J.A.; Holmes, D.S. *Acidithiobacillus ferrooxidans* Metabolism: From Genome Sequence to Industrial Applications. *BMC Genom.* **2008**, *9*, 597. [CrossRef] [PubMed]
53. Osbourn, A.E.; Field, B. Operons. *Cell. Mol. Life Sci.* **2009**, *66*, 3755–3775. [CrossRef] [PubMed]
54. Moinier, D.; Byrne, D.; Amouric, A.; Bonnefoy, V. The Global Redox Responding RegB/RegA Signal Transduction System Regulates the Genes Involved in Ferrous Iron and Inorganic Sulfur Compound Oxidation of the Acidophilic *Acidithiobacillus ferrooxidans*. *Front. Microbiol.* **2017**, *8*, 1277. [CrossRef]
55. Neira, G.; Cortez, D.; Jil, J.; Holmes, D.S. AciDB 1.0: A Database of Acidophilic Organisms, Their Genomic Information and Associated Metadata. *Bioinformatics* **2020**, *36*, 4970–4971. [CrossRef]
56. Bailey, T.L.; Williams, N.; Misleh, C.; Li, W.W. MEME: Discovering and Analyzing DNA and Protein Sequence Motifs. *Nucleic Acids Res.* **2006**, *34*, 369–373. [CrossRef]
57. Crooks, G.; Hon, G.; Chandonia, J.; Brenner, S. WebLogo: A Sequence Logo Generator Gavin. *Genome Res.* **2004**, *14*, 1188–1190. [CrossRef]
58. Schneider, T.D.; Stephens, R.M. Sequence Logos: A New Way to Display Consensus Sequences. *Nucleic Acids Res.* **1990**, *18*, 6097–6100. [CrossRef] [PubMed]
59. Nguyen, N.T.T.; Contreras-moreira, B.; Castro-mondragon, J.A.; Santana-garcia, W.; Ossio, R.; Robles-espinoza, C.D.; Bahin, M.; Collombet, S.; Vincens, P.; Thieffry, D.; et al. RSAT 2018: Regulatory Sequence Analysis Tools 20th Anniversary. *Nucleic Acids Res.* **2018**, *46*, 209–214. [CrossRef]
60. Upadhyay, A.A.; Fleetwood, A.D.; Adebali, O.; Finn, R.D.; Zhulin, I.B. Cache Domains That Are Homologous to, but Different from PAS Domains Comprise the Largest Superfamily of Extracellular Sensors in Prokaryotes. *PLoS Comput. Biol. J.* **2016**, *12*, e1004862. [CrossRef]
61. Davis, M.C.; Kesthely, C.A.; Franklin, E.A.; MacLellan, S.R. The Essential Activities of the Bacterial Sigma Factor. *Can. J. Microbiol.* **2017**, *63*, 89–99. [CrossRef]
62. Liu, X.; Matsumura, P. An Alternative Sigma Factor Controls Transcription of Flagellar Class-III Operons in *Escherichia coli:* Gene Sequence, Overproduction, Purification and Characterization. *Gene* **1995**, *164*, 81–84. [CrossRef] [PubMed]
63. Xia, J.L.; Peng, A.A.; He, H.; Yang, Y.; Liu, X.D.; Qiu, G.Z. A New Strain *Acidithiobacillus albertensis* BY-05 for Bioleaching of Metal Sulfides Ores. *Trans. Nonferrous Met. Soc. China* **2007**, *17*, 168–175. (In English) [CrossRef]
64. Valdés, J.H.; Pedroso, I.; Quatrini, R.; Hallberg, K.B.; Valenzuela, P.D.T.; Holmes, D.S. Insights into the Metabolism and Ecophysiology of Three Acidithiobacilli by Comparative Genome Analysis. *Adv. Mater. Res.* **2007**, *20–21*, 439–442. [CrossRef]
65. Valdés, J.; Pedroso, I.; Quatrini, R.; Holmes, D.S. Comparative Genome Analysis of *Acidithiobacillus ferrooxidans*, *A. Thiooxidans* and *A. caldus*: Insights into Their Metabolism and Ecophysiology. *Hydrometallurgy* **2008**, *94*, 180–184. [CrossRef]
66. Zhang, S.; Yan, L.; Xing, W.; Chen, P.; Zhang, Y.; Wang, W. *Acidithiobacillus ferrooxidans* and Its Potential Application. *Extremophiles* **2018**, *22*, 563–579. [CrossRef]
67. Rowe, O.F.; Sánchez-España, J.; Hallberg, K.B.; Johnson, D.B. Microbial Communities and Geochemical Dynamics in an Extremely Acidic, Metal-Rich Stream at an Abandoned Sulfide Mine (Huelva, Spain) Underpinned by Two Functional Primary Production Systems. *Environ. Microbioly* **2007**, *9*, 1761–1771. [CrossRef]
68. Wang, R.; Lin, J.Q.; Liu, X.M.; Pang, X.; Zhang, C.J.; Yang, C.L.; Gao, X.Y.; Lin, C.M.; Li, Y.Q.; Li, Y.; et al. Sulfur Oxidation in the Acidophilic Autotrophic *Acidithiobacillus* spp. *Front. Microbiol.* **2019**, *10*, 3290. [CrossRef]
69. Wang, Z.B.; Li, Y.Q.; Lin, J.Q.; Pang, X.; Liu, X.M.; Liu, B.Q.; Wang, R.; Zhang, C.J.; Wu, Y.; Lin, J.Q.; et al. The Two-Component System RsrS-RsrR Regulates the Tetrathionate Intermediate Pathway for Thiosulfate Oxidation in *Acidithiobacillus caldus*. *Front. Microbiol.* **2016**, *7*, 1755. [CrossRef] [PubMed]
70. Vignais, P.M.; Billoud, B. Occurrence, Classification, and Biological Function of Hydrogenases: An Overview. *Chem. Rev.* **2007**, *107*, 4206–4272. [CrossRef]
71. Vignais, P.M.; Elsen, S.; Colbeau, A. Transcriptional Regulation of the Uptake [NiFe]Hydrogenase Genes in *Rhodobacter capsulatus*. *Biochem. Soc. Trans.* **2005**, *33*, 28–32. [CrossRef]
72. Mackintosh, M.E. Nitrogen Fixation by *Thiobacillus ferrooxidans*. *J. Gen. Microbiol.* **1978**, *105*, 215–218. [CrossRef]
73. Bonato, P.; Alves, L.R.; Osaki, J.H.; Rigo, L.U.; Pedrosa, F.O.; Souza, E.M.; Zhang, N.; Schumacher, J.; Buck, M.; Wassem, R.; et al. The NtrY–NtrX Two-Component System Is Involved in Controlling Nitrate Assimilation in *Herbaspirillum seropedicae* Strain SmR1. *FEBS J.* **2016**, *283*, 3919–3930. [CrossRef] [PubMed]
74. Leigh, J.A.; Dodsworth, J.A. Nitrogen Regulation in Bacteria and Archaea. *Annu. Rev. Microbiol.* **2007**, *61*, 349–377. [CrossRef]
75. Drepper, T.; Groß, S.; Yakunin, A.F.; Hallenbeck, P.C.; Masepohl, B.; Klipp, W. Role of GlnB and GlnK in Ammonium Control of Both Nitrogenase Systems in the Phototrophic Bacterium *Rhodobacter capsulatus*. *Microbiology* **2003**, *149*, 2203–2212. [CrossRef]

76. Levicán, G.; Ugalde, J.A.; Ehrenfeld, N.; Maass, A.; Parada, P. Comparative Genomic Analysis of Carbon and Nitrogen Assimilation Mechanisms in Three Indigenous Bioleaching Bacteria: Predictions and Validations. *BMC Genom.* **2008**, *9*, 581. [CrossRef]
77. Ishida, M.L.; Assumpção, M.C.; Machado, H.B.; Benelli, E.M.; Souza, E.M.; Pedrosa, F.O. Identification and Characterization of the Two-Component NtrY/NtrX Regulatory System in *Azospirillum brasilense*. *Braz. J. Med. Biol. Res.* **2002**, *35*, 651–661. [CrossRef] [PubMed]
78. Nakamura, S.; Minamino, T. Flagella-Driven Motility of Bacteria. *Biomolecules* **2019**, *9*, 279. [CrossRef]
79. Jyot, J.; Dasgupta, N.; Ramphal, R. FleQ, the Major Flagellar Gene Regulator in *Pseudomonas aeruginosa*, Binds to Enhancer Sites Located Either Upstream or Atypically Downstream of the RpoN Binding Site. *J. Bacteriol.* **2002**, *184*, 5251–5260. [CrossRef]
80. Jacobi, S.; Schade, R.; Heuner, K. Characterization of the Alternative Sigma Factor Σ^{54} and the Transcriptional Regulator FleQ of *Legionella pneumophila*, Which Are Both Involved in the Regulation Cascade of Flagellar Gene Expression. *J. Bacteriol.* **2004**, *186*, 2540–2547. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

microorganisms

MDPI

Article

Revealing the Existence of Diverse Strategies for Phosphorus Solubilization and Acquisition in Plant-Growth Promoting *Streptomyces misionensis* SwB1

Yunzhu Chen [1,†], Zhuangzhuang Gao [1,2,†], Yan Yang [1], Qiang Liu [2], Lijuan Jiang [2], Jingzhen Chen [1], Xiao Zhou [1,2], Luhong Zhang [1,2], Yuena Ji [1], Jia Tu [1], Zhihong Xiao [1], Peiwang Li [1,*] and Changzhu Li [1,*]

[1] State Key Laboratory of Utilization of Woody Oil Resource, Hunan Academy of Forestry, Changsha 410004, China; cyzcarol@163.com (Y.C.); gaozz1129@foxmail.com (Z.G.); yangyanzupei@126.com (Y.Y.); chenjingzhen621@sina.com (J.C.); zx924647234@163.com (X.Z.); 20210100019@csuft.edu.cn (L.Z.); jyn13875854782@163.com (Y.J.); tujiashanito@gmail.com (J.T.); xiaozhihong@hnlky.cn (Z.X.)

[2] College of Life Science and Technology, Central South University of Forestry and Technology, 498 South Shaoshan Road, Changsha 410004, China; liu.qiangcs@163.com (Q.L.); znljiang2542@163.com (L.J.)

* Correspondence: lindan523@163.com (P.L.); lichangzhu2013@aliyun.com (C.L.)

† These authors contributed equally to this work.

Abstract: Phosphorus deficiency poses a significant challenge to plant growth and development, particularly in red soil. To alleviate this limitation, phosphorus-solubilizing bacteria (PSB) play a crucial role by converting insoluble phosphates present in the soil into soluble forms that are accessible to plants. *Cornus wilsoniana* Wangerin is a representative oil crop cultivated in red soil, holding a prominent position within China's forestry economic system. Consequently, it is essential to develop highly stable microbial phosphorus enhancement strategies to manage agricultural phosphorus in red soil regions, thereby maintaining the available phosphorus content necessary for the production of *C. wilsoniana*. In this study, the application of *Streptomyces misionensis* SwB1 bacterial suspension to the rhizosphere of *C. wilsoniana* significantly increased the content of various phosphorus fractions (H_2O-P, $NaHCO_3$-P, NaOH-P, HCl-P) in red soil, with $NaHCO_3$-P content increasing by 4.97 times and NaOH-P content by 3.87 times. Additionally, the genome of *S. misionensis* SwB1 contains 25 phosphorus-solubilizing genes, 13 nitrogen-fixing genes, 17 siderophore production genes, and 11 indole-3-acetic acid (IAA) production genes, indicating its potential for enhancing nutrient availability. Comparative genomic analysis of 15 strains belonging to five species of *Streptomyces* revealed that *S. misionensis* SwB1 possesses an extensive genetic repertoire and complete gene clusters associated with phosphorus solubilization. Furthermore, five phosphorus solubilization pathways of *S. misionensis* SwB1 were summarized: the Pst system, Pit system, siderophore transport, phosphatase synthesis, and organic acid synthesis. Ultimately, the inoculation of *S. misionensis* SwB1 significantly enhanced the growth and biomass accumulation of *C. wilsoniana* at the seedling stage, evidenced by an increase in fresh weight by 81.44%, a rise in net photosynthetic rate by 18.51%, and a surge in the number of root tips by 36.24%. Taken together, our findings support a sophisticated multi-pathway bacteria phosphorus solubilization approach and identified a highly efficient phosphorus-solubilizing strain, *S. misionensis* SwB1, which has the potential to become a microbial fertilizer.

Keywords: phosphate solubilizer; phosphorus mobilization mechanism; comparative genomics; *Cornus wilsoniana*; microbial fertilizer

1. Introduction

Nowadays, ensuring sufficient access to soil nutrients globally poses a significant challenge to the sustainability of modern agriculture [1]. Due to its crucial role in converting solar energy into food, fiber, and other plant products, phosphorus is the major nutrient that can limit plant growth despite its abundance in soils in both inorganic and organic forms [2]. In red soil, which covers approximately 22.7% of China's land area, insoluble phosphorus predominates, posing challenges for absorption and accumulation by plants [3]. Until now, the excessive application of chemical fertilizers and pesticides has directly and indirectly compromised the soil environment and crop yields [4]. Only 5–25% of the applied phosphorus can be absorbed by plants, while the remainder remains in the soil in insoluble forms [5]. Converting insoluble phosphorus into available phosphorus by phosphorus-solubilizing bacteria (PSB) is an effective way to enhance phosphorus utilization efficiency, soil structure, and plant production [6]. In modern agriculture, there is an urgent need for the use of phosphorus-solubilizing bacteria as biofertilizers to enhance crop productivity and augment nutritional security [7].

Phosphorus-solubilizing bacteria utilize various mechanisms to dissolve insoluble phosphorus [8]. These mechanisms include (i) organic acids, which can reduce pH and dissolve insoluble inorganic phosphorus. Gluconic acid, in particular, demonstrates effectiveness in enhancing the solubility of insoluble inorganic phosphorus in acidic soil and mitigating aluminum toxicity to plants [9]. Glucose dehydrogenase (GDH) catalyzes the synthesis of gluconic acid, requiring the combination with pyrrole quinoline quinone (PQQ) to form an active holoenzyme [10]. (ii) Phosphatases, which mineralize organic phosphorus and enhance the availability of soil organic phosphorus [11]. Most bacteria possess at least one of three phosphatases (PhoA, PhoX, and PhoD) exhibiting broad substrate specificity [12]. (iii) The Pst transport system, which is a protein channel responsible for phosphorus transportation and consumes energy [13]. Typically operating in low phosphorus environments, this system efficiently binds and transports inorganic phosphorus into the cytoplasm [14]. The phosphate-specific transport (Pst) system consists of peripheral-binding proteins (PstS), transmembrane proteins (PstA and PstC), and ATP-binding proteins (PstB), and regulatory protein (PhoU) [15].

Cornus wilsoniana Wangerin is a novel woody oil plant. The fruit oil of *C. wilsoniana* possesses a potent hypolipidemic effect, which is attributed to its high concentration of polyunsaturated fatty acids, particularly linoleic acid and γ-linolenic acid [16]. In China, *C. wilsoniana* is commonly cultivated in mountainous regions with red soil characterized by high iron content and low phosphorus availability. However, the production and quality of its fruit oil are significantly constrained by the phosphorus fixation by iron and other elements [17,18]. Thus, efficient and sustainable approaches to provide available phosphorus to *C. wilsoniana* are of great concern.

Our previous study identified a strain of *Streptomyces misionensis* SwB1, which has a highly efficient phosphorus solubilization ability [19]. However, the detailed mechanism of phosphorus solubilization mediated by *Streptomyces misionensis* SwB1 has not yet been completely elucidated. In the present study, genome sequencing and comparative genomic analysis were used to analyze the genetic composition characteristics and phosphorus solubilization mechanism of *Streptomyces misionensis* SwB1. Mechanistically, *Streptomyces misionensis* SwB1 solubilizes phosphorus by secreting organic acid, phosphatase, siderophore, and Pst phosphorus transport system. In addition, *Streptomyces misionensis* SwB1 has a variety of plant growth-promoting genes, which can effectively promote the biomass accumulation, photosynthesis, and root development of *C. wilsoniana*. Our findings support a basis for the research, development, and later application of this strain as a microbial fertilizer.

2. Materials and Methods

2.1. Plant Materials and Microorganisms

Cornus wilsoniana cultivar "Xianglin G1" saplings were provided by the *C. wilsoniana* germplasm repository of the Economic Forest Research Institute of Hunan Academy of Forestry (113°00′ E, 28°11′ N in Changsha, China), where the glasshouse experiment was conducted.

The *Streptomyces misionensis* SwB1 was grown on the International Streptomyces Program (ISP)-2 medium(MDBio, Qingdao, China) [20]. The matured spores were scraped off the colonies and stored in sterile glycerol at −80 °C.

2.2. Whole Genome Sequencing

The cryovial containing *S. misionensis* SwB1 is carefully removed from the −80 °C freezer and promptly thawed in a 37 °C water bath. Subsequently, in a strictly aseptic environment, the thawed bacterial suspension is meticulously transferred into pre-sterilized ISP-2 medium, and a sterile pipette is used to gently mix the suspension to ensure it is uniformly distributed throughout the ISP-2 medium. The conical flask is then carefully placed in a pre-calibrated incubator shaker set at 28 °C and 180 rpm to facilitate the efficient recovery and robust growth of the bacteria. After the *S. misionensis* SwB1 bacteria have been cultured to an appropriate density, the bacterial suspension is centrifuged at room temperature at 14,000× *g* for 5 min. The mycelium was then collected after discarding the medium, and DNA extraction and sequencing were performed. The sequencing was conducted by Frasergen, located at Wuhan, China. The PacBio Sequel II platform was employed for sequencing, followed by data processing using SMRTLINK 11.0.0 software [21,22]. After filtering low-quality and unknown sequences, the *S. misionensis* SwB1 genome was assembled using flye-2.9 and corrected by pilon-1.24 software based on paired reads [23,24].

For the strain *S. misionensis* SwB1 has been deposited in the China Center for Type Culture Collection under the accession number CCTCC NO. M 2023024. Furthermore, the genome sequence of *S. misionensis* SwB1 has been submitted to the NCBI database (GenBank ID: SAMN43297603).

Functional annotation was completed by aligning sequencing reads from the genome with the primary database. Essential databases used for *S. misionensis* SwB1 genome annotation include Non-Redundant Protein Sequence (NR), Clusters of Orthologous Groups (COG), Gene Ontology (GO), SwissProt, and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases [25–27]. Genes involved in soil microbial plant growth-promoting function were searched in the KEGG database (https://www.genome.jp/kegg/mapper/ (accessed on 18 May 2024)).

2.3. Characterization of Growth-Promoting Ability

To investigate the plant growth-promoting properties of *S. misionensis* SwB1, we employed a variety of specialized growth media that target different mechanisms of plant growth promotion. These media were designed to assess the strain's ability to solubilize inorganic and organic phosphorus, produce siderophores, synthesize indole-3-acetic acid (IAA), and promote growth through other means. All chemicals used for the preparation of these media were of analytical grade (MDBio, Qingdao, China). Below, we detail the composition of each medium and the observations made following the cultivation of *S. misionensis* SwB1.

The mycelia of *S. misionensis* SwB1 were added to media supplemented with inorganic phosphorus medium and organic phosphorus medium. After 7 days of culture, the formation of a transparent circle around the colony indicated phosphorus solubilization. The inorganic phosphorus medium included glucose 10 g, $(NH_4)_2SO_4$ 0.5 g, yeast extract

0.5 g, NaCl 0.3 g, KCl 0.3 g, $MgSO_4$ 0.3 g, $MnSO_4$ 0.03 g, $Ca_3(PO_4)_2$ 5 g, AGAR 15.0 g, deionized water 1 L, and pH = 7.0–7.5. The organic phosphorus medium included glucose 10 g, $(NH_4)_2SO_4$ 0.5 g, yeast extract 0.5 g, NaCl 0.3 g, KCl 0.3 g, $MgSO_4$ 0.3 g, $MnSO_4$ 0.03 g, $Ca_3(PO_4)_2$ 1 g, lecithin 0.2 g, AGAR 15.0 g, deionized water 1 L, and pH = 7.0–7.5.

The CAS detection medium for assessing siderophore production was composed of Chromium azurazine S (CAS) 0.0605 g, sodium dihydrogen phosphate dihydrate 0.29525 g, sodium hydrogen phosphate dodecahydrate 1.2135 g, HDTMA 0.0729 g, NaCl 0.0625 g, $FeCl_3 \cdot 6H_2O$ 0.002645 g, AGAR 9.0 g, NH_4Cl 0.125 g, KH_2PO_4 0.0375 g, deionized water 1 L, and pH adjusted to 6.7–6.9. After 7 days of culture, the formation of an orange-red zone around the colony signified siderophore production [28].

For the IAA production assay, the IAA assay medium included L-tryptophan 2.0 g, peptone 10.0 g, beef paste 3.0 g, NaCl 5.0 g, deionized water 1 L, and pH = 7.0–7.4 [29]. After 7 days of cultivation, the culture filtrate was obtained by centrifugation at 4000 rpm for 20 min, 1 mL of the supernatant of the centrifuged medium was taken, and 1 mL of Salkowski reagent was added. After the reaction in the dark for 30 min, the supernatant turned pink, indicating that the strain could produce IAA.

Lastly, the Ashby medium was used to assess the overall plant growth-promoting potential and was composed of KH_2PO_4 0.2 g, NaCl 0.2 g, $MgSO_4$ 0.2 g, K_2SO_4 0.1 g, $CaCO_3$ 5.0 g, glucose 10.0 g, AGAR 15.0 g, deionized water 1 L, and pH = 7.2–7.6, 25 °C. The normal growth of the strain after 7 days of culture indicated that the strain had the ability to fix nitrogen [30].

2.4. Comparative Genomic Analysis

Genomic data for 14 *Streptomyces* strains across 5 species were downloaded from NCBI and annotated using Prokka 1.12 [31]. Homology analysis of the genome was performed using OrthoFinder 2.5.2 [32], followed by drawing a petal map in R and plotting the gene clusters with ChiPlot (https://www.chiplot.online/ (accessed on 22 May 2024)). Evolutionary analysis was performed using MEGA 11 (developed by the Molecular Evolutionary Genetics Analysis Team). The Average Nucleotide Identity (ANI) of each strain was calculated by CompareM 0.1.2 (developed by the DOE Joint Genome Institute).

2.5. Pot Assay

2.5.1. Glasshouse Experimental Design

Spontaneous germination of *C. wilsoniana* saplings with uniform height was transplanted to 2.5 L plastic pots with one sapling per pot under glasshouse conditions. The substrate was sterilized with red soil. Four treatments were set up: red soil + water (CK_0), red soil + *S. misionensis* SwB1 ($SwB1_0$), red soil + water + seedlings (CK), red soil + *S. misionensis* SwB1 + seedlings (SwB1). Fifty milliliters of spore suspension (107 spore mL^{-1}) of *S. misionensis* SwB1 was applied to the base of each sapling (CK, SwB1), and the same volume of water was added to the control. The plants were grown in a greenhouse at 27 °C under 16 h light/8 h dark photoperiod, and relative humidity of 50–80%. All plants were regularly watered without any fertilizers. Each treatment had 10 replicate samplings, and all pots were completely randomized. After 60 days of growth, all plants were harvested.

2.5.2. Determination of Soil Phosphorus Content

Starting from the base of the seedling, carefully dig with a small spade to expose the roots. Gently move away the large soil clumps around the roots, taking care not to disturb the root growth area. Lightly remove the large soil particles attached to the roots. Use a soft-bristled brush to gently brush off the fine soil particles adhering to the roots; these particles are the rhizosphere soil samples. These rhizosphere soils will be used to determine soil phosphorus content, including H_2O-P, $NaHCO_3$-P, NaOH-P, and HCl-P

content were determined. The phosphorus was sequentially extracted by H_2O, $NaHCO_3$, NaOH, and HCl, and the phosphorus content was determined by molybdenum antimony resistance colorimetry [33]. Soil samples were collected every five days and measured. The measurement was performed in triplicates.

2.5.3. Plant Growth Parameter Analyses

The height of the seedling and ground-level stem diameters were measured. The whole plant was weighed for fresh biomass and then dried at 60 °C until constant weight for dry biomass determinations. Each treatment was performed with ten plants.

Three healthy leaves in the upper part of the saplings were selected for the determination of photosynthetic parameters. The net photosynthetic rate, intercellular CO_2 concentration, and stomatal conductance were measured using a portable photosynthesis system Li-6400 [34] at 60% RH, 25 °C air temperature and 400 μmol mol^{-1} CO_2. All measurements were carried out in the morning from 9:00 a.m. to 11:00 a.m. [35]. Each treatment was performed with ten plants.

Root morphological parameters, including total root length, root area, root volume, and the number of tips, were measured using a root automatism scanner (Epson Twain Pro 2.10, Purchased from Wseen, Hangzhou, China) coupled with WinRHIZO™2000 Professional Edition program software. Each treatment was performed with ten plants.

2.6. Statistical Analysis

For the pot assay, data are mean values of independent experimental repetitions. Depending on the experiments, differences among treatments were reanalyzed by unpaired *t*-test or one-way ANOVA, followed by Tukey's multiple range, using GraphPad Prism 9.5.0 and taking $p \leq 0.0001$, $p \leq 0.001$, $p \leq 0.01$ or $p \leq 0.05$, based on unpaired experimental design, as significant.

3. Results and Discussion

3.1. Effects of S. misionensis SwB1 on Phosphorus-Solubilizing in Red Soil

In a previous study, we isolated a total of 55 culturable phosphorus-solubilizing bacteria from the rhizosphere soil of *C. wilsoniana*. Among these isolates, we identified a strain of *S. misionensis* SwB1, which exhibited the highest efficiency in phosphorus solubilization and demonstrated significant potential to enhance plant growth [19].

According to the Hedley phosphorus classification, inorganic phosphorus components can be categorized into four distinct categories: active inorganic phosphorus (H_2O-P, $NaHCO_3$-P), medium stable inorganic phosphorus (NaOH-P), stable inorganic phosphorus (HCl-P), and residual phosphorus (Figure 1A) [33]. H_2O-P represents phosphorus in its ionic form within the soil, while $NaHCO_3$-P refers to phosphorus that is adsorbed onto the soil surface and can be readily absorbed and utilized by plants. NaOH-P, on the other hand, is strongly attached to the soil surface through chemisorption. HCl-P is a solid phosphorus compound often coated with an iron and aluminum oxide film for closed storage [36,37].

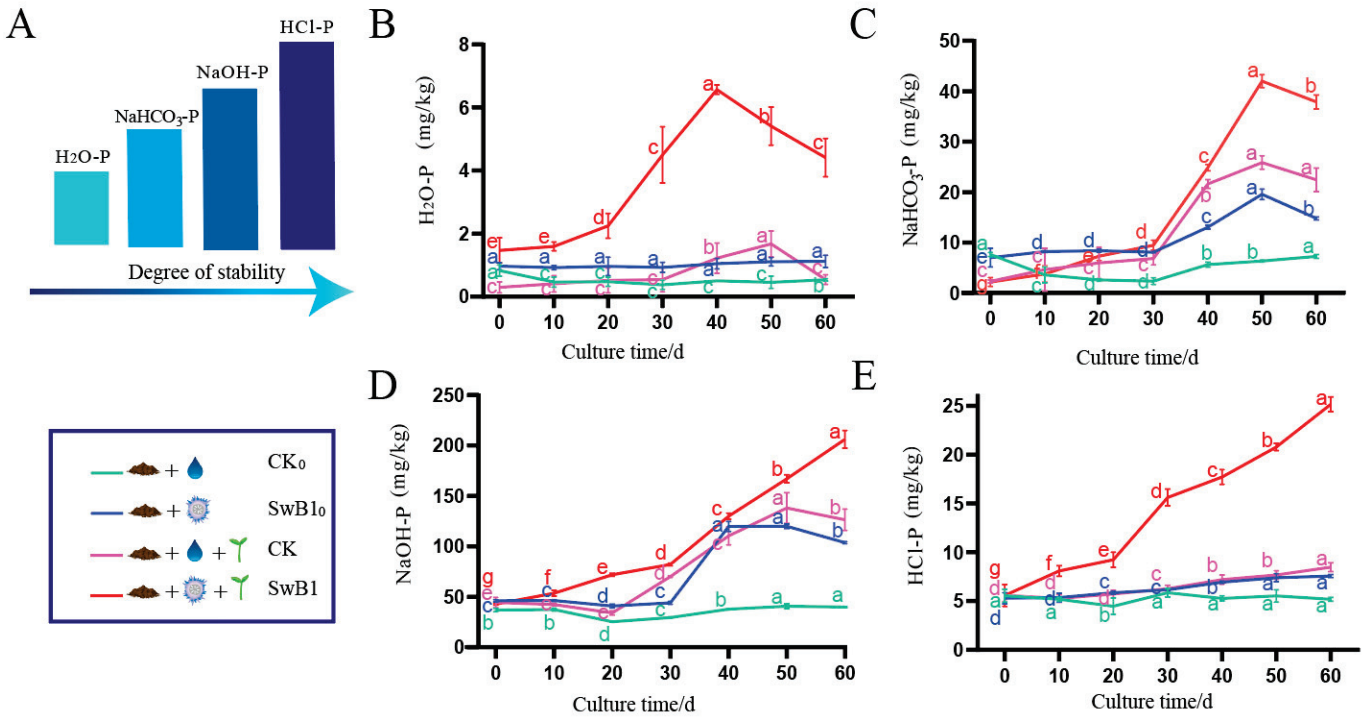


Figure 1. Effect of *S. misionensis* SwB1 on phosphorus content levels in red soil. (**A**) Phosphorus solubility schematic: HCl-P, NaOH-P, $NaHCO_3$-P, and H_2O-P, where darker colors indicate lower solubility. (**B**–**E**) Changes in H_2O-P, $NaHCO_3$-P, NaOH-P, and HCl-P content in soil samples under different treatments. Treatments included: red soil + water (CK_0), red soil + SwB1 ($SwB1_0$), red soil + water + seedlings (CK), and red soil + SwB1 + seedlings (SwB1), with three replicates per treatment. The letter markers indicate the significant differences determined by the ANOVA post hoc Tukey's test, where the same letter denotes no significant difference ($p > 0.05$), and different letters indicate a significant difference ($p \leq 0.05$). The error bars represent the standard error.

To better understand the ability of *S. misionensis* SwB1 to solubilize various forms of phosphorus in red soil, this study measured the change in different levels of phosphorus (H_2O-P, $NaHCO_3$-P, NaOH-P, HCl-P) in red soils subjected to various treatments (Figure 1). After 60 days of incubation, the H_2O-P content varied among the four treatment groups. The SwB1 group exhibited the highest increase at 2.98 times, followed by the CK group at 19.86%, and the $SwB1_0$ group at 15.82%. In contrast, the CK_0 group experienced a decrease of 44.35% (Figure 1B). Similarly, the $NaHCO_3$-P content also showed varying changes; the SwB1 group experienced the highest increase at 4.97 times, followed by the CK group with an increase of 2.75 times, and the $SwB1_0$ group with an increase of 1.11 times. The CK_0 group decreased again, this time by 4.86% (Figure 1C). The NaOH-P content demonstrated significant changes as well, increasing by 3.87 times in the SwB1 group, 1.82 times in the CK group, and 1.38 times in the $SwB1_0$ group, while the CK_0 group saw a modest increase of 6.64% (Figure 1D). The HCl-P content also changed, with an increase of 3.51 times in the SwB1 group, 51.11% in the CK group, 43.26% in the $SwB1_0$ group and increased by 6.78% in CK_0 group (Figure 1E). Comparisons between the CK_0 and $SwB1_0$ treatment groups revealed that *S. misionensis* SwB1 significantly enhanced phosphorus content across all levels, especially in the $NaHCO_3$-P and NaOH-P fractions. This indicates that the strain possesses the ability to alleviate the low phosphorus dilemma in red soil. Furthermore, comparisons between the CK and SwB1 groups demonstrated that the interaction between *C. wilsoniana* saplings and *S. misionensis* SwB1 strains facilitated the dissolution of insoluble phosphorus in red soil. It is possible that *S. misionensis* SwB1 promotes the growth of *C. wilsoniana* by improving phosphorus solubilization, and the well-grown *C. wilsoniana* also

provides a beneficial living environment for *S. misionensis* SwB1, establishing a mutually beneficial relationship in the environment. Surprisingly, across all treatments, the CK group contributed to a comparable or even higher level of phosphorus content at various levels than the $SwB1_0$ group. This may be attributed to the fact that the root exudates of *C. wilsoniana* seedlings include a wide array of organic acids, phenolic compounds, hormones, and other secondary metabolites [38]. These substances can more effectively solubilize the insoluble phosphorus in the soil. In contrast, the phosphorus-solubilizing ability of phosphate-solubilizing bacteria typically relies on specific enzymes or metabolic pathways, which may not be as diverse as the secretions from *C. wilsoniana* seedlings [39].

3.2. Genomic Characteristics and Functional Annotation

The whole genome sequencing resulted in the assembly of a complete circular genome of *S. misionensis* SwB1(Figure 2), with a total length of 8,633,379 bp and a GC content of 72.45% (Table S1). The genome contains 8100 coding DNA sequence (CDS) genes, 82 tRNA genes, and 18 rRNA genes. Furthermore, a total of 12 gene islands, 62 CRISPR sequences, and 30 gene clusters were predicted within the genome. Notably, 72.89% of the genes were annotated in the COG database, and these genes were categorized into 26 functional classes (Figure 3A). The top five categories included transcription (875 genes), general function prediction only (748 genes), carbohydrate transport and metabolism (598 genes), amino acid transport and metabolism (510 genes), and signal transduction mechanisms (473 genes). Additionally, KEGG classification identified 2676 genes across six categories: cellular processes (CP), environmental information processing (EP), genetic information processing (GP), human diseases (HD), metabolism (MB), and organismal systems (OS) (Figure 3B).

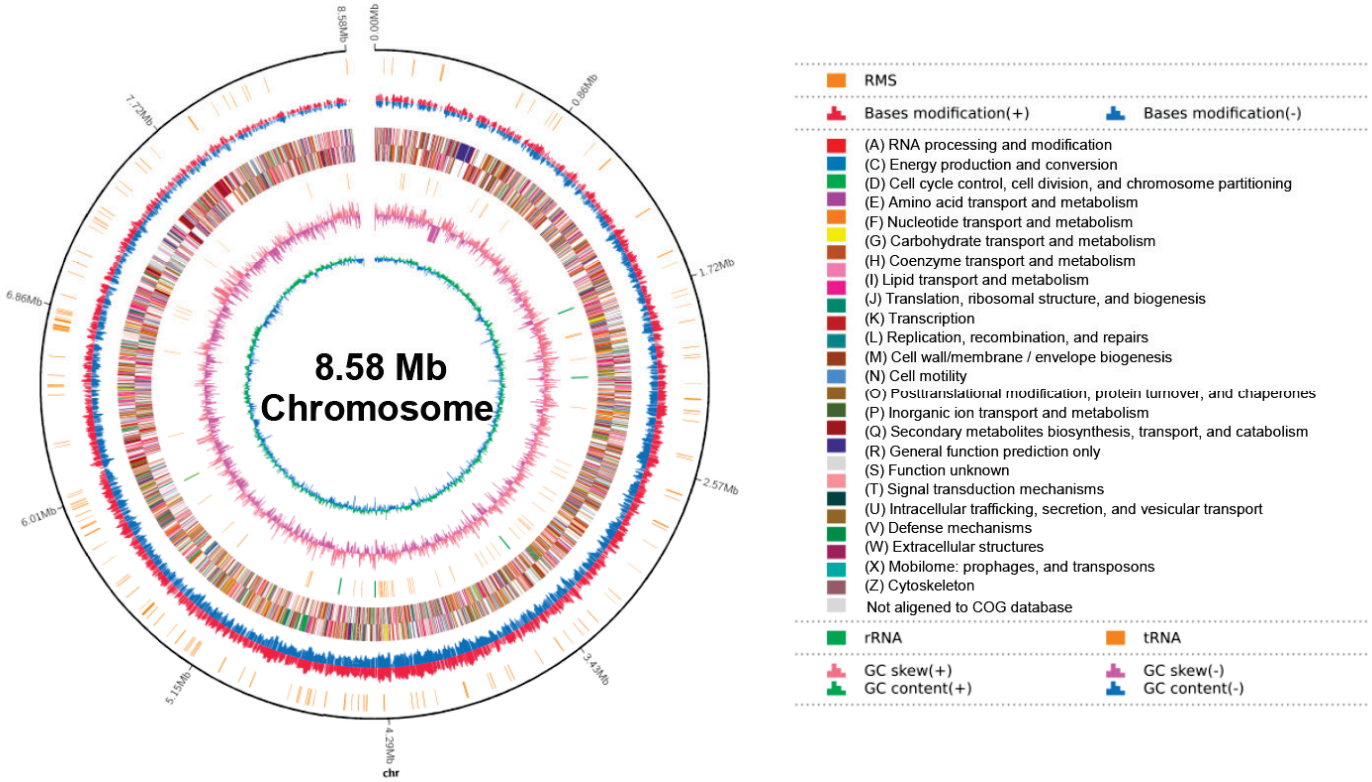


Figure 2. Circular representation of the *S. misionensis* SwB1 chromosome. The rings, labeled from the inside out, Ring 1, GC content (green for positive strand, blue for negative strand); Ring 2, GC skew (pink for positive strand, purple for negative strand); Ring 3, rRNA and tRNA (green and orange vertical lines, respectively); Rings 4 and 5, COG database annotations for positive and negative strands; Ring 6, base modifications (red for positive strand, blue for negative strand); Ring 7, enzymes associated with restriction-modification systems (orange vertical line).

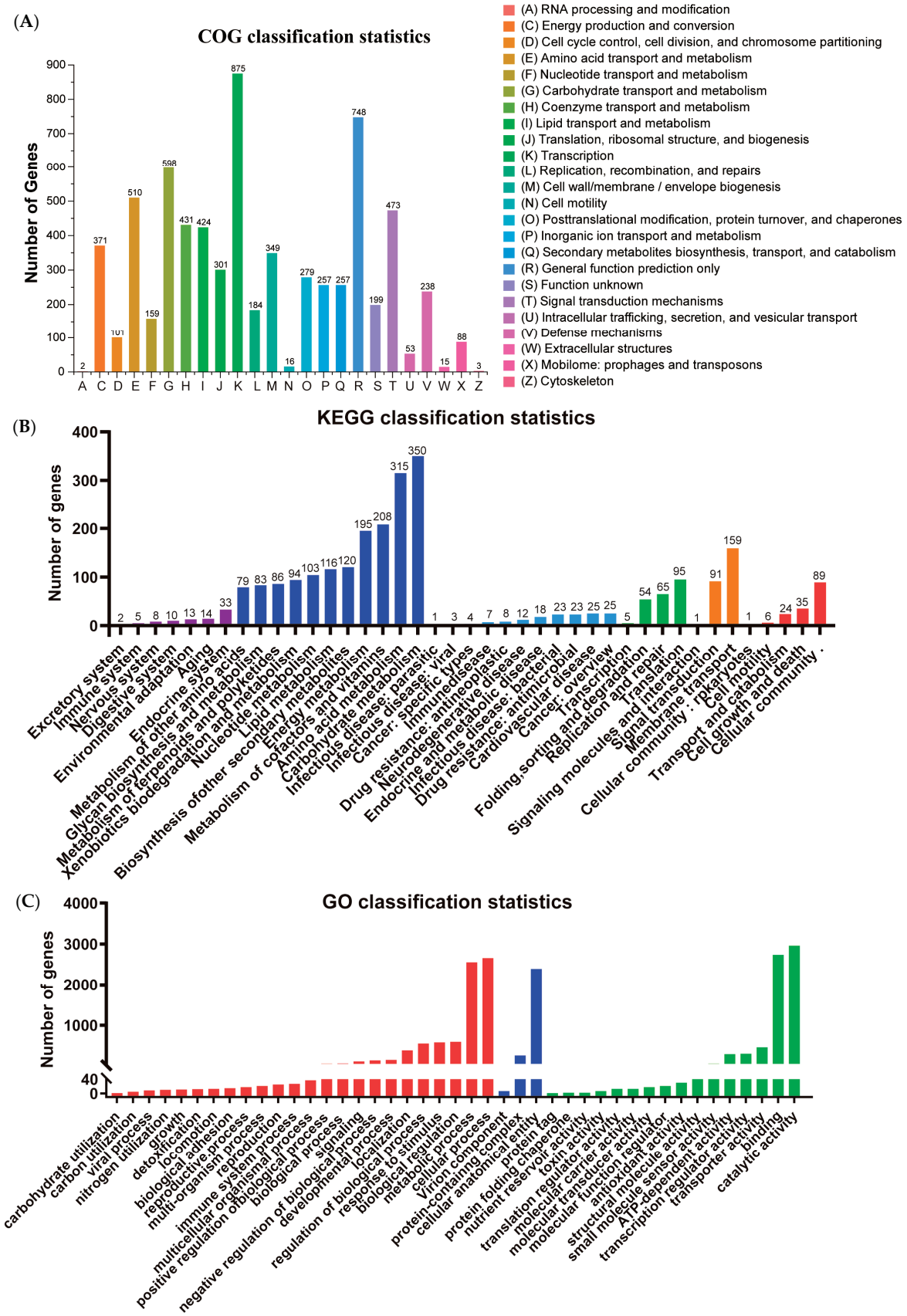


Figure 3. Annotated map of functional classification of *S. misionensis* SwB1 in whole genome. (**A**) COG functional annotation. (**B**) KEGG functional annotation. (**C**) GO function annotation. The *X*-axis represents GO function classifications, while the *Y*-axis indicates the number of genes associated with each classification. The numbers above each bar correspond to the total number of genes within each category.

A total of 4641 genes were annotated into three categories based on Gene Ontology (GO): Biological Process (BP), Cellular Component (CC), and Molecular Function (MF) (Figure 3C). The top five categories identified were catalytic activity (2958 genes), binding (2734 genes), cellular process (2657 genes), metabolic process (2553 genes), and cellular anatomical entity (2389 genes).

3.3. Evolutionary Analysis and Growth-Promoting Gene Analysis

The complete 16S rRNA sequence (1514 bp) of *S. misionensis* SwB1 was subjected to similarity-based searches in the NCBI database, and an evolutionary tree was constructed based on 16S rRNA sequences (Figure S1). The results showed that *S. misionensis* SwB1 is classified as *Streptomyces misionensis* with 100% cladistic confidence. To further elucidate the taxonomic position of *S. misionensis* SwB1, a phylogenetic tree was generated using Average Nucleotide Identity (ANI) analysis based on 15 complete genomes from other *Streptomyces* species (Figure 4A). As depicted in Figure 4A, *S. misionensis* SwB1, along with *Streptomyces misionensis* VOGW01, *Streptomyces misionensis* FNTD01, and *Streptomyces misionensis* 66, formed a monophyletic clade and was deeply nested within the tree, indicating a close evolutionary relationship. In contrast to *S. misionensis* SwB1, which was isolated from the rhizosphere soil, *Streptomyces misionensis* VOGW01, *Streptomyces misionensis* FNTD01, and *Streptomyces misionensis* 66 were isolated from the oasis and the ocean [40–42]. Given the differences in their respective niches, a comparative genome analysis of these strains may provide valuable insights into the specific adaptations and genomic evolution of *S. misionensis* SwB1.

To investigate the growth-promoting capabilities of *Streptomyces*, we conducted a comparative analysis in silico of the strain characteristics associated with phosphorus solubilization, nitrogen fixation, siderophore production, and indole-3-acetic acid (IAA) production among various *Streptomyces* strains (Figure 4A). The results revealed that *S. misionensis* SwB1 possessed the highest number of phosphorus-solubilizing genes (25), whereas *Streptomyces phaeochromogenes* JAUSYT01 exhibited the greatest number of nitrogen-fixing genes (17). Furthermore, *Streptomyces phaeochromogenes* LIQZ01 had the most siderophore production genes (22), and *Streptomyces koyangensis* SCSIO 5802 contained the highest number of IAA-producing genes (14). Moreover, the results indicated that the other *Streptomyces* strains possessed a greater number of genes related to nitrogen fixation, siderophore production, and IAA production in comparison to *S. misionensis* SwB1. In contrast, *S. misionensis* SwB1 demonstrated more advantageous characteristics regarding phosphorus solubilization. This result explained that *S. misionensis* SwB1 showed high efficiency in phosphorus solubilization in the red soil phosphorus transformation experiment (Figure 1).

The genomic annotation of *S. misionensis* SwB1 revealed the presence of genes associated with phosphorus solubilization, nitrogen fixation, siderophore production, and IAA production (Table 1 and Tables S2–S4). Consequently, this study investigated the role of *S. misionensis* SwB1 in promoting plant growth through experiments conducted on Petri plates. The capability to solubilize phosphorus and fix nitrogen may alter soil nutrient composition and structure, thereby enhancing the availability of soil nutrients [43]. In the phosphate solubilizing assay, the formation of a clear halo surrounding the colony indicates that *S. misionensis* SwB1 possesses the ability to solubilize both organic and inorganic phosphorus (Figure 4B,C). *S. misionensis* SwB1 was able to grow well in an Ashby medium, demonstrating the strain's capability to fix nitrogen (Figure 4D). Furthermore, the secretion of siderophores and IAA has the potential to promote plant growth and enhance agronomic traits as well as stress resistance [44]. It also formed an orange halo on CAS agar medium, indicating siderophore production (Figure 4E). After 7 days of incubation with IAA detection medium inoculated with *S. misionensis* SwB1, the medium

turned pink, whereas the control medium without *S. misionensis* SwB1 exhibited no color change, indicating the production of IAA (Figure 4F). In summary, the plant-promoting capabilities of *S. misionensis* SwB1, including phosphorus solubilization, nitrogen fixation, siderophores production, and IAA production, are consistent with the gene annotation results, thereby highlighting the potential of this strain as a microbial fertilizer. Previous studies have demonstrated that *Streptomyces misionensis* exhibits anticancer, insecticidal, and antibacterial activities [45–48]. However, this study represents the first investigation to highlight the plant growth-promoting properties, particularly its exceptional capability in phosphorus solubilization.

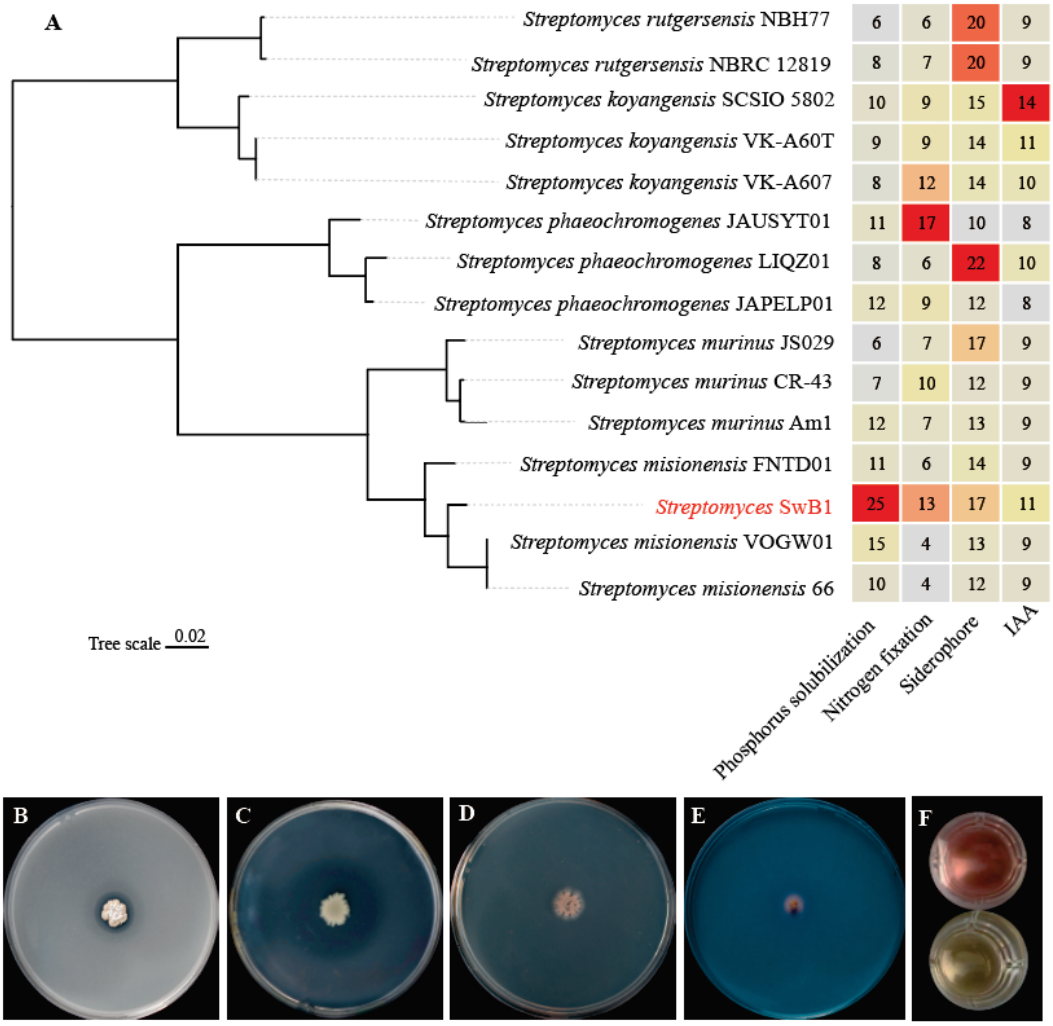


Figure 4. Characterization of plant growth-promoting the ability of *S. misionensis* SwB1. (**A**) phylogenetic relationships and growth-promoting genes among 15 *Streptomyces* strains. The cells in the heatmap are displayed in different colors, with hues closer to red indicating higher values at that position, and hues closer to gray indicating lower values. The values in the cells represent the number of growth-promoting genes in each strain. The text in red denotes the strains used in this study. (**B**) Inorganic phosphate solubilization. (**C**) Organic phosphate solubilization. (**D**) Siderophore production. (**E**) Nitrogen fixation. (**F**) IAA production.

Table 1. Phosphorus-solubilizing genes of *S. misionensis* SwB1.

Gene Name	Gene Annotation	Functional Classification	Gene Number
pstA	2-amino-1-hydroxymethyl phosphonate dioxygenase	Phosphorus transport	1
pstB	Phosphinothricin acetyltransferase	Phosphorus transport	1
pstC	Phosphinothricin acetyltransferase	Phosphorus transport	1
pstS	N-acetylglucosamine PTS system EIICBA or EIICB component	Phosphorus transport	1
phoU	Phosphate-specific transport system accessory protein PhoU	Phosphorus transport	1
pit	N-acetylglucosamine PTS system EIICBA or EIICB component	Phosphorus transport	1
phoA	Alkaline phosphatase H	Phosphatase synthesis	1
phoD	Alkaline phosphatase D	Phosphatase synthesis	2
phoX	Secreted phosphatase, PhoX family	Phosphatase synthesis	1
phoE	Broad specificity phosphatase PhoE	Phosphatase synthesis	10
gcd	Glucose dehydrogenase, PQQ-dependent	Gluconic acid synthesis	1
pqqB	Pyrroloquinoline quinone biosynthesis protein B	Gluconic acid synthesis	1
pqqC	Pyrroloquinoline-quinone synthase	Gluconic acid synthesis	1
pqqD	Pyrroloquinoline quinone biosynthesis protein D	Gluconic acid synthesis	1
pqqE	PqqA peptide cyclase	Gluconic acid synthesis	1

3.4. Comparative Genomics Analysis and Phosphorus Solubilization Gene Clusters

The mineralization of organic phosphorus and dissolution of inorganic phosphorus are critical mechanisms through which rhizosphere bacteria enhance soil AP content [49]. Twenty-five genes associated with phosphorus transport, phosphatase synthesis, and gluconic acid production were identified in the *S. misionensis* SwB1 genome (Table 1). Genes involved in phosphorus transport include *pstA*, *pstB*, *pstC*, *pstS*, *phoU*, and *pit*, which are part of the phosphate special transport system (Pst) and phosphate inorganic transport system (Pit) [50,51]. Among these, four paralogous genes are associated with phosphatase activity: *phoA*, *phoD*, *phoX*, and *phoE*. Additionally, gluconic acid synthesis includes the gluconate dehydrogenase gene (*gcd*) and cofactor synthesis genes: *pqqB*, *pqqC*, *pqqD*, and *pqqE* [52,53]. To investigate the plant growth-promoting genes of *S. misionensis* SwB1 further, a comparison was made with those of 15 Streptomyces strains from the NCBI database and previous studies. These 15 *Streptomyces* strains collectively contain 3195 core genes, accounting for 41.3% to 62.3% of their genome (Figure S2). The Pst phosphorus transport system, gluconic acid synthesis pathway, and phosphatase genes were analyzed across these 15 *Streptomyces* strains to explore their roles in phosphorus solubilization.

The Pst phosphorus transport system, comprising four proteins (PstA, PstB, PstC, and PstS), serves as the primary mechanism for phosphorus absorption in microorganisms under low-phosphorus conditions [54]. The Pst phosphorus transport system is anchored in the inner cell membrane by two multi-spanning proteins, PstC, and PstA, forming a transmembrane (TM) channel that facilitates phosphorus transport across the cytoplasmic membrane [55]. PstB is responsible for ATP hydrolysis, which energizes the TM channel, enabling it to function as a phosphorus transporter [56,57]. PstS is a peripheral binding protein with a high affinity for phosphorus, responsible for sensing and binding phosphorus in the periplasmic space [58]. In this study, fifteen strains of *Streptomyces*, including *Streptomyces murinus* and *Streptomyces misionensis*, possess a complete Pst phosphorus transport system (Figure 5A). Additionally, the homology and arrangement of genes in the Pst phosphorus transport system were entirely consistent among the four *Streptomyces*

misionensis strains, including *S. misionensis* SwB1. However, some strains of *Streptomyces phaeochromogenes*, along with both *Streptomyces rutgersensis* and *Streptomyces koyangensis*, were deficient in the *pstA* gene. The *pstA* gene encodes an integral membrane protein with six transmembrane helices, and its disruption or loss decreases the protein's affinity for membrane insertion, consequently impacting the Pst system, thereby reducing phosphorus transport and affecting the phosphorus uptake rate in microorganisms [59,60].

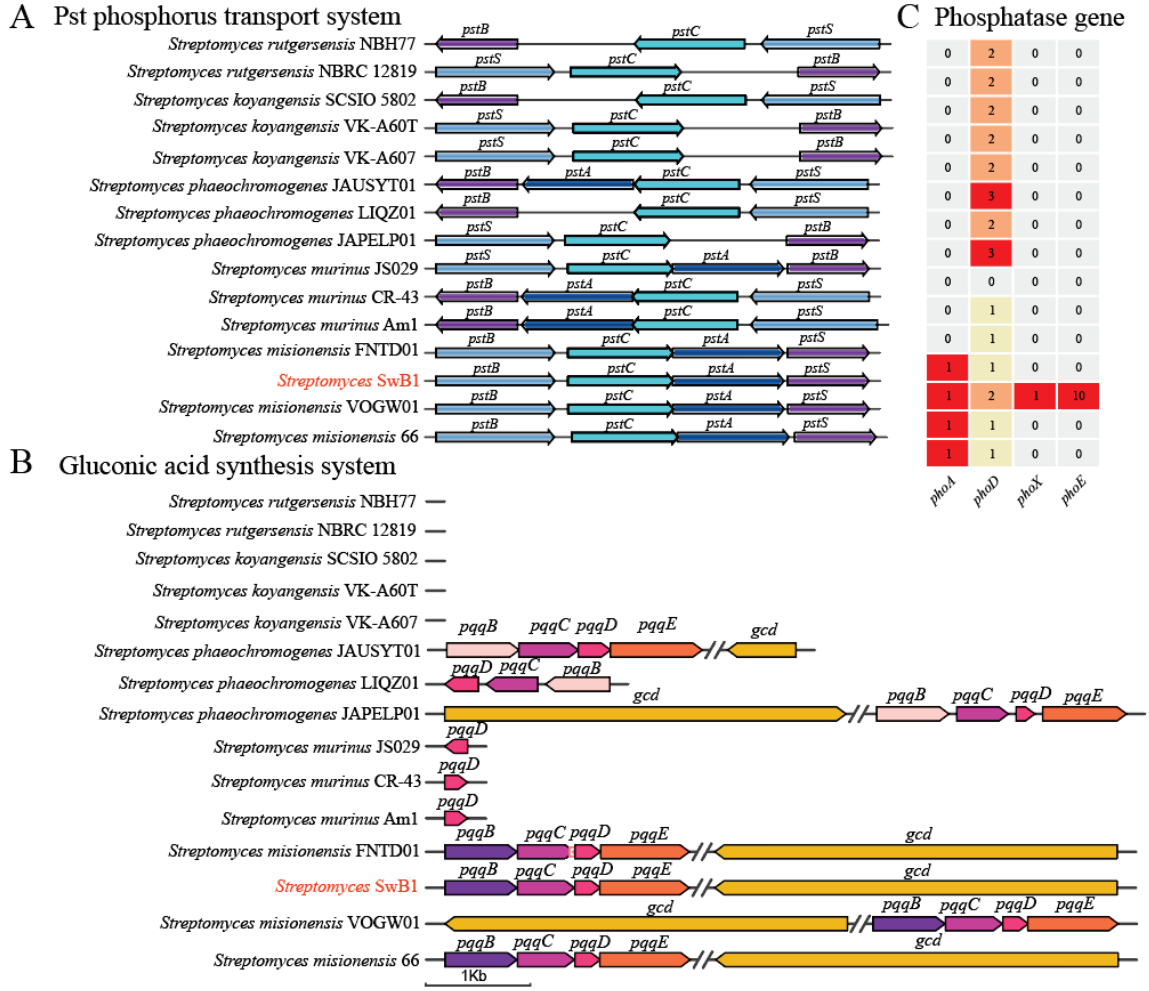


Figure 5. Comparison of phosphorolytic gene clusters and phosphatase genes among 15 *Streptomyces* strains. (**A**) Pst phosphorus transport system gene cluster. The length of the arrows corresponds to the length of the respective genes. The arrow's direction indicates the gene's orientation (5′ to 3′ from left to right). Each arrow is labeled with the corresponding gene name, and different genes are represented by different colors. The text in red denotes the strains used in this study. (**B**) Gluconic acid synthesis gene cluster. The text in red denotes the strains used in this study. (**C**) Heat map illustrating the types and quantities of phosphatase genes. The numbers in the heat map represent the quantity of genes. Darker shades correspond to a higher number of genes represented. The cells in the heatmap are displayed in different colors, with hues closer to red indicating higher values at that position, and hues closer to gray indicating lower values. The values in the cells represent the number of growth-promoting genes in each strain.

The gluconic acid synthesis system regulates the production of gluconate dehydrogenase and its coenzymes [61]. The *gcd* gene is responsible for regulating the production of glucose dehydrogenase in bacteria, and Pyrroloquinoline quinone (PQQ) serves as a coenzyme for glucose dehydrogenase. The synthesis of PQQ in bacteria requires the participation of the *pqqB*, *pqqC*, *pqqD*, and *pqqE* genes [62–64]. Strains of *Streptomyces misionensis* and *Streptomyces phaeochromogenes* possess a complete gluconic acid synthesis

system. Interestingly, the number of genes in the gluconic acid synthesis system of the four *Streptomyces misionensis* strains was found to be completely consistent; however, the orientation of the PQQ gene in *Streptomyces misionensis* VOGW01 was found to be reversed. Strains of *Streptomyces phaeochromogenes* have a homologous PQQ synthesis gene, but the lengths of the *gcd* genes differed, with the *gcd* gene in *Streptomyces phaeochromogenes* LIQZ01 completely absent (Figure 5B), which may hinder gluconic acid synthesis for phosphorus dissolution [62]. Strains of *Streptomyces rutgersensis* and *Streptomyces koyangensis* may lack a complete gluconic acid synthesis system, which results in the loss of the ability to synthesize glucose dehydrogenase. This leads to the inability of the strain to convert to gluconic acid, thereby reducing its capacity to secrete organic acid-soluble phosphorus. These two bacterial strains had lost the *pstA* gene, which impaired their normal phosphorus uptake capability. Subsequently, they also lost the ability to synthesize gluconate, yet this additional loss did not impose further survival stress on them, possibly because their compromised phosphorus acquisition already placed them at a competitive disadvantage, rendering the impact of this further functional loss relatively limited.

Soil phosphatase is closely associated with soil phosphorus availability, functioning to hydrolyze phosphate groups of organophosphorus and promoting phosphorus absorption by plants [11]. The genes *phoA*, *phoD*, and *phoX* are the most common conserved phosphatase genes in bacteria, which hydrolyze phosphate monolipids to release phosphate ions [65]. This study compared the phosphatase genes of 15 *Streptomyces* strains, all of which contained *phoD* genes except *Streptomyces murinus* JS0295. Additionally, four strains of *Streptomyces misionensis* contained *phoA* genes. Notably, *S. misionensis* SwB1 possesses three common phosphatase genes—*phoD*, *phoX*, and *phoA*—along with 10 *phoE* genes, which have been annotated as a broadly specific phosphatase. This variety of phosphatase genes not only highlights the genetic diversity within *S. misionensis* SwB1 but also emphasizes the overall type and quantity of phosphorus-solubilizing genes present, which are crucial as they reflect the plant growth-promoting ability of microbials [29]. Furthermore, the abundance of *phoE* genes may significantly enhance the phosphorus solubilization ability of the *S. misionensis* SwB1 strain, which exhibits the most extensive variety of phosphorus solubilizing genes. This extensive gene repertoire and a complete gene cluster (Table 1) are essential for the strain's remarkable capability to solubilize phosphorus effectively.

3.5. Effects of S. misionensis SwB1 Application on Physiological and Biochemical Regulation of C. wilsoniana

In recent years, microbial inoculation has emerged as a promising method for enhancing soil properties and increasing crop production [66].This study demonstrated that the *S. misionensis* SwB1 strain exhibits excellent phosphorus solubilization ability. The inoculation of *S. misionensis* SwB1 on the roots of *C. wilsoniana* significantly enhances phosphorus solubilization (Figure 1). Genome annotation identified numerous phosphorus solubilization genes, as well as nitrogen fixation genes, siderophore production genes, and IAA production genes in the *S. misionensis* SwB1 strain (Table 1). To further validate the plant growth-promoting ability of the *S. misionensis* SwB1 strain, a pot experiment was conducted using red soil as the substrate to assess its effect on the growth of *C. wilsoniana* seedlings.

The SwB1 group caused a significant increase in seedling height (49.41%, $p \leq 0.05$) and ground diameter (30.61%) compared to the CK group (Figure 6A–C). Additionally, *S. misionensis* SwB1 application significantly enhanced the biomass, with fresh weight increasing by 115% and dry weight by 81.44% compared to the CK group (Figure 6D,E). Overall, these results indicate that *S. misionensis* SwB1 promotes the growth and nutrient accumulation of *C. wilsoniana* seedlings.

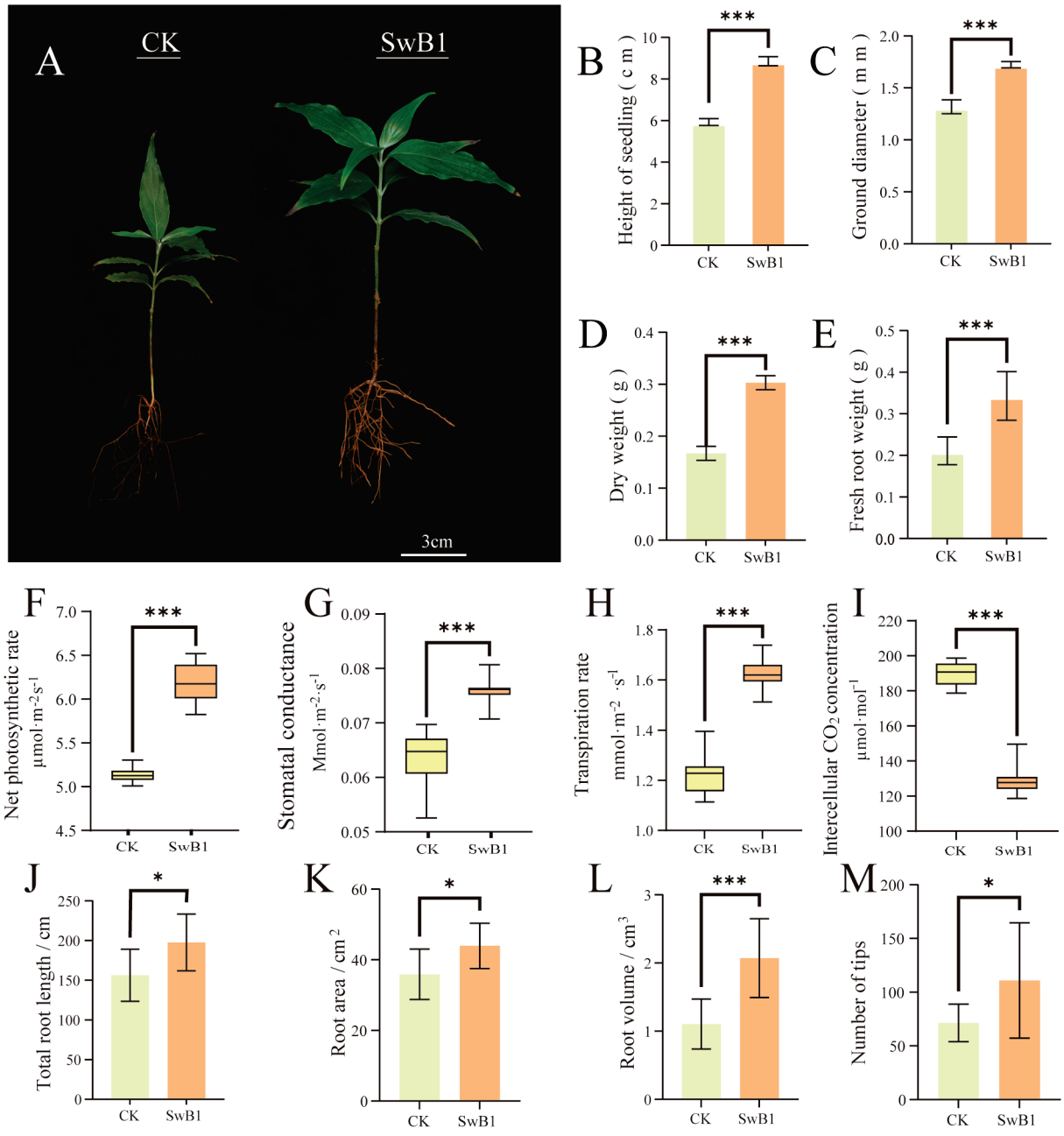


Figure 6. Effect of *S. misionensis* SwB1 on the morphological indices of *C. wilsoniana*. (**A**) Comparison of morphological characteristics of *C. wilsoniana* saplings. (**B**) Height of seedlings. (**C**) Ground diameter. (**D**) Dry weight. (**E**) Fresh weight. (**F**) Net photosynthetic rate. (**G**) Stomatal conductance. (**H**) Transpiration rate. (**I**) Intercellular CO_2 concentration. (**J**) Total root length. (**K**) Root area. (**L**) Root volume. (**M**) The number of root tips. *, $0.01 < p \leq 0.05$; ***, $p \leq 0.001$; the error bar represents the standard error (SE).

S. misionensis SwB1 positively influenced both the leaf area and photosynthetic physiology of the seedlings. As the leaf area of *C. wilsoniana* increased, the net photosynthetic rate in the SwB1 group rose by 18.51% compared to the CK group (Figure 6F). Correspondingly, the leaf transpiration rate and stomatal conductance were significantly higher in the SwB1 group ($p \leq 0.05$), with increases of 32.79% and 34.21%, respectively (Figure 6G,H). Additionally, the intercellular CO_2 concentration in the leaves decreased by 34.21% compared to the CK group ($p \leq 0.05$) (Figure 6I), which was consistent with the increase in the net photosynthetic rate [37]. These results demonstrate that *S. misionensis* SwB1 significantly enhances

photosynthesis in *C. wilsoniana* leaves, leading to a greater accumulation of photosynthetic products and promoting overall plant growth.

The efficiency of nutrient absorption in plants is closely linked to root development, which enhances nutrient uptake and promotes growth. Scanning the root system of *C. wilsoniana* seedlings revealed that *S. misionensis* SwB1 promotes root development. Compared to the CK group, SwB1-treated seedlings exhibited a 26.28% increase in total root length (Figure 6J), with significant improvements in root area and volume—22.49% and 88.18% increases, respectively (Figure 6K,L). The number of root tips also increased significantly, with a 36.24% rise in the SwB1 group (Figure 6M). Therefore, *S. misionensis* SwB1 enhances the root system development of *C. wilsoniana*, induces root meristematic activity, and expands the contact area between roots and soil, thereby improving nutrient absorption and promoting the growth of the above-ground parts of the plant.

3.6. Proposing a High-Efficiency Dissolved Phosphorus Model for S. misionensis SwB1

Different phosphorus-solubilizing bacteria use various mechanisms to solubilize phosphorus [67]. We have compiled a study on *S. misionensis* SwB1, which sheds light on its phosphorus solubilization pathway. Our findings indicate that this strain exhibits diverse modes of phosphorus enhancement, which include a Pst system, Pit system, siderophore transport system, phosphatase secretion system, and gluconic acid secretion system (Figure 7).

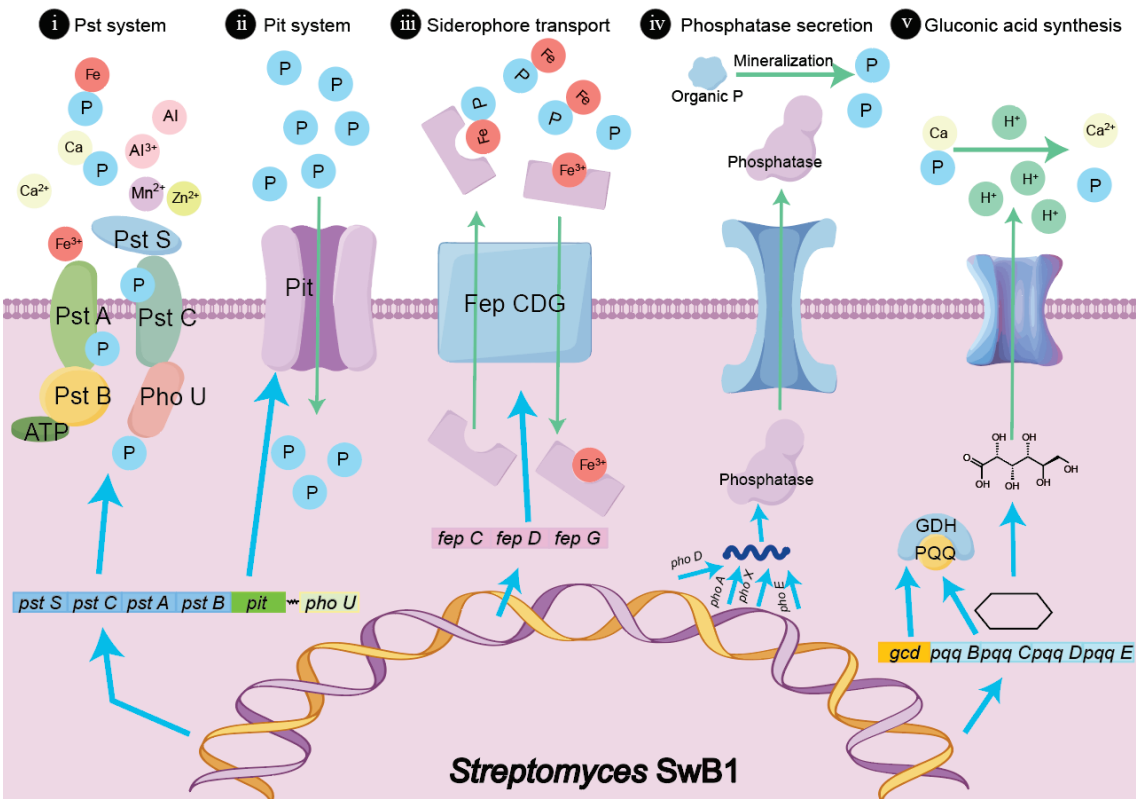


Figure 7. Illustrated schematic of the pivotal mechanism by which *S. misionensis* SwB1 facilitates phosphorus solubilization.

The genome of the *S. misionensis* SwB1 strain encompasses five distinct phosphorus solubilization systems: (i) The Pst system comprises genes for peripheral-binding proteins (PstS), transmembrane proteins (PstA and PstC), ATP-binding proteins (PstB), and regulatory proteins (PhoU), facilitating phosphate transport from iron, aluminum, and calcium phosphates under low-phosphorus conditions [68]. (ii) The *pit* gene encodes a rapid phosphorus transport system that operates optimally in high phosphate ion concen-

trations, allowing for cellular uptake through specialized channels [69]. (iii) Genes *fepC*, *fepD*, and *fepG* form the siderophore transport system, which is crucial for phosphorus dissolution in red soil; siderophores, which have a high affinity for iron, help prevent phosphorus immobilization by forming complexes with iron in iron phosphate [37,70]. (iv) Four phosphatase genes (*PhoA*, *PhoD*, *PhoX*, and *PhoE*) enhance phosphorus solubilization in red soil by converting organic phosphate into available phosphorus (AP), vital for plant growth [71]. (v) Gluconic acid, produced through the interaction of glucose dehydrogenase (GDH) and pyrroloquinoline quinone (PQQ), facilitates phosphorus dissolution [72,73]. The *gcd* gene catalyzes gluconic acid production, along with the genes synthesizing PQQ (pqqB, pqqC, pqqD, and pqqE) identified in the genome. The microbial release of gluconic acid lowers soil pH, enhances the breakdown of insoluble inorganic phosphorus, and increases available phosphorus content. The combined effects of these mechanisms enhance *S. misionensis* SwB1's phosphorus solubilization capacity, improving phosphorus utilization by *C. wilsoniana* cultivated in red soil.

4. Conclusions

In this study, *Streptomyces misionensis* SwB1, isolated from *C. wilsoniana*, was shown to significantly increase phosphorus content in red soil through its symbiotic relationship with *C. wilsoniana*. Genome sequencing revealed that*S. misionensis* SwB1 contains numerous phosphorus solubilizing genes, as well as genes related to nitrogen fixation, siderophore production, and IAA production, which were confirmed through plate verification. Comparative genomic analysis showed that *S. misionensis* SwB1 possesses a remarkable capacity for phosphorus solubilization, attributable to its complete phosphorus transport and gluconic acid synthesis systems, along with a higher number of phosphatase genes. *S. misionensis* SwB1 significantly enhanced nutrient uptake by *C. wilsoniana*, thereby improving plant growth and productivity through beneficial effects on physiological and biochemical parameters. To harness this microbial-mediated phosphorus solubilization as a sustainable strategy for boosting plant productivity and soil health, further studies are needed to elucidate the role of microorganisms in the growth regulation network of plants.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/microorganisms13020378/s1, Figure S1. Phylogenetic analysis of *S. misionensis* SwB1. The bold font is SwB1. The " T " in the upper right corner of the strain represents the model strain. The percentage at the branch indicates branch confidence. Figure S2. Petal map of the genome of 15 strains. The strains in red font are the strains isolated in this study. The middle number is the number of genes shared by all strains. Each petal represents a unique gene; Table S1. *S. misionensis SwB1* genomic characterization information. Table S2. Nitrogen fixation gene of *S. misionensis SwB1*. Table S3. Siderophore production gene of *S. misionensis SwB1*. Table S4. IAA-producing gene of *S. misionensis SwB1*.

Author Contributions: P.L. and C.L. design the study. Y.C., L.J. and Z.G. performed the experiments. Y.Y. and Q.L. analyzed the genome data with the assistance of J.C., X.Z. and L.Z. conducted the data analysis. Y.C. and Z.G. drafted the manuscript under the guidance of Y.J., J.T. and Z.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the State Key Laboratory of Woody Oil Resources Construction Project (2024PT1001), the Young Scientists Fund of the National Natural Science Foundation of China (32401626), the Natural Science Foundation of Changsha Municipal Natural Science Foundation (kq2208099, kq2208098), Key Research and Development Plan of Hunan Province (2022NK2033), the Natural Science Foundation of Hunan Province (2024JJ7283), the Innovation Demonstration Project of Chenzhou City (2022sfq53), the Science and Technology Plan Major Project of Changsha City (kq2301004), the Natural Science Foundation of Hunan Province (2021JJ31143).

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: The data presented in this study are openly available inGenBank, reference number SAMN43297603.

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

References

1. Upadhyay, S.K.; Rajput, V.D.; Kumari, A.; Espinosa-Saiz, D.; Menendez, E.; Minkina, T.; Dwivedi, P.; Mandzhieva, S. Plant growth-promoting rhizobacteria: A potential bio-asset for restoration of degraded soil and crop productivity with sustainable emerging techniques. *Environ. Geochem. Health* **2023**, *45*, 9321–9344. [CrossRef]
2. Li, H.P.; Han, Q.Q.; Liu, Q.M.; Gan, Y.N.; Rensing, C.; Rivera, W.L.; Zhao, Q.; Zhang, J.L. Roles of phosphate-solubilizing bacteria in mediating soil legacy phosphorus availability. *Microbiol. Res.* **2023**, *272*, 127375. [CrossRef]
3. Yu, H.; Wu, X.; Zhang, G.; Zhou, F.; Harvey, P.R.; Wang, L.; Fan, S.; Xie, X.; Li, F.; Zhou, H.; et al. Identification of the Phosphorus-Solubilizing Bacteria Strain JP233 and Its Effects on Soil Phosphorus Leaching Loss and Crop Growth. *Front. Microbiol.* **2022**, *13*, 892533. [CrossRef] [PubMed]
4. An, N.; Zhang, L.; Liu, Y.; Shen, S.; Li, N.; Wu, Z.; Yang, J.; Han, W.; Han, X. Biochar application with reduced chemical fertilizers improves soil pore structure and rice productivity. *Chemosphere* **2022**, *298*, 134304. [CrossRef]
5. Batool, S.; Iqbal, A. Phosphate solubilizing rhizobacteria as alternative of chemical fertilizer for growth and yield of Triticum aestivum (Var. Galaxy 2013). *Saudi J. Biol. Sci.* **2019**, *26*, 1400–1410. [CrossRef]
6. Lin, L.; Li, C.; Ren, Z.; Qin, Y.; Wang, R.; Wang, J.; Cai, J.; Zhao, L.; Li, X.; Cai, Y.; et al. Transcriptome profiling of genes regulated by phosphate-solubilizing bacteria Bacillus megaterium P68 in potato (*Solanum tuberosum* L.). *Front. Microbiol.* **2023**, *14*, 1140752. [CrossRef]
7. Wang, C.; Pan, G.; Lu, X.; Qi, W. Phosphorus solubilizing microorganisms: Potential promoters of agricultural and environmental engineering. *Front. Bioeng. Biotechnol.* **2023**, *11*, 1181078. [CrossRef] [PubMed]
8. Lidbury, I.; Hitchcock, A.; Groenhof, S.R.M.; Connolly, A.N.; Moushtaq, L. New insights in bacterial organophosphorus cycling: From human pathogens to environmental bacteria. *Adv. Microb. Physiol.* **2024**, *84*, 1–49. [CrossRef] [PubMed]
9. Li, X.L.; Zhao, X.Q.; Dong, X.Y.; Ma, J.F.; Shen, R.F. Secretion of Gluconic Acid From Nguyenibacter sp. L1 Is Responsible for Solubilization of Aluminum Phosphate. *Front. Microbiol.* **2021**, *12*, 784025. [CrossRef]
10. Rasul, M.; Yasmin, S.; Suleman, M.; Zaheer, A.; Reitz, T.; Tarkka, M.T.; Islam, E.; Mirza, M.S. Glucose dehydrogenase gene containing phosphobacteria for biofortification of Phosphorus with growth promotion of rice. *Microbiol. Res.* **2019**, *223–225*, 1–12. [CrossRef] [PubMed]
11. Lidbury, I.; Scanlan, D.J.; Murphy, A.R.J.; Christie-Oleza, J.A.; Aguilo-Ferretjans, M.M.; Hitchcock, A.; Daniell, T.J. A widely distributed phosphate-insensitive phosphatase presents a route for rapid organophosphorus remineralization in the biosphere. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2118122119. [CrossRef] [PubMed]
12. Garaycochea, S.; Altier, N.A.; Leoni, C.; Neal, A.L.; Romero, H. Abundance and phylogenetic distribution of eight key enzymes of the phosphorus biogeochemical cycle in grassland soils. *Environ. Microbiol. Rep.* **2023**, *15*, 352–369. [CrossRef]
13. Liang, X.; Hu, X.; Wang, X.; Wang, J.; Fang, Y.; Li, Y. Characterization of the phosphate-specific transport system in Cronobacter sakazakii BAA-894. *J. Appl. Microbiol.* **2017**, *123*, 710–723. [CrossRef] [PubMed]
14. Wang, Y.; Sun, Z.; Zhao, Q.; Yang, X.; Li, Y.; Zhou, H.; Zhao, M.; Zheng, H. Whole-genome analysis revealed the growth-promoting and biological control mechanism of the endophytic bacterial strain Bacillus halotolerans Q2H2, with strong antagonistic activity in potato plants. *Front. Microbiol.* **2023**, *14*, 1287921. [CrossRef]
15. Ding, J.; Yang, W.; Liu, X.; Zhao, Q.; Dong, W.; Zhang, C.; Liu, H.; Zhao, Y. Unraveling the rate-limiting step in microorganisms' mediation of denitrification and phosphorus absorption/transport processes in a highly regulated river-lake system. *Front. Microbiol.* **2023**, *14*, 1258659. [CrossRef]
16. He, Z.; Chao, H.; Zhou, X.; Ni, Q.; Hu, Y.; Yu, R.; Wang, M.; Li, C.; Chen, J.; Chen, Y.; et al. A chromosome-level genome assembly provides insights into Cornus wilsoniana evolution, oil biosynthesis, and floral bud development. *Hortic. Res.* **2023**, *10*, uhad196. [CrossRef] [PubMed]
17. Long, H.; Wasaki, J. Effects of Phosphate-solubilizing Bacteria on Soil Phosphorus Fractions and Supply to Maize Seedlings Grown in Lateritic Red Earths and Cinnamon Soils. *Microbes Environ.* **2023**, *38*, ME22075. [CrossRef]
18. Yin, M.; Li, X.; Liu, Q.; Tang, F. Rice husk ash addition to acid red soil improves the soil property and cotton seedling growth. *Sci. Rep.* **2022**, *12*, 1704. [CrossRef]

19. Gao, Z.; Li, P.; Li, C.; Tang, R.; Wang, M.; Chen, J.; Yang, Y.; He, Z.; Xiao, Z.; Ma, Y.; et al. Identification, functional annotation, and isolation of phosphorus-solubilizing bacteria in the rhizosphere soil of Swida wilsoniana (Wanger) Sojak. *Appl. Soil Ecol.* **2024**, *194*, 105207. [CrossRef]
20. Suriyachadkun, C.; Chunhachart, O.; Srithaworn, M.; Tangchitcharoenkhul, R.; Tangjitjareonkun, J. Zinc-Solubilizing Streptomyces spp. as Bioinoculants for Promoting the Growth of Soybean (*Glycine max* (L.) Merrill). *J. Microbiol. Biotechnol.* **2022**, *32*, 1435–1446. [CrossRef] [PubMed]
21. Zhang, T.; Li, H.; Ma, S.; Cao, J.; Liao, H.; Huang, Q.; Chen, W. The newest Oxford Nanopore R10.4.1 full-length 16S rRNA sequencing enables the accurate resolution of species-level microbial community profiling. *Appl. Environ. Microbiol.* **2023**, *89*, e0060523. [CrossRef]
22. Krajaejun, T.; Patumcharoenpol, P.; Rujirawat, T.; Kittichotirat, W.; Tangphatsornruang, S.; Lohnoo, T.; Yingyong, W. PacBio long read-assembled draft genome of Pythium insidiosum strain Pi-S isolated from a Thai patient with pythiosis. *BMC Res. Notes* **2023**, *16*, 271. [CrossRef]
23. Buttler, J.; Drown, D.M. Accuracy and Completeness of Long Read Metagenomic Assemblies. *Microorganisms* **2022**, *11*, 96. [CrossRef] [PubMed]
24. 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]
25. O'Leary, N.A.; Wright, M.W.; Brister, J.R.; Ciufo, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; et al. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **2016**, *44*, D733–D745. [CrossRef]
26. Galperin, M.Y.; Kristensen, D.M.; Makarova, K.S.; Wolf, Y.I.; Koonin, E.V. Microbial genome analysis: The COG approach. *Brief. Bioinform.* **2019**, *20*, 1063–1070. [CrossRef]
27. Cai, Z.; Xie, Z.; Huang, L.; Wang, Z.; Pan, M.; Yu, X.; Xu, S.; Luo, J. Full-length transcriptome analysis of Adiantum flabellulatum gametophyte. *PeerJ* **2022**, *10*, e13079. [CrossRef] [PubMed]
28. Murakami, C.; Tanaka, A.R.; Sato, Y.; Kimura, Y.; Morimoto, K. Easy detection of siderophore production in diluted growth media using an improved CAS reagent. *J. Microbiol. Methods* **2021**, *189*, 106310. [CrossRef] [PubMed]
29. Wang, Y.; Zhao, Q.; Sun, Z.; Li, Y.; He, H.; Zhang, Y.; Yang, X.; Wang, D.; Dong, B.; Zhou, H.; et al. Whole-genome analysis revealed the growth-promoting mechanism of endophytic bacterial strain Q2H1 in potato plants. *Front. Microbiol.* **2022**, *13*, 1035901. [CrossRef] [PubMed]
30. Shi, Z.; Guo, X.; Lei, Z.; Wang, Y.; Yang, Z.; Niu, J.; Liang, J. Screening of high-efficiency nitrogen-fixing bacteria from the traditional Chinese medicine plant Astragalus mongolicus and its effect on plant growth promotion and bacterial communities in the rhizosphere. *BMC Microbiol.* **2023**, *23*, 292. [CrossRef] [PubMed]
31. Tun, T.; Baleivanualala, S.C.; Erdmann, M.B.; Gimenez, G.; Aung, H.L. An extensively drug-resistant Pseudomonas aeruginosa isolate from Myanmar. *J. Glob. Antimicrob. Resist.* **2024**, *36*, 1–3. [CrossRef]
32. Emms, D.M.; Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **2019**, *20*, 238. [CrossRef] [PubMed]
33. Damo, J.L.C.; Ramirez, M.D.A.; Agake, S.I.; Pedro, M.; Brown, M.; Sekimoto, H.; Yokoyama, T.; Sugihara, S.; Okazaki, S.; Ohkama-Ohtsu, N. Isolation and Characterization of Phosphate Solubilizing Bacteria from Paddy Field Soils in Japan. *Microbes Environ.* **2022**, *37*, ME21085. [CrossRef] [PubMed]
34. Liu, M.; Yang, C.; Mu, R. Effect of soil water-phosphorus coupling on the photosynthetic capacity of *Robinia pseudoacacia* L. seedlings in semi-arid areas of the Loess Plateau, China. *Environ. Monit. Assess.* **2023**, *195*, 932. [CrossRef] [PubMed]
35. Zhang, G.; Liu, Q.; Zhang, Z.; Ci, D.; Zhang, J.; Xu, Y.; Guo, Q.; Xu, M.; He, K. Effect of Reducing Nitrogen Fertilization and Adding Organic Fertilizer on Net Photosynthetic Rate, Root Nodules and Yield in Peanut. *Plants* **2023**, *12*, 2902. [CrossRef]
36. Dinkler, K.; Li, B.; Guo, J.; Hulsemann, B.; Becker, G.C.; Muller, J.; Oechsner, H. Adapted Hedley fractionation for the analysis of inorganic phosphate in biogas digestate. *Bioresour. Technol.* **2021**, *331*, 125038. [CrossRef] [PubMed]
37. Cui, K.; Xu, T.; Chen, J.; Yang, H.; Liu, X.; Zhuo, R.; Peng, Y.; Tang, W.; Wang, R.; Chen, L.; et al. Siderophores, a potential phosphate solubilizer from the endophyte *Streptomyces* sp. CoT10, improved phosphorus mobilization for host plant growth and rhizosphere modulation. *J. Clean. Prod.* **2022**, *367*, 133110. [CrossRef]
38. Santoro, V.; Schiavon, M.; Celi, L. Role of soil abiotic processes on phosphorus availability and plant responses with a focus on strigolactones in tomato plants. *Plant Soil* **2023**, *494*, 1–49. [CrossRef]
39. Vassileva, M.; Mendes, G.; Deriu, M.; Benedetto, G.; Flor-Peregrin, E.; Mocali, S.; Martos, V.; Vassilev, N. Fungi, P-Solubilization, and Plant Nutrition. *Microorganisms* **2022**, *10*, 1716. [CrossRef] [PubMed]
40. Tarazona, G.; Schleissner, C.; Rodriguez, P.; Perez, M.; Canedo, L.M.; Cuevas, C. Streptenols F-I Isolated from the Marine-Derived *Streptomyces misionensis* BAT-10-03-023. *J. Nat. Prod.* **2017**, *80*, 1034–1038. [CrossRef]

41. Pylro, V.S.; Dias, A.C.F.; Andreote, F.D.; Varani, A.M.; Andreote, C.C.F.; Ribeiro, I.A.F.; Kitano, I.T.; Bernardo, E.R.A. Draft Genomic Sequences of *Streptomyces misionensis* ACT66 and Streptomyces albidoflavus ACT77, Bacteria with Potential Application for Phytopathogen Biocontrol. *Microbiol. Resour. Announc.* **2019**, *8*, e00949-19. [CrossRef] [PubMed]
42. Saadouli, I.; Zendah El Euch, I.; Trabelsi, E.; Mosbah, A.; Redissi, A.; Ferjani, R.; Fhoula, I.; Cherif, A.; Sabatier, J.M.; Sewald, N.; et al. Isolation, Characterization and Chemical Synthesis of Large Spectrum Antimicrobial Cyclic Dipeptide (l-leu-l-pro) from *Streptomyces misionensis* V16R3Y1 Bacteria Extracts. A Novel (1)H NMR Metabolomic Approach. *Antibiotics* **2020**, *9*, 270. [CrossRef] [PubMed]
43. Ren, H.; Huang, B.; Fernández-García, V.; Miesel, J.; Yan, L.; Lv, C. Biochar and Rhizobacteria Amendments Improve Several Soil Properties and Bacterial Diversity. *Microorganisms* **2020**, *8*, 502. [CrossRef]
44. Khanna, K.; Kohli, S.K.; Sharma, N.; Kour, J.; Devi, K.; Bhardwaj, T.; Dhiman, S.; Singh, A.D.; Sharma, N.; Sharma, A.; et al. Phytomicrobiome communications: Novel implications for stress resistance in plants. *Front. Microbiol.* **2022**, *13*, 912701. [CrossRef] [PubMed]
45. Mondal, S.; Ravishankar Rai, V. Molecular profiling and anti-infective potency of endophytic actinomycetes inhabiting Madhuca insignis Radlk., from Western Ghats of India. *J. Genet. Eng. Biotechnol.* **2021**, *19*, 36. [CrossRef]
46. Raimundi, E.A.; Dominguez, E.; Salles, F.F. Description of a new species, nymphs of two known species and a new record of Miroculis/ Edmunds, 1963 (Ephemeroptera: Leptophlebiidae) from Brazil. *Zootaxa* **2021**, *4963*, 384–392. [CrossRef] [PubMed]
47. Maiti, P.K.; Mandal, S. Streptomyces cupreus sp. nov., an antimicrobial producing actinobacterium isolated from Himalayan soil. *Arch. Microbiol.* **2021**, *203*, 1601–1609. [CrossRef]
48. Law, J.W.; Ser, H.L.; Duangjai, A.; Saokaew, S.; Bukhari, S.I.; Khan, T.M.; Ab Mutalib, N.S.; Chan, K.G.; Goh, B.H.; Lee, L.H. *Streptomyces colonosanans* sp. nov., A Novel Actinobacterium Isolated from Malaysia Mangrove Soil Exhibiting Antioxidative Activity and Cytotoxic Potential against Human Colon Cancer Cell Lines. *Front. Microbiol.* **2017**, *8*, 877. [CrossRef]
49. Pan, C.; Sun, C.; Qu, X.; Yu, W.; Guo, J.; Yu, Y.; Li, X. Microbial community interactions determine the mineralization of soil organic phosphorus in subtropical forest ecosystems. *Microbiol. Spectr.* **2024**, *12*, e0135523. [CrossRef] [PubMed]
50. Huang, X.; Zeng, Z.; Chen, Z.; Tong, X.; Jiang, J.; He, C.; Xiang, T. Deciphering the potential of a plant growth promoting endophyte Rhizobium sp. WYJ-E13, and functional annotation of the genes involved in the metabolic pathway. *Front. Microbiol.* **2022**, *13*, 1035167. [CrossRef]
51. Liu, S.; Zeng, J.; Yu, H.; Wang, C.; Yang, Y.; Wang, J.; He, Z.; Yan, Q. Antimony efflux underpins phosphorus cycling and resistance of phosphate-solubilizing bacteria in mining soils. *ISME J.* **2023**, *17*, 1278–1289. [CrossRef]
52. Xie, J.G.; Yan, Z.Q.; Wang, G.F.; Xue, W.Z.; Li, C.; Chen, X.W.; Chen, D.F. A Bacterium Isolated From Soil in a Karst Rocky Desertification Region Has Efficient Phosphate-Solubilizing and Plant Growth-Promoting Ability. *Front. Microbiol.* **2021**, *11*, 625450. [CrossRef]
53. Joshi, S.; Gangola, S.; Jaggi, V.; Sahgal, M. Functional characterization and molecular fingerprinting of potential phosphate solubilizing bacterial candidates from Shisham rhizosphere. *Sci. Rep.* **2023**, *13*, 7003. [CrossRef] [PubMed]
54. Hussein, F.B.; Venkiteshwaran, K.; Mayer, B.K. Cell surface-expression of the phosphate-binding protein PstS: System development, characterization, and evaluation for phosphorus removal and recovery. *J. Environ. Sci.* **2020**, *92*, 129–140. [CrossRef]
55. Zhao, F.; Lin, X.; Cai, K.; Jiang, Y.; Ni, T.; Chen, Y.; Feng, J.; Dang, S.; Zhou, C.Z.; Zeng, Q. Biochemical and structural characterization of the cyanophage-encoded phosphate-binding protein: Implications for enhanced phosphate uptake of infected cyanobacteria. *Environ. Microbiol.* **2022**, *24*, 3037–3050. [CrossRef]
56. Vuppada, R.K.; Hansen, C.R.; Strickland, K.A.P.; Kelly, K.M.; McCleary, W.R. Phosphate signaling through alternate conformations of the PstSCAB phosphate transporter. *BMC Microbiol.* **2018**, *18*, 8. [CrossRef]
57. Overbeck, T.J.; Welker, D.L.; Hughes, J.E.; Steele, J.L.; Broadbent, J.R. Transient MutS-Based Hypermutation System for Adaptive Evolution of Lactobacillus casei to Low pH. *Appl. Environ. Microbiol.* **2017**, *83*, e01120-17. [CrossRef]
58. Neznansky, A.; Blus-Kadosh, I.; Yerushalmi, G.; Banin, E.; Opatowsky, Y. The Pseudomonas aeruginosa phosphate transport protein PstS plays a phosphate-independent role in biofilm formation. *FASEB J.* **2014**, *28*, 5223–5233. [CrossRef]
59. Zhang, Y.; Zhang, Y.; Li, P.; Wang, Y.; Wang, J.; Shao, Z.; Zhao, G. GlnR positive transcriptional regulation of the phosphate-specific transport system pstSCAB in Amycolatopsis mediterranei U32. *Acta Biochim. Biophys. Sin.* **2018**, *50*, 757–765. [CrossRef]
60. Hudek, L.; Premachandra, D.; Webster, W.A.; Brau, L. Role of Phosphate Transport System Component PstB1 in Phosphate Internalization by Nostoc punctiforme. *Appl. Environ. Microbiol.* **2016**, *82*, 6344–6356. [CrossRef] [PubMed]
61. Wagh, J.; Shah, S.; Bhandari, P.; Archana, G.; Kumar, G.N. Heterologous expression of pyrroloquinoline quinone (pqq) gene cluster confers mineral phosphate solubilization ability to Herbaspirillum seropedicae Z67. *Appl. Microbiol. Biotechnol.* **2014**, *98*, 5117–5129. [CrossRef] [PubMed]
62. An, R.; Moe, L.A. Regulation of Pyrroloquinoline Quinone-Dependent Glucose Dehydrogenase Activity in the Model Rhizosphere-Dwelling Bacterium Pseudomonas putida KT2440. *Appl. Environ. Microbiol.* **2016**, *82*, 4955–4964. [CrossRef] [PubMed]
63. Toyama, H.; Lidstrom, M.E. pqqA is not required for biosynthesis of pyrroloquinoline quinone in Methylobacterium extorquens AM1. *Microbiology* **1998**, *144*, 183–191. [CrossRef]

64. Martins, A.M.; Latham, J.A.; Martel, P.J.; Barr, I.; Iavarone, A.T.; Klinman, J.P. A two-component protease in Methylorubrum extorquens with high activity toward the peptide precursor of the redox cofactor pyrroloquinoline quinone. *J. Biol. Chem.* **2019**, *294*, 15025–15036. [CrossRef] [PubMed]
65. Neha; Bhardwaj, Y.; Reddy, B.; Dubey, S.K. Organic Farming Favors phoD-Harboring Rhizospheric Bacterial Community and Alkaline Phosphatase Activity in Tropical Agroecosystem. *Plants* **2023**, *12*, 68. [CrossRef] [PubMed]
66. Dhawi, F. The Role of Plant Growth-Promoting Microorganisms (PGPMs) and Their Feasibility in Hydroponics and Vertical Farming. *Metabolites* **2023**, *13*, 247. [CrossRef] [PubMed]
67. Pan, L.; Cai, B. Phosphate-Solubilizing Bacteria: Advances in Their Physiology, Molecular Mechanisms and Microbial Community Effects. *Microorganisms* **2023**, *11*, 2904. [CrossRef] [PubMed]
68. Liang, J.L.; Feng, S.W.; Lu, J.L.; Wang, X.N.; Li, F.L.; Guo, Y.Q.; Liu, S.Y.; Zhuang, Y.Y.; Zhong, S.J.; Zheng, J.; et al. Hidden diversity and potential ecological function of phosphorus acquisition genes in widespread terrestrial bacteriophages. *Nat. Commun.* **2024**, *15*, 2827. [CrossRef]
69. Dai, Z.; Liu, G.; Chen, H.; Chen, C.; Wang, J.; Ai, S.; Wei, D.; Li, D.; Ma, B.; Tang, C.; et al. Long-term nutrient inputs shift soil microbial functional profiles of phosphorus cycling in diverse agroecosystems. *ISME J.* **2020**, *14*, 757–770. [CrossRef]
70. Feigl, V.; Medgyes-Horvath, A.; Kari, A.; Torok, A.; Bombolya, N.; Berkl, Z.; Farkas, E.; Fekete-Kertesz, I. The potential of Hungarian bauxite residue isolates for biotechnological applications. *Biotechnol. Rep.* **2024**, *41*, e00825. [CrossRef]
71. Liang, J.L.; Liu, J.; Jia, P.; Yang, T.T.; Zeng, Q.W.; Zhang, S.C.; Liao, B.; Shu, W.S.; Li, J.T. Novel phosphate-solubilizing bacteria enhance soil phosphorus cycling following ecological restoration of land degraded by mining. *ISME J.* **2020**, *14*, 1600–1613. [CrossRef] [PubMed]
72. Pang, F.; Li, Q.; Solanki, M.K.; Wang, Z.; Xing, Y.X.; Dong, D.F. Soil phosphorus transformation and plant uptake driven by phosphate-solubilizing microorganisms. *Front. Microbiol.* **2024**, *15*, 1383813. [CrossRef] [PubMed]
73. You, M.; Fang, S.; MacDonald, J.; Xu, J.; Yuan, Z.C. Isolation and characterization of Burkholderia cenocepacia CR318, a phosphate solubilizing bacterium promoting corn growth. *Microbiol. Res.* **2020**, *233*, 126395. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

MDPI AG
Grosspeteranlage 5
4052 Basel
Switzerland
Tel.: +41 61 683 77 34

Microorganisms Editorial Office
E-mail: microorganisms@mdpi.com
www.mdpi.com/journal/microorganisms



Disclaimer/Publisher's Note: The title and front matter of this reprint are at the discretion of the Guest Editor. The publisher is not responsible for their content or any associated concerns. The statements, opinions and data contained in all individual articles are solely those of the individual Editor and contributors and not of MDPI. MDPI disclaims responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.





Academic Open
Access Publishing

mdpi.com

ISBN 978-3-7258-9256-3