International Journal of
Molecular Sciences

MDPI

Article

The First High-Quality Genome Assembly of Freshwater Pearl Mussel *Sinohyriopsis cumingii*: New Insights into Pearl Biomineralization

Zhiyi Bai [1,2,3,†], Ying Lu [4,5,†], Honghui Hu [1,3], Yongbin Yuan [1,3], Yalin Li [4,5], Xiaojun Liu [1], Guiling Wang [1,2,3], Dandan Huang [1], Zhiyan Wang [1,3], Yingrui Mao [1,3], He Wang [1,3], Liangbiao Chen [4,5,*] and Jiale Li [1,2,3,*]

1 Key Laboratory of Freshwater Aquatic Genetic Resources, Ministry of Agriculture and Rural Affairs, Shanghai Ocean University, Shanghai 201306, China
2 Shanghai Engineering Research Center of Aquaculture, Shanghai Ocean University, Shanghai 201306, China
3 Shanghai Collaborative Innovation Center of Aquatic Animal Breeding and Green Aquaculture, Shanghai Ocean University, Shanghai 201306, China
4 International Research Center for Marine Biosciences, Ministry of Science and Technology, Shanghai Ocean University, Shanghai 201306, China; yinglu@shou.edu.cn (Y.L.)
5 Key Laboratory of Exploration and Utilization of Aquatic Genetic Resources, Ministry of Education, Shanghai Ocean University, Shanghai 201306, China
* Correspondence: lbchen@shou.edu.cn (L.C.); jlli@shou.edu.cn (J.L.)
† These authors contributed equally to this work.

Abstract: China leads the world in freshwater pearl production, an industry in which the triangle sail mussel (*Sinohyriopsis cumingii*) plays a pivotal role. In this paper, we report a high-quality chromosome-level genome assembly of *S. cumingii* with a size of 2.90 Gb—the largest yet reported among bivalves—and 89.92% anchorage onto 19 linkage groups. The assembled genome has 37,696 protein-coding genes and 50.86% repeat elements. A comparative genomic analysis revealed expansions of 752 gene families, mostly associated with biomineralization, and 237 genes under strong positive selection. Notably, the *fibrillin* gene family exhibited gene family expansion and positive selection simultaneously, and it also exhibited multiple high expressions after mantle implantation by transcriptome analysis. Furthermore, RNA silencing and an in vitro calcium carbonate crystallization assay highlighted the pivotal role played by one fibrillin gene in calcium carbonate deposition and aragonite transformation. This study provides a valuable genomic resource and offers new insights into the mechanism of pearl biomineralization.

Keywords: pearl; *Sinohyriopsis cumingii*; genome; *fibrillin* gene family; biomineralization

check for updates

Citation: Bai, Z.; Lu, Y.; Hu, H.; Yuan, Y.; Li, Y.; Liu, X.; Wang, G.; Huang, D.; Wang, Z.; Mao, Y.; et al. The First High-Quality Genome Assembly of Freshwater Pearl Mussel *Sinohyriopsis cumingii*: New Insights into Pearl Biomineralization. *Int. J. Mol. Sci.* **2024**, *25*, 3146. https://doi.org/10.3390/ijms25063146

Academic Editor: Giacomina Brunetti

Received: 6 January 2024
Revised: 31 January 2024
Accepted: 11 February 2024
Published: 9 March 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Freshwater and marine pearl culture, which has been practiced since ancient times, has evolved into the world's preeminent aquaculture industry [1]. Since 1984, China has maintained its position as the world's leading freshwater pearl producer. Today, China is responsible for approximately 90% of total annual global production, with the annual production of freshwater pearls arriving at four hundred tons in the last five years, the value of production exceeding seven billion dollars, and several hundred thousand people working in this sector [2]. Due to its outstanding pearl production performance, characterized by high yields and superior-quality pearl formation (big, round, and colorful), the triangle sail mussel *Sinohyriopsis cumingii* (Lea, 1852) has emerged as the favored species for freshwater pearl production in China. *S. cumingii* inhabits large lakes, rivers, and estuaries across China [3,4]. The abundant genetic resources of *S. cumingii* may be seen as the foundation for its high-quality pearl products. Generally, seawater pearls are of better quality than freshwater pearls in some terms, such as luster, while freshwater pearls have thicker nacre, exhibiting their advantages in yield. Furthermore, the color of

freshwater pearls is relatively simple compared with seawater ones, resulting in the low value of freshwater pearls. Improving the quality of pearls is therefore a key concern of the freshwater pearl industry.

Understanding the mechanism of freshwater pearl biomineralization can be seen as the theoretical basis for improving pearl culture technology, including the use of genetic breeding to improve aquacultural growing conditions; however, such improvements also involve considerable challenges. The process of pearl—or shell—formation is a remarkable biomineralization phenomenon involving intricate organic matrices [5–7]. This process entails the conversion of ions from the environment into solid minerals, followed by the ordered growth of crystals with the participation of biological cells, ultimately leading to the formation of distinctive biological minerals [8,9]. Numerous organic matrices regulate the deposited calcium carbonate, influencing the arrangement and growth of crystals for nucleation [10]. Gene duplications within carbonic anhydrase (CA), von Willebrand factor A (VWA), and chitin-binding (CB) domain-containing protein families in mollusks have been proposed as key events after their divergence from other lophotrochozoan lineages. These are known to play roles in molluscan shell matrix proteins (SMPs) and in influencing the transition from ancestral exoskeletons to mineralized shells [11]. Notably, the nacre, a convergent carbonate mineral structure, exhibits limited homology, or the absence of similarity, in nacre-associated protein repertoires across bivalves, gastropods, and cephalopods, highlighting evolutionary plasticity [12,13].

Whole-genome sequencing may greatly empower the most fundamental inquiries in biology and evolution; specifically, it may prove instrumental in advancing genetic improvement efforts across various shellfish species. For instance, the oyster (*Crassostrea gigas*) genome has elucidated environmental adaptation and shell formation in bivalves [14]. Two scallop (*Chlamys farreri* and *Patinopecten yessoensis*) genomes have provided insight into the growth and development of bilaterian evolution [15,16]. The pearl oyster (*Pinctada fucata*) genome has shed light on biomineralization in nacre formation [17]. The Nautilus (*Nautilus pompilius*) genome has proven to be a valuable resource in the study of eye evolution and biomineralization in cephalopods [18]. At the present time, genomic resources for freshwater bivalves are limited. Until recently, published genome sequences have been confined to the five freshwater mussels (*Dreissena rostriformis*, *Venustaconcha ellipsiformi*, *Megalonaias nervosa*, *Potamilus streckersoni*, and *Margaritifera margaritifera*) [19–23]. It is especially noteworthy that the whole-genome sequence for the freshwater pearl mussel is still not available. This genome gap has greatly limited our understanding of pearl biomineralization.

In this paper, we report our genome assembly of the triangle sail mussel and a subsequent comparative genome analysis. We present evidence of molecular adaptation and evolution in gene content, which lies behind the biomineralization process. In particular, the research described here shows that the *fibrillin* family contributes especially to the remarkable pearl-production ability of this species. New insights into our understanding of molecular mechanisms are also described. These genomic resources not only advance our comprehension of the mechanism of pearl formation but also lay a solid foundation for future genetic improvements and innovations in culture technology.

2. Results

2.1. De Novo Sequencing and Genomic Characterization of S. cumingii

High-depth genome sequencing of a single female *S. cumingii* individual was performed to generate a high-quality reference genome. Using the Illumina sequencing technology, 153.49 Gb of Illumina PE150 data, 265.49 Gb of whole-genome shotgun data, and 206.21 Gb of 10× Genomics data were obtained in total (Table S1). Using Pacbio sequencing, we generated 86.49 Gb of Pacbio HiFi data (Table S1). After quality control, we obtained 111.35 Gb, 213.70 Gb, 206.20 Gb, and 86.49 Gb of clean data for Illumina PE150, whole-genome shotgun, 10× Genomics, and Pacbio HiFi, respectively. We performed a survey to estimate the genome size and heterozygosity of *S. cumingii* using Illumina

(San Diego, CA, USA) PE150 data. The genome had an estimated size of close to 2.91 Gb and exhibited high heterozygosity (0.92%) (Figure S1, Table S2). The genome assembled with the Illumina and 10× Genomics data consisted of 8268 scaffolds with a scaffold N50 length of 3.19 Mb and a total length of 3.38 Gb, which contained 15,982 contigs with a contig N50 length of 736.17 Kb (Table S3). The total length of the genome was larger than the estimated size. Hence, we also performed genome assembly using Pacbio HiFi data. Finally, the genome assembled consisted of 1808 contigs with a contig N50 of 5.30 Mb and a total length of 2.90 Gb (Table 1); this was consistent with the estimated genome size.

Table 1. Statistics of Assembly for *S. cumingii* using Pacbio HIFI data.

Section	Value
Number of contigs	1808
Contig N50 (bp)	5,295,426
Contig N90 (bp)	1,144,919
Total length of contigs (bp)	2,904,942,185

Using simple sequence repeats (SSRs) and single-nucleotide polymorphisms (SNPs) on two genetic maps, around 2.61 Gb of scaffolds (corresponding to 89.92% of the genome) were anchored onto 19 linkage groups (Figure 1, Tables S4 and S5). BUSCO analysis revealed a high level of completeness, identifying 253 (99.22%) complete genes and 1 (0.39%) fragmented gene in 255 BUSCOs in eukaryotes and 920 (96.44%) complete and 22 (2.31%) fragmented genes in a total of 954 BUSCOs in metazoans (Figure S2, Tables S6 and S7). The results show that the assembly covered most regions of the genome.

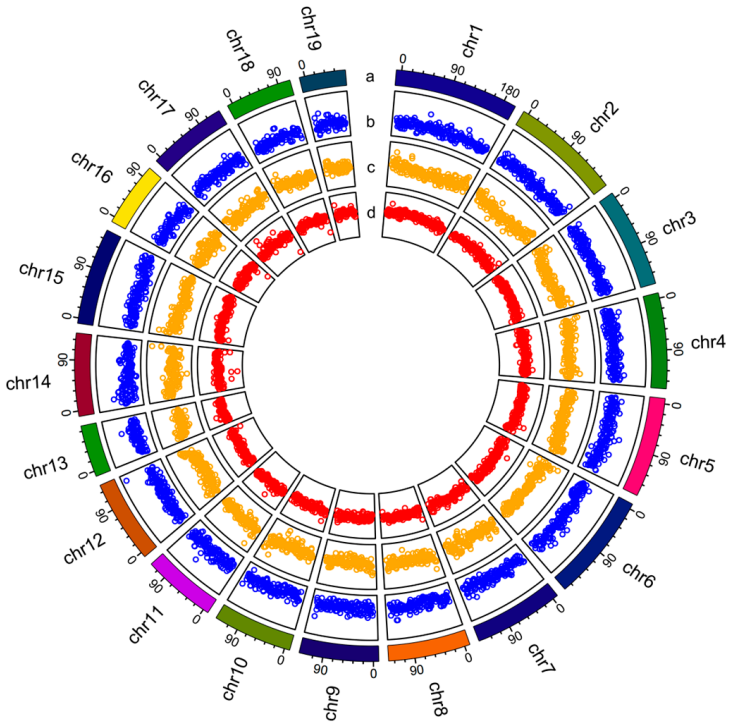


Figure 1. Diagram and genomic landscape of the freshwater pearl mussel *S. cumingii*. From outer to inner circles: a represents the 19 haploid chromosomes at the Mb scale; b represents gene density (blue points) on each chromosome; c represents GC content (orange points) across the genome; and d represents repeat density (red points), drawn in 1 Mb sliding windows.

2.2. Repeat Element Identification and Genome Annotation

Repeat element analysis showed that the repeat elements identified in *S. cumingii* constituted 50.86% of the whole genome (Figure 1 and Table 2) and specifically included

DNA transposons (13.21%), retroelements (12.38%), and unclassified repeat sequences (23.08%) (Table 2). Long terminal repeat (LTR) elements represented the majority of the confirmed interspersed repeats, accounting for 6.86% of the genome (Table 2).

Table 2. Statistics of repeat elements for *S. cumingii* assembly.

Element	Number of Elements	Length Occupied (bp)	Percentage
Retroelements	781,650	359,759,912	12.38%
SINEs	0	0	0.00%
Penelope	18,874	7,315,525	0.25%
LINEs	258,394	160,402,819	5.52%
CRE/SLACS	0	0	0.00%
L2/CR1/Rex	83,861	48,684,191	1.68%
R1/LOA/Jockey	38,560	33,760,425	1.16%
R2/R4/NeSL	3247	596,989	0.02%
RTE/Bov-B	69,197	46,843,959	1.61%
L1/CIN4	1401	864,374	0.03%
LTR elements	523,256	199,357,093	6.86%
BEL/Pao	3768	1,264,899	0.04%
Ty1/Copia	22	6808	0.00%
Gypsy/DIRS1	152,335	87,172,551	3.00%
Retroviral	0	0	0.00%
DNA transposons	400,216	383,834,032	13.21%
hobo-Activator	64,261	23,061,802	0.79%
Tc1-IS630-Pogo	34,110	12,365,403	0.43%
En-Spm	0	0	0.00%
MuDR-IS905	0	0	0.00%
PiggyBac	3422	1,122,485	0.04%
Tourist/Harbinger	0	0	0.00%
Other (Mirage, P-element, Transib)	6018	2,447,569	0.08%
Rolling-circles	0	0	0.00%
Unclassified	3,228,829	670,454,699	23.08%
Total interspersed repeats		1,414,048,643	48.68%
Small RNA	0	0	0.00%
Satellites	1	413	0.00%
Simple repeats	922,774	60,734,831	2.09%
Low complexity	54,102	2,669,708	0.09%
Total number of elements	5,387,572	1,477,453,595	100.00%

In addition, a total of 37,696 protein-coding genes, with an average CDS length of 1821 bp, were identified (Table 3). All the genes were aligned to public databases, including GO (Figure S3), COG (Figure S4), and KEGG (Figure S5) for functional annotation. A total of 21,246 (60.99%) genes could be mapped to at least one database.

Table 3. Statistics of gene model features for *S. cumingii*.

Section	Results
Genome size (bp)	2,904,942,185
Repeat sequence (bp)	1,477,453,595
Number of genes	37,696
Gene average length (CDS)	1820.50
Gene average length (DNA)	37,988.60
Exon number per gene	6.77
Exon average length (bp)	218.78
Genome GC content (%)	36.07%

2.3. Gene Family Identification and Phylogenetic Analyses

Among the 13 species, a total of 38,275 gene families were identified; of these, 3802 were present in all species, and 1094 were only found in *S. cumingii* (Table S8). In addition, 130 single-copy orthologous genes were shared among all species. Comparative genomic analysis also showed that 7675 gene families were shared among *S. cumingii*, *P. fucata*, *C. gigas*, *C. farreri*, and *D. rostriformis* (Figure S6). Specifically, *S. cumingii* shared 102 gene families with *P. fucata* and 453 with *D. rostriformis* (Figure S6). According to the results of the KEGG pathway enrichment analysis, the 1094 *S. cumingii*-specific gene families were mainly enriched in 78 KEGG pathways (Table S9), which were mostly related to the immune system and disease resistance (Figure S7). However, some of them were also enriched in the "ECM-receptor interaction", "glycosaminoglycan biosynthesis-keratan sulfate", and "calcium signaling" pathways, which are associated with biomineralization (Table S9).

The phylogenetic tree demonstrated that *S. cumingii* is most closely related to *D. rostriformis* and that these two species diverged about 18.5 million years ago (Mya). The bivalve clade, including mussels, scallops, and oysters, is distantly related to *L. gigantea* and *O. bimaculoides*, and it diverged from these two species around 65 Mya and 75 Mya, respectively (Figure 2).

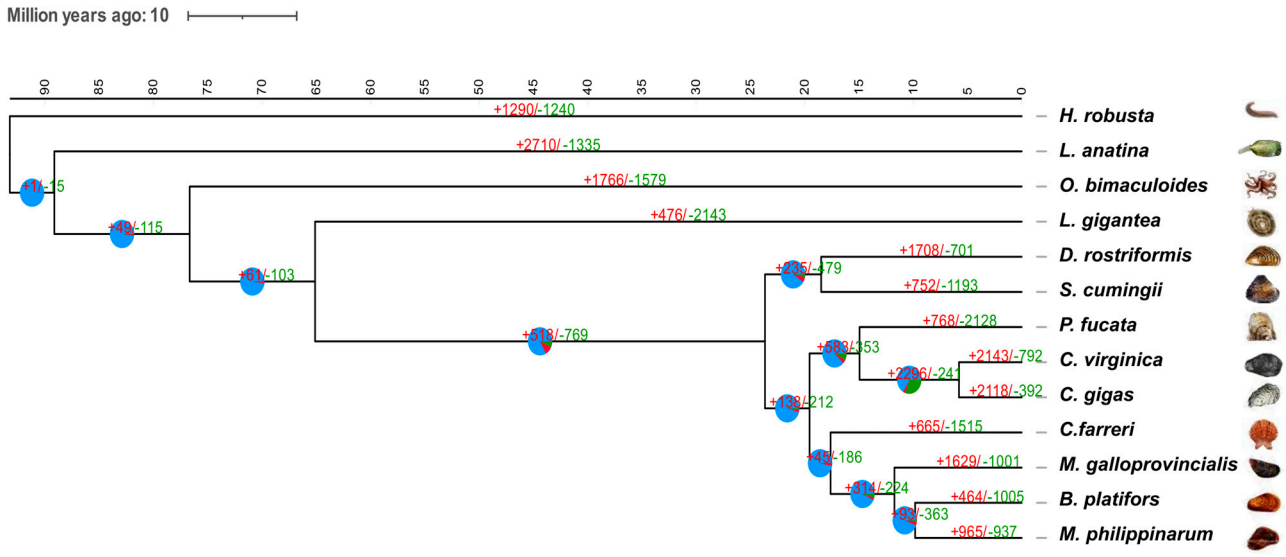


Figure 2. The phylogenetic relationships of *S. cumingii* with other species. The numbers of gene expansion (+) and contraction (−) are shown on the branches. The divergence times are dated and displayed below the phylogenetic tree.

2.4. Expansion and Contraction of Gene Families and Positively Selected Genes in the S. cumingii Genome

We performed an expansion and contraction of gene families in the *S. cumingii* genome via a probabilistic model with separated birth and death rates using CAFÉ version 4.2.1 software. In the genome of *S. cumingii*, it was found that 752 and 1193 gene families were significantly expanded and contracted, respectively ($p < 0.05$) (Figure 2). Specifically, the genes in the expanded families were significantly enriched in 84 KEGG pathways (Table S10), some of which were related to the immune system and disease resistance (Figure S8). In particular, some gene families were enriched in biomineralization-related pathways such as "calcium signaling" and "glycosaminoglycan biosynthesis-chondroitin sulfate/dermatan sulfate" (Table S10). The genes in the contracted families were significantly enriched in 18 KEGG pathways (Table S11), which included those related to the biosynthesis and metabolism of some organic compounds (Figure S9).

Then we used the Yang–Neilsen method in the CodeML program in PAML version 4.7a software to calculate ratios of nonsynonymous to synonymous substitutions and identify the positive selection genes. A total of 237 (1.70%) genes were found to be under

positive selection (Table S12). KEGG pathway enrichment analysis further showed that these genes were significantly enriched in 17 KEGG pathways (Table S13), which were mostly metabolic pathways (Figure S10). Notably, the characteristic genes (*Hcu_0920.t1* and *Hcu_0920.t33*), annotated as *fibrillin*, were also included in the expanded gene families (OG0000322 and OG0000925).

2.5. Fibrillin Family Genes in S. cumingii's Genome

A total of 146 fibrillin genes were identified in the genome of *S. cumingii* (Table S14). According to the topological structure of the phylogenetic tree, the fibrillin proteins of *S. cumingii* were roughly divided into three subgroups, each of which had a certain amount of fibrillins, among which subgroup 1 (yellow) had the largest amount of fibrillin proteins (Figure S11). At the same time, the conserved domain analysis of fibrillin proteins showed that they contain several calcium-binding epidermal growth factor (EGF-CA) domains (Figure S12). RNA-seq showed that *fibrillin* genes generally exhibited diverse tissue expression patterns in *S. cumingii*. The expression patterns of *fibrillin* gene family members were mainly divided into four groups, of which two groups showed a low expression level, and some members were even not expressed in all tissues. The expression levels of the other two groups were relatively high, and most of them were expressed in all tissues. Overall, the expression patterns of the *fibrillin* gene family had poor tissue specificity, with high expression levels in the mantle, ax foot, and heart (Figure 3A). Meanwhile, we analyzed the expression of *fibrillin* genes in sac tissues after mantle implantation (Figure 3B). The expression of *fibrillin* genes was relatively stable at all times and was roughly divided into six groups, with half of the family members showing high expression at all times. Overall, the expression level of the fibrillin gene family was relatively high from day 4 to day 14 after mantle implantation.

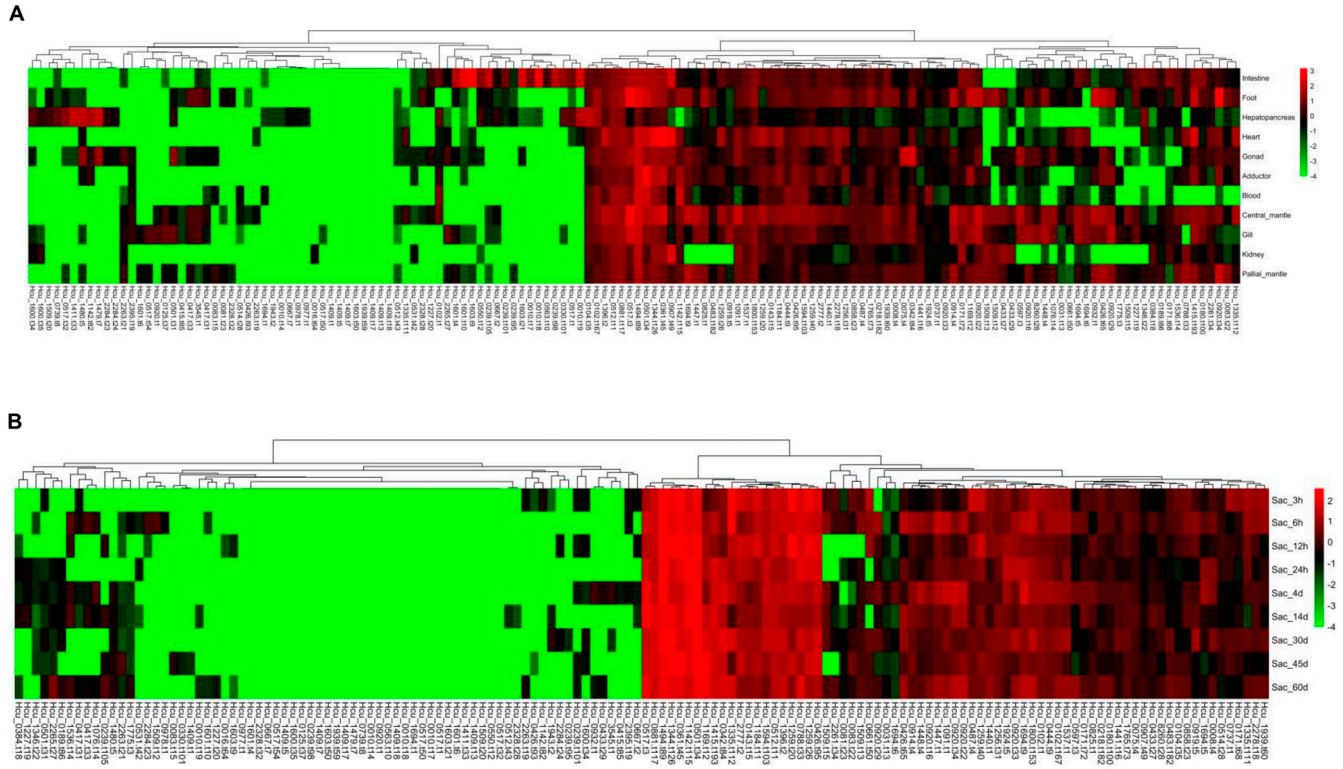


Figure 3. The expression profiles of *fibrillin* genes based on RNA-seq: (**A**) the expression profiles of *fibrillin* genes among eleven different tissues; (**B**) the expression profiles of *fibrillin* genes on sac tissue at different times after mantle implantation.

2.6. Function Analysis of One Key Fibrillin Gene Associated with Biomineralization in Sinohyriopsis cumingii

After the *fibrillin* gene was inhibited by RNA interference, its expression level was significantly downregulated, accounting for 75.1% of the original relative expression ($p < 0.01$, Figure 4A). SEM observations revealed that, in general, nacre growth was dominated by hexagonal aragonite flakes with smooth surfaces and nucleation sites. The disturbed aragonite flakes were irregular, with shapes that ceased to be hexagonal and gradually became round (Figure 4C). Furthermore, the surface of the normal prism layer was flat, and the crystals were closely connected through the organic matrix (Figure 4C). However, the surface of calcium carbonate crystals after interference was rougher, with great size differences, and the organic matrix between crystals appeared hollow (Figure 4C).

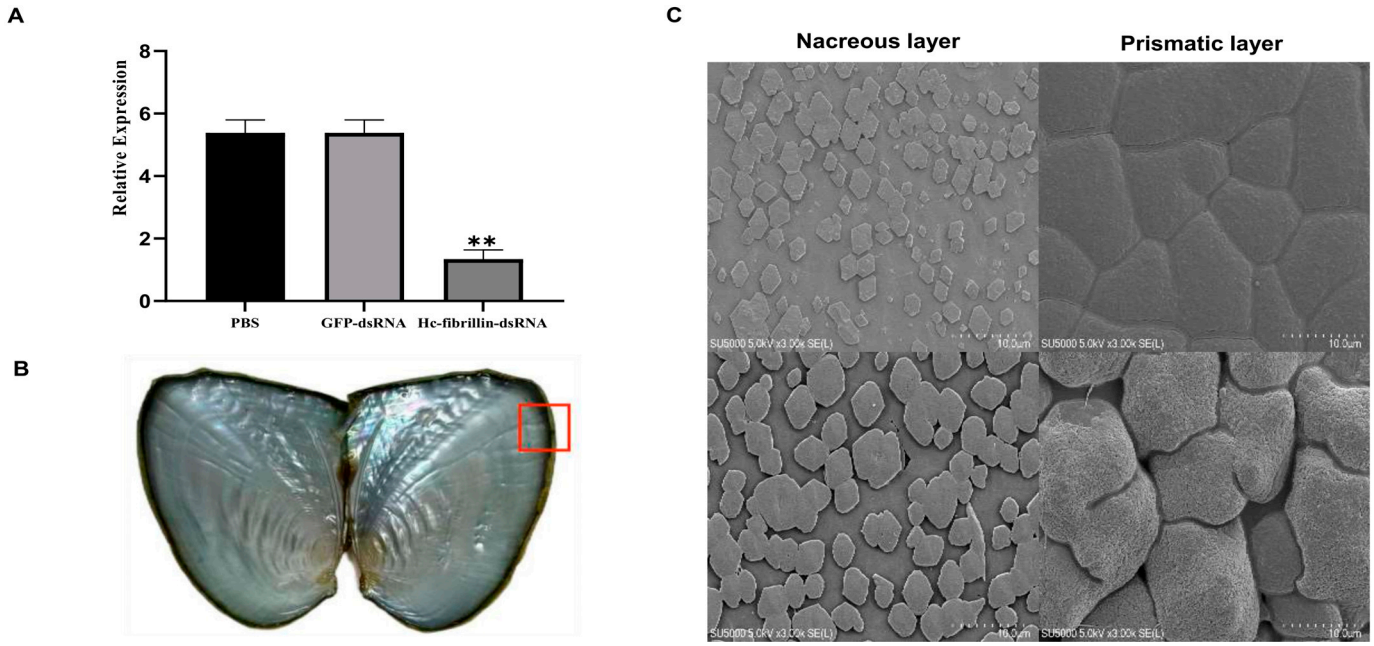


Figure 4. RNA interference analysis of the *fibrillin* gene in the biomineralization of *S. cumingii*: (**A**) Relative expression of the *fibrillin* gene in the mantle after dsRNA injection. Note: "**" indicates a significant difference ($p < 0.01$). (**B**) Inner shell of *S. cumingii* after RNA interference, red box represented location of shell pieces for SEM. (**C**) Microstructure of nacreous and prismatic layers observed after inhibition of the fibrillin gene (bar = 10 μm). Up, PBS and GFP; down, dsRNA-fibrillin.

The peptide comprising the EGF-CA domain was synthesized, and its effects on $CaCO_3$ crystallization were investigated. Crystals in the saturated $Ca(HCO_3)_2$ solution without peptide exhibited smooth rhomboid shapes (Figure 5A,B). Raman spectroscopy showed that characteristic peaks of intensity were at around 267, 1089, and 2952 cm^{-1} (Figure 5C). When crystals were saturated in $Ca(HCO_3)_2$ solution with 40 μg/mL peptide, calcium carbonate tended to be oval-shaped (Figure 5D,E). Raman spectroscopy showed that characteristic peaks of intensity were at approximately 254, 713, 1087, and 2956 cm^{-1} (Figure 5F); this was considered typical of aragonite crystals. Crystals in the saturated $Ca(HCO_3)_2$ solution with 80 μg/mL peptide appeared like radial needles (Figure 5G,H). When crystals were further examined by Raman spectroscopy, the characteristic peaks of intensity were at approximately 257, 713, 1087, and 2960 cm^{-1} (Figure 5I).

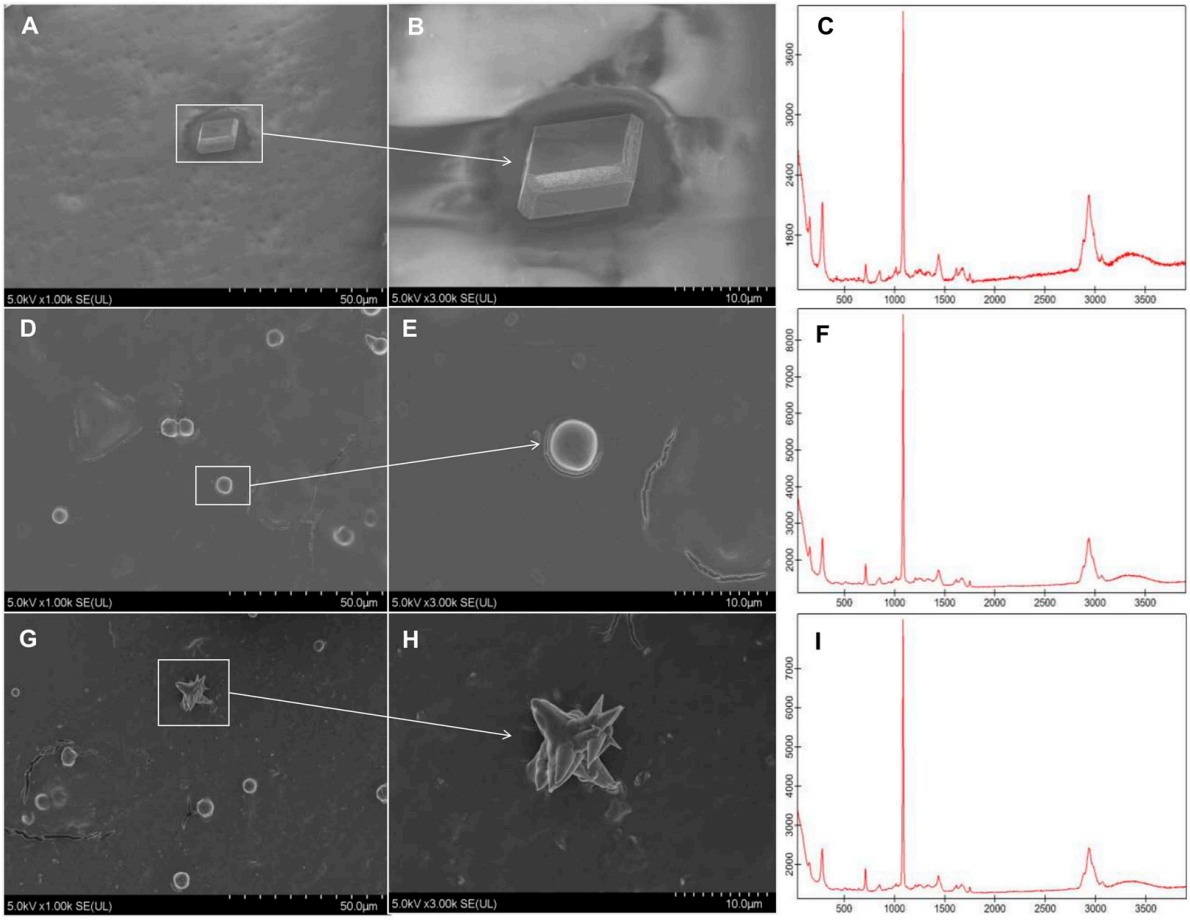


Figure 5. In vitro calcium carbonate crystallization analysis of the EGF-CA peptide: (**A**–**C**) represent SEM images and Raman spectroscopy of saturated $Ca(HCO_3)_2$ solution without peptide, respectively; (**D**–**F**) represent SEM images and Raman spectroscopy of $Ca(HCO_3)_2$ solution with 40 μg/mL EGF-CA peptides, respectively; (**G**–**I**) represent SEM images and Raman spectroscopy of $Ca(HCO_3)_2$ solution with 80 μg/mL EGF-CA peptides, respectively. The *X*-axis of Raman spectroscopy represents the Raman shift, while the *Y*-axis represents the spectral intensity.

3. Discussion

In this study, the genome of a female *S. cumingii* individual was subjected to high-depth sequencing to obtain a high-quality genome assembly. The total size of the assembled genome was 2.90 Gb with a contig N50 length of 5.30 Mb, and 89.92% of the genome sequence (2.61 Gb) was anchored onto 19 linkage groups. This represents the largest bivalve genome ever published in a database [14–17,19–46]. In contrast to the genome assembly techniques previously used for other freshwater bivalves [19–23], for the present study we used Pacbio HiFi technology, which was more conducive to our genome assembly of *S. cumingii*. Hence, we obtained a higher-quality genome assembly of *S. cumingii*, which exhibited a greater N50 length. The repeat elements identified in *S. cumingii* constituted 50.86% of the whole genome, suggesting a highly dynamic range of repeat content (9.7–62.0%) in mollusks [47]. Furthermore, the proportion of repeat content in *S. cumingii* was found to be higher than in most bivalves, and this may help to explain why *S. cumingii* has the largest genome size among bivalves. The repeat elements included 13.21% of DNA transposons and 12.38% of retroelements. These two repeat elements mediate gene duplication in the genome of the organism [48,49]. Higher proportions of DNA transposons and retroelements indicated that there might be a large number of gene families expanding rapidly during *S. cumingii* genome evolution. Moreover, this genome was predicted to contain a

total of 37,681 protein-coding genes. The number of genes in *S. cumingii* is comparable to that of *D. rostriformis* [19]; both species are freshwater bivalves with a reference genome.

Lineage specificity and the expansion of gene families play the most important roles in phenotypic diversity and in evolutionary adaptation to the environment [50]. In the genome of *S. cumingii*, the lineage-specific gene families were significantly enriched in pathways including "calcium signaling", "glycosaminoglycan biosynthesis-keratan sulfate", "glycosaminoglycan biosynthesis-chondroitin sulfate/dermatan sulfate", and "ECM-receptor interaction". The expanded gene families were mainly enriched in the "calcium signaling" and "glycosaminoglycan biosynthesis-chondroitin sulfate/dermatan sulfate" pathways. The composition of shells and pearls, both of which are products of calcium metabolism, is driven by the deposition of calcium carbonate, a complex process that is highly controlled by calcium signaling [51]. The lineage specificity and expansion of "calcium signaling" pathways confirmed the common sense that the formation of freshwater pearl shells and pearls relies greatly on calcium metabolism. Glycosaminoglycans with sulfonate groups could cooperate with acidic glycoproteins containing carboxyl groups to enrich calcium ions and participate in the nucleation of crystals [5]. Sulfotransferase (CHST), a key enzyme in the biosynthesis of glycosaminoglycans, plays an important role in the process of shell biomineralization by catalyzing the transfer of sulfonic acid groups to produce glycosaminoglycans with a rich negative charge [52]. Hence, it is speculated that the evolutionary impetus for the expansion of the *CHST* family in the genome of *S. cumingii* might also be due to the processes required for rapid biomineralization. Some extracellular matrix (ECM) proteins, such as collagen and VWAP, have been proven to be involved in molluscan biomineralization [17]. In bivalves, ECMs are also important in the composition of the blood circulation system, where amorphous calcium is abundant. The specific families of "ECM-receptor interaction" pathways may indicate that matrix proteins enter the shell through the circulatory system, or through granulocytes and exosomes, to participate in mineralization, thus supporting the cellular model of Mollusca mineralization [14,53,54]. However, some biomineral-related gene families previously found to be expanded in the pearl oyster genome [17], such as chitin synthases (CHSs), *chitinase*, VWA-containing proteins (VWAPs), and tyrosinase (Tyr), were not expanded in *S. cumingii*, suggesting that a diversity of mechanisms is involved in pearl formation in both seawater and freshwater mussels.

Through comparative genome analysis, we identified a significantly expanded family that was annotated as *fibrillin*. Fibrillins are crucial components of extracellular matrices and play a role in providing structural support [55,56]. We found more *fibrillin* genes in the genomes of *S. cumingii* and *P. fucata* than in those of *O. bimaculoides*, a cephalopod without mineralized shells; this implies that these genes play a role in shell formation. Fibrillin proteins contain many EGF-CA domains, which have been confirmed to exhibit a high affinity for calcium [57–60]. Based on RNA-seq data, the constant expression indicated that fibrillin genes continuously form microfibrils and then affect chitin network formation to play important roles in organic scaffold construction following crystal deposition [61], thus contributing to biomineralization during pearl formation. In the expanded *fibrillin* gene family, one gene was under positive selection, and its expression represented upregulation during pearl formation; this was used for preliminary functional analysis. We discovered that silencing the *fibrillin* gene led to irregular growth in mineralization layers, which has been previously observed in other shell matrix proteins [62,63]. Through an in vitro crystallization assay using the EGF-CA peptide, we observed that this peptide significantly altered the morphology of prismatic calcite. Despite this, the peptide with a higher concentration also induced the formation of spiculate crystals, commonly known as vaterite. Raman spectroscopy verified that these crystals were similar to calcite. The vaterite constitutes an unstable transitional phase in spherical shell formation and can be easily transferred to aragonite or calcite. A previous study suggested that the soluble organic matrix induces vaterite formation in the otolith [64]. Several shell matrix proteins in seawater pearl oysters have also been found to induce and stabilize vaterite formation [63,65,66]. Our findings

suggest that *fibrillin* genes are crucial for the deposition of calcium carbonate and the formation of amorphous crystals during the initial stages of biomineralization. Indeed, the excellent biomineralization ability of the mussel probably results from the massive expansion of the *fibrillin* gene family during the evolution of *S. cumingii*.

4. Materials and Methods

4.1. Sample Preparation and Sequencing

A healthy three-year-old female triangle sail mussel, *S. cumingii* (Figure 6), was sampled at the Chongming aquaculture base of Shanghai Ocean University (Chongming District, Shanghai, China) for genome sequencing. Genomic DNA was extracted from the mussel using the TIANamp Marine Animals DNA Kit (TIANGEN, Beijing, China) following the manufacturer's instructions. The genome sequencing of *S. cumingii* involved the creation of three distinct sequencing libraries, following established procedures: one library with a 350 bp insert size; another with 20 kb inserts; and a third with inserts ranging from 15 to 18 kb. The 350 bp insert-size library underwent sequencing using the whole-genome shotgun method on the Illumina platform. The library with inserts exceeding 20 kb was sequenced using the 10× Genomics sequencing platform. Finally, the library with insert sizes ranging from 15 to 18 kb was sequenced on the Pacbio Sequel II/IIe platform. Genome Size Estimation and de novo Genome Assembly.

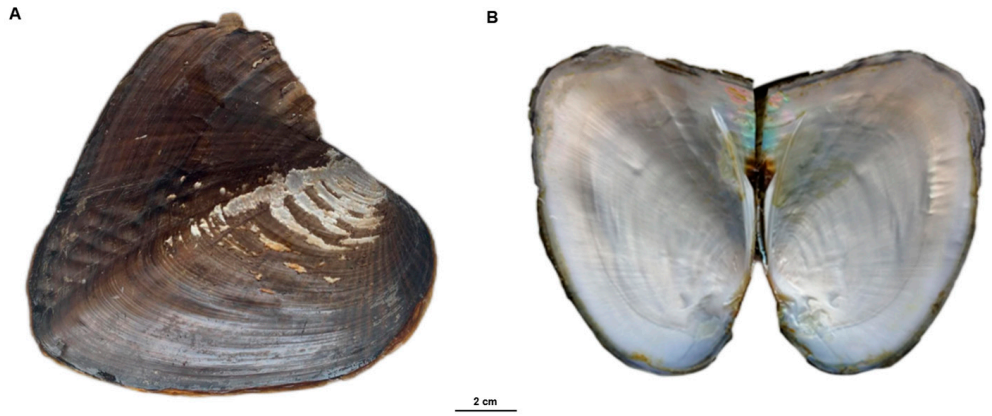


Figure 6. Photograph of a three-year-old triangle sail mussel, *S. cumingii* ((**A**): surface; (**B**): inner shell; bar = 2 cm).

The raw data of Illumina and 10× Genomics reads were filtered using fastp version version 0.23.2 software (https://github.com/OpenGene/fastp, accessed on 10 June 2023) to remove the low-quality sequences and adapter-contained reads. Further, ccs version 6.4.0 software (https://github.com/PacificBiosciences/ccs, accessed on 15 July 2023) was used to perform the quality control of Pacbio HiFi raw reads. The genome size was estimated by calculating the rate of K-mer number and peak depth using jellyfish version 2.2.7 [67] software. The Illumina and 10× Genomics reads were assembled into scaffolds using Supernova version 2.1.1 [68]; then, the scaffolds were anchored onto the chromosomes using data obtained from two *S. cumingii* genetic maps containing 492 SSRs and 4330 SNPs [69,70]. The Pacbio HiFi reads were also used for haploid genome assembly using HiFiasm version 0.16.1 with default parameters [71]. We then used the newly developed program Khaper to select primary contigs and filter redundant sequences from the initial assembly [72]. Finally, the filtering contigs were also anchored onto the chromosomes using the molecular markers in the genetic maps. The Benchmarking Universal Single-Copy Orthologs (BUSCO version 5.22) program [73] was used to evaluate the assembly's completeness by estimating the core genes based on the eukaryote odb10 and metazoan odb10 databases.

4.2. Repetitive Sequence Identification and Genome Annotation

Repetitive sequences in the genome assembly were identified through both de novo identification and homologous sequence alignments. RepeatModule version 1.0.11 was used to build a repeat library via de novo identification with default parameters. The Repbase database [74] was analyzed in RepeatMasker version 4.0.9 to identify repetitive sequences based on homology [75].

Based on the de novo methods, gene models were obtained using EVidenceModeler version 1.1.1 [76]. The predicted protein-coding genes were aligned with public databases, including the Clusters of Orthologous Groups (COGs) database [77] and the Kyoto Encyclopedia of Genes and Genomes (KEGG) [78], using BLASTP version 2.15.0 with a default E-value threshold (1×10^{-5}). The Gene Ontology (GO) database was also used for function annotation in Blast2GO version 2.5 [79].

4.3. Gene Family Identification and Phylogenetic Analysis

The protein sequence sets of 9 bivalves and 3 other invertebrate species were retrieved from the NCBI database for gene family analysis; these included *Bathymodiolus platifrons*, *Chlamys farreri*, *Crassotrea gigas*, *C. virginica*, *D. rostriformis*, *Helobdella robusta*, *Lingula anatina*, *Lottia gigantea*, *Mytilus galloprovincialis*, *Modiolus philippinarum*, *Octopus bimaculoides*, and *P. fucata*. OrthoFinder version 2.3.3 [80] was used to assign gene family clusters with default parameters. The enrichment of *S. cumingii*-specific gene families in KEGG pathways was also analyzed using KOBAS version 1.2.0 [81].

A phylogenetic tree was inferred using the 130 shared single-copy orthologs identified in OrthoFinder. Multiple alignments of each protein sequence were performed using MUSCLE version 3.8.31 [82], and the conserved blocks obtained from the alignments were selected in Gblocks version 0.91b [83]. All alignments were subsequently concatenated to form a super-protein sequence and used to generate the maximum likelihood (ML) phylogenetic tree in RAxML version 8.2.12 [84] with 1000 bootstrap replicates. Divergence times were estimated using the MCMCTree program in the PAML package version 4.7a [85] based on the topological structure and comparison matrix. The reference divergence time was obtained from the TimeTree database [86].

4.4. Gene Family Expansion and Contraction, and Positive Selection Analysis

Expansion and contraction of the conserved homolog clusters were determined in CAFE version 4.2.1 [87] using a probabilistic model with separated birth and death rates and a *p*-value threshold of 0.01. Conserved coding DNA sequence (CDS) alignments of each single-copy gene family were extracted for further identification of positively selected genes. The ratios of nonsynonymous to synonymous substitutions (Ka/Ks) were estimated for each single-copy orthologous gene using the Yang–Neilsen method in the CodeML program [88] with the branch-site model implemented in PAML version 4.7a. A likelihood ratio test was conducted, and a false discovery rate correction was performed for multiple comparisons. Genes with a rate of Ka/Ks > 1 and a corrected *p*-value < 0.05 were defined as positively selected genes. The enrichment of expanded, contracted, and positively selected genes in KEGG pathways was also analyzed using KOBAS.

4.5. Identification of Fibrillin Gene Family Members and Phylogenetic Analysis

Fibrillin proteins were identified among the protein sequences of *S. cumingii*'s genome based on a homology search conducted via BLASTP [89] against ten *C. gigas* fibrillin protein sequences obtained from the MolluscDB [90] database. Then, all candidate protein sequences were confirmed on the NCBI Conserved Domain Database (CDD) [91] and Simple Modular Architecture Research Tool (SMART) [92] databases. In addition, the preliminary candidate fibrillin genes were further identified using the hmmsearch tool in HMMER version 3.3.2 [93] with the Hidden Markov Model (HMM) profile of EGF_CA (PF07645). The same workflow was also applied to identify fibrillin protein sequences in *C. gigas*, *C. farreri*, *D. rostriformis*, *O. bimaculoides*, and *P. fucata*. The fibrillin amino acid

sequences identified in these four species were aligned using MUSCLE. A phylogenetic tree was constructed using the ML method in IQ-Tree version 1.6.12 [94]. Bootstrapping with 1000 replications was applied to estimate the support rate of branch nodes. The best-fitting model (VT + R7) was selected via ModelFinder in IQ-Tree according to the Bayesian information criterion (BIC). Fibrillins obtained from two species belonging to the phylum Platyhelminthes (*Schistosoma haematobium* and *Schistosoma mansoniwere*) were used as outgroups. The physicochemical parameters, molecular weights (kDa), and isoelectric points (pI) of the fibrillins were calculated in ExPASy version 3.0 [95].

4.6. Transcriptome Analysis

Samples of 11 tissues, namely the intestine, foot, hepatopancreas, heart, gonad, adductor, blood, mantle, gill, kidney, and pallial line, were collected from three individuals. Sac tissues were also collected from three individuals at different time points after mantle implantation, i.e., at 3, 6, 12, and 24 h, and at 4, 14, 30, 45, and 60 days. All samples were stored in the RNAstore (TIANGEN, Beijing, China) for transportation. Total mRNA was extracted with TRIzol reagent (Invitrogen, Waltham, MA, USA) according to the manufacturer's instructions, and three repeat tissues were then mixed for a reverse transcription reaction. Illumina RNA-seq libraries were prepared and sequenced on an Illumina HiSeq2500 platform with a PE150 model. After quality control, high-quality reads were mapped onto the genome of *S. cumingii* using Hisat2 version 2.2.1 [96]. The gene expression level was determined by calculating the transcripts per kilobase of the exon model per million mapped reads (TPM) using Featurecounts (version 2.0.3) [97].

4.7. RNA Silencing Assay

Double-stranded RNA (dsRNA) was generated for the RNAi assay via PCR amplification of the *fibrillin* gene under positive selection (Hcu_0920.t1) with a T7 promoter sequence (GGATCCTAATACGACTCACTATAGGG) attached to the primers (forward: GAAAATAATGGTGGATGCG; reverse: CAGAAAAAAGAGCCGATAGT). After PCR amplification, in vitro transcription was conducted using the T7 High Efficiency Transcription Kit (TransGen Biotech, Beijing, China). Phosphate-buffered saline (PBS) and a fragment of the green fluorescence protein (GFP) sequence from the pEGFP-N1 plasmid were used as blank control and negative control dsRNA, respectively. Ten one-year-old mussels from each group were used as the experimental samples. The dsRNA was diluted to 60 μg/100 μL using RNase-free water and was then injected into the adductor. Total RNA from the mantle tissue was extracted 7 days after injection, and qRT-PCR was performed to analyze expression performance. Then, the shell pieces were washed and dried (Figure 4B), and the nacre and prismatic layers of the shell were observed by scanning electron microscopy (SEM).

4.8. In Vitro Calcium Carbonate Crystallization Assay

The peptide of the conserved calcium-binding epidermal growth factor domain (EGF-CA) (DIDECAKYASKICQNGKCLNTNPSYTCECYNGYVPDDKNMTCK) was synthesized by Shanghai Qiangyao Biological Co., Ltd. (Shanghai, China) (Figure S13). The peptide was diluted to concentrations of 40 μg/mL and 80 μg/mL before being mixed with a saturated solution of calcium bicarbonate ($Ca(HCO_3)_2$). In the blank control groups, a reaction system consisting of 20 μL $Ca(HCO_3)_2$ solution was used. In the experimental group, the peptides were divided into three concentrations, and 10 μL of different concentrations of peptide was added to 10 μL of $Ca(HCO_3)_2$ solution. The crystallization reaction was carried out on siliconized slides under airtight conditions for 48 h. Crystal morphology was identified by SEM, and crystal types were determined by testing the spectral intensity of crystals using Raman spectroscopy, with Raman shifts ranging from 0 to 4000 cm^{-1}.

5. Conclusions

In this study, we created a high-quality genome assembly of *S. cumingii* with a size of 2.90 Gb and a contig N50 length of 5.30 Mb. It is the first genome of the Sinohyriopsis genus. A substantial 89.92% of sequences (2.61 Gb) were successfully anchored onto 19 linkage groups. The annotation revealed 37,681 protein-coding genes, and it highlighted the prevalence of repeat elements, explaining why the genome size of *S. cumingii* is so large. A comparative genomics analysis uncovered 752 expanded gene families and 237 genes exhibiting positive selection. Notably, our study provides compelling evidence demonstrating that the remarkable biomineralization ability of this species is driven by the expansion of the *fibrillin* gene. In light of these results, we suggest that freshwater pearl production could be improved using *S. cumingii*. The genome assembly achieved in this study serves as a valuable resource for a wide range of genomic, biological, and ecological investigations into *S. cumingii*. Additionally, it lays the groundwork for practical developments in the freshwater pearl industry, such as molecular breeding and innovations in culture technology.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms25063146/s1.

Author Contributions: J.L., L.C., Z.B. and Y.L. (Ying Lu) conceived and designed the experiment. Z.B., Y.L. (Ying Lu) and H.H. wrote, revised, and edited the paper. H.H. and Y.Y. performed the experiments. H.H. and Y.L. (Yalin Li) analyzed the data. X.L., G.W., D.H., Z.W., Y.M. and H.W. contributed materials/analysis tools. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Key Research and Development Program of China (2022YFD240010, 2018YFD0901406, 2018YFD0900101) and the earmarked fund for the China Agriculture Research System (CARS-49).

Data Availability Statement: The Whole Genome Shotgun project of *S. cumingii* has been deposited in the NCBI Sequence Read Archive (SRA) database under Bioproject ID: PRJNA909938. The genome assembly data have been deposited at DDBJ/ENA/GenBank under the accession number GCA_028554795.2.

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

References

1. Taylor, J.; Strack, E. Pearl Production. In *The Pearl Oyster*; Southgate, P.C., Lucas, J.S., Eds.; Elsevier: Amsterdam, The Netherlands, 2008; pp. 273–302; ISBN 9780444529763.
2. Li, J.; Wang, D.; Bai, Z.; Guan, Y.; Wu, C.; Chen, L. Report on the Development of Freshwater Pearl Culture Industry in China. *China Aquac.* **2019**, 23–29.
3. Li, J.; Wang, G.; Bai, Z. Genetic Variability in Four Wild and Two Farmed Stocks of the Chinese Freshwater Pearl Mussel (*Hyriopsis cumingii*) Estimated by Microsatellite DNA Markers. *Aquaculture* **2009**, *287*, 286–291. [CrossRef]
4. Griffith, A.W.; Gobler, C.J. Harmful Algal Blooms: A Climate Change Co-Stressor in Marine and Freshwater Ecosystems. *Harmful Algae* **2020**, *91*, 101590. [CrossRef]
5. Addadi, L.; Joester, D.; Nudelman, F.; Weiner, S. Mollusk Shell Formation: A Source of New Concepts for Understanding Biomineralization Processes. *Chem.-Eur. J.* **2006**, *12*, 980–987. [CrossRef]
6. Marin, F.; Luquet, G.; Marie, B.; Medakovic, D. Molluscan Shell Proteins: Primary Structure, Origin, and Evolution. *Curr. Top. Dev. Biol.* **2007**, *80*, 209–276. [CrossRef]
7. Furuhashi, T.; Schwarzinger, C.; Miksik, I.; Smrz, M.; Beran, A. Molluscan Shell Evolution with Review of Shell Calcification Hypothesis. *Comp. Biochem. Physiol.-B Biochem. Mol. Biol.* **2009**, *154*, 351–371. [CrossRef]
8. Kenneth, S.; Wilbur, K.M. Biomineralization: Cell Biology and Mineral Deposition. *Q. Rev. Biol.* **1989**, *1*, 257–264.
9. Lowenstam, H.A.; Weiner, S. *On Biomineralization*; Oxford University Press: Oxford, UK, 1989.
10. Bevelander, G.; Nakahara, H. An Electron Microscope Study of the Formation of the Nacreous Layer in the Shell of Certain Bivalve Molluscs. *Calcif. Tissue Res.* **1969**, *3*, 84–92. [CrossRef] [PubMed]
11. Zhao, R.; Takeuchi, T.; Luo, Y.J.; Ishikawa, A.; Kobayashi, T.; Koyanagi, R.; Villar-Briones, A.; Yamada, L.; Sawada, H.; Iwanaga, S.; et al. Dual Gene Repertoires for Larval and Adult Shells Reveal Molecules Essential for Molluscan Shell Formation. *Mol. Biol. Evol.* **2018**, *35*, 2751–2761. [CrossRef] [PubMed]

12. Jackson, D.J.; McDougall, C.; Woodcroft, B.; Moase, P.; Rose, R.A.; Kube, M.; Reinhardt, R.; Rokhsar, D.S.; Montagnani, C.; Joubert, C.; et al. Parallel Evolution of Nacre Building Gene Sets in Molluscs. *Mol. Biol. Evol.* **2010**, *27*, 591–608. [CrossRef] [PubMed]
13. Marin, F. Mollusc Shellomes: Past, Present and Future. *J. Struct. Biol.* **2020**, *212*, 107583. [CrossRef]
14. Zhang, G.; Fang, X.; Guo, X.; Li, L.; Luo, R.; Xu, F.; Yang, P.; Zhang, L.; Wang, X.; Qi, H.; et al. The Oyster Genome Reveals Stress Adaptation and Complexity of Shell Formation. *Nature* **2012**, *490*, 49–54. [CrossRef]
15. Li, Y.; Sun, X.; Hu, X.; Xun, X.; Zhang, J.; Guo, X.; Jiao, W.; Zhang, L.; Liu, W.; Wang, J.; et al. Scallop Genome Reveals Molecular Adaptations to Semi-Sessile Life and Neurotoxins. *Nat. Commun.* **2017**, *8*, 1721. [CrossRef]
16. Wang, S.; Zhang, J.; Jiao, W.; Li, J.; Xun, X.; Sun, Y.; Guo, X.; Huan, P.; Dong, B.; Zhang, L.; et al. Scallop Genome Provides Insights into Evolution of Bilaterian Karyotype and Development. *Nat. Ecol. Evol.* **2017**, *1*, 120. [CrossRef]
17. Du, X.; Fan, G.; Jiao, Y.; Zhang, H.; Guo, X.; Huang, R.; Zheng, Z.; Bian, C.; Deng, Y.; Wang, Q.; et al. The Pearl Oyster *Pinctada Fucata martensii* Genome and Multi-Omic Analyses Provide Insights into Biomineralization. *GigaScience* **2017**, *6*, gix059. [CrossRef]
18. Zhang, Y.; Mao, F.; Mu, H.; Huang, M.; Bao, Y.; Wang, L.; Wong, N.K.; Xiao, S.; Dai, H.; Xiang, Z.; et al. The Genome of *Nautilus pompilius* Illuminates Eye Evolution and Biomineralization. *Nat. Ecol. Evol.* **2021**, *5*, 927–938. [CrossRef] [PubMed]
19. Calcino, A.D.; De Oliveira, A.L.; Simakov, O.; Schwaha, T.; Zieger, E.; Wollesen, T.; Wanninger, A. The Quagga Mussel Genome and the Evolution of Freshwater Tolerance. *DNA Res.* **2019**, *26*, 411–422. [CrossRef] [PubMed]
20. Renaut, S.; Guerra, D.; Hoeh, W.R.; Stewart, D.T.; Bogan, A.E.; Ghiselli, F.; Milani, L.; Passamonti, M.; Breton, S. Genome Survey of the Freshwater Mussel *Venustaconcha ellipsiformis* (Bivalvia: Unionida) Using a Hybrid de Novo Assembly Approach. *Genome Biol. Evol.* **2018**, *10*, 1637–1646. [CrossRef] [PubMed]
21. Rogers, R.L.; Grizzard, S.L.; Titus-McQuillan, J.E.; Bockrath, K.; Patel, S.; Wares, J.P.; Garner, J.T.; Moore, C.C. Gene Family Amplification Facilitates Adaptation in Freshwater Unionid Bivalve *Megalonaias nervosa*. *Mol. Ecol.* **2020**, *30*, 1155–1173. [CrossRef] [PubMed]
22. Smith, C.H. A High-Quality Reference Genome for a Parasitic Bivalve with Doubly Uniparental inheritance (Bivalvia: Unionida). *Genome Biol. Evol.* **2021**, *13*, evab029. [CrossRef] [PubMed]
23. Gomes-Dos-Santos, A.; Lopes-Lima, M.; Machado, A.M.; Marcos Ramos, A.; Usié, A.; Bolotov, I.N.; Vikhrev, I.V.; Breton, S.; Castro, L.F.C.; Da Fonseca, R.R.; et al. The Crown Pearl: A Draft Genome Assembly of the European Freshwater Pearl Mussel *Margaritifera margaritifera* (Linnaeus, 1758). *DNA Res.* **2021**, *28*, dsab002. [CrossRef]
24. Sun, J.; Zhang, Y.; Xu, T.; Zhang, Y.; Mu, H.; Zhang, Y.; Lan, Y.; Fields, C.J.; Hui, J.H.L.; Zhang, W.; et al. Adaptation to Deep-Sea Chemosynthetic Environments as Revealed by Mussel Genomes. *Nat. Ecol. Evol.* **2017**, *1*, 121. [CrossRef]
25. Uliano-Silva, M.; Dondero, F.; Dan Otto, T.; Costa, I.; Lima, N.C.B.; Americo, J.A.; Mazzoni, C.J.; Prosdocimi, F.; Rebelo, M.d.F. A Hybrid-Hierarchical Genome Assembly Strategy to Sequence the Invasive Golden Mussel, *Limnoperna fortunei*. *GigaScience* **2018**, *7*, gix128. [CrossRef]
26. Li, C.; Liu, X.; Liu, B.; Ma, B.; Liu, F.; Liu, G.; Shi, Q.; Wang, C. Draft Genome of the Peruvian Scallop *Argopecten purpuratus*. *GigaScience* **2018**, *7*, giy031. [CrossRef]
27. Powell, D.; Subramanian, S.; Suwansa-Ard, S.; Zhao, M.; O'Connor, W.; Raftos, D.; Elizur, A.; Kohara, Y. The Genome of the Oyster Saccostrea Offers Insight into the Environmental Resilience of Bivalves. *DNA Res.* **2018**, *25*, 655–665. [CrossRef]
28. Ran, Z.; Li, Z.; Yan, X.; Liao, K.; Kong, F.; Zhang, L.; Cao, J.; Zhou, C.; Zhu, P.; He, S.; et al. Chromosome-Level Genome Assembly of the Razor Clam *Sinonovacula constricta* (Lamarck, 1818). *Mol. Ecol. Resour.* **2019**, *19*, 1647–1658. [CrossRef]
29. Yan, X.; Nie, H.; Huo, Z.; Ding, J.; Li, Z.; Yan, L.; Jiang, L.; Mu, Z.; Wang, H.; Meng, X.; et al. Clam Genome Sequence Clarifies the Molecular Basis of Its Benthic Adaptation and Extraordinary Shell Color Diversity. *iScience* **2019**, *19*, 1225–1237. [CrossRef] [PubMed]
30. Bai, C.M.; Xin, L.S.; Rosani, U.; Wu, B.; Wang, Q.C.; Duan, X.K.; Liu, Z.H.; Wang, C.M. Chromosomal-Level Assembly of the Blood Clam, *Scapharca* (Anadara) *broughtonii*, Using Long Sequence Reads and Hi-C. *GigaScience* **2019**, *8*, giz067. [CrossRef] [PubMed]
31. Thai, B.T.; Lee, Y.P.; Gan, H.M.; Austin, C.M.; Croft, L.J.; Trieu, T.A.; Tan, M.H. Whole Genome Assembly of the Snout Otter Clam, *Lutraria rhynchaena*, Using Nanopore and Illumina Data, Benchmarked Against Bivalve Genome Assemblies. *Front. Genet.* **2019**, *10*, 1158. [CrossRef]
32. Wang, X.; Xu, W.; Wei, L.; Zhu, C.; He, C.; Song, H.; Cai, Z.; Yu, W.; Jiang, Q.; Li, L.; et al. Nanopore Sequencing and De Novo Assembly of a Black-Shelled Pacific Oyster (*Crassostrea gigas*) Genome. *Front. Genet.* **2019**, *10*, 1211. [CrossRef] [PubMed]
33. Li, R.; Zhang, W.; Lu, J.; Zhang, Z.; Mu, C.; Song, W.; Migaud, H.; Wang, C.; Bekaert, M. The Whole-Genome Sequencing and Hybrid Assembly of *Mytilus coruscus*. *Front. Genet.* **2020**, *11*, 440. [CrossRef] [PubMed]
34. Kenny, N.J.; McCarthy, S.A.; Dudchenko, O.; James, K.; Betteridge, E.; Corton, C.; Dolucan, J.; Mead, D.; Oliver, K.; Omer, A.D.; et al. The Gene-Rich Genome of the Scallop *Pecten maximus*. *GigaScience* **2020**, *9*, giaa037. [CrossRef]
35. Wei, M.; Ge, H.; Shao, C.; Yan, X.; Nie, H.; Duan, H.; Liao, X.; Zhang, M.; Chen, Y.; Zhang, D.; et al. Chromosome-Level Clam Genome Helps Elucidate the Molecular Basis of Adaptation to a Buried Lifestyle. *iScience* **2020**, *23*, 101148. [CrossRef]
36. Dong, Y.; Zeng, Q.; Ren, J.; Yao, H.; Lv, L.; He, L.; Ruan, W.; Xue, Q.; Bao, Z.; Wang, S.; et al. The Chromosome-Level Genome Assembly and Comprehensive Transcriptomes of the Razor Clam (*Sinonovacula constricta*). *Front. Genet.* **2020**, *11*, 664. [CrossRef]
37. Yang, J.L.; Feng, D.D.; Liu, J.; Xu, J.K.; Chen, K.; Li, Y.F.; Zhu, Y.T.; Liang, X.; Lu, Y. Chromosome-Level Genome Assembly of the Hard-Shelled Mussel *Mytilus coruscus*, a Widely Distributed Species from the Temperate Areas of East Asia. *GigaScience* **2021**, *10*, giab024. [CrossRef]

38. Teng, W.; Xie, X.; Nie, H.; Sun, Y.; Liu, X.; Yu, Z.; Zheng, J.; Liu, H.; Li, D.; Zhang, M.; et al. Chromosome-Level Genome Assembly of *Scapharca kagoshimensis* Reveals the Expanded Molecular Basis of Heme Biosynthesis in Ark Shells. *Mol. Ecol. Resour.* **2022**, *22*, 295–306. [CrossRef]
39. Peñaloza, C.; Gutierrez, A.P.; Eöry, L.; Wang, S.; Guo, X.; Archibald, A.L.; Bean, T.P.; Houston, R.D. A Chromosome-Level Genome Assembly for the Pacific Oyster *Crassostrea gigas*. *GigaScience* **2021**, *10*, giab020. [CrossRef]
40. Bao, Y.; Zeng, Q.; Wang, J.; Zhang, Z.; Zhang, Y.; Wang, S.; Wong, N.K.; Yuan, W.; Huang, Y.; Zhang, W.; et al. Genomic Insights into the Origin and Evolution of Molluscan Red-Bloodedness in the Blood Clam *Tegillarca granosa*. *Mol. Biol. Evol.* **2021**, *38*, 2351–2365. [CrossRef] [PubMed]
41. Zeng, Q.; Liu, J.; Wang, C.; Wang, H.; Zhang, L.; Hu, J.; Bao, L.; Wang, S. High-Quality Reannotation of the King Scallop Genome Reveals No 'Gene-Rich' Feature and Evolution of Toxin Resistance. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 4954–4960. [CrossRef] [PubMed]
42. Takeuchi, T.; Suzuki, Y.; Watabe, S.; Nagai, K.; Masaoka, T.; Fujie, M.; Kawamitsu, M.; Satoh, N.; Myers, E.W. A High-Quality, Haplotype-Phased Genome Reconstruction Reveals Unexpected Haplotype Diversity in a Pearl Oyster. *DNA Res.* **2022**, *29*, dsac035. [CrossRef]
43. Takeuchi, T.; Kawashima, T.; Koyanagi, R.; Gyoja, F.; Tanaka, M.; Ikuta, T.; Shoguchi, E.; Fujiwara, M.; Shinzato, C.; Hisata, K.; et al. Draft Genome of the Pearl Oyster *Pinctada fucata*: A Platform for Understanding Bivalve Biology. *DNA Res.* **2012**, *19*, 117–130. [CrossRef] [PubMed]
44. Takeuchi, T.; Koyanagi, R.; Gyoja, F.; Kanda, M.; Hisata, K.; Fujie, M.; Goto, H.; Yamasaki, S.; Nagai, K.; Morino, Y.; et al. Bivalve-Specific Gene Expansion in the Pearl Oyster Genome: Implications of Adaptation to a Sessile Lifestyle. *Zool. Lett.* **2016**, *2*, 3. [CrossRef] [PubMed]
45. Murgarella, M.; Puiu, D.; Novoa, B.; Figueras, A.; Posada, D.; Canchaya, C. A First Insight into the Genome of the Filter-Feeder Mussel *Mytilus galloprovincialis*. *PLoS ONE* **2016**, *11*, e0151561. [CrossRef]
46. Mun, S.; Kim, Y.J.; Markkandan, K.; Shin, W.; Oh, S.; Woo, J.; Yoo, J.; An, H.; Han, K. The Whole-Genome and Transcriptome of the Manila Clam (*Ruditapes philippinarum*). *Genome Biol. Evol.* **2017**, *9*, 1487–1498. [CrossRef] [PubMed]
47. Yang, Z.; Zhang, L.; Hu, J.; Wang, J.; Bao, Z.; Wang, S. The Evo-Devo of Molluscs: Insights from a Genomic Perspective. *Evol. Dev.* **2020**, *22*, 409–424. [CrossRef]
48. Brosius, J. Retroposons-Seeds of Evolution. *Science* **1991**, *251*, 753. [CrossRef]
49. Juretic, N.; Hoen, D.R.; Huynh, M.L.; Harrison, P.M.; Bureau, T.E. The Evolutionary Fate of MULE-Mediated Duplications of Host Gene Fragments in Rice. *Genome Res.* **2005**, *15*, 1292–1297. [CrossRef]
50. Harris, R.M.; Hofmann, H.A. Seeing Is Believing: Dynamic Evolution of Gene Families. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 1252–1253. [CrossRef]
51. Fang, Z.; Yan, Z.; Li, S.; Wang, Q.; Cao, W.; Xu, G.; Xiong, X.; Xie, L.; Zhang, R. Localization of Calmodulin and Calmodulin-like Protein and Their Functions in Biomineralization in *P. fucata*. *Prog. Nat. Sci.* **2008**, *18*, 405–412. [CrossRef]
52. Fernandez, M.; Arriagada, K.; Arias, J. SEM Localization of Proteoglycans in Abalone Shell (*Haliotis rufescens*). *Microsc. Microanal.* **2007**, *13*, 1462–1463. [CrossRef]
53. Schwaner, C.; Farhat, S.; Haley, J.; Pales Espinosa, E.; Allam, B. Transcriptomic, Proteomic, and Functional Assays Underline the Dual Role of Extrapallial Hemocytes in Immunity and Biomineralization in the Hard Clam *Mercenaria mercenaria*. *Front. Immunol.* **2022**, *13*, 838530. [CrossRef]
54. Mount, A.S.; Wheeler, A.P.; Paradkar, R.P.; Snider, D. Hemocyte-Mediated Shell Mineralization in the Eastern Oyster. *Science* **2004**, *304*, 297–300. [CrossRef] [PubMed]
55. Smaldone, S.; Ramirez, F. Fibrillin Microfibrils in Bone Physiology. *Matrix Biol.* **2016**, *52–54*, 191–197. [CrossRef]
56. Sakai, L.Y.; Keene, D.R. Fibrillin Protein Pleiotropy: Acromelic Dysplasias. *Matrix Biol.* **2019**, *80*, 6–13. [CrossRef]
57. Handford, P.A.; Baron, M.; Mayhew, M.; Willis, A.; Beesley, T.; Brownlee, G.G.; Campbell, I.D. The First EGF-like Domain from Human Factor IX Contains a High-Affinity Calcium Binding Site. *EMBO J.* **1990**, *9*, 475–480. [CrossRef] [PubMed]
58. Huang, L.H.; Cheng, H.; Sweeney, W.V.; Pardi, A.; Tam, J.P. Sequence-Specific 1H NMR Assignments, Secondary Structure, and Location of the Calcium Binding Site in the First Epidermal Growth Factor Like Domain of Blood Coagulation Factor IX. *Biochemistry* **1991**, *30*, 7402–7409. [CrossRef]
59. Liu, Y.; Annis, D.S.; Mosher, D.F. Interactions among the Epidermal Growth Factor-like Modules of Thrombospondin-1. *J. Biol. Chem.* **2009**, *284*, 22206–22212. [CrossRef] [PubMed]
60. Valcarce, C.; Selander-Sunnerhagen, M.; Tamlitz, A.M.; Drakenberg, T.; Bjork, I.; Stenflo, J. Calcium Affinity of the NH_2-Terminal Epidermal Growth Factor-like Module of Factor, X. Effect of the γ-Carboxyglutamic Acid-Containing Module. *J. Biol. Chem.* **1993**, *268*, 26673–26678. [CrossRef]
61. Jin, C.; Zhao, J.; Pu, J.; Liu, X.; Li, J. Hichin, a Chitin Binding Protein Is Essential for the Self-Assembly of Organic Frameworks and Calcium Carbonate during Shell Formation. *Int. J. Biol. Macromol.* **2019**, *135*, 745–751. [CrossRef]
62. Zhang, X.; Xia, Z.; Liu, X.; Li, J. The Novel Matrix Protein Hic7 of *Hyriopsis cumingii* Participates in the Formation of the Shell and Pearl. *Comp. Biochem. Physiol. Part-B Biochem. Mol. Biol.* **2021**, *256*, 110640. [CrossRef]
63. Zhang, X.; Yin, Z.; Ma, Z.; Liang, J.; Zhang, Z.; Yao, L.; Chen, X.; Liu, X.; Zhang, R. Shell Matrix Protein N38 of *Pinctada fucata*, Inducing Vaterite Formation, Extends the DING Protein to the Mollusca World. *Mar. Biotechnol.* **2022**, *24*, 531–541. [CrossRef]

64. Lakshminarayanan, R.; Chi-Jin, E.O.; Loh, X.J.; Kini, R.M.; Valiyaveettil, S. Purification and Characterization of a Vaterite-Inducing Peptide, Pelovaterin, from the Eggshells of *Pelodiscus sinensis* (Chinese Soft-Shelled Turtle). *Biomacromolecules* **2005**, *6*, 1429–1437. [CrossRef]
65. Natoli, A.; Wiens, M.; Schröder, H.C.; Stifanic, M.; Batel, R.; Soldati, A.L.; Jacob, D.E.; Müller, W.E.G. Bio-Vaterite Formation by Glycoproteins from Freshwater Pearls. *Micron* **2010**, *41*, 359–366. [CrossRef]
66. Yan, Y.; Yang, D.; Yang, X.; Liu, C.; Xie, J.; Zheng, G.; Xie, L.; Zhang, R. A Novel Matrix Protein, PfY2, Functions as a Crucial Macromolecule during Shell Formation. *Sci. Rep.* **2017**, *7*, 6021. [CrossRef]
67. Guillaume, M.; Carl, K. A Fast, Lock-free Approach for Efficient Parallel Counting of Occurrences of K-mers. *Bioinformatics* **2011**, *27*, 764–770. [CrossRef]
68. Coombe, L.; Warren, R.L.; Jackman, S.D.; Yang, C.; Vandervalk, B.P.; Moore, R.A.; Pleasance, S.; Coope, R.J.; Bohlmann, J.; Holt, R.A.; et al. Assembly of the Complete *Sitka spruce* Chloroplast Genome Using 10X Genomics' GemCode Sequencing Data. *PLoS ONE* **2016**, *11*, 0163059. [CrossRef]
69. Bai, Z.Y.; Han, X.K.; Liu, X.J.; Li, Q.Q.; Li, J. Le Construction of a High-Density Genetic Map and QTL Mapping for Pearl Quality-Related Traits in *Hyriopsis cumingii*. *Sci. Rep.* **2016**, *6*, 32608. [CrossRef]
70. Bai, Z.; Han, X.; Luo, M.; Lin, J.; Wang, G.; Li, J. Constructing a Microsatellite-Based Linkage Map and Identifying QTL for Pearl Quality Traits in Triangle Pearl Mussel (*Hyriopsis cumingii*). *Aquaculture* **2015**, *437*, 102–110. [CrossRef]
71. Cheng, H.; Concepcion, G.T.; Feng, X.; Zhang, H.; Li, H. Haplotype-Resolved de Novo Assembly Using Phased Assembly Graphs with Hifiasm. *Nat. Methods* **2021**, *18*, 170–175. [CrossRef] [PubMed]
72. Zhang, X.; Chen, S.; Shi, L.; Gong, D.; Zhang, S.; Zhao, Q.; Zhan, D.; Vasseur, L.; Wang, Y.; Yu, J.; et al. Haplotype-Resolved Genome Assembly Provides Insights into Evolutionary History of the Tea Plant Camellia Sinensis. *Nat. Genet.* **2021**, *53*, 1250–1259. [CrossRef] [PubMed]
73. Waterhouse, R.M.; Seppey, M.; Simao, F.A.; Manni, M.; Ioannidis, P.; Klioutchnikov, G.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol. Biol. Evol.* **2018**, *35*, 543–548. [CrossRef]
74. Bao, W.; Kojima, K.K.; Kohany, O. Repbase Update, a Database of Repetitive Elements in Eukaryotic Genomes. *Mob. DNA* **2015**, *6*, 4–9. [CrossRef]
75. Tarailo-Graovac, M.; Chen, N. Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. *Curr. Protoc. Bioinform.* **2009**, *25*, 1–14. [CrossRef]
76. Haas, B.J.; Salzberg, S.L.; Zhu, W.; Pertea, M.; Allen, J.E.; Orvis, J.; White, O.; Robin, C.R.; Wortman, J.R. Automated Eukaryotic Gene Structure Annotation Using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **2008**, *9*, R7. [CrossRef]
77. Tatusov, R.L.; Fedorova, N.D.; Jackson, J.D.; Jacobs, A.R.; Kiryutin, B.; Koonin, E.V.; Krylov, D.M.; Mazumder, R.; Mekhedov, S.L.; Nikolskaya, A.N.; et al. The COG Database: An Updated Version Includes Eukaryotes. *BMC Bioinform.* **2003**, *4*, 41. [CrossRef]
78. Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [CrossRef]
79. Conesa, A.; Götz, S.; García-Gómez, J.M.; Terol, J.; Talón, M.; Robles, M. Blast2GO: A Universal Tool for Annotation, Visualization and Analysis in Functional Genomics Research. *Bioinformatics* **2005**, *21*, 3674–3676. [CrossRef]
80. Emms, D.M.; Kelly, S. OrthoFinder: Solving Fundamental Biases in Whole Genome Comparisons Dramatically Improves Orthogroup Inference Accuracy. *Genome Biol.* **2015**, *16*, 157. [CrossRef]
81. Xie, C.; Mao, X.; Huang, J.; Ding, Y.; Wu, J.; Dong, S.; Kong, L.; Gao, G.; Li, C.Y.; Wei, L. KOBAS 2.0: A Web Server for Annotation and Identification of Enriched Pathways and Diseases. *Nucleic Acids Res.* **2011**, *39*, 316–322. [CrossRef]
82. Edgar, R.C. MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [CrossRef]
83. Castresana, J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Mol. Biol. Evol.* **2000**, *17*, 540–552. [CrossRef]
84. Stamatakis, A. RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [CrossRef]
85. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [CrossRef]
86. Kumar, S.; Stecher, G.; Suleski, M.; Hedges, S.B. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol. Biol. Evol.* **2017**, *34*, 1812–1819. [CrossRef]
87. De Bie, T.; Cristianini, N.; Demuth, J.P.; Hahn, M.W. CAFE: A Computational Tool for the Study of Gene Family Evolution. *Bioinformatics* **2006**, *22*, 1269–1271. [CrossRef]
88. Bielawski, J.P.; Baker, J.L.; Mingrone, J. Inference of Episodic Changes in Natural Selection Acting on Protein Coding Sequences via CODEML. *Curr. Protoc. Bioinform.* **2016**, *2016*, 1–32. [CrossRef]
89. 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]
90. Liu, F.; Li, Y.; Yu, H.; Zhang, L.; Hu, J.; Bao, Z.; Wang, S. MolluscDB: An Integrated Functional and Evolutionary Genomics Database for the Hyper-Diverse Animal Phylum Mollusca. *Nucleic Acids Res.* **2021**, *49*, D988–D997. [CrossRef]
91. Lu, S.; Wang, J.; Chitsaz, F.; Derbyshire, M.K.; Geer, R.C.; Gonzales, N.R.; Gwadz, M.; Hurwitz, D.I.; Marchler, G.H.; Song, J.S.; et al. CDD/SPARCLE: The Conserved Domain Database in 2020. *Nucleic Acids Res.* **2020**, *48*, D265–D268. [CrossRef]

92. Letunic, I.; Khedkar, S.; Bork, P. SMART: Recent Updates, New Developments and Status in 2020. *Nucleic Acids Res.* **2021**, *49*, D458–D460. [CrossRef]
93. Eddy, S.R. Profile Hidden Markov Models. *Bioinformatics* **1998**, *14*, 755–763. [CrossRef]
94. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; Von Haeseler, A.; Lanfear, R.; Teeling, E. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534. [CrossRef]
95. Artimo, P.; Jonnalagedda, M.; Arnold, K.; Baratin, D.; Csardi, G.; De Castro, E.; Duvaud, S.; Flegel, V.; Fortier, A.; Gasteiger, E.; et al. ExPASy: SIB Bioinformatics Resource Portal. *Nucleic Acids Res.* **2012**, *40*, 597–603. [CrossRef]
96. Kim, D.; Langmead, B.; Salzberg, S.L. HISAT: A Fast Spliced Aligner with Low Memory Requirements. *Nat. Methods* **2015**, *12*, 357–360. [CrossRef] [PubMed]
97. Liao, Y.; Smyth, G.K.; Shi, W. FeatureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features. *Bioinformatics* **2014**, *30*, 923–930. [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.