Journal of *Marine Science and Engineering*

MDPI

Article

Identification and Phylogenetic Analysis of Chitin Synthase Genes from the Deep-Sea Polychaete *Branchipolynoe onnuriensis* Genome

Hyeongwoo Choi [1], Sang Lyeol Kim [2,3], Man-Ki Jeong [4], Ok Hwan Yu [3] and Seongil Eyun [1,*]

1 Department of Life Science, Chung-Ang University, Seoul 06974, Korea; creo9447@cau.ac.kr
2 Ocean Science and Technology School, Korea Maritime University, Busan 49111, Korea; boyis20c@kiost.ac.kr
3 Marine Ecosystem and Biological Research Centre, Korea Institute of Ocean Science and Technology, Busan 49111, Korea; ohyu@kiost.ac.kr
4 Department of Smart Fisheries Resources Management, Chonnam National University, Yeosu 59626, Korea; jmgdeux@chonnam.ac.kr
* Correspondence: eyun@cau.ac.kr

Abstract: Chitin, one of the most abundant biopolymers in nature, is a crucial material that provides sufficient rigidity to the exoskeleton. In addition, chitin is a valuable substance in both the medical and industrial fields. The synthesis of chitin is catalyzed by chitin synthase (*CHS*) enzymes. Although the chitin synthesis pathway is highly conserved from fungi to invertebrates, *CHSs* have mostly only been investigated in insects and crustaceans. Especially, little is known about annelids from hydrothermal vents. To understand chitin synthesis from the evolutionary view in a deep-sea environment, we first generated the whole-genome sequencing of the parasitic polychaete *Branchipolynoe onnuriensis*. We identified seven putative *CHS* genes (*BonCHS1-BonCHS7*) by domain searches and phylogenetic analyses. This study showed that most crustaceans have only a single copy or two gene copies, whereas at least two independent gene duplication events occur in *B. onnuriensis*. This is the first study of *CHS* obtained from a parasitic species inhabiting a hydrothermal vent and will provide insight into various organisms' adaptation to the deep-sea hosts.

Keywords: chitin synthase; polychaete; *Branchipolynoe*; host–parasite interaction

Citation: Choi, H.; Kim, S.L.; Jeong, M.-K.; Yu, O.H.; Eyun, S. Identification and Phylogenetic Analysis of Chitin Synthase Genes from the Deep-Sea Polychaete *Branchipolynoe onnuriensis* Genome. *J. Mar. Sci. Eng.* **2022**, *10*, 598. https://doi.org/10.3390/jmse10050598

Academic Editor: Nguyen Hong Nguyen

Received: 24 March 2022
Accepted: 22 April 2022
Published: 28 April 2022

1. Introduction

Chitin, a linear polymer of β-(1,4)-*N*-acetyl-D-glucosamine (GlcNac), is the second most abundant biopolymer in nature, followed by cellulose, with more than 100 billion tons synthesized annually [1–5]. It is found in various organisms, ranging from fungi to various invertebrates, and provides them with sufficient rigidity to support their shape and structure [1]. In arthropods, chitin plays a crucial role in forming new cuticles during molting and is a component of the intestine peritrophic matrix in insects, which supports digestion [1,6,7]. In nematodes, chitin components are found in the eggshell and pharynx [8]. Furthermore, in Lophotrochozoa, chitin forms the radula and shell in mollusks [9–11], beak in cephalopods [12], and chaetae in annelids [13]. Due to its diverse function, chitin is attracting attention as a raw material for various fields, such as the pharmaceutical and biotechnological industries [14].

Chitin is polymerized by an enzyme called chitin synthase (*CHS*, chitin 4-β-*N*-acetylglucosaminyltransferase; EC 2.4.1.16), which is generally characterized by three functional domains: A, B, and C [15]. Domain A, composed of several transmembrane helixes, is located at the N-terminal, and this domain sequence may vary between species. Domain B (chitin_synth_2), the catalytic core that contains two highly conserved motifs ("EDR" and "QRRRW"), is in the middle of the gene. Domain C is located at the C-terminal, with approximately seven transmembrane helices, and has the conserved motif "WGTRE" [1].

Generally, insects have two *CHS* genes (*CHS1* and *CHS2*). *CHS1* is responsible for cuticle formation in the epidermis, while *CHS2* is involved in chitin synthesis in the peritrophic membrane of the intestine [1,6]. A *CHS* gene knockdown study in the crustacean *Lepeophtheirus salmonis* showed the formation of an abnormal appendage, which eventually led to death, suggesting the multifunctional role of *CHS* [16]. However, interestingly, compared to the ecdysozoans, which have only a single or two gene copies located in the same chromosome, numerous *CHS* genes have been identified in lophotrochozoans [1]. For example, 31 *CHS* genes were identified in the brachiopod *Lingula anatina* [17]. In addition, four and five *CHS* genes were identified in the shallow-sea polychaetes *Capitella teleta* and *Dimorphilus gyrociliatus*, respectively, whereas 19 and 12 *CHS* genes were significantly expanded in the deep-sea polychaetes *Paraescarpia echinospica* and *Lamellibrachia luymesi* [18]. These findings suggest that the *CHS* gene duplication event occurs lineage-specifically. However, *CHSs* have mostly been explored only in arthropods; so far, little is known about *CHSs* in annelids. Studies are gradually investigating the poorly explored realm of the lophotrochozoan *CHS*, but data on the evolutionary process of *CHS* gene expansion are still lacking. In addition, no phylogenetic analysis, including deep-sea parasitic polychaetes, has been reported.

In this study, we performed whole-genome sequencing (WGS) of parasitic polychaete *Branchipolynoe onnuriensis* collected from bivalves living in a hydrothermal vent [19] and identified seven *CHS* genes (*BonCHS1–BonCHS7*) belonging to the glycosyltransferase 2 (GT2) family. This is the first study of *CHSs* from deep-sea parasitic polychaetes. We also analyzed the relationship of *B. onnuriensis CHS* genes with those from lophotrochozoans. In addition, we expanded on the classification of lophotrochozoan *CHS* gene groups, in order to obtain information about their *CHS* gene family expansion and categorized them into five different subgroups. Our results will provide important information for those who study the chitin synthesis mechanism in deep-sea parasitic polychaetes in the future.

2. Materials and Methods

2.1. Sample Collection and Next-Generation Sequencing

An individual parasitic polychaete *Branchipolynoe onnuriensis* was separated from its host *Gigantidas vrijenhoeki* (class Bivalvia) using a video-guided hydraulic grab (Oktopus, Germany) around the Onnuri Vent Field (OVF, 11°14′55.92″ S, 66°15′15.10″ E; depth of 2014.5 m) during a Korea Institute of Ocean Science and Technology (KIOST) expedition along the Central Indian Ridge (CIR) in 2019 [20]. Immediately after being collected, the sample was stored in 95% ethanol at –20 °C, until DNA extraction in the laboratory. Genomic DNA was extracted using the QIAGEN Blood & Cell Culture DNA Mini Kit (QIAGEN, Hilden, Germany), according to the manufacturer's instructions. A paired-end library was constructed using the TruSeq DNA Nano 550 bp kit (Illumina, Inc., San Diego, CA, USA), with an insert size of 550 bp, and 150 bp sequencing was performed using the Novaseq6000 platform (Illumina).

2.2. Data Filtering and De Novo Genome Assembly

Adaptor sequences and low-quality reads that were lower than the mean quality score of 20 were removed. In addition, reads shorter than 120 bp or with unknown bases (n) were filtered using Trim Galore (ver. 0.6.6) [21]. The cleaned reads were obtained with the following parameters: `-quality 20 -length 120 -max_n 0`.

After quality control, de novo assembly, using 21-, 31-, and 51-mers to build an initial *de Bruijn* graph, was performed with SPAdes (ver. 3.14.0) [22]. Finally, quality assessment software for genome assembly, QUAST (ver. 5.0.2) were utilized to obtain diverse metrics, such as the number of contigs, a large contig, the total length, N50, or L50, without a reference genome [23].

2.3. Gene Prediction and Identification of the Chitin Synthase Gene

The genome structure of *B. onnuriensis* was annotated using ab initio gene prediction with Augustus (ver. 3.4.0) using a generalized hidden Markov model [24].

To extract the putative CHS sequences, we combined the basic local alignment search tool (BLAST) searches using National Center for Biotechnology Information (NCBI) and domain predictions. According to Zakrzewski et al. (2014) [7], since lophotrochozoans have four subgroups of *CHS* genes in type 2, we assumed that there would be at least one gene in each group (A, B, C, and D). First, we mined five *CHS* genes corresponding to each group from the same polychaete species, three *CHS* genes from *Owenia fusiformis* (group A, accession no. AHX26704.1; group D, accession no. AHX26707.1; type 1, accession no. AHX26703.1), and two *CHS* genes from *Sabellaria alveolata* (group B, accession no. AHX26717.1; group C, accession no. AHX26711.1) from NCBI and used them as queries to search homologous genes in our sample. In addition, we performed BLAST searches against the customized database with an *E*-value cut-off of $< 1 \times 10^{-50}$ and a length of > 300 amino acids (aa) [25]. Next, domain searches of each obtained gene were carried out using the simple modular architecture research tool (SMART) [26]. We identified seven putative genes and named them *BonCHS1–BonCHS7*. To confirm the putative BonCHSs, we performed BLAST searches against the Carbohydrate-Active enZymes Database (CAZy; https://bcb.unl.edu/dbCAN2/download/CAZyDB.09242021.fa, accessed on 1 June 2022) which contains enzymes that synthesize or break down carbohydrates and glycoconjugates, with an *E*-value cut-off of $< 1 \times 10^{-100}$ [27]. In order to obtain comparable E-values, the database size of 1.58×10^{11} (using the "`-dbsize`" option) was set to be equivalent to the size of the non-redundant (NR) protein database at NCBI.

2.4. Phylogenetic Analysis of Chitin Synthase

We conducted two phylogenetic analyses, i.e., the lophotrochozoan and metazoan trees. In the lophotrochozoan phylogenetic tree, 52 protein sequences, including seven putative *BonCHS1-BonCHS7*, were retrieved from 20 species (Table 1). In the metazoan *CHS* gene tree, 74 protein sequences were retrieved from 33 species (Table 1). Multiple sequence alignments were performed with MAFFT (ver. 7.475) [28]. We also used IQ-TREE (ver. 2.2.0) to select the best substitution model via Bayesian information criterion (BIC) [29]. The best evolutionary model of LG + I + G4 and LG + F + I + G4 was selected to construct the maximum likelihood (ML) for metazoan and lophotrochozoan data sets using RAxML-NG (ver. 0.9.0) [30]. In addition, each branch was supported by 1000 bootstrap replicates in the ML tree. The Bayesian tree was constructed using MrBayes (ver. 3.2.4), using the LG + I + G4 and LG + F + I + G4 models for metazoan and lophotrochozoan datasets. A total of four chains (three heated and one cold chain), for sampling all 5×10^2 generations, were carried out in two independent analyses. We performed 1×10^6 generations of MCMC analysis, and the first 25% trees as burn-in and incorporated with the ML tree [31]. Finally, each metazoan and lophotrochozoan phylogenetic tree was visualized using FigTree (ver. 1.4.4).

Table 1. Gene list of *chitin synthases* used for phylogenetic analysis.

Taxon (Phylum: Class)	Species	Gene Name	Type	Accession No.
Annelida *: Polychaeta	*Owenia fusiformis*	*OfuCHS1*	Type 2, group A	AHX26704.1
		OfuCHS2	Type 2, group C	AHX26705.1
		OfuCHS3	Type 2, group C	AHX26706.1
		OfuCHS4	Type 2, group D	AHX26707.1
		OfuCHS5	Type 2, group D	AHX26713.1
		OfuCHS6	Type 1	AHX26703.1
	Sabellaria alveolata	*SalCHS1*	Type 2, group B	AHX26717.1
		SalCHS2	Type 2, group C	AHX26711.1
		SalCHS3	Type 1	AHX26710.1
	Dimorphilus gyrociliatus	*DgyCHS1*	Type 2, group A	CAD5118528.1
		DgyCHS2	Type 2, group B	CAD5114651.1

Table 1. *Cont.*

Taxon (Phylum: Class)	Species	Gene Name	Type	Accession No.
	Platynereis dumerilii	*PduCHS1*	Type 2, group B	AHX26708.1
		PduCHS2	Type 2, group B	AHX26709.1
		PduCHS3	Type 2, group C	AHX26716.1
	Capitella teleta	*CteCHS1*	Type 2, group A	ELU08572.1
		CteCHS2	Type 2, group C	ELT98539.1
		CteCHS3	Type 2, group D	ELT92724.1
		CteCHS4	Type 2, group D	ELT92107.1
Annelida *: Annelida *incertae sedis*	*Myzostoma cirriferum*	*MciCHS1*	Type 2, group B	AHX26714.1
Mollusca *: Bivalvia	*Lottia gigantea*	*LgiCHS1*	Type 2, group A	XP_009061726.1
		LgiCHS2	Type 2, group A	XP_009061725.1
		LgiCHS3	Type 2, group A	XP_009061724.1
		LgiCHS4	Type 2, group B	XP_009063632.1
		LgiCHS5	Type 2, group C	XP_009047936.1
		LgiCHS6	Type 2, group D	XP_009066852.1
		LgiCHS7	Type 2, group D	XP_009066854.1
		LgiCHS8	Type 2, group D	XP_009051436.1
		LgiCHS9	Type 2, group D	XP_009051165.1
	Mytilus edulis	*MedCHS1*	Type 2, group A	CAG2205753.1
	Tegillarca granosa	*TgrCHS1*	Type 2, group A	AON76719.1
	Atrina rigida	*AriCHS1*	Type 2, group A	AAY86556.1
	Pinctada fucata	*PfuCHS1*	Type 2, group A	BAF73720.1
Mollusca *: Gastropoda	*Elysia marginata*	*EmaCHS1*	Type 2, group A	GFS24687.1
		EmaCHS2	Type 2, group C	GFR89942.1
		EmaCHS3	Type 2, group D	GFR83755.1
		EmaCHS4	Type 2, group D	GFR70591.1
		EmaCHS5	Type 1	GFS00558.1
		EmaCHS6	Type 1	GFR82903.1
	Crassostrea gigas	*CgiCHS1*	Type 2, group A	XP_034323514.1
Mollusca *: Polyplacophora	*Leptochiton asellus*	*LasCHS1*	Type 2, group A	AHX26699.1
		LasCHS2	Type 2, group C	AHX26700.1
Arthropoda: Insecta	*Helicoverpa zea*	*HzeCHS1*	Type 2, group 1	ADX66429.1
		HzeCHS2	Type 2, group 2	ADX66427.1
	Ostrinia furnacalis	*OfurCHS1*	Type 2, group 1	ACF53745.1
		OfurCHS2	Type 2, group 2	ABB97082.1
	Manduca sexta	*MseCHS1*	Type 2, group 1	AAL38051.2
		MseCHS2	Type 2, group 2	AAX20091.1
Arthropoda: Copepoda	*Lepidopterous salmonis*	*LsaCHS1*	Type 2, group 1	AYN59157.1
		LsaCHS2	Type 2, group 2	AYN59158.1
Chordata: Actinopterygii	*Danio rerio*	*DreCHS1*	Type 2 deuterostome	AJW72838.1
	Esox lucius	*EluCHS1*	Type 2 deuterostome	XP_010887243.2
	Megalops cyprinoides	*McyCHS1*	Type 2 deuterostome	XP_036403039.1
Chordata: Ascidiacea	*Ciona robusta*	*CroCHS1*	Type 2 deuterostome	BBB15954.1
Chordata: Amphibia	*Xenopus laevis*	*XlaCHS1*	Type 2 deuterostome	XP_018120159.2
Choanoflagellatea	*Salpingoeca rosetta*	*SroCHS1*	Type 1	EGD80959.1
	Monosiga brevicollis	*MbrCHS1*	Type 1	XP_001743227.1

Table 1. *Cont.*

Taxon (Phylum: Class)	Species	Gene Name	Type	Accession No.
Porifera: Calcarea	*Sycon ciliatum*	*SciCHS1*	Type 1	AHX26712.1
	Leucosolenia complicata	*LcoCHS1*	Type 1	AHX26702.1
		LcoCHS2	Type 1	AHX26701.1
Cnidaria: Hexacorallia	*Nematostella vectensis*	*NveCHS1*	Type 1	EDO41482.1
		NveCHS2	Type 1	EDO44996.1
	Stylophora pistillata	*SpiCHS1*	Type 1	PFX15170.1
		SpiCHS2	Type 1	PFX17869.1
	Hydra vulgaris	*HvuCHS1*	Type 1	XP_004207525.2
		HvuCHS2	Type 1	XP_012554922.1
Fungi: Eurotiomycetes	*Aspergillus fumigatus*	*AfuCHS1*	Fungi group	P54267.2
Fungi: Sordariomycetes	*Neurospora crassa*	*NcrCHS1*	Fungi group	P30588.2

* Lophotrochozoan phyla.

3. Results and Discussion

3.1. Data Filtering and Genome Assembly

Using Illumina paired-end sequencing, we generated 39.82 Gb raw reads from the parasitic polychaete *Branchipolynoe onnuriensis*. A stringent quality filter process (with Phred quality scores of 20 or more, see Materials and Methods) was applied; then, a total of 37.58 Gb (94.36%) filtered reads remained. After the filtering process, *de novo* assemblies of the genome sequences, using the software package SPAdes (ver. 3.14.0), were performed. Our initial genome assembly comprised 14,816 contigs, with a total length of 196,561,892 bp. The largest contig was 210,881 bp long, with an N50 length of 12,818 bp. Although the data obtained were insufficient for downstream analysis, since there are no genomic data available in *Brachipolynoe* spp. or parasitic polychaetes, we performed gene prediction to identify CHS protein-coding regions (see Table 2 for general information).

Table 2. General information from next-generation sequencing to gene annotation in *Branchipolynoe onnuriensis*.

Sequencing	Number of reads before filtering	263,730,178
	Mean quality score	35.47
	Percentage of $\geq$ Q30 (%)	90.95
	Number of bases (Gb)	37.08
Data filtering	Number of reads after filtering	250,683,082
Assembly	Number of contigs (> 10,000 bp)	14,816
	Length of *N*50 (bp)	12,818
	Total length of contigs (bp)	196,561,892
	Length of the largest contig (bp)	210,881
	GC content (%)	43.71
Gene prediction	Number of predicted genes	353,344

3.2. Gene Prediction and Chitin Synthase Search

Gene structure prediction was conducted using the ab initio method, which yielded 353,344 protein-coding genes. To extract *CHSs* from *B. onnuriensis*, we performed sequence similarity searches and extracted the top five best genes in each group (type 1; groups A, B, C, and D in type 2) and investigated the sequences thoroughly (Table 3). Two genes were identified as type 1; four in groups A, B, and D; and five in group C. The first top-hit gene in each group was assumed to be the *BonCHS* genes belonging to the corresponding group. However, in group B, the third top-hit gene, *g91735.t1*, was considered a candidate *CHS*

gene, because the first and second genes, *g38534.t* and *g45117.t1*, belonged to groups D and A, respectively. In addition, phylogenetic analysis for more sensitive identification showed no outlier (Figure 1). Therefore, we determined these five genes as putative *BonCHS* genes.

In addition, the number of genes in each group was determined to be one, except for group C. For example, the first top-hit gene in group A was included in group A, and all three genes belong to groups B, C, and D. However, in group C, the fourth and fifth top-hit genes belong to groups D and A, respectively. Since information about which group the second and third top-hit genes belong to is unknown, we first added two genes to the phylogenetic tree. As a result, both genes were included in the same clade of group C. Thus, the number of genes in group C was determined to be three. Consequently, from the similarity search and phylogenetic analysis, we extracted seven different *CHS* genes from *B. onnuriensis*: *BonCHS1-BonCHS7*.

3.3. Protein Domain Search, Identification of the GT2 Family, and Multiple Sequence Alignments

The seven *BonCHS* genes (*BonCHS1-BonCHS7*) were subjected to predict the domain structures using the SMART web server (http://smart.embl-heidelberg.de, accessed on 23 March 2022). We found that seven BonCHS sequences have chitin_synth_2 domain (Pfam domain: PF03142), except for BonCHS3 and BonCHS5. We supposed that three genes (*BonCHS3*, *BonCHS4*, and *BonCHS5*) in group C were only partially assembled, due to the limitation of Illumina short-read sequencing and lower coverage depth. However, the BLAST searches against the NCBI and UniProt web server showed that *BonCHSs* with the top-hit was *CHS* genes of the lophotrochozoan species, except for *BonCHS6* (Table 4). Furthermore, multiple sequence alignment was performed using 45 amino acid sequences obtained from lophotrochozoans. The two unique motifs, "EDR" (associated with catalytic function) and "QRRRW" (conferring processivity to CHS), were found to be highly conserved in all annelids and mollusks, suggesting their significance in chitin synthesis (Figure 2) [7,32].

Table 3. BLAST result with *E*-value cut-off threshold of $< 1 \times 10^{-50}$ and length > 300 aa.

Database Type		Top Genes	Length (aa)	*E*-Value	Identified Group	Gene Name	Species
Type 1		*g58373.t1*	838	0	Type 1	*BonCHS7*	*Owenia fusiformis*
		g38534.t1	321	2×10^{-51}	Group D		
Type 2	Group A	*g45117.t1*	733	2×10^{-135}	Group A	*BonCHS1*	*O. fusiformis*
		g38534.t1	321	1×10^{-127}	Group D		
		g120019.t1	304	7×10^{-91}	Group C		
		g91735.t1	755	1×10^{-83}	Group B		
	Group B	*g38534.t1*	321	$2 \times 10v^{120}$	Group D	*BonCHS2*	*Sabellaria alveolata*
		g45117.t1	733	2×10^{-106}	Group A		
		g91735.t1	755	4×10^{-95}	Group B		
		g120019.t1	304	5×10^{-76}	Group C		
	Group C	*g20614.t1*	464	0	Group C	*BonCHS3*	*S. alveolata*
		g120019.t1	304	2×10^{-142}	Group C	*BonCHS4*	
		g86068.t1	472	3×10^{-137}	Group C	*BonCHS5*	
		g38534.t1	321	8×10^{-129}	Group D		
		g45117.t1	733	3×10^{-106}	Group A		
	Group D	*g38534.t1*	321	3×10^{-137}	Group D	*BonCHS6*	*O. fusiformis*
		g45117.t1	733	2×10^{-122}	Group A		
		g91735.t1	755	5×10^{-84}	Group B		
		g120019.t1	304	3×10^{-69}	Group C		

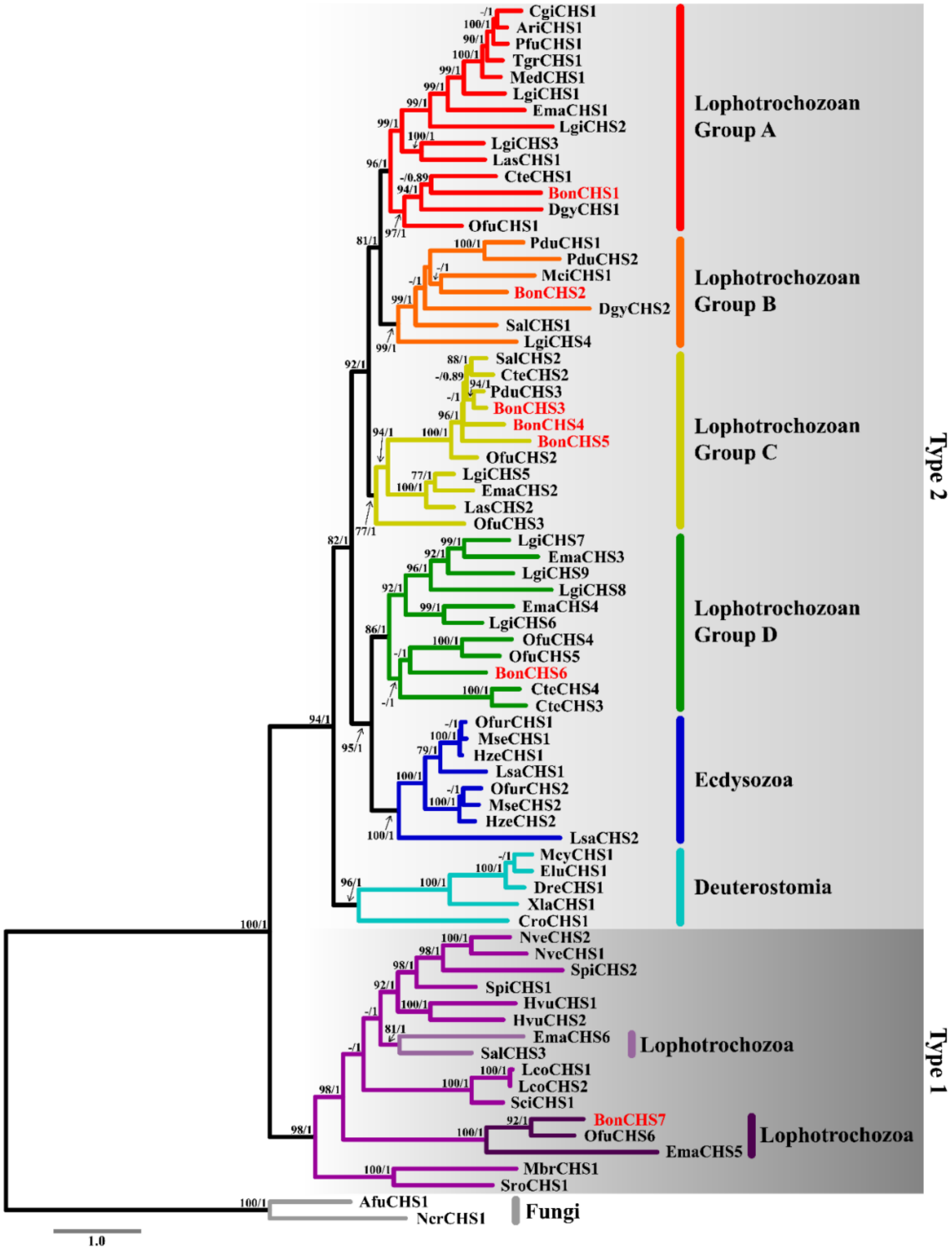

Figure 1. Maximum likelihood (ML) phylogeny of types 1 and 2 chitin synthase (*CHS*) genes: 74 amino acid sequences from 31 metazoans and two fungi. The fungi sequences were used as an outgroup. The gene names are abbreviated and listed in Table 1. *CHS* genes from *Branchipolynoe onnuriensis* are colored in red. Numbers in each node are supporting values and shown only > 70 and > 0.7 from the ML and Bayesian inference. The scale bar is provided at the bottom of the tree and represents the amino acid substitutions per site.

The similarity searches against the CAZy database showed their inclusion in the GT2 family (Table 5). For all genes, the *E*-value was $< 1 \times 10^{-130}$ and their identities ranged from 37.02% to 88.46%. Although BonCHS3 and BonCHS5 were not confirmed by the domain searches, their *E*-values showed 0 and 4.67×10^{-132}, with an identity of 76.46% and 53.22%. Note that our analysis failed to find the "EDR" and "QRRRW" motifs from BonCHS7. Thus, *BonCHS3*, *BonCHS5*, and *BonCHS7* were excluded from the evolutionary patterns.

3.4. Phylogenetic Analysis of Chitin Synthase

In order to confirm the ortholog relationship and understand the molecular evolutionary history, we conducted phylogenetic analysis, including all type 1 and 2 *CHS* genes from the NCBI (Figure 1). As suggested by Zakrzewski et al. (2014) [7], type 1 *CHS* genes generally exist in all metazoans, and *BonCHS7* was found to be closely related to *OfuCHS6* from *O. fusiformis* and *EmaCHS5* from *Elysia marginata*. However, since *E. marginata EmaCHS6* and *S. alveolata SalCHS3* form another clade, type 1 lophotrochozoan CHS is considered a paraphyletic group.

To understand the evolutionary relationship of lophotrochozoan *CHS* genes, we reconstructed a ML phylogenetic tree, with CHS protein sequences from seven annelid species, two gastropods, one polyplacophora, and five bivalves (Figure 3). Five deuterostome sequences were used as an outgroup. Type 2 *CHS* genes mainly consist of four groups (groups A, B, C, and D). In each group, annelid and mollusk clades are clearly separated, with well-supported values of $\geq$ 87% and 1 from the ML and Bayesian inference, which suggests that the lophotrochozoan *CHS* gene duplication event occurred before the divergence of annelids and mollusks [7]. Except for group C, annelid and mollusk *CHS* genes formed a monophyletic clade in all groups. This means that the *O. fusiformis* group C *CHS* gene has undergone a more complex evolutionary process. In the same context, in group C, we found three *BonCHS* genes. Since these genes (*BonCHS3-BonCHS5*) originated from different contigs, they are more likely to result from the gene duplication events, rather than be isoforms. However, not all polychaetes have increased gene copies in group C. For example, two genes were identified from *Platynereis dumerilii* in group B, two and three genes in *O. fusiformis* and *B. onnuriensis* in group C, and two genes in *O. fusiformis* and *C. teleta* in group D, but with no copies in group A. Even for the same taxon, polychaeta, gene duplication did not occur in the same group, which appears to be a species-specific event. Moreover, several *CHS* copies were also found in mollusks (*L. gigantidas* and *E. marginata*). Considering that two types of *CHS* genes with different functions exist in ecdysozoans (component of the exoskeleton and peritrophic matrix), all four different types of CHS of lophotrochozoans may play a different function. Additionally, since *B. onnuriensis* was collected by chance from its host, *Gigantidas vrijenhoeki*, there was a limit to describing their ecological characteristics, except for their habitat and parasitism. However, we obtained evidence of a gene duplication event in group C, and it is best to say that it is due to the two factors mentioned above. To determine the underlying mechanisms and functions of lophotrochozoan enzymes, gene and protein characterization studies are required.

Table 4. The top-hit BLAST results against NCBI and UniProt database.

Database	Query ID	Species	Database ID	Identity (%)	*E*-Value
NCBI	BonCHS1	*Lamellibrachia satsuma*	KAI0208509.1	46.94	2×10^{-156}
	BonCHS2	*Lamellibrachia satsuma*	KAI0242735.1	56.45	1×10^{-92}
	BonCHS3	*Platynereis dumerilii*	AHX26716.1	73.04	0
	BonCHS4	*Platynereis dumerilii*	AHX26716.1	88.09	8×10^{-132}
	BonCHS5	*Sabellaria alveolata*	AHX26711.1	52.52	6×10^{-112}
	BonCHS6	*Homalodisca vitripennis*	KAG8240581.1	64.47	1×10^{-154}
	BonCHS7	*Owenia fusiformis*	CAH1788656.1	51.09	1×10^{-170}

Table 4. *Cont.*

Database	Query ID	Species	Database ID	Identity (%)	*E*-Value
UniProt	BonCHS1	*Capitella teleta*	R7UXD6	46.10	4.3×10^{-161}
	BonCHS2	*Lottia gigantea*	V4B948	44.1	1.7×10^{-121}
	BonCHS3	*Capitella teleta*	R7TXS7	70.30	3×10^{-155}
	BonCHS4	*Capitella teleta*	R7TXS7	87.00	9.3×10^{-135}
	BonCHS5	*Lingula unguis*	A0A1S3IM62	48.10	7×10^{-109}
	BonCHS6	*Bombyx mori*	H9J0C4	66.20	5.7×10^{-153}
	BonCHS7	*Lingula unguis*	A0A1S3IM62	48.10	0

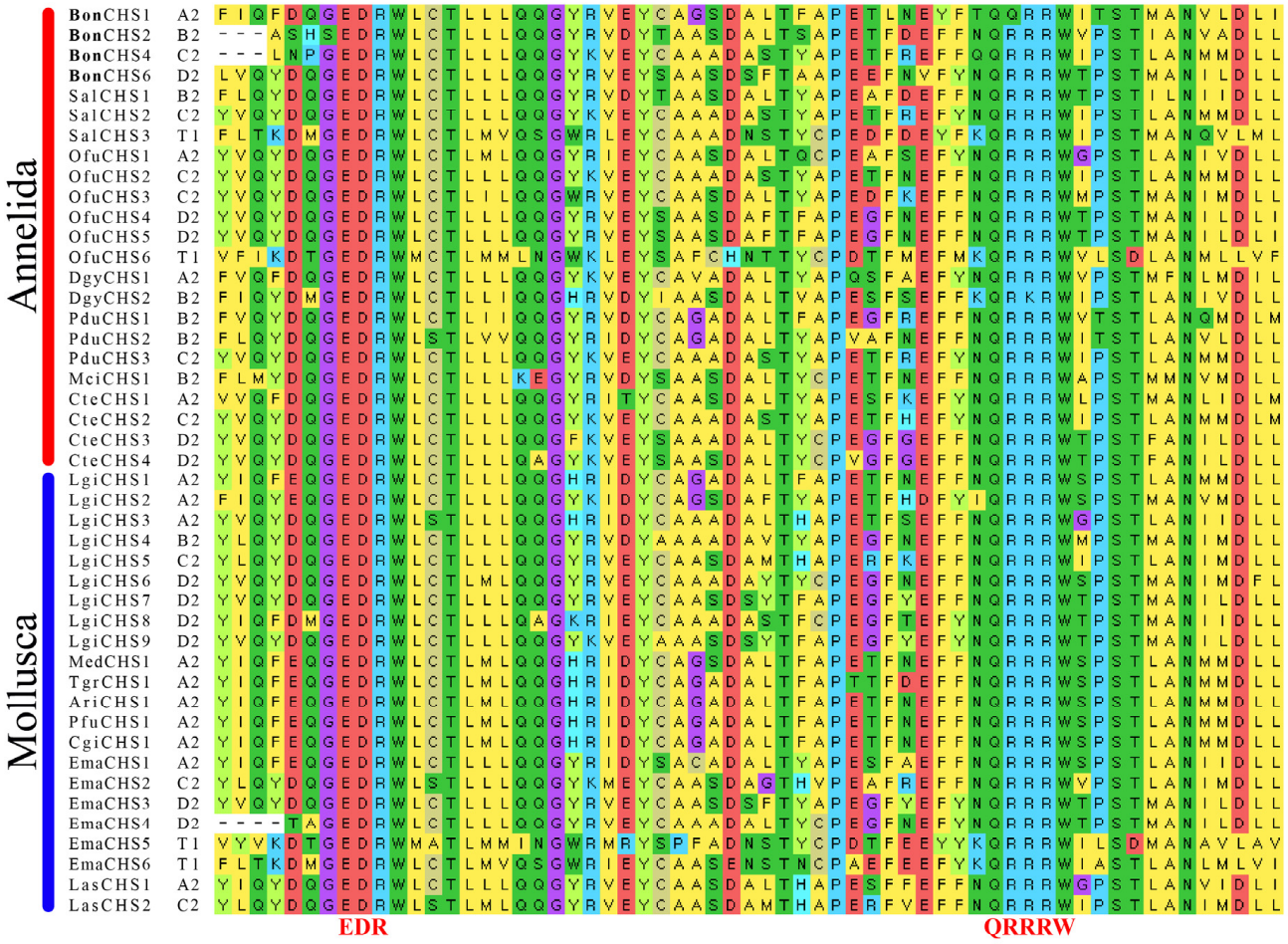

Figure 2. Multiple sequence alignment of CHSs from lophotrochozoan species—23 *CHS* genes from annelids and 22 from mollusks were used. Gene types are marked next to the name (A2, B2, C2, and D2 for groups A, B, C, and D in type 2 and T1 for type 1). Two highly conservative motifs (EDR and QRRRW) are indicated in bold red. The color code is followed by physicochemical properties.

Table 5. The top-hit BLAST results against CAZy database.

Query ID	Species	Database ID	Enzyme Class	Identity (%)	*E*-Value
BonCHS1	*Macandrevia cranium*	AHX26715.1	GT2	42.40	1.54×10^{-154}
BonCHS2	*Myzostoma cirriferum*	AHX26714.1	GT2	37.02	1.71×10^{-141}
BonCHS3	*Platynereis dumerilii*	AHX26716.1	GT2	76.46	0
BonCHS4	*Platynereis dumerilii*	AHX26716.1	GT2	88.46	2.10×10^{-140}
BonCHS5	*Sabellaria alveolate*	AHX26711.1	GT2	53.22	4.67×10^{-132}
BonCHS6	*Bombyx mori*	AFC69002.1	GT2	66.25	4.35×10^{-153}
BonCHS7	*Owenia fusiformis*	AHX26703.1	GT2	50.23	0

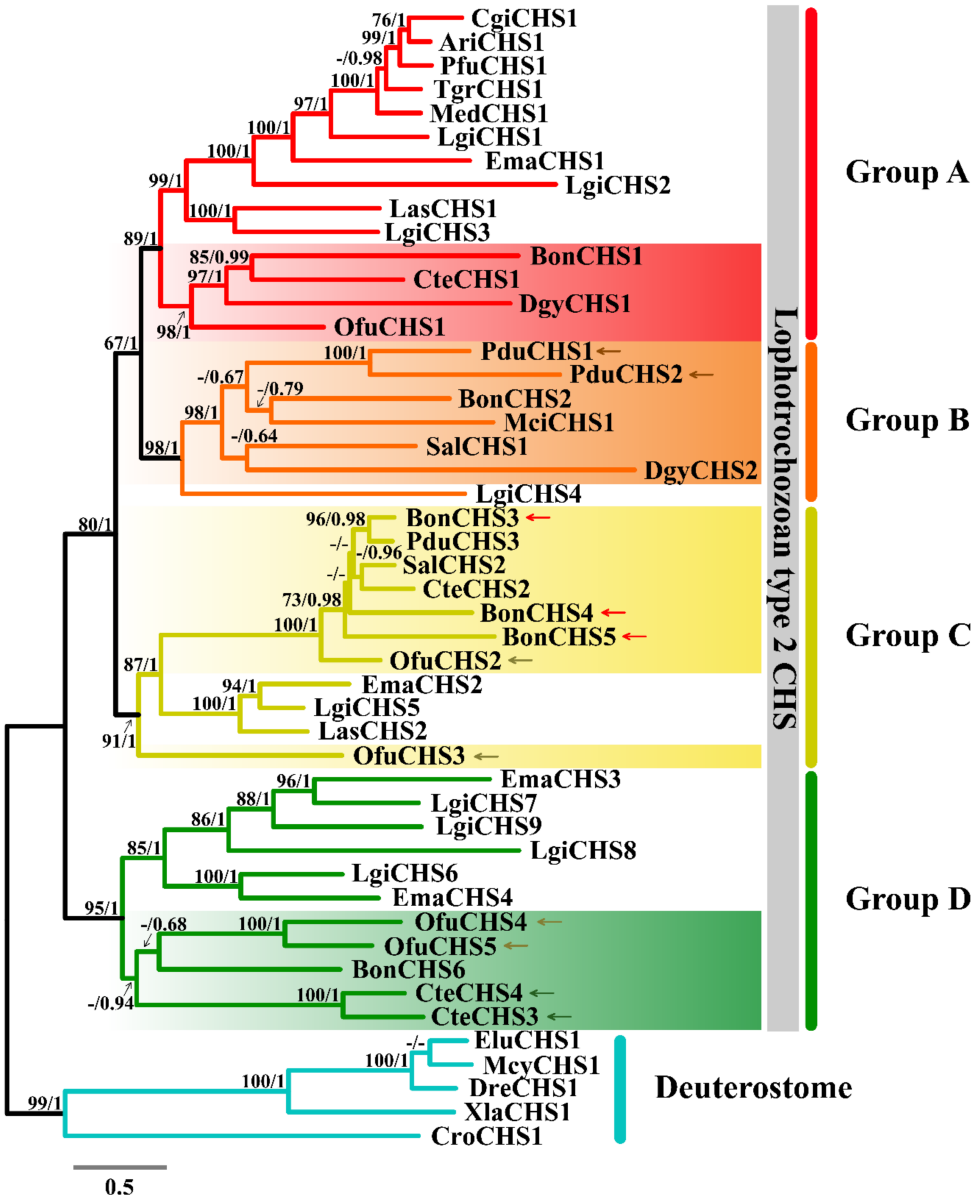

Figure 3. Maximum likelihood phylogeny of chitin synthase genes—48 genes from 15 lophotrochozoan and five deuterostome species are included. The gene names are abbreviated and listed in Table 1. Genes belonging to the polychaetes are colored in each group (red in group A, orange in group B, yellow in group C, and green in group D). Deuterostome sequences are used as an outgroup. In each node, supporting values for ML and Bayesian inference are shown in this order. The nodes supporting values of <60 are indicated with "-". The arrows indicate annelids with several copies (orange for *Platynereis dumerilii*, yellow for *Owenia fusiformis*, green for *Capitellateleta*, and red for *Branchipolynoe onnuriensis*). The clades with the gradient boxes represent the polychaete species in each group. The scale bar represents the amino acid substitutions per site.

4. Conclusions

Chitin, a natural polysaccharide, is the second-largest substance on earth and valuable for many industries. However, compared to the ecdysozoan *CHSs*, which are relatively well-researched, little is known about the lophotrochozoan *CHS* gene. Therefore, in this study, we collected the parasitic polychaete *B. onnuriensis* living in the deep-sea and conducted WGS to investigate the evolutionary aspect of *CHSs*. As a first step toward understanding the role of lophotrochozoan enzymes, we successfully identified seven *CHS* genes (*BonCHS1-BonCHS7*) and classified them into five groups. Because of the lower coverage depth and

limitation of short-read sequencing, the *B. onnuriensis* genome may have more than seven *CHS* genes. In addition, it is a common phenomenon that lophotrochozoans have several *CHS* genes, especially species living in deep-sea polychaetes. *B. onnuriensis* was found from the host recently; thus, their evolutionary and ecological aspects had remained largely unexplored, except that they survive in the deep-sea and are parasitic. We speculated that the *CHS* gene duplication event might be involved in a harsh environment or parasitic life. Although further research is needed on which tissues the *CHS* gene in group C is expressed and its function, this study suggests the possibility that *CHS* genes in group C are duplicated, which may play a key role in adaptation to parasitic life in harsh environments.

Author Contributions: Conceptualization: S.E.; sample and data collection: S.L.K. and O.H.Y.; data analysis: H.C., S.L.K., M.-K.J., O.H.Y. and S.E.; writing—original draft: H.C., S.L.K., M.-K.J., O.H.Y. and S.E.; writing—review and editing: H.C. and S.E. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Chung-Ang University Research Scholarship Grants in 2021 and Korea Polar Research Institute (KOPRI, PE22900).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All sequencing data are archived at NCBI Sequence Read Archive (SRA) database, under BioProject number PRJNA817330.

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

References

1. Zhang, X.; Yuan, J.; Li, F.; Xiang, J. Chitin synthesis and degradation in crustaceans: A genomic view and application. *Mar. Drugs* **2021**, *19*, 153. [CrossRef] [PubMed]
2. Uddowla, M.H.; Kim, A.R.; Park, W.; Kim, H.W. CDNAs encoding chitin synthase from shrimp (*Pandalopsis japonica*): Molecular characterization and expression analysis. *J. Aquac. Res. Dev.* **2015**, *6*, 298. [CrossRef]
3. Abo Elsoud, M.M.; El Kady, E.M. Current trends in fungal biosynthesis of chitin and chitosan. *Bull. Natl. Res. Cent.* **2019**, *43*, 59. [CrossRef]
4. Tharanathan, R.N.; Kittur, F.S. Chitin-the undisputed biomolecule of great potential. *Crit. Reveiws Food Sci. Nutr.* **2003**, *43*, 61–87. [CrossRef]
5. Lee, S.; Kang, H.A.; Eyun, S. Evolutionary analysis and protein family classification of chitin deacetylases in *Cryptococcus neoformans*. *J. Microbiol.* **2020**, *58*, 805–811. [CrossRef]
6. Xin, T.; Li, Z.; Chen, J.; Wang, J.; Zou, Z.; Xia, B. Molecular characterization of chitin synthase gene in *Tetranychus cinnabarinus* (Boisduval) and its response to sublethal concentrations of an insecticide. *Insects* **2021**, *12*, 501. [CrossRef]
7. Zakrzewski, A.; Weigert, A.; Helm, C.; Adamski, M.; Adamska, M.; Bleidorn, C.; Raible, F.; Hausen, H. Early divergence, broad distribution, and high diversity of animal chitin synthases. *Genome Biol. Evol.* **2014**, *6*, 316–325. [CrossRef]
8. Veronico, P.; Gray, L.J.; Jones, J.T.; Bazzicalupo, P.; Arbucci, S.; Cortese, M.R.; di Vito, M.; de Giorgi, C. Nematode chitin synthases: Gene structure, expression and function in *Caenorhabditis elegans* and the plant parasitic nematode *Meloidogyne artiellia*. *Mol. Genet. Genom.* **2001**, *266*, 28–34. [CrossRef]
9. Weiss, I.M.; Schönitzer, V.; Eichner, N.; Sumper, M. The chitin synthase involved in marine bivalve mollusk shell formation contains a myosin domain. *FEBS Lett.* **2006**, *580*, 1846–1852. [CrossRef]
10. Vortsepneva, E.; Tzetlin, A. General morphology and ultrastructure of the radula of *Testudinalia testudinalis* (O.F. Müller, 1776) (Patellogastropoda, Gastropoda). *J. Morphol.* **2019**, *280*, 1714–1733. [CrossRef]
11. Agbaje, O.B.A.; Shir, I.B.; Zax, D.B.; Schmidt, A.; Jacob, D.E. Biomacromolecules within bivalve shells: Is chitin abundant? *Acta Biomater.* **2018**, *80*, 176–187. [CrossRef] [PubMed]
12. Tan, Y.; Hoon, S.; Guerette, P.A.; Wei, W.; Ghadban, A.; Hao, C.; Miserez, A.; Waite, J.H. Infiltration of chitin by protein coacervates defines the squid beak mechanical gradient. *Nat. Chem. Biol.* **2015**, *11*, 488–495. [CrossRef] [PubMed]
13. Picken, L.; Lotmar, W. Oriented protein in chitinous structures. *Nature* **1950**, *165*, 599–600. [CrossRef]
14. Merzendorfer, H. The cellular basis of chitin synthesis in fungi and insects: Common principles and differences. *Eur. J. Cell Biol.* **2011**, *90*, 759–769. [CrossRef] [PubMed]
15. Coutinho, P.M.; Deleury, E.; Davies, G.J.; Henrissat, B. An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.* **2003**, *328*, 307–317. [CrossRef]

16. Braden, L.; Michaud, D.; Igboeli, O.O.; Dondrup, M.; Hamre, L.; Dalvin, S.; Purcell, S.L.; Kongshaug, H.; Eichner, C.; Nilsen, F.; et al. Identification of critical enzymes in the *Salmon louse* chitin synthesis pathway as revealed by RNA interference-mediated abrogation of infectivity. *Int. J. Parasitol.* **2020**, *50*, 873–889. [CrossRef]
17. Luo, Y.-J.; Takeuchi, T.; Koyanagi, R.; Yamada, L.; Kanda, M.; Khalturina, M.; Fujie, M.; Yamasaki, S.-I.; Endo, K.; Satoh, N. The *Lingula* genome provides insights into brachiopod evolution and the origin of phosphate biomineralization. *Nat. Commun.* **2015**, *6*, 8301. [CrossRef]
18. Sun, Y.; Sun, J.; Yang, Y.; Lan, Y.; Ip, J.C.-H.; Wong, W.C.; Kwan, Y.H.; Zhang, Y.; Han, Z.; Qiu, J.-W.; et al. Genomic signatures supporting the symbiosis and formation of chitinous tube in the deep-sea tubeworm *Paraescarpia echinospica*. *Mol. Biol. Evol.* **2021**, *38*, 4116–4134. [CrossRef]
19. Kim, S.L.; Choi, H.; Eyun, S.; Kim, D.; Yu, O.H. A new *Branchipolynoe* (Aphroditiformia: Polynoidae) scale worm from the Onnuri Deep-sea hydrothermal vent field, northern Central Indian Ridge. *Zool. Stud.* **2022**, *61*, 21.
20. Ryu, T.; Kim, J.G.; Lee, J.; Yu, O.H.; Yum, S.; Kim, D.; Woo, S. First transcriptome assembly of a newly discovered vent mussel, Gigantidas vrijenhoeki, at Onnuri Vent Field on the northern Central Indian Ridge. *Mar. Genom.* **2021**, *57*, 100819. [CrossRef]
21. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **2011**, *17*, 3. [CrossRef]
22. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [CrossRef] [PubMed]
23. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [CrossRef] [PubMed]
24. Stanke, M.; Morgenstern, B. AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **2005**, *33*, W465–W467. [CrossRef]
25. 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]
26. Schultz, J.; Copley, R.R.; Doerks, T.; Ponting, C.P.; Bork, P. SMART: A web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.* **2000**, *28*, 231–234. [CrossRef]
27. Cantarel, B.L.; Coutinho, P.M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for Glycogenomics. *Nucleic Acids Res.* **2009**, *37*, D233–D238. [CrossRef]
28. 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]
29. 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]
30. 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]
31. Ronquist, F.; Teslenko, M.; van der Mark, P.; Ayres, D.L.; Darling, A.; Höhna, S.; Larget, B.; Liu, L.; Suchard, M.A.; Huelsenbeck, J.P. MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **2012**, *61*, 539–542. [CrossRef] [PubMed]
32. Yu, H.-Z.; Li, N.-Y.; Xie, Y.-X.; Zhang, Q.; Wang, Y.; Lu, Z.-J. Identification and functional analysis of two chitin synthase genes in the common cutworm, *Spodoptera litura*. *Insects* **2020**, *11*, 253. [CrossRef] [PubMed]