



The Draft Genome of Chinese Endemic Species *Phascolosoma esculenta* (Sipuncula, Phascolosomatidae) Reveals the Phylogenetic Position of Sipuncula

Shengping Zhong^{1,2*†}, Xiaowan Ma^{3†}, Yan Jiang^{2†}, Ying Qiao³, Longyan Zhao¹, Lianghua Huang¹, Guoqiang Huang¹, Yongzhen Zhao⁴, Yonghong Liu^{1*†} and Xiuli Chen^{4*}

OPEN ACCESS

Edited by:

Sonia Andrade,
University of São Paulo, Brazil

Reviewed by:

Yosuke Seto,
Japanese Foundation for Cancer
Research, Japan
Qin Zhang,
Guangxi University for Nationalities,
China

*Correspondence:

Shengping Zhong
shpzhong@foxmail.com
Yonghong Liu
yonghongliu@sclsio.ac.cn
Xiuli Chen
chenxiuli2001@163.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 01 April 2022

Accepted: 26 May 2022

Published: 22 July 2022

Citation:

Zhong S, Ma X, Jiang Y, Qiao Y,
Zhao L, Huang L, Huang G, Zhao Y,
Liu Y and Chen X (2022) The Draft
Genome of Chinese Endemic Species
Phascolosoma esculenta (Sipuncula,
Phascolosomatidae) Reveals the
Phylogenetic Position of Sipuncula.
Front. Genet. 13:910344.
doi: 10.3389/fgene.2022.910344

¹Institute of Marine Drugs, Guangxi University of Chinese Medicine, Nanning, China, ²Guangxi Engineering Technology Research Center for Marine Aquaculture, Guangxi Institute of Oceanology Co., Ltd., Beihai, China, ³Key Laboratory of Tropical Marine Ecosystem and Bioresource, Fourth Institute of Oceanography, Ministry of Natural Resources, Beihai, China, ⁴Guangxi Key Laboratory of Aquatic Genetic Breeding and Healthy Aquaculture, Guangxi Academy of Fishery Sciences, Nanning, China

Keywords: *Phascolosoma esculenta*, sipuncula, genome, nanopore, phylogenomics

INTRODUCTION

Lophotrochozoa, the sister taxon of Ecdysozoa, is an ecologically and morphologically diverse clade of protostome animals within the Spiralia (Bleidorn, 2019). Lophotrochozoans represent more than one-third of known marine animals and include the segmented annelids (e.g., Errantia and Sedentaria worms) and the shell bearing molluscs (e.g., oysters and scallops) (Luo et al., 2018). Annelida, also known as the ringed or segmented worms, is a major phylum within the superphylum Lophotrochozoa (Shen et al., 2009). The annelids play important ecological roles in the marine ecosystem, are comprised of more than 21,000 described species, and constitute the dominant benthic macrofauna in all the major oceans (Weigert et al., 2014). Despite their ecological importance, the phylogenetic positions among annelid groups have not been fully resolved and their deep-level evolutionary relationships are still poorly understood (Struck et al., 2011; Bleidorn, 2019). Over the last decade, phylogenomic analyses have suggested that Sipuncula should be included within annelid groups, together with Echiura, even though historically they were considered separate phyla given their lack of segmentation (Dordel et al., 2010; Schulze and Kawauchi, 2021).

Sipuncula (known as peanut or star worms) are a clade of unsegmented, coelomate marine worms, which inhabit marine sediments and occur worldwide from the intertidal zones to the deep sea (Schulze et al., 2005; Lemer et al., 2015). Sipuncula has about 150 species in 17 genera and a fossil record dating back to the Early Cambrian (Huang et al., 2004; Zhong et al., 2020). The fossil evidence indicates that sipunculans may have undergone evolutionary stasis over the last 520 million years (Huang et al., 2004), and the monophyly of Sipuncula taxon has been supported by morphological and molecular data (Schulze et al., 2007; Mwinyi et al., 2009). However, the phylogenetic position of Sipuncula has changed drastically many times over several decades (Dordel et al., 2010; Weigert and Bleidorn, 2016). Previous cladistic analyses based on morphology proposed a close sister taxon relationship between Sipuncula and Mollusca, because of the presence of the “molluscan cross” organization of micromeres during spiral cleavage (Scheltema, 1993). Due to the lack of segmentation and body appendages, Sipuncula were once regarded as their own separate phyla, closely related to Annelida, but not part of Annelida (Rouse and Fauchald, 1997). In contrast to these morphology-based studies, phylogenomic analyses of large molecular datasets, such as transcriptome (Weigert et al., 2014), expressed sequence tags (ESTs) (Struck et al., 2011) and MicroRNAs (Sperling



FIGURE 1 | Harvest *P. esculenta* and the genomics feature of *P. esculenta*. **(A)** Harvest *P. esculenta* from local aquaculture farms. **(B)** A K-mer analysis of the genome sequencing reads for the *P. esculenta* using GenomeScope v2.0.

et al., 2009) favor the inclusion of Sipuncula into annelid taxon. The placement of non-segmented Sipuncula within the segmented Annelida taxon implies that patterns of segmentation within annelids have been evolutionarily labile (Struck et al., 2007), and Sipuncula may have secondarily lost segmentation, as is the case with non-segmented Echiura. Although, the placement of Sipuncula within Annelida has been supported by phylogenomic analyses (Weigert and Bleidorn, 2016), it is debated whether Sipuncula is a deeply-nested Annelid or the sister taxon of Annelids (Andrade et al., 2015; Weigert et al., 2016). Therefore, a robust genomic reconstruction of Sipuncula phylogenetic position within Annelida is needed. However, until now, genome data from Sipuncula has not been published.

Phascolosoma esculenta, a Sipuncula species restricted to the coastal zone of southeast China, is a valuable fisheries resource for both nutrition and ingredients of Chinese traditional medicines (Zhao et al., 2019). As a burrowing organism thriving in the intertidal zone ecosystem, *P. esculenta* has evolved an extraordinary resilience to harsh and dynamically changing intertidal stresses (Su et al., 2010; Shen et al., 2021). Although, *P. esculenta* has high tolerance of abiotic stress, including temperature, salinity, and hypoxia, the molecular mechanism bolstering *P. esculenta* against intertidal stresses has not been thoroughly investigated (Meng et al., 2022). Hence, decoding the genome of *P. esculenta* will allow for a more thorough understanding of Sipuncula's ecological adaptation and phylogenetic evolution. In this report, we provided a draft genome of *P. esculenta* using Oxford Nanopore Technologies. We assembled the genome sequences into 1,446 contigs with a total length of 1.71 Gb and a contig N50 length of 2.49 Mb. Furthermore, a total of 1,688 gene families were identified as species-specific of *P. esculenta*, and 1,032 gene families were significantly expanded in the *P. esculenta* genome. The availability of the

first genome data from Sipuncula offers better insight into deep-level evolutionary analysis of Annelida, and also provides a valuable resource for the analysis of ecological adaptation and the molecular mechanism of *P. esculenta* against intertidal stresses.

MATERIALS AND METHODS

Sampling, Library Construction, and Sequencing

A healthy individual of *P. esculenta* (body weight 4.87 g) was obtained from local aquaculture farms (Figure 1A) (Beihai, Guangxi Province, China, 21.473645 N, 109.469912 E). The introvert, trunk, intestine and nephridia were collected, immediately frozen and stored in liquid nitrogen until extraction of the genomic DNA and total RNA. Total genomic DNA was extracted from the muscle tissue of the trunk sample using the QIAamp DNA Mini Kit (QIAGEN, Hilden, Germany). Approximately 1 µg of genomic DNA was used for constructing the Nanopore 20 kb insert library; the large size fraction of DNA (>20 kb) was selected by automated gel electrophoresis (BluePippin, Sage Science); the sequencing library was prepared by standard ligation sequencing kit (SQK-LSK109, Oxford Nanopore Technologies); and the constructed library was sequenced using the Oxford Nanopore MinION platform at BGI Genomics Co., Ltd., Shenzhen, China. Meanwhile, a paired-end DNA sequencing library with insert size 350 bp was constructed and sequenced using the BGISEQ-500 platform according to the manufacturer's protocol. Finally, total RNAs were extracted from introvert, intestine and nephridia samples using RNAiso kit (TaKaRa, Dalian, China). The RNA sequencing library with insert size 300 bp was constructed and sequenced (2 × 150 bp paired-end) using the BGI DNBseq platform.

TABLE 1 | Summary statistics of genome assembly and gene prediction of *P. esculenta*.

Summary Statistics of genome assembly	
Total length of genome (Gbp)	1.71
Contig N50 size (Mbp)	2.49
Contig number	1,446
The length of largest contig (Mbp)	12.99
Proportion of BUSCO in genome model (%)	98.5
Summary statistics of gene prediction	
Protein-coding gene number	41,469
The length of largest protein-coding gene (bp)	38,466
Mean transcript length (bp)	1,270
Mean exons length (bp)	201
Mean exons number per gene	6.3
Proportion of BUSCO in proteins model (%)	99.4

Genome Size Estimation, Genome Assembly and Polishing

The genomic DNA data sequenced by BGISEQ-500 platform was filtered by fastp v0.23.0 (Chen et al., 2018). The k-mer analysis was conducted to estimate the genome size and heterozygosity of *P. esculenta* using Jellyfish v2.3.0 (Marçais and Kingsford, 2011) and Genomescope v2.0 (Ranallo-Benavidez et al., 2020), with the k-mer length set to 19. The Nanopore long read was corrected and then the preliminarily genome was assembled using NextDenovo v2.5.0 assembler (<https://github.com/Nextomics/NextDenovo>). The preliminarily genome was polished using NextPolish v1.4.0 software (Hu et al., 2020) to correct base errors caused by Nanopore sequencing. Finally, the redundancy in the polished genome was resolved using Purge Dups v. 1.2.5 (Guan et al., 2020), and the assembly completeness was evaluated by Benchmarking Universal Single-Copy Orthologs (BUSCO) software v5.2.2 (Manni et al., 2021) based on the metazoa_odb10 lineages.

Genome Annotation

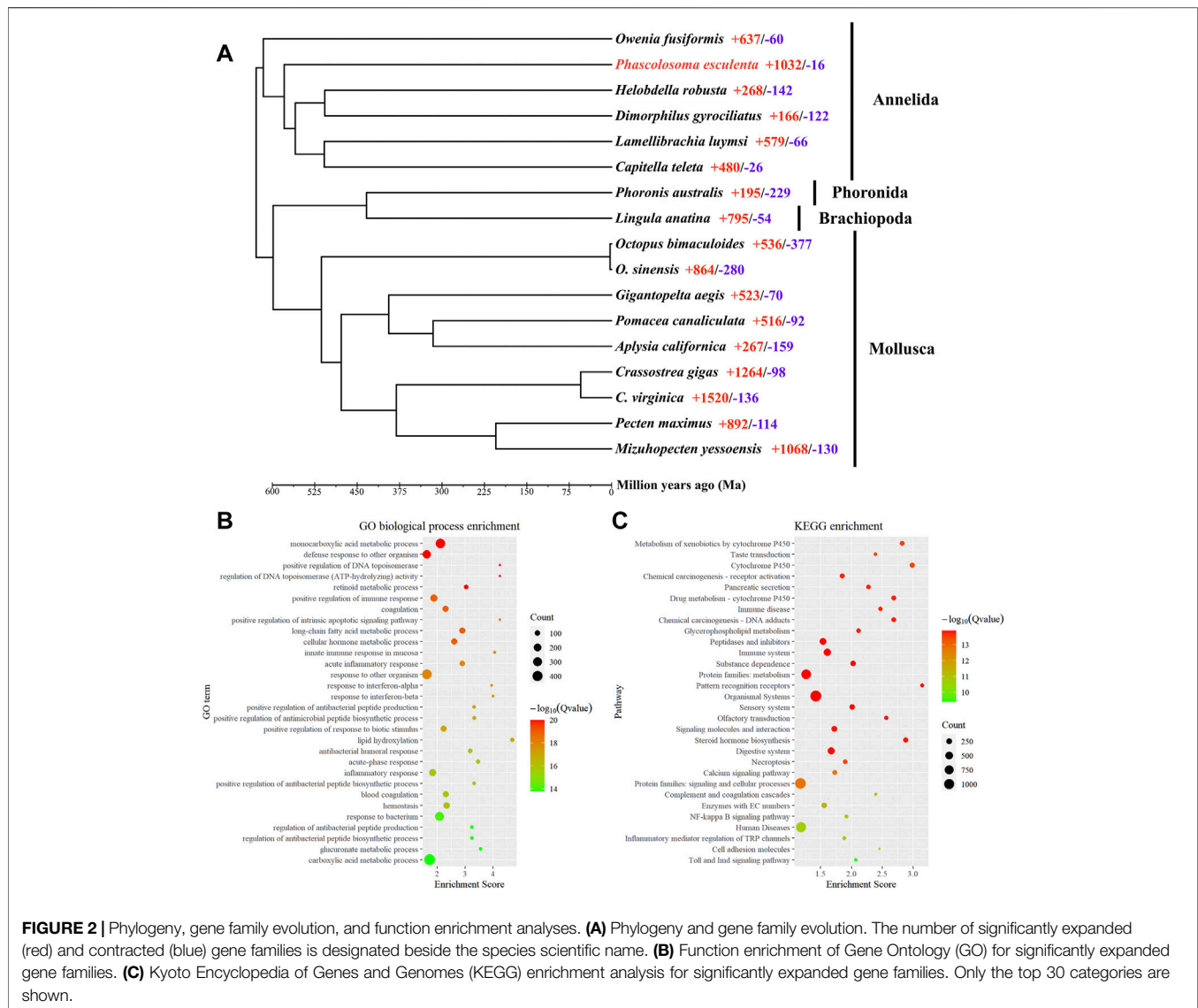
Firstly, the comprehensive transposable elements (TEs) in the genome were detected using EDTA package v2.0.0 (Ou et al., 2019) and the genome were further masked using RepeatMasker v4.1.2 (<http://www.repeatmasker.org>), based on the TE library generated by the EDTA package. Secondly, a combination of *ab initio*, homology, and transcript-based prediction methods were used to detect the protein-coding genes in the genome. The RNA sequencing data of three tissues were used for transcript-based prediction. *De novo* transcript sequences were reconstructed using the Trinity platform v2.13.2 (Haas et al., 2013), and genome guided transcript sequences were performed using HISAT v2.2.1 and StringTie v2.2.0 (Pertea et al., 2016). A combination of the *de novo* and genome-guided transcript was aligned to the genome by PASA software v2.5.2 (Haas et al., 2008) to obtain transcript-based predicted protein. For the homology-based method, the protein data of Annelida, including *Owenia fusiformis*, *Helobdella robusta*, *Dimorphilus gyrociliatus*,

Lamellibrachia luymisi, *Capitella teleta*, and *Enchytraeus crypticus* from Genbank were used as reference data to predict the protein-coding genes in the *P. esculenta* genome by GeMoMa program v1.7.1 (Keilwagen et al., 2018). For the *ab initio*-based method, gene prediction was performed using BRAKER2 (Brůna et al., 2021) with protein and RNA-seq forms of evidence. Finally, the gene structures in the *P. esculenta* genome were predicted by a combination of the three methods with EvidenceModeler v1.1.1 (Haas et al., 2008), and the predicted protein-coding genes' functional annotation was performed using BLASTP v2.12.0 + on a local high performance server (PowerEdge T630, Dell Technologies), with the E-value cutoff of 1e-5, to search the homologous sequences from the public databases, including UniProtKB/Swiss-Prot, UniProtKB/TremBL, Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) protein databases. The gene annotation integrity was evaluated by BUSCO software v5.2.2 based on the metazoa_odb10 lineages.

The transfer RNAs (tRNAs) were identified using tRNAscan-SE v2.0.6 (Chan et al., 2021); microRNAs were identified using miRNature v1.0 (Velandia-Huerto et al., 2021); and the transfer RNAs (tRNAs) were identified using RNAmmer v1.2 (Lagesen et al., 2007). The other small nuclear RNAs (snRNAs) were identified by searching against the Rfam database using Infernal v1.1.2 (Nawrocki and Eddy, 2013).

Phylogenomics and Gene Family Evolution

To investigate the phylogenetic status of Sipuncula, the protein data of 16 Lophotrochozoa species including *Mizuhopecten yessoensis*, *Pecten maximus*, *Crassostrea gigas*, *C. virginica*, *Aplysia californica*, *Pomacea canaliculata*, *Gigantopelta aegis*, *Octopus bimaculoides*, *O. sinensis*, *Lingula anatine*, *Phoronis australis*, *C. teleta*, *L. luymisi*, *D. gyrociliatus*, *H. robusta*, and *O. fusiformis* were retrieved from the NCBI genome database (<https://www.ncbi.nlm.nih.gov/genome/>), and the longest transcript of each gene was selected for identification of the orthologous gene. The orthologous gene clusters of *P. esculenta* and 16 related species were identified using OrthoFinder v2.5.4 (Emms and Kelly, 2019); the single-copy protein sequences were aligned by MUSCLE v3.8.31 with default parameters; and the ambiguously aligned positions were trimmed by trimAl v1.4.1 with default parameters. The alignments of the single-copy sequences were concatenated into continuous super protein sequences, and then the super protein sequences were used to construct the phylogenetic tree using RAXML v8.2.12 (Stamatakis, 2014) under Gamma + LG + F amino acid substitution model with 1,000 bootstrap replicates. Analysis of the likelihood for gene family gain and loss of *P. esculenta* and 16 related species was performed by CAFE v4.2.1 (Han et al., 2013) with $p < 0.05$. Divergence time between Lophotrochozoa species was estimated using r8s v1.71, with the divergence time of *M. yessoensis* and *P. maximus* obtained from the TimeTree website (<http://www.timetree.org>) for calibration. The GO and KEGG enrichment analyses were carried out using TBtools (Chen et al., 2020) to analyze the significant expansion in gene family function from *P. esculenta*.



RESULTS AND DISCUSSION

For genome assembly of *P. esculenta*, approximately 81.26 Gb clean Nanopore long reads, with average read length of 20,978 bp and 112.40 Gb clean illumina short reads with 96.11% Q20, were generated (**Supplementary Table S1**). To estimate the main genome characteristics of *P. esculenta*, the k-mer-based method based on illumina short data was applied. The estimated genome size of *P. esculenta* is about 1,465.56 Mb, the repeat content and the heterozygous rate of the genome were about 57.84 and 2.62%, respectively (**Figure 1B**, **Supplementary Table S2**). The estimated genome size is smaller than *Phascolosoma scolops*, which was about 1760.04 Mb by flow cytometer analysis (Adachi et al., 2016).

The initial assembly yielded a total length of 1.99 Gb, comprising 2,206 contigs with a contig N50 length of 2.04 Mb. Due to high heterozygosity of the *P. esculenta* genome, the initial genome assembly was larger than the estimated genome size by

k-mer-based method. After correcting base errors of the initial genome assembly, we resolved the redundancy of genome assembly by Purge_Dups. The final genome assembly was 1.71 Gb in total length, comprising 1,446 contigs with a contig N50 of 2.49 Mb, and the largest contig was 12.99 Mb in length (**Table 1**). The BUSCO analysis was then performed to evaluate the completeness of the final genome assembly, and the result showed that the completeness of this assembled genome was 98.5% (95.0% complete BUSCOs plus 3.5% fragmented BUSCOs) (**Supplementary Table S3**). The overall genome completeness of *P. esculenta* is one of the highest among the published annelid genomes (Li et al., 2019; Sun et al., 2021), which indicates that the genome integrity is high in the final genome assembly. To further validate the assembly completeness, the BGISEQ-500 short read data were mapped to the final assembly with BWA v0.7.17, and the mapping rate was 94.37%.

A total of 60.15% of the *P. esculenta* genome (1.03 Gb) were identified as repetitive elements, similar to the estimate in the

tubeworm *Paraescarpia echinospica* genome (55.10%) (Sun et al., 2021). The *P. esculenta* genome is among the highest percentage of repetitive sequences among the published annelid genomes, and the most abundant transposable elements were terminal inverted repeats (TIRs, 30.18% of the genome), followed by long terminal repeats (LTRs, 18.95% of the genome) and helitron (5.97%) (Supplementary Table S4). Meanwhile, a total of 41,469 genes in the *P. esculenta* genome were predicted as protein-coding genes by a combination of three gene identifying methods with EvidenceModeler. BUSCO analysis of these protein-coding genes showed that 99.4% of metazoan core conserved genes were detected in the *P. esculenta* gene set, with 98.7% and 0.7% being identified as complete and fragmented, respectively (Supplementary Table S3). Approximately 87.11% of the predicted protein-coding genes were successfully annotated by at least one of the public databases: Swiss-Prot (66.34%), TremBL (87.10%), KEGG (37.13%), and GO (43.56%) (Supplementary Table S5). Furthermore, approximately 0.58 Mb of the genome were annotated as non-coding RNAs, including 1,699 microRNAs, 2,894 tRNAs, 120 rRNAs, and 170 snRNAs (Supplementary Table S6).

To reveal the phylogenetic relationships between *P. esculenta* and other lophotrochozoan species, OrthoFinder was applied for identification of the orthologous genes. A total of 4,445 orthologous genes clusters shared by all species were identified, including 245 single-copy orthologous in a 1:1:1 manner. A total of 1,688 orthogroups (6,971 genes) were identified as species-specific orthogroups of *P. esculenta* (Supplementary Table S7). The phylogenetic tree was constructed with 245 single-copy orthologous genes of 73,344 amino acid sites using RAxML, and the result showed that *P. esculenta* (Sipuncula), *O. fusiformis* (Oweniidae), and other annelid species including *L. luymisi* (Siboglinidae), *C. teleta* (Capitellidae), *D. gyrocoliatius* (*Dinophilidae*) and *H. robusta* (Glossiphoniidae) were clustered and constituted the annelid clade. *P. esculenta* (Sipuncula) was found clustered within Annelida, and together with *O. fusiformis* (Oweniidae) taxa, occupied the basal branch of the annelid clade (Supplementary Figure S1). Our phylogenomics analysis result indicated that Sipuncula is in the basal branching position of Annelida, which is congruent with previous phylogenomics studies using mitogenome (Weigert et al., 2016) and transcriptomic (Weigert et al., 2014; Weigert and Bleidorn, 2016) analysis.

Computational analysis of gene family evolution between *P. esculenta* and other lophotrochozoan species revealed that 1,032 and 16 gene families were significantly expanded and contracted in the *P. esculenta* genome, respectively (Figure 2A). Superoxide dismutases (SODs) have an important functional role in protecting cells against oxidative damage induced by environmental stress (Finkel and Holbrook, 2000; Surai et al., 2019). Genomic expansions of Cu/Zn superoxide dismutase (Cu/Zn-SOD) genes and Mn superoxide dismutase (Mn-SOD) in *P. esculenta* genome were

revealed by likelihood analysis. Most annelid genomes contain less than three copies of the SOD gene (Sun et al., 2021), but *P. esculenta* genome has 12 copies of the Cu/Zn-SOD gene and 15 copies of the Mn-SOD gene, the largest number of copies among the published annelid genomes. Significant expansion of SODs in the *P. esculenta* genome may play an important role in overcoming oxidative damage induced by intertidal stress and could be essential for *P. esculenta* to inhabit the intertidal zone. Moreover, the GO and KEGG enrichment analyses with 1,032 significantly expanded gene families were performed to clarify the molecular biological function of gene family evolution in *P. esculenta*. GO enrichment analysis revealed that these expanded gene families are mainly involved in defense responses against other organisms (GO: 0098542, GO:0051707), positive regulation of immune response (GO:0050778, GO:0002227), blood coagulation (GO:0050817, GO:0007596), positive regulation of the apoptotic signaling pathway (GO:1902231), and antibacterial humoral response (GO:0019731) (Figure 2B). Meanwhile, KEGG enrichment analysis showed that these expanded gene families were significantly enriched in 91 pathways, including immune and defense associated pathways, such as metabolism of xenobiotics by cytochrome P450, the immune system, pattern recognition receptors, complement and coagulation cascades, and the Toll and Imd signaling pathway (Figure 2C). Several gene families related to defense pathways—including immune responses, apoptosis and anti-oxidation—were also expanded in the *C. gigas* genome. The expansion of key immune and defense genes indicates the sophisticated genomic adaptations of the oyster in inhabiting a highly stressful environment (Zhang et al., 2012). The defense related gene family cytochrome P450 (CYPs) were also significantly expanded in the *Daphnia pulex* genome, which provides important insight into the adaptation of *D. pulex* to environmental changes (Baldwin et al., 2009). So too the GO and KEGG enrichment analysis results in this study suggest that genomic expansions of immune and defense associated gene families offer important insight into the ecological adaptation of *P. esculenta* to environmental stresses in the intertidal zone.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

AUTHOR CONTRIBUTIONS

SZ, YL, and XC conceived the study. SZ, XM, YJ, and YQ performed bioinformatics analysis. SZ, LZ, and GH collected samples. LH, GH and YZ extracted DNA and RNA. SZ, XM, and YJ wrote the manuscript. All authors approved the final submission.

FUNDING

This work was financially supported by the Scientific Research and Technology Development Program of Guangxi (Grant No. AB16380105), the Innovation Driven Development Foundation of Guangxi (AD19245135 and AD19245161), the China Postdoctoral Science Foundation (2021M701798), the Traditional Chinese Medicine inheritance and Innovation team of Guangxi University of Chinese Medicine (Grant No. 2022A007), the Development Program of High-level Talent Team under Qihuang Project of Guangxi University of Chinese Medicine (Grant No. 2021004), the Team Research Funds for Institute of marine drugs Guangxi University of Chinese Medicine (Grant No. 2018ZD005-A10 and 2018ZD005-A20), the National Modern Agricultural

Industry Technology System Guangxi Innovation Team (Grant No. nycytxgxcxt-14-06), the General Program of Guangxi Zhuang Autonomous Region (Grant No. 2021GXNSFAA220030 and 2020GXNSFBA297126), the Youth Foundation of Guangxi Zhuang Autonomous Region (2022GXNSFBA035166) and the PhD research startup foundation of Guangxi University of Chinese Medicine (Grant No. 2019BS018).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.910344/full#supplementary-material>

REFERENCES

- Adachi, K., Kuramochi, T., Takaki, Y., Ohnishi, K., Yoshinaga, T., and Okumura, S.-i. (2016). Genome Size of Spoon and Peanut Worms, and the Validity of Frozen Samples for Flow Cytometry Analysis. *Fish Genet. Breed. Sci.* 45, 25–31.
- Andrade, S. C. S., Novo, M., Kawachi, G. Y., Worsaae, K., Pleijel, F., Giribet, G., et al. (2015). Articulating “Archiannelids”: Phylogenomics and Annelid Relationships, with Emphasis on Meiofaunal Taxa. *Mol. Biol. Evol.* 32 (11), 2860–2875. doi:10.1093/molbev/msv157
- Baldwin, W. S., Marko, P. B., and Nelson, D. R. (2009). The Cytochrome P450 (CYP) Gene Superfamily in *Daphnia pulex*. *BMC Genomics* 10 (1), 169. doi:10.1186/1471-2164-10-169
- Bleidorn, C. (2019). Recent Progress in Reconstructing Lophotrochozoan (Spiralian) Phylogeny. *Org. Divers. Evol.* 19 (4), 557–566. doi:10.1007/s13127-019-00412-4
- Brúna, T., Hoff, K. J., Lomsadze, A., Stanke, M., and Borodovsky, M. (2021). BRAKER2: Automatic Eukaryotic Genome Annotation with GeneMark-Ep+ and AUGUSTUS Supported by a Protein Database. *NAR Genomics Bioinforma.* 3 (1), lqaa108. doi:10.1093/nargab/lqaa108
- Chan, P. P., Lin, B. Y., Mak, A. J., and Lowe, T. M. (2021). tRNAscan-SE 2.0: Improved Detection and Functional Classification of Transfer RNA Genes. *Nucleic Acids Res.* 49 (16), 9077–9096. doi:10.1093/nar/gkab688
- Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant* 13 (8), 1194–1202. doi:10.1016/j.molp.2020.06.009
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: an Ultra-fast All-In-One FASTQ Preprocessor. *Bioinformatics* 34 (17), i884–i890. doi:10.1093/bioinformatics/bty560
- Dordel, J., Fisse, F., Purschke, G., and Struck, T. H. (2010). Phylogenetic Position of Sipuncula Derived from Multi-Gene and Phylogenomic Data and its Implication for the Evolution of Segmentation. *J. Zoological Syst. Evol. Res.* 48 (3), 197–207. doi:10.1111/j.1439-0469.2010.00567.x
- Emms, D. M., and Kelly, S. (2019). OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics. *Genome Biol.* 20 (1), 238. doi:10.1186/s13059-019-1832-y
- Finkel, T., and Holbrook, N. J. (2000). Oxidants, Oxidative Stress and the Biology of Ageing. *Nature* 408 (6809), 239–247. doi:10.1038/35041687
- Guan, D., McCarthy, S. A., Wood, J., Howe, K., Wang, Y., and Durbin, R. (2020). Identifying and Removing Haplotypic Duplication in Primary Genome Assemblies. *Bioinformatics* 36 (9), 2896–2898. doi:10.1093/bioinformatics/btaa025
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De Novo transcript Sequence Reconstruction from RNA-Seq Using the Trinity Platform for Reference Generation and Analysis. *Nat. Protoc.* 8 (8), 1494–1512. doi:10.1038/nprot.2013.084
- Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., et al. (2008). Automated Eukaryotic Gene Structure Annotation Using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* 9 (1), R7. doi:10.1186/gb-2008-9-1-r7
- Han, M. V., Thomas, G. W. C., Lugo-Martinez, J., and Hahn, M. W. (2013). Estimating Gene Gain and Loss Rates in the Presence of Error in Genome Assembly and Annotation Using CAFE 3. *Mol. Biol. Evol.* 30 (8), 1987–1997. doi:10.1093/molbev/mst100
- Hu, J., Fan, J., Sun, Z., and Liu, S. (2020). NextPolish: a Fast and Efficient Genome Polishing Tool for Long-Read Assembly. *Bioinformatics* 36 (7), 2253–2255. doi:10.1093/bioinformatics/btz891
- Huang, D.-Y., Chen, J.-Y., Vannier, J., and Saiz Salinas, J. I. (2004). Early Cambrian Sipunculan Worms from Southwest China. *Proc. R. Soc. Lond. B* 271 (1549), 1671–1676. doi:10.1098/rspb.2004.2774
- Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S. O., and Grau, J. (2018). Combining RNA-Seq Data and Homology-Based Gene Prediction for Plants, Animals and Fungi. *BMC Bioinforma.* 19 (1), 189. doi:10.1186/s12859-018-2203-5
- Lagesen, K., Hallin, P., Rodland, E. A., Stærfeldt, H.-H., Rognes, T., and Ussery, D. W. (2007). RNAMmer: Consistent and Rapid Annotation of Ribosomal RNA Genes. *Nucleic Acids Res.* 35 (9), 3100–3108. doi:10.1093/nar/gkm160
- Lemer, S., Kawachi, G. Y., Andrade, S. C. S., González, V. L., J. Boyle, M. M., and Giribet, G. (2015). Re-evaluating the Phylogeny of Sipuncula through Transcriptomics. *Mol. Phylogenetics Evol.* 83, 174–183. doi:10.1016/j.ympev.2014.10.019
- Li, Y., Tassia, M. G., Waits, D. S., Bogantes, V. E., David, K. T., and Halanych, K. M. (2019). Genomic Adaptations to Chemosymbiosis in the Deep-Sea Seep-Dwelling Tubeworm *Lamellibrachia Luymesii*. *BMC Biol.* 17 (1), 91. doi:10.1186/s12915-019-0713-x
- Luo, Y.-J., Kanda, M., Koyanagi, R., Hisata, K., Akiyama, T., Sakamoto, H., et al. (2018). Nemertean and Phoronid Genomes Reveal Lophotrochozoan Evolution and the Origin of Bilaterian Heads. *Nat. Ecol. Evol.* 2 (1), 141–151. doi:10.1038/s41559-017-0389-y
- Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., and Zdobnov, E. M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol. Biol. Evol.* 38 (10), 4647–4654. doi:10.1093/molbev/msab199
- Marçais, G., and Kingsford, C. (2011). A Fast, Lock-free Approach for Efficient Parallel Counting of Occurrences of K-Mers. *Bioinformatics* 27 (6), 764–770. doi:10.1093/bioinformatics/btr011
- Meng, J., Gao, X., Luo, S., Lin, C., Du, C., Hou, C., et al. (2022). Cloning, Functional Characterization and Response to Cadmium Stress of the Thioredoxin-like Protein 1 Gene from *Phascolosoma Esculenta*. *Ijms* 23 (1), 332. doi:10.3390/ijms23010332
- Mwinyi, A., Meyer, A., Bleidorn, C., Lieb, B., Bartolomaeus, T., and Podsiadlowski, L. (2009). Mitochondrial Genome Sequence and Gene Order of Sipunculus Nudus Give Additional Support for an Inclusion of Sipuncula into Annelida. *BMC Genomics* 10 (1), 27. doi:10.1186/1471-2164-10-27
- Nawrocki, E. P., and Eddy, S. R. (2013). Infernal 1.1: 100-fold Faster RNA Homology Searches. *Bioinformatics* 29 (22), 2933–2935. doi:10.1093/bioinformatics/btt509

- Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J. R. A., Hellinga, A. J., et al. (2019). Benchmarking Comprehensible Element Annotation Methods for Creation of a Streamlined, Comprehensive Pipeline. *Genome Biol.* 20 (1), 275. doi:10.1186/s13059-019-1905-y
- Perete, M., Kim, D., Perete, G. M., Leek, J. T., and Salzberg, S. L. (2016). Transcript-level Expression Analysis of RNA-Seq Experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* 11 (9), 1650–1667. doi:10.1038/nprot.2016.095
- Ranallo-Benavidez, T. R., Jaron, K. S., and Schatz, M. C. (2020). GenomeScope 2.0 and Smudgeplot for Reference-free Profiling of Polyploid Genomes. *Nat. Commun.* 11 (1), 1432. doi:10.1038/s41467-020-14998-3
- Rouse, G. W., and Fauchald, K. (1997). Cladistics and Polychaetes. *Zool. Scr.* 26 (2), 139–204. doi:10.1111/j.1463-6409.1997.tb00412.x
- Scheltema, A. H. (1993). Aplousophora as Progenetic Aculiferans and the Coelomate Origin of Mollusks as the Sister Taxon of Sipuncula. *Biol. Bull.* 184 (1), 57–78. doi:10.2307/1542380
- Schulze, A., Cutler, E. B., and Giribet, G. (2007). Phylogeny of Sipunculan Worms: A Combined Analysis of Four Gene Regions and Morphology. *Mol. Phylogenetics Evol.* 42 (1), 171–192. doi:10.1016/j.ympev.2006.06.012
- Schulze, A., Cutler, E. B., and Giribet, G. (2005). Reconstructing the Phylogeny of the Sipuncula. *Hydrobiologia* 535–536 (1), 277–296. doi:10.1007/s10750-004-4404-3
- Schulze, A., and Kawachi, G. Y. (2021). How Many Sipunculan Species Are Hiding in Our Oceans? *Diversity* 13 (2), 43. doi:10.3390/d13020043
- Shen, W., Liu, C., Ni, J., Gao, X., Ni, J., Wang, J., et al. (2021). Effects of Low Temperature Stress on the Morphology and Hsp70 and Hsp90 Gene Expression of *Phascolosoma Esculenta*. *J. Ocean. Univ. China* 20 (1), 159–168. doi:10.1007/s11802-021-4475-z
- Shen, X., Ma, X., Ren, J., and Zhao, F. (2009). A Close Phylogenetic Relationship between Sipuncula and Annelida Evidenced from the Complete Mitochondrial Genome Sequence of *Phascolosoma Esculenta*. *BMC Genomics* 10 (1), 136. doi:10.1186/1471-2164-10-136
- Sperling, E. A., Vinther, J., Moy, V. N., Wheeler, B. M., Sémon, M., Briggs, D. E. G., et al. (2009). MicroRNAs Resolve an Apparent Conflict between Annelid Systematics and Their Fossil Record. *Proc. R. Soc. B* 276 (1677), 4315–4322. doi:10.1098/rspb.2009.1340
- Stamatakis, A. (2014). RAxML Version 8: a Tool for Phylogenetic Analysis and Post-analysis of Large Phylogenies. *Bioinformatics* 30 (9), 1312–1313. doi:10.1093/bioinformatics/btu033
- Struck, T. H., Paul, C., Hill, N., Hartmann, S., Hösel, C., Kube, M., et al. (2011). Phylogenomic Analyses Unravel Annelid Evolution. *Nature* 471 (7336), 95–98. doi:10.1038/nature09864
- Struck, T. H., Schult, N., Kusen, T., Hickman, E., Bleidorn, C., McHugh, D., et al. (2007). Annelid Phylogeny and the Status of Sipuncula and Echiura. *BMC Evol. Biol.* 7 (1), 57. doi:10.1186/1471-2148-7-57
- Su, X., Du, L., Li, Y., Li, Y., Zhou, J., and Li, T. (2010). Cloning and Expression of HSP70 Gene of Sipuncula *Phascolosoma Esculenta*. *Fish Shellfish Immunol.* 28 (3), 461–466. doi:10.1016/j.fsi.2009.12.014
- Sun, Y., Sun, J., Yang, Y., Lan, Y., Ip, J. C.-H., Wong, W. C., et al. (2021). Genomic Signatures Supporting the Symbiosis and Formation of Chitinous Tube in the Deep-Sea Tubeworm *Paraescarpia Echinospica*. *Mol. Biol. Evol.* 38 (10), 4116–4134. doi:10.1093/molbev/msab203
- Surai, P. F., Kochish, I. I., Fisinin, V. I., and Kidd, M. T. (2019). Antioxidant Defence Systems and Oxidative Stress in Poultry Biology: An Update. *Antioxidants* 8 (7), 235. doi:10.3390/antiox8070235
- Velandia-Huerto, C. A., Fallmann, J., and Stadler, P. F. (2021). miRNA-Computational Detection of microRNA Candidates. *Genes* 12 (3), 348. doi:10.3390/genes12030348
- Weigert, A., and Bleidorn, C. (2016). Current Status of Annelid Phylogeny. *Org. Divers Evol.* 16 (2), 345–362. doi:10.1007/s13127-016-0265-7
- Weigert, A., Golombek, A., Gerth, M., Schwarz, F., Struck, T. H., and Bleidorn, C. (2016). Evolution of Mitochondrial Gene Order in Annelida. *Mol. Phylogenetics Evol.* 94, 196–206. doi:10.1016/j.ympev.2015.08.008
- Weigert, A., Helm, C., Meyer, M., Nickel, B., Arendt, D., Hausdorf, B., et al. (2014). Illuminating the Base of the Annelid Tree Using Transcriptomics. *Mol. Biol. Evol.* 31 (6), 1391–1401. doi:10.1093/molbev/msu080
- Zhang, G., Fang, X., Guo, X., Li, L., Luo, R., Xu, F., et al. (2012). The Oyster Genome Reveals Stress Adaptation and Complexity of Shell Formation. *Nature* 490 (7418), 49–54. doi:10.1038/nature11413
- Zhao, Y., Zhang, Q., and Zhong, S. (2019). The Complete Mitochondrial Genome of Chinese Endemic Species *Phascolosoma Esculenta* (Sipuncula, Phascolosomatidae) from Beibu Bay. *Mitochondrial DNA Part B* 4 (1), 380–381. doi:10.1080/23802359.2018.1547169
- Zhong, S., Huang, L., Liu, Y., Huang, G., and Chen, X. (2020). The First Complete Mitochondrial Genome of Siphonosoma from *Siphonosoma Cumanense* (Sipuncula, Sipunculidae). *Mitochondrial DNA Part B* 5 (1), 957–958. doi:10.1080/23802359.2020.1721364

Conflict of Interest: Author SZ and YJ were employed by Guangxi Institute of Oceanology Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhong, Ma, Jiang, Qiao, Zhao, Huang, Huang, Zhao, Liu and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.