



The Chromosome-Level Genome Assembly and Comprehensive Transcriptomes of the Razor Clam (*Sinonovacula constricta*)

Yinghui Dong^{1†}, Qifan Zeng^{2†}, Jianfeng Ren^{3†}, Hanhan Yao¹, Liyuan Lv¹, Lin He¹, Wenbin Ruan¹, Qinggang Xue¹, Zhenmin Bao^{2,4}, Shi Wang^{2,5,6*} and Zhihua Lin^{1*}

¹ Key Laboratory of Aquatic Germplasm Resource of Zhejiang, College of Biological and Environmental Sciences, Zhejiang Wanli University, Ningbo, China, ² MOE Key Laboratory of Marine Genetics and Breeding, College of Marine Life Sciences, Ocean University of China, Qingdao, China, ³ Key Laboratory of Exploration and Utilization of Aquatic Genetic Resources, Ministry of Education, College of Fisheries and Life Science, Shanghai Ocean University, Shanghai, China, ⁴ Laboratory for Marine Fisheries Science and Food Production Processes, Pilot National Laboratory for Marine Science and Technology, Qingdao, China, ⁵ Laboratory for Marine Biology and Biotechnology, Pilot National Laboratory for Marine Science and Technology, Qingdao, China, ⁶ The Sars-Fang Centre, Ocean University of China, Qingdao, China

OPEN ACCESS

Edited by:

Shaojun Liu,
Hunan Normal University, China

Reviewed by:

Gen Hua Yue,
Temasek Life Sciences
Laboratory, Singapore
Yong-Hua Sun,
Institute of Hydrobiology (CAS), China

*Correspondence:

Shi Wang
swang@ouc.edu.cn
Zhihua Lin
zhihua9988@126.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 March 2020

Accepted: 01 June 2020

Published: 07 July 2020

Citation:

Dong Y, Zeng Q, Ren J, Yao H, Lv L,
He L, Ruan W, Xue Q, Bao Z, Wang S
and Lin Z (2020) The
Chromosome-Level Genome
Assembly and Comprehensive
Transcriptomes of the Razor Clam
(*Sinonovacula constricta*).
Front. Genet. 11:664.
doi: 10.3389/fgene.2020.00664

Keywords: razor clam, *Sinonovacula constricta*, genome, transcriptome, stress response

INTRODUCTION

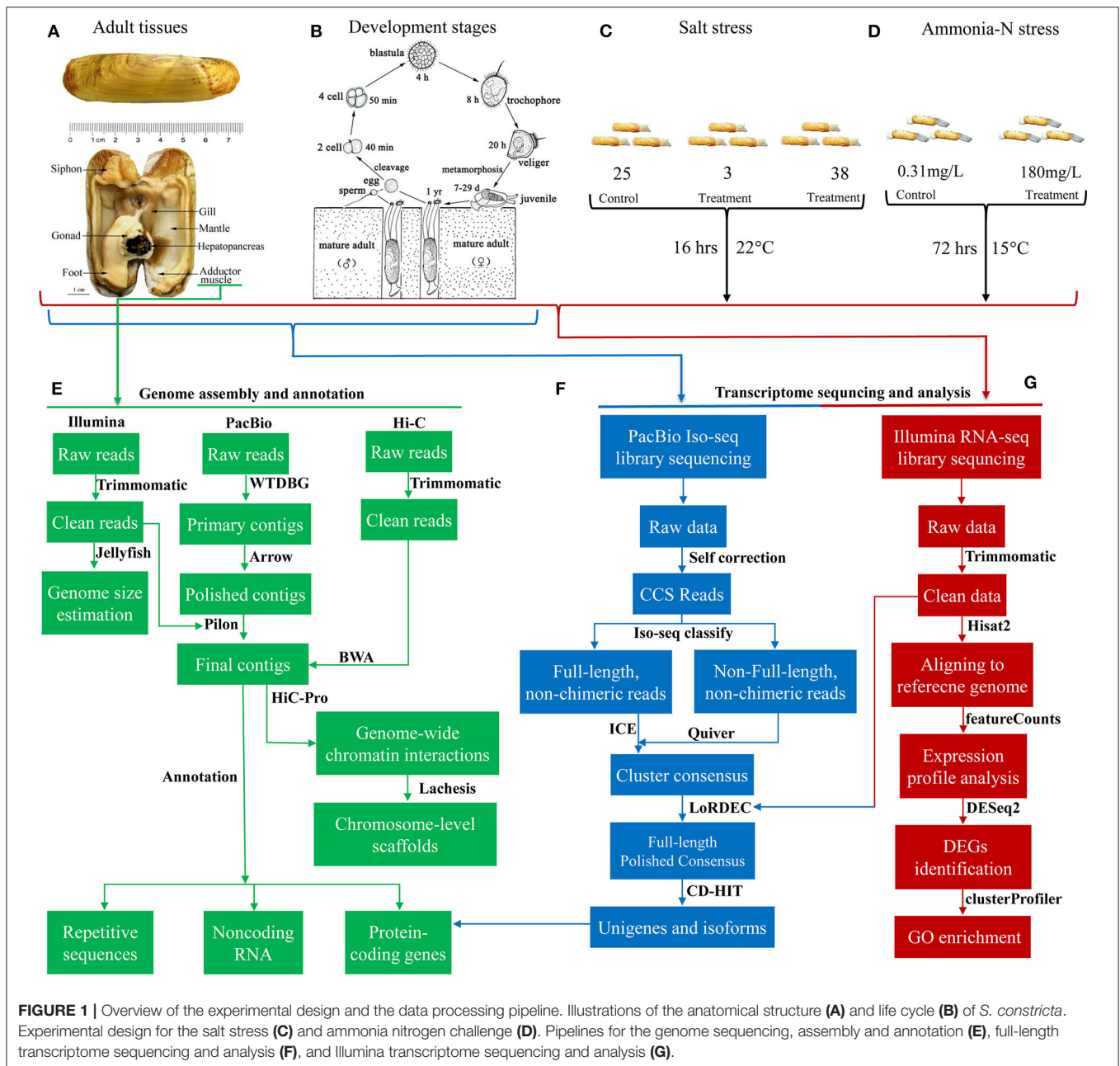
The Chinese razor clam *Sinonovacula constricta* (Lamarck, 1818) is a member of the *Solenidae* family of bivalve molluscs, recognizing for its straight razor-like shape and fragile shells (Morton, 1984). It is widely distributed in the intertidal zone along the west Pacific Ocean and engages in a pelago-benthic life cycle (Wang and Xu, 1997). As adaptation to a deep-burrowing lifestyle, the razor clam is characterized by smooth shells, muscular foot, and elongated siphons (Morton, 1984). Benefit from its relatively short production cycle and high productive efficiency, the razor clam has become one of most important maricultured bivalve species in Asia with over 800,000 metric tons of production in 2017 (FAO, 2019).

As living in the estuarine and intertidal region, the razor clam faces tremendous exposure to extreme environmental stresses such as drastic salinity fluctuation, highly variable temperature, high concentration of ammonia nitrogen and hydrogen sulfide (Morton, 1984). Unlike oysters, mussels and most clams with thick and sealed shells for protecting their soft bodies, the razor clam with two thin and unclosed shells has chosen a survival strategy of deep-burrowing lifestyle with high tolerance of a broad range of salinities (5–45 psu) (Morton, 1984; Peng et al., 2019). Therefore, it is an ideal model with which to investigate the adaptive mechanisms of a deep-burrowing lifestyle. Despite that increased genomic sequences (Dong et al., 2019; Ran et al., 2019) and transcriptomic data (Niu et al., 2013) have been generated, full-spectrum spatial-temporal transcriptomes are still insufficient for exploring its unique biology and adaptive evolution.

Here, we generated the high-quality chromosomal-level genome assembly and comprehensive transcriptomes of *S. constricta* and investigated the transcriptomic changes under salinity and ammonia stresses. These genomic resources will lay a prime foundation for future studies of deep-burrowing lifestyle-related adaptive evolution and genetic improvement in commercial breeding programs of razor clam.

DATA

The experimental design was illustrated in **Figure 1**. In total, 386.2 Gb clean data were used for assembling the genome of *S. constricta*, including 129.73 Gb Illumina reads, 101.79 Gb Pacbio reads, and 154.68 Gb Hi-C reads (**Table 1**). The genome size was estimated to be 1,244.27 Mb,



with a heterozygous ratio of 1.53% and a repeat rate of 53.12%, which was consistent with the results of a previous study (Ran et al., 2019). By using the WTDBG pipeline, the resulting 1,331.97 Mb assembly was obtained with a contig N50 of 678,857 bp (Table 1). The contigs were clustered into 19 linkage groups using Lachesis (Figure 2A), which was in concordance with the karyotype analysis (Wang et al., 1998). The final assembly was 1,331.28 Mb in length and the scaffold N50 reached over 57.99 Mb. The *S. constricta* genome exhibited a remarkable level of macrosynteny to the ancestral bilaterian linkage groups with a conservation index (CI) of 0.71, indicating the high accuracy of the Hi-C clustering (Figure 2B; Simakov et al., 2013; Wang et al., 2017). The integrity of the genomic assembly was evaluated

by alignment of the Illumina reads. In summary, 88.90% of the genomic sequences were covered by 93.93% of the total reads (Table S1). The Core Eukaryotic Genes Mapping Approach (CEGMA) analysis (Parra et al., 2007) and Benchmarking Universal Single-Copy Orthologs (BUSCO version 3) analysis also revealed a high-level of completeness by identifying 227 of the 248 core eukaryotic genes (91.53%) and 868 of the 978 near-universal single-copy metazoan orthologs (88.7%), respectively (Tables S2, S3).

Genomic annotation identified 50.71% of the assembled genome as repetitive sequences. Repetitive sequences were dominated by tandem repeats (15.39%) and followed by DNA transposons (14.38%) and LTR retrotransposons (10.84%)

TABLE 1 | Summary statistics of genome assembly of *Sinonovacula constricta*.

Section	Results	
A. SEQUENCES		
	Clean data (Gb)	Depth (X)
Illumina	129.73	104.62
PacBio	101.79	82.01
Hi-C	154.68	124.74
Total	386.20	311.37
B. ASSEMBLY		
Number of scaffolds	7,932	
Total bases (bp)	1,332,277,427	
Contig N50 (bp)	678,857	
Scaffold N50 (bp)	57,991,182	
C. GENES		
Number of genes	26,270	
Complete BUSCOs	88.75%	
Number of functional annotated genes	26,140	
D. REPETITIVE SEQUENCES		
	Repeat size (bp)	% of genome
DNA	191,499,094	14.38
LINE	71,938,692	5.40
LTR	144,451,530	10.84
SINE	5,528,172	0.42
Tandem repeat	204,889,587	15.38
Other	157,232	0.01
Unknown	56,940,582	4.27

(Table S4). Using a combination of *de novo* prediction, homologous prediction, and alignment of transcriptomic data, 26,270 protein-coding genes were annotated (Table S5). A total of 26,140 (99.5%) genes could be functionally annotated by at least one database (Table S6).

The gene set of *S. constricta* were compared to fifteen eumetazoan species, including *Homo sapiens*, *Branchiostoma floridae*, *Drosophila melanogaster*, *Apis mellifera*, *Helobdella robusta*, *Capitella teleta*, *Octopus bimaculoides*, *Lottia gigantea*, *Aplysia californica*, *Crassostrea gigas*, *Crassostrea virginica*, *Saccostrea glomerata*, *Patinopecten yessoensis*, *Chlamys farreri*, and *Nematostella vectensis*. The proteins from all 16 species were assigned to 39,058 families with 337 strict single-copy orthologs. Maximum likelihood phylogenetic inference of these single-copy orthologs revealed that *S. constricta* diverged from the lineage that led to oysters and scallops ~494 million years ago (Mya; Figure 2C). Among the 12,945 gene families identified in the *S. constricta* genome, 193 and 31 gene families were significantly expanded and contracted, respectively (Figure 2C). Notably, the cytoskeletal protein alpha tubulin (*TUA*) and the motor protein dynein heavy chain (*DYH*) families were rapidly expanded in the *S. constricta* genome (Figure 2D, Figures S1, S2). These families play vital roles in the microtubule architecture and bending movement of cilia (Mohri et al., 2012). The razor clam has a highly developed ciliary system for gill filtering,

food-particle retention and water pumping (Morton, 1984). Adjoining cilia generate effective beats through coordinated wavelike movements. The pumping rate of the ciliary system in the gill and mantle cavity can be adjusted to generate powerful currents that facilitate the principal sorting and retaining of suspended particles in the labial palps. Effluxes can also be ejected from the pedal gape to flush away sources of irritation detected by sensory tentacles (Morton, 1984). The transcriptomic data revealed that high levels of *TUA* and *DYH* genes are expressed in the gills (Figures 2E,F), implicating that the expansion of these genes might be an adaptation to the deep-burrowing lifestyle.

Molecular adaptations of *S. constricta* to salt fluctuations were analyzed by comparing the whole transcriptomic alterations under the salinity of 3, 25, and 38 psu (Figure S3). A total of 462 upregulated and 655 downregulated genes were identified under elevated salinity. Functional analysis showed that these differentially expressed genes (DEGs) were significantly enriched in chitin metabolism, immune response, scavenger receptor activity, and ATPase activity coupled to transmembrane substance movement. When facing the decreased salinity, *S. constricta* enhanced the expression of 898 genes and suppressed 826 genes. Enrichment analysis revealed that these DEGs were mainly related to transition metal ion binding and oxidation-reduction process. Tolerance of razor clams to the ammonia nitrogen stress were investigated by transcriptome analysis (Figure S4). In the gill, 1,029 and 386 genes were upregulated and downregulated after ammonia challenge, respectively. Functional analysis indicated that DEGs were significantly enriched in the regulation of nitrogen metabolism and nitrogen transport. A smaller number of DEGs were identified in the liver, including 248 upregulated and 58 downregulated genes, respectively. GO and KEGG analysis showed that the DEGs were significant enrichments in response to stress, exopeptidase activity, and copper ion binding.

MATERIALS AND METHODS

Genomic DNA Preparation, PacBio, and Illumina Sequencing

An individual *S. constricta* was collected from the brood stock of the genetic breeding center of Zhejiang Wanli University. Genomic DNA was extracted from the adductor muscle with phenol-chloroform as described (Green and Sambrook, 2012). High molecular-weight genomic DNA was sheared into fragments of ~30 kb using a Covaris ultrasonicator (Covaris Inc., Woburn, MA, USA). The fragments were enzymatically repaired and converted into SMRTbell™ template library following the manufacturer's instructions. Size-selection was performed to enrich the DNA fragments longer than 10 kb for sequencing on a Pacific Biosciences (PacBio) Sequel Single Molecule Real Time (SMRT) platform. A paired-end Illumina library with an insert size of 350 bp was prepared using Illumina Genomic DNA sample preparation kits and sequenced on an Illumina Xten system. Adductor muscle tissue of a razor clam from the same population was collected to construct a Hi-C library. The specimen was fixed with 1% formaldehyde and the genomic DNA

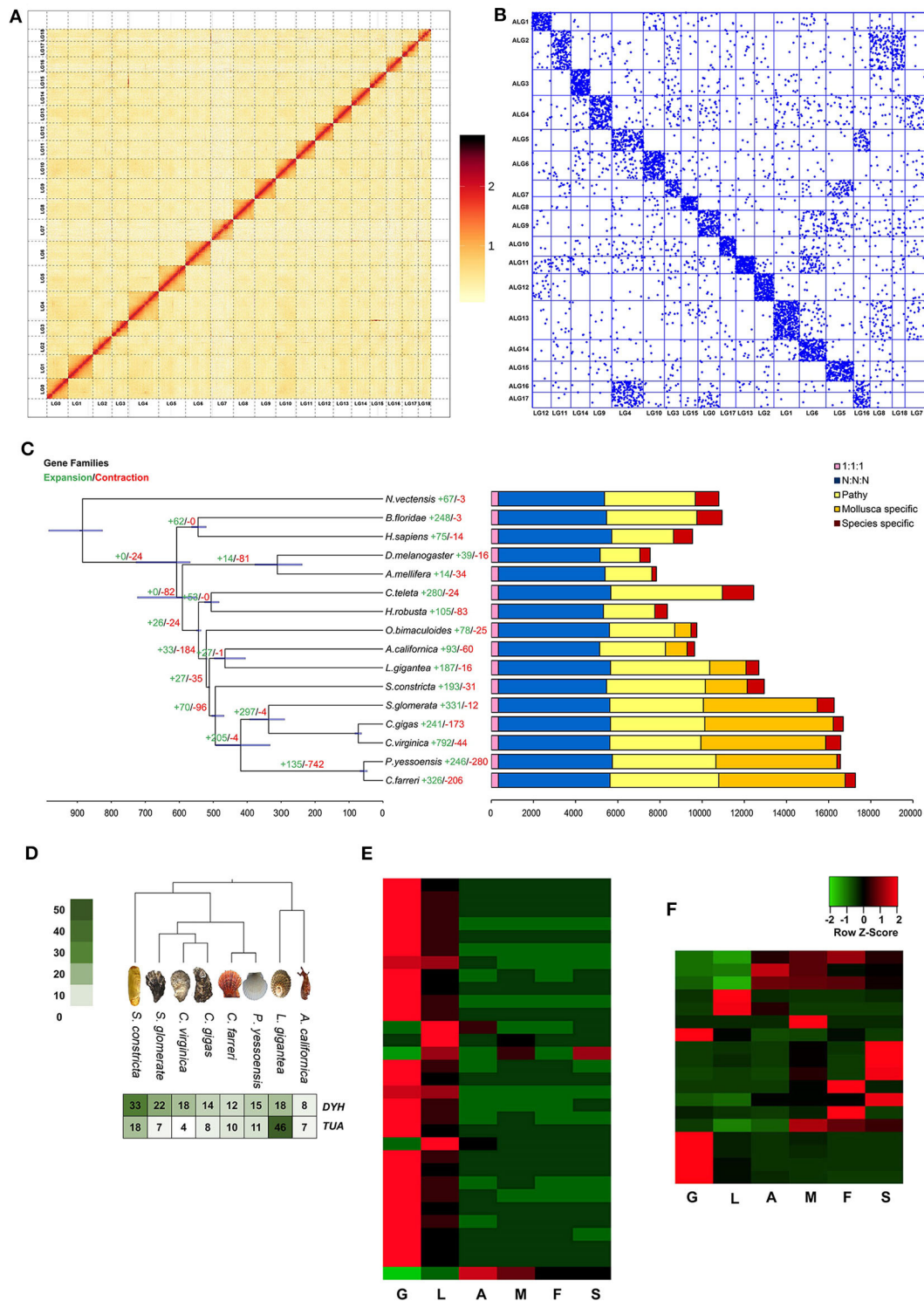


FIGURE 2 | (A) Phylogenetic tree and number of shared orthologs among *S. constricta* and other animal species. Numbers of gene families undergoing expansion and contraction for each lineage are indicated as red and green, respectively. **(B)** The comparison of the copy numbers of dynein heavy chain (*DYH*) and alpha tubulin (*TUA*) genes in 8 molluscan species. **(C–D)** The tissue-wide expression patterns of *DYH* genes and *TUA* genes. G, gill; L, liver; A, adductor muscle; M, mantle; F, foot; S, siphon. **(E)** Hi-C interaction heatmap of *S. constricta*. **(F)** Chromosome-based macro-synteny between *S. constricta* and the 17 presumed bilaterian ALGs retrieved from Simakov et al. (2013).

was cross-linked, digested with the restriction enzyme *Mbo*I, labeled with a biotinylated residue and end repaired. The library was sequenced on an Illumina NovaSeq platform.

Estimation of Genome Size and Sequencing Coverage

The Illumina reads were first trimmed to remove adaptors and reads with >10% ambiguous or >20% low-quality bases using the Trimmomatic package (Bolger et al., 2014). The distribution of the 17-mer frequency was estimated using the clean reads. Genome size was estimated according to the formula: genome size = k-mer number/peak depth (Varshney et al., 2011).

De novo Genome Assembly and Quality Assessment

PacBio long reads were assembled using the WTDBG package (Ruan and Li, 2019). The consensus calling of preceding assembly was conducted with wtDBG-cns to reduce sequencing errors and subsequently polishing by Arrow (SMRT Link v 5.1.0). To ensure high accuracy of the genome assembly, Illumina paired-end clean reads were aligned to the assembly using BWA. Post-processing error was corrected and conflicts of the assembly were resolved via Pilon (Walker et al., 2014).

The HiC reads were truncated at the junctions and aligned to the polished genome using BWA (version 0.7.17) with default parameters. Only unique aligned reads with a mapping quality >20 were further processed. After filtering invalid interaction pairs using HiC-Pro (v.2.8.0) (Servant et al., 2015), the valid pairs were used to evaluate interaction strength among whole genome contigs. Lachesis (version 2e27abb) was used to cluster and anchor the contigs to the chromosomes using agglomerative hierarchical clustering (Burton et al., 2013).

To assess the integrity of the genome assembly, Illumina reads were mapped to the contigs using BWA (version 0.6.2). Genome completeness was also evaluated using the Core Eukaryotic Genes Mapping Approach (CEGMA) (Parra et al., 2007) and Benchmarking Universal Single-Copy Orthologs (BUSCO version 3) (Waterhouse et al., 2017).

Repetitive Sequence Annotation

Repetitive sequences in the genome assembly were identified through *ab initio* prediction and homology-based searches. Repeat families in the *S. constricta* genome were identified *de novo* using RepeatScout (version 1.0.5) and Repeat Modeler (version 1.0.11). Full length long terminal repeat (LTR) retrotransposons were also identified using LTR-finder (version 1.0.2) (Xu and Wang, 2007) with the parameters “-E -C.” Tandem repeats were screened using Tandem Repeats Finder (TRF version 4.09) (Benson, 1999) with the parameters “match = 2, mismatching penalty = 7, indel penalty = 7, match probability = 80, indel probability = 10, minimum alignment score = 50, maximum period size = 500.” The predicted repetitive sequences along with the RepBase database (Bao et al., 2015) were used for homology-based searches using Repeatmasker (version 4.5.0) with the parameters “-a -nolow -no_is -norna -parallel 32 -small -xsmall -poly -e ncbi -pvalue 0.0001” (Tarailo-Graovac and Chen, 2009).

Protein-Coding Gene Prediction and Annotation

Protein-coding genes were annotated based on *de novo* prediction, homology-based searches, and transcriptome-assisted methods. The protein sequences of 11 species (Table S5) were aligned to the genome assembly using TBLASTN with the parameters “-evalue 1e-5.” The gene structures were predicted using GeneWise (version 2.4.1) (Doerks et al., 2002). The Illumina RNA-seq reads of eight tissues and eight developmental stages were aligned to the genome assembly using Tophat (version 2.1.1) (Trapnell et al., 2009) and gene models were generated using Cufflinks (version 2.1.0) (Trapnell et al., 2010) with the parameter “-multi-read-correct.” The resulting GTF file and the PacBio Iso-seq transcripts were used to model gene structures with the PASA pipeline (version 2.0.2) (Haas et al., 2008). *De novo* gene prediction packages, including Augustus (version 2.5.5) (Stanke et al., 2006), glimmerHMM (version 3.01) (Majoros et al., 2004), SNAP (version 2006-07-28) (Leskovec and Susic, 2016), Geneid (version 1.4) (Parra et al., 2000), and Genscan (version 3.1) (Burge and Karlin, 1997) were used to predict genes with repeat-masked genome sequences. All gene model evidences were integrated using EvidenceModeler (version 1.1.1) (Haas et al., 2008).

Functional annotation was performed by aligning the predicted protein sequences to public databases, including KEGG, SwissProt and NCBI-NR, using BLASTP with an E-value threshold of 1e-5. InterProScan (v.4.8) (Jones et al., 2014) was also used to identify motifs and domains obtained from searching the Pfam, InterPro, and Gene Ontology (GO) databases.

Noncoding RNA Prediction and Annotation

The noncoding RNA genes, including miRNA, rRNA, snRNA, and tRNA, were annotated in the *S. constricta* genome. Transfer RNA were predicted by tRNAscan-SE 2.0 (Lowe and Chan, 2016) with eukaryote parameters. The miRNA and snRNA were screened using INFERNAL 1.1.2 against the Rfam database (version 14.1) (Kalvari et al., 2018) with default parameters (Table S7).

Gene Family Identification and Phylogenetic Analysis

Fifteen Eumetazoan species were selected for gene family analysis. All data were downloaded from either NCBI or Ensembl. The longest protein sequence was selected to represent a gene with multiple alternative splicing isoforms. Gene family clusters from all 16 species were first assigned using OrthoMCL (version 2.0.9) (Li et al., 2003) with an inflation value of 1.5. Gene family expansion and contraction under a maximum likelihood framework were analyzed using CAFE (version 3) (De Bie et al., 2006).

Phylogenetic inferences of the 16 species were conducted using 337 single-copy orthologs. Multiple sequences were aligned for the protein sequences of each ortholog gene using MUSCLE (version 3.8.31) (Edgar, 2004). The alignments of all the orthologs were then concatenated into a super alignment

matrix. RAxML (version 8.2.12) (Stamatakis, 2014) was used to infer the alignment matrix via a maximum likelihood method with the substitution model PROTGAMMAAUTO. Node support was provided by bootstrapping with 100 replicates. The time when species diverged was estimated using MCMCTree in the PAML package (version 4.7a) (Yang, 1997) with the parameters of “burn-in = 1,000, sample-number = 1,000,000, sample-frequency = 2.”

Transcriptomic Sequencing and Analysis

Embryos (eggs, four cells, blastulae, gastrulae), larvae (trochophore, D-shaped, umbo larvae and juvenile) and adults of *S. constricta* were collected for RNA sequencing. Eight tissues/organs (gill, liver, foot, mantle, adductor muscle, siphon, ovary, and testis) were dissected from three individuals. All the samples were flash frozen in liquid nitrogen and stored at -80°C until use. Full-length RNA was sequenced using mixed RNA from samples of the eight development stages and eight adult tissues. Three libraries with different insert lengths (e.g., 1–2k, 2–3k, and 3–6k) were constructed and sequenced on a PacBio Sequel platform. The full-length RNA transcriptome was analyzed using the SMRT Link software (v4.0.0). After redundant sequence clustering using the ICE algorithm, consensus sequence building using the DAGCon algorithm and polishing with Quiver. The transcripts were then polished and corrected using Illumina reads with LoRDEC (Salmela and Rivals, 2014). Finally, the corrected transcripts were further clustered with CD-HIT (version 4.6) (Li and Godzik, 2006).

To identify genes and pathways involved in the tolerance to salt fluctuations and ammonia. *S. constricta* were subjected to salt stress for 16 h under 22°C at low-salinity (3 psu), high-salinity (38 psu), and normal-salinity (25 psu). Three replicate tanks were set for each group with 10 individuals for each replicate. Gills were dissected from three individuals of each replicate and stored at -80°C . For the ammonia challenge, two groups of *S. constricta* were subjected to total ammonia nitrogen (TAN) at the concentration of 180 and 0.31 mg/L for 72 h (15°C , 23‰ sea water and pH 8.17). Three replicate tanks were set for each group with 80 individuals per tank. Gill and liver samples were dissected from three individuals of each replicate and stored at -80°C . Total mRNA were extracted using TRIzol (Omega Bio-Tek Inc., Norcross, GA, USA). RNA libraries were constructed and sequenced on an Illumina X Ten system. Raw reads were then filtered using Trimmomatic (Bolger et al., 2014) (Table S8). The clean reads were aligned to the indexed reference genome using Hisat2 version 2.0.5 (Kim et al., 2015). Read numbers mapped to each gene were counted using featureCounts version 1.5.0 (Liao et al., 2014). Differentially expressed genes (DEG) between different experiment groups were identified using the DESeq2 R package version 1.16.1 with an adjusted $P < 0.05$ (Love et al., 2014). Gene Ontology (GO) enrichment of DEG was analyzed using the cluster Profiler 3.4.4 R package with

corrected $P < 0.05$ being taken as significantly enriched GO terms (Yu et al., 2012).

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the NCBI under accession numbers WSYO00000000.1. The raw data of the PacBio, Illumina, and Hi-C sequencing for genome assembly are deposited at the NCBI Sequence Read Archive (SRA) with the accession numbers SRR10605359–SRR10605381, SRR10583055, and SRR10586373. The raw data of transcriptome sequencing are available at the NCBI SRA with the accession numbers SRR10753889–SRR10753895, SRR2162883, SRR2162887, SRR2162892, SRR2162895, SRR2162898, SRR2162902, SRR9937008–SRR9937013, SRR10097413–SRR10097424, SRR9959746–SRR9959754, and SRR9943679–SRR9943690.

ETHICS STATEMENT

All experimental procedures were approved by the Institutional Animal Care and Use Committee (IACUC) of Zhejiang Wanli University, China.

AUTHOR CONTRIBUTIONS

YD, ZL, and ZB conceived the project. HY, WR, and LH conducted the environmental stress and collected the samples. QZ, SW, JR, and LL performed the genome assembly, annotation, transcriptome analysis and other bioinformatics analysis. YD, JR, QZ, SW, and QX wrote and revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was financially supported by the National Key Research and Development Program of China (No. 2018YFD0901405), Zhejiang Major Program of Science and Technology (No. 2016C02055-9), Modern Agro-industry Technology Research System (No. CARS-49), Ningbo Major Project of Science and Technology (No. 2019B10005), and Taishan Scholar Project Fund of Shandong Province of China (to SW).

ACKNOWLEDGMENTS

This manuscript has been released as a pre-print at bioRxiv (Dong et al., 2019).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00664/full#supplementary-material>

REFERENCES

- Bao, W., Kojima, K., and Kohany, O. (2015). Repbase update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* 6:11. doi: 10.1186/s13100-015-0041-9
- Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580. doi: 10.1093/nar/27.2.573
- Bolger, A., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* 268, 78–94. doi: 10.1006/jmbi.1997.0951
- Burton, J., Adey, A., Patwardhan, R., Qiu, R., Kitzman, J., and Shendure, J. (2013). Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions. *Nat. Biotechnol.* 31, 1119–1125. doi: 10.1038/nbt.2727
- De Bie, T., Cristianini, N., Demuth, J., and Hahn, M. (2006). CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* 22, 1269–1271. doi: 10.1093/bioinformatics/btl097
- Doerks, T., Copley, R., Schultz, J., Ponting, C., and Bork, P. (2002). Systematic identification of novel protein domain families associated with nuclear functions. *Genome Res.* 12, 47–56. doi: 10.1101/gr.203201
- Dong, Y., Zeng, Q., Ren, J., Yao, H., Ruan, W., Lv, L., et al. (2019). The chromosomal-level genome assembly and comprehensive transcriptomes of Chinese razor clam (*Sinonovacula constricta*) with deep-burrowing life style and broad-range salinity adaptation. *BioRxiv [Preprint]* 735142. doi: 10.1101/735142
- Edgar, R. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- FAO (2019). *FAO Yearbook. Fishery and Aquaculture Statistics 2017*. Available online at: <http://www.fao.org/fishery/statistics/yearbook/en> (accessed June 24, 2020).
- Green, M., and Sambrook, J. (2012). *Molecular Cloning: A Laboratory Manual. 4th Edn. Vol. II*. New York, NY: Cold Spring Harbor Laboratory Press.
- 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:R7. doi: 10.1186/gb-2008-9-1-r7
- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031
- Kalvari, I., Nawrocki, E. P., Argasinska, J., Quinones-Olvera, N., Finn, R. D., Bateman, A., et al. (2018). Non-coding RNA analysis using the RFAM database. *Curr. Protoc. Bioinformatics* 62:e51. doi: 10.1002/cpbi.51
- Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317
- Leskovec, J., and Soscic, R. (2016). SNAP: A general purpose network analysis and graph mining library. *ACM Trans. Intell. Syst. Technol.* 8:8361. doi: 10.1145/2898361
- Li, L., Stoeckert, C. J. Jr., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503
- Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158
- Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8
- Lowe, T. M., and Chan, P. P. (2016). tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* 44, W54–W57. doi: 10.1093/nar/gkw413
- Majoros, W. H., Pertea, M., and Salzberg, S. L. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* 20, 2878–2879. doi: 10.1093/bioinformatics/bth315
- Mohri, H., Inaba, K., Ishijima, S., and Baba, S. A. (2012). Tubulin-dynein system in flagellar and ciliary movement. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* 88, 397–415. doi: 10.2183/pjab.88.397
- Morton, B. (1984). The functional morphology of *Sinonovacula constricta* with a discussion on the taxonomic status of the Novaculininae (Bivalvia). *J. Zool. Lond.* 202, 299–325. doi: 10.1111/j.1469-7998.1984.tb05085.x
- Niu, D., Wang, L., Sun, F., Liu, Z., and Li, J. (2013). Development of molecular resources for an intertidal clam, *Sinonovacula constricta*, using 454 transcriptome sequencing. *PLoS ONE* 8:e67456. doi: 10.1371/journal.pone.0067456
- Parra, G., Blanco, E., and Guigo, R. (2000). GeneID in *Drosophila*. *Genome Res.* 10, 511–515. doi: 10.1101/gr.10.4.511
- Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23, 1061–1067. doi: 10.1093/bioinformatics/btm071
- Peng, M., Liu, X., Niu, D., Ye, B., Lan, T., Dong, Z., et al. (2019). Survival, growth and physiology of marine bivalve (*Sinonovacula constricta*) in long-term low-salt culture. *Sci. Rep.* 9:2819. doi: 10.1038/s41598-019-39205-2
- Ran, Z., Li, Z., Yan, X., Liao, K., Kong, F., Zhang, L., et al. (2019). Chromosome-level genome assembly of the razor clam *Sinonovacula constricta* (Lamarck, 1818). *Mol. Ecol. Resour.* 19, 1647–1658. doi: 10.1111/1755-0998.13086
- Ruan, J., and Li, H. (2019). Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* 17, 155–158. doi: 10.1038/s41592-019-0669-3
- Salmela, L., and Rivals, E. (2014). LoRDEC: accurate and efficient long read error correction. *Bioinformatics* 30, 3506–3514. doi: 10.1093/bioinformatics/btu538
- Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C. J., Vert, J. P., et al. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* 16, 259. doi: 10.1186/s13059-015-0831-x
- Simakov, O., Marletaz, F., Cho, S. J., Edsinger-Gonzales, E., Havlak, P., Hellsten, U., et al. (2013). Insights into bilaterian evolution from three spiralian genomes. *Nature* 493, 526–531. doi: 10.1038/nature11696
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Res.* 34, W435–W439. doi: 10.1093/nar/gkl200
- Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinforma.* 25:4.10.1–4.10.14. doi: 10.1002/0471250953.bi0410s25
- Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111. doi: 10.1093/bioinformatics/btp120
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621
- Varshney, R. K., Chen, W., Li, Y., Bharti, A. K., Saxena, R. K., Schlueter, J. A., et al. (2011). Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat. Biotechnol.* 30, 83–89. doi: 10.1038/nbt.2022
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* 9:e112963. doi: 10.1371/journal.pone.0112963
- Wang, J. X., Zhao, X. F., Zhou, L. H., and Xiang, J. H. (1998). Chromosome study of *Sinonovacula constricta* (Bivalvia). *Oceanol. Limnol. Sinica* 29, 191–196.

- Wang, S., Zhang, J., Jiao, W., Li, J., Xun, X., Sun, Y., et al. (2017). Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nat. Ecol. Evol.* 1:0120. doi: 10.1038/s41559-017-0120
- Wang, W. X., and Xu, Z. Z. (1997). Larval swimming and postlarval drifting behavior in the infaunal bivalve *Sinonovacula constricta*. *Mar. Ecol. Prog. Ser.* 148, 71–81. doi: 10.3354/meps148071
- Waterhouse, R. M., Seppey, M., Simao, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., et al. (2017). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* 35, 543–548. doi: 10.1093/molbev/msx319
- Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* 35, W265–W268. doi: 10.1093/nar/gkm286
- Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* 13, 555–556. doi: 10.1093/bioinformatics/13.5.555
- Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omic*s 16, 284–287. doi: 10.1089/omi.2011.0118

Conflict of Interest: The 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.

Copyright © 2020 Dong, Zeng, Ren, Yao, Lv, He, Ruan, Xue, Bao, Wang and Lin. 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.