Acta Oceanol. Sin., 2018, Vol. 37, No. 8, P. 38–45
DOI: 10.1007/s13131-018-1232-2
http://www.hyxb.org.cn
E-mail: hyxbe@263.net

Transcriptome assembly of *Modiolus modiolus* and comparative analysis with *Bathymodiolus platifrons*

MENG Jie[1, 2, 3, 4], YANG Mei[5, 6], XU Fei[1, 3, 4], LI Xinzheng[5, 6, 7*], LI Li[1, 2, 3, 4*]

[1] CAS Key Laboratory of Experimental Marine Biology, Institute of Oceanology, Chinese Academy of Sciences (CAS), Qingdao 266071, China

[2] Laboratory for Marine Fisheries and Aquaculture, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266071, China

[3] National and Local Joint Engineering Laboratory of Ecological Mariculture, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China

[4] Center for Ocean Mega-Science, Chinese Academy of Sciences, Qingdao 266071, China

[5] Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China

[6] University of Chinese Academy of Sciences, Beijing 100039, China

[7] Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266071, China

Received 18 August 2017; accepted 13 March 2018

Abstract

The genetic basis for bivalves' adaptation and evolution is not well understood. Even few studies have focused on the mechanism of molluscan molecular evolution between the coastal intertidal zone and deep-sea environment. In our studies, we first conducted the transcritpome assembly of *Modiolus modiolus* mussels living in coastal intertidal zones. Also, we conducted transcriptome comparison analyses between *M. modiolus* and *Bathymodiolus platifrons* living in hydrothermal vents and cold methane/sulfide-hydrocarbon seeps. De novo assemblies of the clean reads yielded a total of 182 476 and 156 261 transcripts with N50 values of 1 769 and 1 545 in *M. modiolus* and *B. platifrons*. A total of 27 868 and 23 588 unigenes were identified, which also displayed the similar GO representation patterns. Among the 10 245 pairs of putative orthologs, we identified 26 protein-coding genes under strong positive selection (Ka/Ks>1) and 12 genes showing moderate positive selection (0.5<Ka/Ks<1). Most of those genes are predicted to be involved in stress resistance. Overall, our study first provides the transcriptomic database for *M. modiolus*. Transcriptome comparison illustrates the genome evolution between *M. modiolus* and *B. platifrons*, and provides an important foundation for future studies on these two species.

Key words: mollusc, transcriptome comparision, positive selection, stress adaptation

Citation: Meng Jie, Yang Mei, Xu Fei, Li Xinzheng, Li Li. 2018. Transcriptome assembly of *Modiolus modiolus* and comparative analysis with *Bathymodiolus platifrons*. Acta Oceanologica Sinica, 37(8): 38–45, doi: 10.1007/s13131-018-1232-2

1 Introduction

Bivalves, which were comprised of 30 000 extant species, are an important component of the ecosystem and biodiversity (Saavedra and Bachere, 2006). They were widely spread from the intertidal coastal areas to hydrothermal vents and cold seeps (Egas et al., 2012; Li et al., 2013). However, the genetic basis for their different adaptations is not well understood (Dame, 2011). *Modiolus modiolus* is a benthic marine organism, which filter feeds in near-shore habitats. As the important intertidal coastal habitat shellfish, its transcriptome sequence has not been conducted. *Bathymodiolus platifrons* are phylogenetically close to *M. modiolus* and belonged to the same family—Mytilidae. *Bathymodiolus platifrons* is a highly specialized animal inhabiting hydrothermal vent and cold seep ecosystems (Barry et al., 2002) and its genome sequences were also completed. These results have provided good data sets for the further genome comparision analysis. In our study, we first conduct the transcriptome analysis of *M. modiolus* which provided good data sources for further analysis. Also, we conducted the transcriptome comparision between *M. modiolus* and *B. platifrons* living in different environments, which provide valuable information to understand their different environmental adaptation mechanism.

In previous studies of marine invertebrate, the adaptive research has been conducted using single markers or candidate genes (Riesgo et al., 2012). Next-generation sequencing technology enabled analysis of large quantities of sequence data efficiently and cost-effectively (Schuster, 2008; Wang et al., 2009), which provided an efficient way to identify adaptive genes and

Foundation item: The Strategic Priority Research Program of the Chinese Academy of Sciences under contract No. XDB06010101; the Technological Innovation Project financially supported by Qingdao National Laboratory for Marine Science and Technology under contract No. 2015ASKJ02-03; Shandong Provincial Natural Science Foundation, China under contract No. ZR2016DQ13; the Earmarked Fund for Modern Agro-industry Technology Research System under contract No. CARS-48; the Taishan Scholars Climbing Program of Shandong; the project funded by China Postdoctoral Science Foundation.
*Corresponding author, E-mail: lixzh@qdio.ac.cn; lili@qdio.ac.cn

explain the adaptive evolution process. Recently, marine bivalves' genomic databases have been obtained, including *Crassostrea gigas* (Zhang et al., 2012), *B. platifrons*, *M. philippinarum* (Sun et al., 2017) and *Patinopecten yessoensis* (Wang et al., 2017), etc. Most of these publications mainly focused on the responses to multiple stressors including periodic hypoxia, hyposalinity, temperature fluctuations, and pollution. For evolutionary analysis, Zhao et al. (2014) conducted a comparative transcriptome analysis of two oysters, *C. gigas* and *C. hongkongensis*, and explained their adaptations and evolutionary mechanisms for dealing with hypo-osmotic conditions (Zhao et al., 2014). Wang et al. (2013) performed the first large-scale transcriptome comparison between the two scallop species, *Chlamys farreri* and *P. yessoensis*, and identified fast evolving genes, which played an important role in their speciation and local adaptation (Wang et al., 2013). For the evolutionary mechanisms of molluscans living in coastal intertidal zones and deep sea environments, there are also some studies. For example, Zheng et al. (2017) have conducted transcriptome comparision among *B. platifrons*, *B. manusensis*, *M. kurilensis* and *Perna viridis*. The results indicated that some immune responsive genes were positively selected and more highly expressed in the deep-sea mussels, which may be related with their endosymbiosis (Wang and Sun, 2017).

In this study, we first performed *de novo* transcriptome sequencing of *M. modiolus* using the Illumina sequence platform. Also, according to transcriptome comparison between *M. modiolus* and *B. platifrons*, 38 putative fast-evolving genes were identified, which may explain their different evolutionary mechanisms. This is the first time that the transcriptome of *M. modiolus* has been sequenced and will provide transcriptome resources for this mollusk. Additionally, in comparison with *B. platifrons*, we may use this transcriptome to find orthologous genes under potential positive selection between two species. This will help us to explain the different mechanisms for adaptation to hydrothermal vent and cold seep ecosystems versus coastal intertidal environments.

2 Materials and methods

2.1 *Sample collection and RNA isolation for Illumina sequencing*

Liquid nitrogen-frozen samples *B. platifrons* were provided by Li Xinzheng from the Institute of Oceanology, Chinese Academy of Science. These were originally sampled from a cold seep located at a depth of 996.9 m (27°47.44′N, 126°53.802 9′E). *Modiolus modiolus* specimens were collected in Dalian, Liaoning Province, China. These samples were collected and immediately frozen in liquid nitrogen and then transferred and stored at –80°C. For these two species, various tissues (including gills, mantles and adductor muscle) were mixed equally and ground in liquid nitrogen. Total RNA was isolated using Trizol reagent (Invitrogen). RNA purity, concentration, and integrity were checked using a NanoPhotometer® spectrophotometer (IMPLEN, CA, USA), Qubit® RNA Assay Kit (Life Technologies, CA, USA), and Bioanalyzer 2100 system (Agilent Technologies, CA, USA).

2.2 *Illumina sequencing*

RNA (3 μg) per sample was used for RNA sample preparation. Sequencing libraries were generated using Illumina TruSeq™ RNA Sample Preparation Kit (Illumina, San Diego, USA) and index codes were added to each sample. The mRNA was purified from total RNA and the fragmentation was carried out using divalent cations under elevated temperature. First strand cDNA was synthesized using random oligonucleotides and the second strand cDNA synthesis was performed using DNA Polymerase I and RNase H. Remaining overhangs were converted into blunt ends via exonuclease/polymerase activities and enzymes were removed. After adenylation of the 3′ ends of DNA fragments, Illumina PE adapter oligonucleotides were ligated to prepare for hybridization. Illumina PCR Primer Cocktail in a 10-cycle PCR was conducted to obtain DNA fragments with ligated adaptor molecules. Products were purified and quantified on the Agilent Bioanalyzer 2100 system. The clustering of the index-coded samples was performed using TruSeq PE Cluster Kit v3-cBot-HS (Illumina) on a cBot Cluster Generation System. After cluster generation, the library preparations were sequenced.

2.3 *Quality control and de novo assembly*

Clean data were obtained by removing reads from raw data, containing adapter sequences, or with more than 10% known nucleotides, or with low quality reads (more than 50% base with quality Qphred$\leqslant$5) using NGS QC toolkit package (Version 2.3). The Q20, Q30, GC-content and sequence duplication level of the clean data were calculated. Transcriptome assembly was accomplished using Trinity software (2.4.2669) and the parameters were set as "seqType=fq, min_contig_length=100, min_kmer_cov=2", with the rest being default parameter (Grabherr et al., 2011). All the sequences from two transcriptomes were then taken into further process of redundancy removing using CD-HIT-EST v4.6 62 with a sequence identity threshold of 99% in every 1 000 bp.

2.4 *Gene functional annotation*

Gene function annotation was conducted based on these databases: Pfam (protein family); Nt (NCBI non-redundant nucleotide sequences); Swiss-Prot (amanually annotated and reviewed protein sequence database); KOG/COG (clusters of orthologous groups of proteins); KO (KEGG ortholog database); and GO (gene ontology). Gene expression levels were estimated by RSEM for each sample. Clean data were mapped back onto the assembled transcriptome, and the read count for each gene was obtained (Li and Dewey, 2011). We applied a sensitive HMM scanning method on known Pfam functional protein domains to classify the gene families (Sun et al., 2017). Heatmap analysis was conducted with R script.

2.5 *GO and KEGG enrichment analysis*

GO enrichment analysis DEGs was implemented by the GOseq R packages based Wallenius non-central hyper-geometric distribution (Young et al., 2010). KOBAS software 2.0 was used for KEGG enrichment analysis (Mao et al., 2005). KEGG is a database resource for understanding high-level functions produced by genome sequencing and other high-throughput experimental technologies (Kanehisa et al., 2008).

2.6 *Ka/Ks analysis*

We used the BLAST-base (OrthoMCL) method (Li et al., 2003) to identify putative orthologs between the two species. We retained only those ortholog pairs that matched the same proteins to avoid the inclusion of paralogs. The CDSs of orthologous were aligned for further analysis. The ratio of the number of nonsynonymous substitutions per nonsynonymous site (Ka) to the number of synonymous substitutions per synonymous site (Ks) was used to test for positive selection using PAML-CODEML method (Yang, 2007). The rates of Ka to Ks between putatively orthologous coding regions were calculated based on the maximum-likelihood method using KaKs_Calculator 2.0. The orthologs with a Ks rate less than 0.1 were excluded from further analysis.

3 Results and discussion

3.1 *Overall de novo assembly of the transcriptome for M. modiolus and B. platifrons*

The cDNA libraries representing the different tissues (gills, mantles, and adductor) of *M. modiolus* and *B. platifrons* were constructed and then pooled for sequences. Sequencing of the tissue transcriptomes using the Illumina HiSeq 2000 platform in paired-end mode with a read length of 125 bp resulted in a total of 7.04 Gb and 5.97 Gb clean data in *M. modiolus* and *B. platifrons* (Table 1), respectively. After filtration, 46 904 896 and 39 817 358 clean reads were obtained and over 90% and 85% of them exceeded Q20 and Q30, indicating high quality of the sequencing data. *De novo* assemblies of the clean reads yielded a total of 182 476 and 156 261 transcripts with N50 values of 1 769 and 1 545 in *M. modiolus* and *B. platifrons*. ESTscan and BLAST search of the protein databases also resulted in the prediction of 137 763 and 119 880 coding transcripts (Table 1). The raw sequencing data have been submitted to NCBI under accession number SRR5043294. The statistics for the *de novo* assemblies and functional annotations are displayed in Table 1 and Fig. 1.

Modiolus modiolus has not been sequenced, which may be a bottleneck for further research into its ecology (Dinesen and Morton, 2014). In our study, we sequenced *M. modiolus* using transcriptome methods and more than two-thirds of the annotated unique sequences were matched to the known species. These results provided abundant sequence information for further studies of *M. modiolus*. *Bathymodiolus platifrons* transcriptome has already been published in previous studies (Wong et al., 2015). However, in our study, we obtained more than 1.6-fold numbers of transcripts for the mixture of different tissues used for sequencing, though we did not obtain more annotated transcripts. This may be because of the poor genomics database for marine bivalves. Finally, it should be noted that in hydrothermal vent and cold seep ecosystems, many bacteria are parasitic on *B. platifrons* (Nakamura-Kusakabe et al., 2016). In order to obtain clean sequences for *B. platifrons*, we also performed strict raw data quality control to remove contamination by pathogen genomes.

3.2 *Functional analysis comparison of dominant transcripts in two species*

We constructed functional dominant transcripts according to

Table 1. Summary of assembly and functional annotation of *B. platifrons* and *M. modiolus* transcriptome

Name	*B. platifrons*	*M. modiolus*
De novo assembly by Trinity		
Clean base	5.97 Gb	7.04 Gb
Clean reads	39 817 358	46 904 896
Q20/Q30/%	97.12/92.88	97.78/94.02
Total number of transcripts	156 261	182 476
Mean length of transcripts/bp	832	885
Transcripts size range/bp	201–41 720	201–32 099
Transcripts N50/bp	1 545	1 769
Mean length of unigene/bp	672	693
Unigene size range/bp	201–41 720	201–32 099
Unigene N50/bp	1 097	1 236
Functional annotation		
Total number of transcripts annotated by public databases	31 418 (26.2%)	37 371 (27.12%)
Total CDS predicted	119 880	137 763
NCBI non-redundant database (e-value<10^{-5})	23 823 (19.87%)	28 218 (20.48%)
SwissProt (e-value<10^{-5})	16 336 (13.62%)	18 438 (13.38%)
Annotated in PFAM	23 525 (19.62%)	27 869 (20.22%)
Annotated in GO	23 588 (19.67%)	27 868 (20.22%)
Annotated in KEGG	11 627 (9.69%)	13 078 (9.49%)

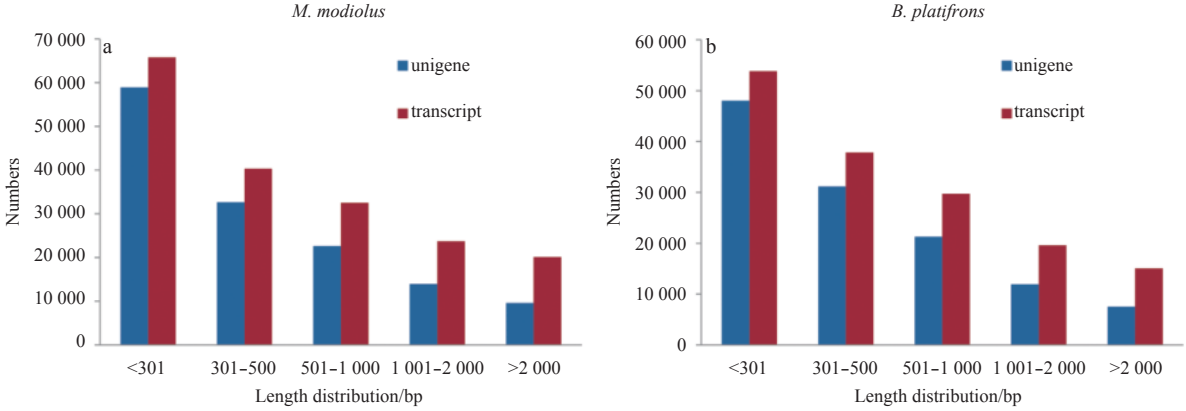

Fig. 1. The length distribution of contigs and coding sequences (CDSs) of *M. modiolus* and *B. platifrons*. Contigs were generated from *de novo* assembly of Illumina sequencing reads.

seven databases. Only 27.12% and 26.20% unique genes in *M. modiolus* and *B. platifrons* database were annotated in at least one database. A total of 27 868 and 23 588 transcripts were assigned with at least one GO term (Level 4) for 464 and 538 GO assignments in *M. modiolus* and *B. platifrons* (Figs S1 and S2). GO classification at Level 2 is shown in Figs S1 and S2. This wide distribution of GO terms further indicates that the transcripts represent a diverse range of functional classes. The top ten enriched GO terms are shown in Fig. 2a. From GO analysis, we can see that the most enriched GO terms in *B. platifrons* are related to the metabolism pathway, including the cellular macromolecule metabolic process, organic cyclic compound metabolic process, cellular nitrogen compound metabolic process, and cellular aromatic compound metabolic process. These enriched terms may be directly related with itsmethane/sulfide-hydrocarbon seeps and organic enrichment living environments. The same distribution was also observed in *M. modiolus*, which may be related with the increasingly polluted coastal environment.

Further, KEGG enrichment analysis was conducted with all annotated sequences. The analysis shows that 9 040 and 8 239 se-

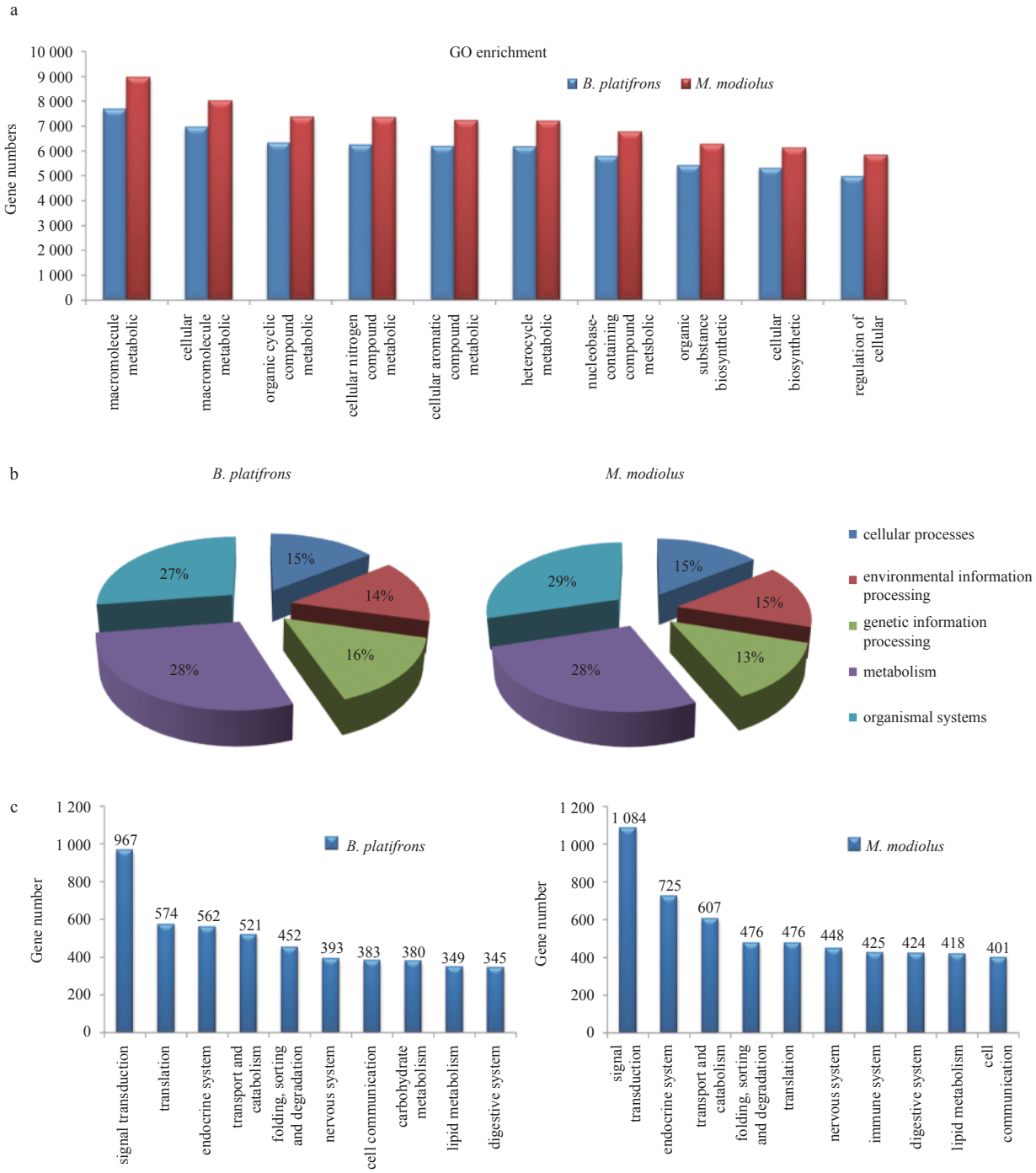

Fig. 2. The GO and KEGG annotation of the transcripts in *B. platifrons* and *M. modiolus*. a. The top ten enriched GO terms in these two species, b. the genes distributed in five KEGG metabolic processes, and c. the top ten enriched KEGG metabolism pathways in these two species.

quences are mapped to 32 metabolic pathways (Hierarchy2) in *M. modiolus* and *B. platifrons* (Figs S3 and S4). Among these, cellular processes (*B. platifrons* 27%, *M. modiolus* 29%) and metabolic processes (*B. platifrons* 28%, *M. modiolus* 28%), had the most unigenes (Fig. 2b). Moreover, in both two species, the "signal transduction pathway" (967 genes in *B. platifrons*, 1 084 genes in *M. modiolus*) was most significantly enriched. We propose that these pathways may be developed to deal with complicated environmental pressures (Fig. 2c). However, the "immune system" metabolic pathways varied between the two species, which may be related with their different living environments. *Bathymodiolus platifrons* is capable of acquiring chemo autotrophic bacteria as its major nutritional food source (Wong et al., 2015). It remains unclear how *Bathymodiolus* mussels distinguish pathogens from symbionts and how pathogens trigger immune responses (Bettencourt et al., 2007). However, *M. modiolus* living in coastal areas are exposed to constant challenge by invasive and pathogenic microbes. It has an open circulatory immune system with hemolymph serum containing diverse immune proteins, including soluble lectins, lysosomal enzymes and various antimicrobial peptides (Canesi et al., 2002). Overall, these annotations are useful to identify functional genes and specific biological processes in these two species.

3.3 *Identification of putative orthologs and analysis of Ka/Ks*

Aassessing the ratio of substitution rates at nonsynonymous and synonymous sites can help to identify genes under positive selection (Vitti et al., 2013). In our results, we searched for orthologs between these two species and found 10 245 putative orthologous genes according to OrthoMCL method (Li et al., 2003). Only 38 protein showed positive selection (dN/dS>0.5). A total of 26 genes (0.12%) had dN/dS>1 suggestive of signs of strong positive selection and 12 (0.079%) genes had 0.5<dN/dS<1 representing signatures of moderate positive selection (Kavembe et al., 2015) (Table 2). When Ka/Ks<0.1, the 8 661 orthologous genes are likely to be experiencing selection constraints.

3.4 *Analysis of candidate genes under positive and conserved selection*

The two species *B. platifrons* and *M. modiolus* have different inhabitations and possess specific adaptations to variable environmental factors between coastal intertidal area and deep sea, such as salinity, temperature, pH, heavy metals, and bacteria (Jones et al., 2006; Duperron et al., 2011). GO analysis was used to analyze the gene categories. The 38 genes were distributed among 15 different GO terms, most of which have physiological functions related to stress response (Fig. 3a). The enrichment GO terms included compound metabolic processes (nitrogen compound metabolic process, cellular aromatic compound metabolic process, and organic substance metabolic process), signal transduction processes (intracellular signal transduction and signal transduction process), and response-to-stimulus processes (pathogenesis and oxidation-reduction process). *Bathymodiolus platifrons* lives in the deep sea and experiences detrimental chemical pollutions, including heavy metal and methane. As a result, this species may have evolved abilities to adapt to the highly toxic chemical environment. Additionally, *B. platifrons* is capable of acquiring chemoautotrophic bacteria as its major nutritional food source and were involved in different immune responses (Fujiwara et al., 2000). The enriched GO terms may indicate that the different living environments have driven the evolution of these two species.

The orthologous genes with Ka/Ks<0.1 was considered to be conserved, and 8 661 orthologous pairs were identified between

Table 2. Genes showing signs of positive selection between *B. platifrons* and *M. modiolus*

Gene ID	Description	dN/dS
OG03779	probable 60S ribosomal protein L37-A	999.00
OG17598	putative fungistatic metabolite	999.00
OG17046	EF-hand calcium-binding domain-containing protein 1	36.03
OG16945	60S ribosomal protein	12.35
OG10206	protein PIF	1.99
OG05168	serum response factor-binding protein 1	1.74
OG18841	basement membrane-specific heparan sulfate proteoglycan core protein	1.39
OG05072	PR domain zinc finger protein 14	1.34
OG18790	cystatin C	1.00
OG17599	putative fungistatic metabolite	1.00
OG16869	heterogeneous nuclear ribonucleoprotein 27C	1.00
OG18737	clusterin	1.00
OG02089	nucleoplasmin-like protein ANO39	1.00
OG11629	low-density lipoprotein receptor-related protein 2	1.00
OG16709	myelin basic protein	1.00
OG16956	no annotation	1.00
OG18823	perlucin-like protein	1.00
OG03027	transcription factor jun-D	1.00
OG16702	cerebellin-2	1.00
OG18752	glial fibrillary acidic protein	1.00
OG16776	no annotation	1.00
OG16787	EF-hand calcium-binding domain-containing protein 1	1.00
OG18891	myelin proteolipid protein	1.00
OG18863	cofilin-2	1.00
OG18718	40S ribosomal protein S11	1.00
OG16693	myelin proteolipid protein	1.00

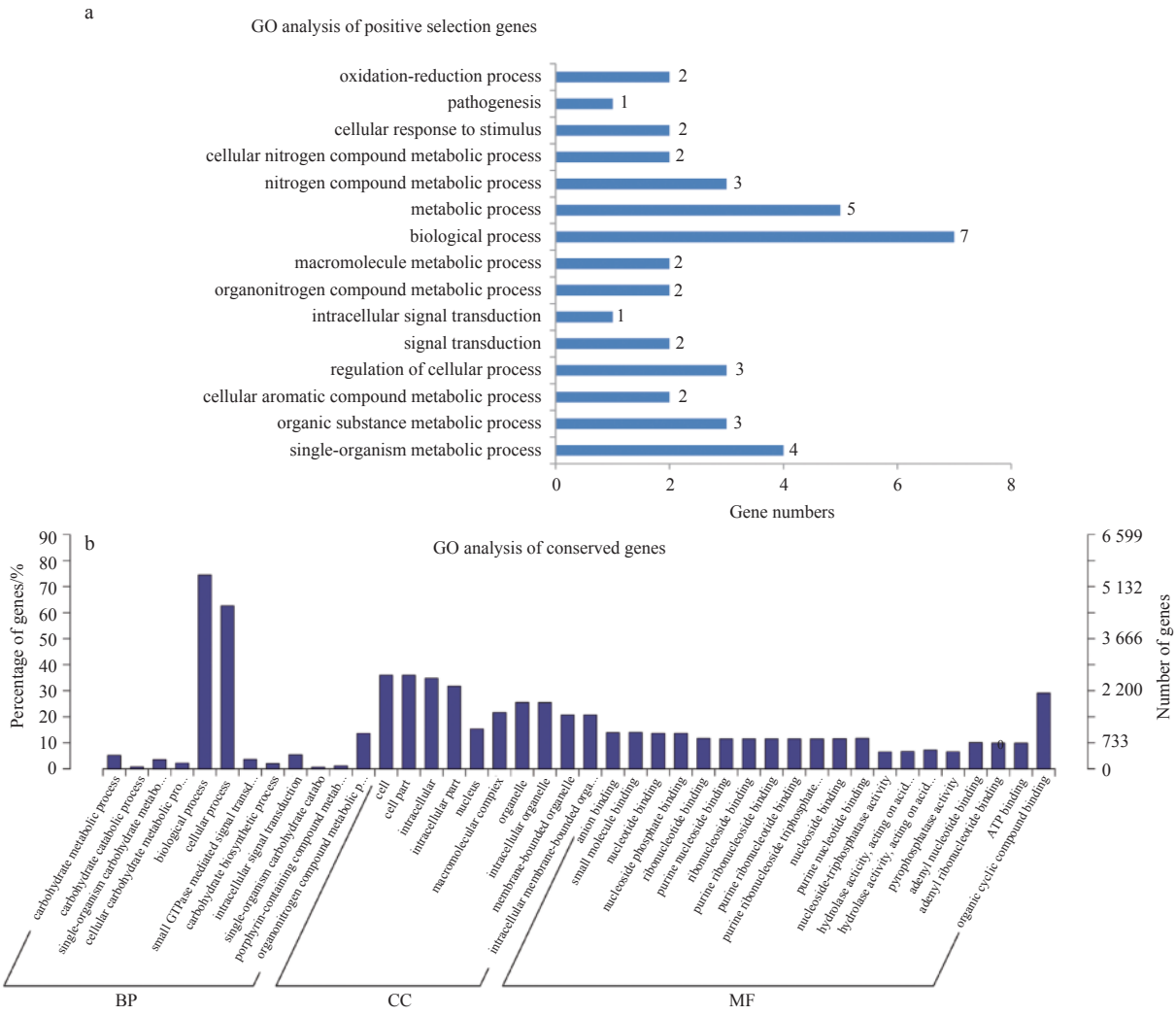

Fig. 3. The GO analysis of positively selected genes and conserved genes in *B. platifrons* and *M. modiolus*. a. GO analysis of positively selected genes and b. GO analysis of conserved genes.

the two species in our results. GO enrichment analysis revealed that 11, 10 and 26 terms were enriched in biological processes, cellular components and molecular function processes, respectively (P-value$\leqslant$0.05) (Fig. 3b). Among the biological processes, carbohydrate metabolism, catabolism, and biosynthetic processes were enriched. Among the molecular function process, nucleotide binding, ribonucleotide binding, and purine nucleotide binding processes were enriched. Among the cellular components, organelles and membrane bound-organelles were enriched. These metabolic pathways are primary processes in many species. For example, carbohydrate metabolism provides energy in nearly all known organisms, and the purine nucleotide binding process is related to nucleotide metabolism. These conserved metabolic pathways showed relative lower Ka/Ks values, indicating that they are subject to strong selection constraints.

3.5 *Immune responsive genes analysis*

One of the most extraordinary adaptation trait of *Bathymodiolin* mussels is their endosymbiosis (Jones et al., 2006). In previous studies, it has been revealed that *B. platifrons* has expanded and specie-specific immune responsive genes, which was the important genome basis for their adaptation under deep sea environment (Sun et al., 2017). However, for few species to be analyzed, it is still unknown whether this adaptation mechanism is species lineage-specific or is broadly conserved in other species. In our result, we conducted the transcriptome comparison analysis using four species, including two deep sea mussels, *B. platifrons* and *B. manusensis*, and two shallow-water mussels, *M. kurilensis* and *M. modiolus*. The transcriptome data of *B. manusensis* and *M. kurilensis* were obtained from previous studies (Zheng et al., 2017). We mainly focused on immune recognition receptors, which played important roles in initiation the immune responses (Toubiana et al., 2013). All these molecules were found in these four species, which were identified with previous studies (Fig. 4) (Zheng et al., 2017). Heatmap analysis of genes numbers revealed that two deep-sea bathymodiolin mussels and two shallow-water mussels clustered into two branches respectively. This further confirms the different immune systems between deep-sea and shallow water mussels may be related with their different living environments.

4 Conclusions

Our study represents the first transcriptome profile in *M. modiolus*. According to comparative transcriptome analysis with *B. platifrons*, our results provide new insights into the molecular mechanisms underpinning unique adaptations to coastal inter-

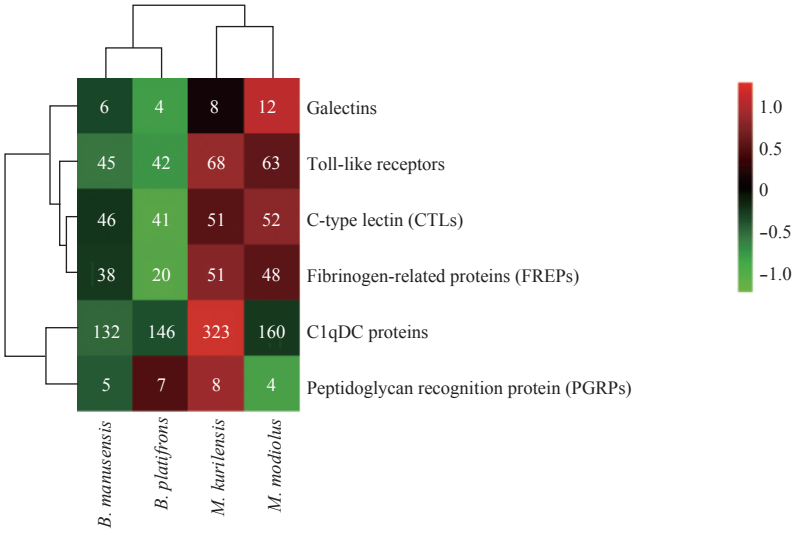

Fig. 4. Heatmap analysis of the representative immune recognition receptors numbers in four mussels. The genes numbers were identified from transcriptome data.

tidal environments or deep sea hydrothermal vent and cold seep environments. Selection analysis revealed that strong positive selection in genes is related to stress responses, indicating that the different living environments have driven the evolution of these two species. Our study provides transcriptomic resources for future genetic or genomic studies on *M. modiolus* and *B. platifrons*.

Acknowledgements

The authors thank the scientists and crew of the R/V *Kexue* for their assistance in specimen collecting.

Data Accessibility

The transcriptome data were submitted to NCBI database under SRR5043294 (PRJNA353979). All data underlying the findings are fully available without restriction.

References

Barry J P, Buck K R, Kochevar R K, et al. 2002. Methane-based symbiosis in a mussel, *Bathymodiolus platifrons*, from cold seeps in Sagami Bay, Japan. Invertebrate Biology, 121(1): 47–54

Bettencourt R, Roch P, Stefanni S, et al. 2007. Deep sea immunity: unveiling immune constituents from the hydrothermal vent mussel *Bathymodiolus azoricus*. Marine Environmental Research, 64(2): 108–127

Canesi L, Gallo G, Gavioli M, et al. 2002. Bacteria-hemocyte interactions and phagocytosis in marine bivalves. Microscopy Research and Technique, 57(6): 469–476

Dame R F. 2011. Ecology of Marine Bivalves: An Ecosystem Approach. 2nd ed. Boca Raton: CRC Press

Dinesen G E, Morton B. 2014. Review of the functional morphology, biology and perturbation impacts on the boreal, habitat-forming horse mussel *Modiolus modiolus* (Bivalvia: Mytilidae: Modiolinae). Marine Biology Research, 10(9): 845–870

Duperron S, Guezi H, Gaudron S M, et al. 2011. Relative abundances of methane- and sulphur-oxidising symbionts in the gills of a cold seep mussel and link to their potential energy sources. Geobiology, 9(6): 481–91

Egas C, Pinheiro M, Gomes P, et al. 2012. The Transcriptome of *Bathymodiolus azoricus* gill reveals expression of genes from endosymbionts and free-living deep-sea bacteria. Marine Drugs, 10(8): 1765–1783

Fujiwara Y, Takai K, Uematsu K, et al. 2000. Phylogenetic characterization of endosymbionts in three hydrothermal vent mussels: influence on host distributions. Marine Ecology Progress Series, 208: 147–155

Grabherr M G, Haas B J, Yassour M, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology, 29(7): 644–652

Jones W J, Won Y J, Maas P A Y, et al. 2006. Evolution of habitat use by deep-sea mussels. Marine Biology, 148(4): 841–851

Kanehisa M, Araki M, Goto S, et al. 2008. KEGG for linking genomes to life and the environment. Nucleic Acids Research, 36: D480–D484

Kavembe G D, Franchini P, Irisarri I, et al. 2015. Genomics of adaptation to multiple concurrent stresses: insights from comparative transcriptomics of a cichlid fish from one of earth's most extreme environments, the Hypersaline Soda Lake Magadi in Kenya, East Africa. Journal of Molecular Evolution, 81(3–4): 90–109

Li Bo, Dewey C N. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics, 12: 323

Li Li, Stoeckert C J Jr, Roos D S. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Research, 13(19): 2178–2189

Li Qi, Zhao Xuelin, Kong Lingfeng, et al. 2013. Transcriptomic response to stress in marine bivalves. ISJ-Invertebrate Survival Journal, 10(1): 84–93

Mao Xizeng, Cai Tao, Olyarchuk J G, et al. 2005. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. Bioinformatics, 21(19): 3787–3793

Nakamura-Kusakabe I, Nagasaki T, Kinjo A, et al. 2016. Effect of sulfide, osmotic, and thermal stresses on taurine transporter mRNA levels in the gills of the hydrothermal vent-specific mussel *Bathymodiolus septemdierum*. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology, 191: 74–79

Riesgo A, Andrade S C S, Sharma P P, et al. 2012. Comparative description of ten transcriptomes of newly sequenced invertebrates and efficiency estimation of genomic sampling in non-model taxa. Frontiers in Zoology, 9: 33

Saavedra C, Bachère E. 2006. Bivalve genomics. Aquaculture, 256(1–4): 1–14

Schuster S C. 2008. Next-generation sequencing transforms today's biology. Nature Methods, 5(1): 16–18

Sun Jin, Zhang Yu, Xu Ting, et al. 2017. Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. Nature Ecology & Evolution, 1(5): 0121

Toubiana M, Gerdol M, Rosani U, et al. 2013. Toll-like receptors and MyD88 adaptors in *Mytilus*: complete cds and gene expression levels. Developmental & Comparative Immunology, 40(2): 158–166

Vitti J J, Grossman S R, Sabeti P C. 2013. Detecting natural selection in genomic Data. Annual Review of Genetics, 47(1): 97–120

Wang Zhong, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics, 10(1): 57–63

Wang Shan, Hou Rui, Bao Zhenmin, et al. 2013. Transcriptome sequencing of Zhikong scallop (*Chlamys farreri*) and comparative transcriptomic analysis with Yesso scallop (*Patinopecten yessoensis*). PLoS One, 8(5): e63927

Wang Hailiang, Sun Li. 2017. Comparative metagenomics reveals insights into the deep-sea adaptation mechanism of the microorganisms in Iheya hydrothermal fields. World Journal of Microbiology Biotechnology, 33: 86

Wang Shi, Zhang Jinbo, Jiao Wenqian, et al. 2017. Scallop genome provides insights into evolution of bilaterian karyotype and development. Nature Ecology & Evolution, 1(5): e0120

Wong Y H, Sun Jin, He Lisheng, et al. 2015. High-throughput transcriptome sequencing of the cold seep mussel *Bathymodiolus platifrons*. Scientific Reports, 5: 16597

Yang Ziheng. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution, 24(8): 1586–1591

Young M D, Wakefield M J, Smyth G K, et al. 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biology, 11(2): R14

Zhang Guofang, Fang Xiaodong, Guo Ximing, et al. 2012. The oyster genome reveals stress adaptation and complexity of shell formation. Nature, 490(7418): 49–54

Zhao Xuelin, Yu Hong, Kong Lingfeng, et al. 2014. Comparative transcriptome analysis of two oysters, *Crassostrea gigas* and *Crassostrea hongkongensis* provides insights into adaptation to hypo-osmotic conditions. PLoS One, 9(11): e111915

Zheng Ping, Wang Minxiao, Li Chaolun, et al. 2017. Insights into deep-sea adaptations and host-symbiont interactions: a comparative transcriptome study on *Bathymodiolus* mussels and their coastal relatives. Molecular Ecology, 26(19): 5133–5148

Supplementary information:

Fig. S1. GO enrichment analysis of annotated genes in *M. modiolus* (GO level 2).

Fig. S2. GO enrichment analysis of annotated genes in *B. platifrons* (GO level 2).

Fig. S3. KEGG enrichment analysis of annotated genes in *M. modiolus*. A represents "cellular processes", B "environmental information processing", C "genetic information processing", D "metabolism processing", and E "organismal systems".

Fig. S4. KEGG enrichment analysis of annotated genes in *B. platifrons*. A represents "cellular processes", B "environmental information processing", C "genetic information processing, D "metabolism processing", and E "organismal systems".

Table S1. GO enrichment analysis of annotated genes in *M. modiolus*.

Table S2. GO enrichment analysis of annotated genes in *B. platifrons*.

The supplementary information is available online at www.hyxb.org.cn/aosen/ch/index.aspx. The supplementary information is published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.