Transcriptome analysis of the freshwater pearl mussel, *Hyriopsis cumingii* (Lea) Uusing Illumina paired-end sequencing to identify genes and markers

Zhang A. J.[1,2,4]; Liu S. L.[2,4]; Zhu J. Y.[1];. Gu Z. M.[2,4]; Zhou Z. M.[2,4]; Zhang G. F.[3]; Lu K. H. [1*]

Received: June 2014 Accepted: December 2014

Abstract

The transcriptome of triangle sail mussel *Hyriopsis cumingii* (Lea) using Illumina paired-end sequencing technology was conducted and analyzed. Equal quantities of total RNA isolated from six tissues, including gonads, hepatopancreas, foot, mantel, gills and adductor muscles, were pooled to construct a cDNA library. A total of 58.09 million clean reads with 98.48 % Q20 bases were generated. Clustering and assembly of these reads produced a non-redundant set of 92,347 unigenes with an average length of 1,150.61 bp. 11,174 unigenes were involved in the molecular function, cellular component and biological process categories by GO (Gene Ontology) analysis. Potential genes and their functions were predicted by KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway mapping and COG (Cluster of Orthologous Groups of proteins) analysis. More than 8 putative genes of interest involved in sex determination/differentiation were identified. Furthermore, 8,014 SSRs and 38,288 SNPs were identified in this transcriptome dataset.

Keywords: *Hyriopsis cumingii*, Illumina paired-end sequencing, Transcriptome, Sex determination, Molecular marker

1-Key Laboratory of Applied Marine Biotechnology, Ministry of Education, School of Marine Sciences, Ningbo University, Ningbo 315211, China

2-Zhejiang Institute of Freshwater Fisheries, Huzhou 313001, China

3-JinHua Polytechnic College, JinHua 321000, China

4- Agriculture Ministry Key Laboratory of Healthy Freshwater Aquaculture, Huzhou 313001, China

* Corresponding author's Email: lukaihong@nbu.edu.cn

[Downloaded from jifro.ir on 2026-06-25]

[DOR: 20.1001.1.15622916.2016.15.1.38.2]

Introduction

The triangle sail mussel, *Hyriopsis cumingii* is the most important mussel in commercial freshwater pearl production of China that has been exported to many countries since the 1970s because of its domestic rapid growth rate, high pearl yields and long research history (Qiu *et al*., 1998; Dong *et al.*, 1999; Fei *et al*., 2004; Li *et al.*, 2004; Hong *et al*., 2006). Nevertheless, the quality of pearls is declining with the scare of mussel culture enlarging rapidly and pearl yields accelerating sharply. To date, a lot of research on this triangle sail mussel has made considerable progress in the improvement of the pearl quality, and it's considered that overexploitation and change of water quality are the main factors that have led to the decline in pearl quality nowadays (Luo *et al.*, 2007). Except for these external factors, the characteristics of *H. cumingii* itself, such as age and sex, also affect the pearl quality (Xu *et al.*, 2011). As for the sex, though some research showed that there were significant differences in the total weights, grain weights and grain sizes of the pearls between female and male mussels at about 3- and 4-year-old (Zhao *et al.*, 2013), sex determination of juvenile mussels especially less than 1-year-old is very difficult and few studies on sex chromosomes and the molecular mechanism involved in sex determination/differentiation in this species have been performed. In recent years, RNA sequencing (RNA-Seq) has been used for advanced research in many areas, including resequencing, microRNA expression profiling, DNA methylation and de novo transcriptome sequencing, as a powerful and cost-efficient tool (Hale *et al.*, 2009; Meyer *et al.*, 2009; Wang *et al.*, 2010; Gao *et al.*, 2013; Liao *et al*., 2013). There is no denying that RNA-Seq greatly expands our understanding of the complexity of gene expression, regulation and networks in model and non-model organisms. There are three main sequencing platforms for RNA-Seq at present, the 454 system, Illumina Genome Analyzer and Applied Biosystems' SOLiD (Collins *et al*., 2008; Parchman *et al*., 2010). For pearl oysters or mussels, previous RNA-Seq has focused primarily on marine species using Roche-454 pyro sequencing or Illumina sequencing technology (Joubert *et al*., 2010; McGinty *et al*., 2012), such as *Pinctada martensi, P. margaritifera, P. maxima* and *Pteria penguin*. As for freshwater mussels, to the best of our knowledge, only tissues secreting purple and white nacre in *H. cumingii* have been sequenced using Roche-454 massive parallel pyro sequencing technology recently, for the purpose of identifying genes involved in the nacre coloring (Bai

[Downloaded from jifro.ir on 2026-06-25]

[DOR: 20.1001.1.15622916.2016.15.1.38.2]

et al., 2013). Although Illumina Genome Analyzer sequencing technology has a shorter read length compared with the 454 system, experimental advances are likely to increase the applicability of this sequencing, for the reason that it can compensate for the lack of a reference genome during de novo sequence assembly and the costs are lower.

In this study, we utilized Illumina paired-end sequencing technology to characterize the transcriptome of *H. cumingii*, following a cDNA library preparation after pooling total RNA from six tissues and organs, including gonads, hepatopancreas, foot, mantel, gills and adductor muscles, and provided genes of the molecular mechanisms involved in sex determination/differentiation for future studies. This transcriptome dataset provides a valuable and exhaustive resource for functional genomics and biological research in *H. cumingii*. The SSR and SNP markers identified here provide material basis for future quantitative trait loci (QTL) analysis and genetic linkage, and will be necessary for accelerating breeding programs to improve pearl quality in the future.

Materials and methods

Tissue samples

Live freshwater pearl mussels used in this study were collected from Weiwang Pearl Farm of Jinhua, Zhejiang Province, China, which reproduced in May and reached an average shell length of 62 mm in December 2013. Nine mussels from three selected populations were sampled, and six tissue samples were isolated from each mussel for RNA extraction, including gonads, hepatopancreas, foot, mantel, gills and adductor muscles. The tissue samples were immediately stored in liquid nitrogen at −80°C.

RNA extraction and quality controls

For Illumina paired-end sequencing, total RNA of each tissue sample was extracted from these materials using TRIzol Reagent (Invitrogen, USA) according to the manufacturer's protocol. The quality and quantity of total RNA was then checked using gel electrophoresis system (Bio-Rad, USA), Spectrophotometer (Thermo, USA) and Bioanalyzer (Agilent, USA). Only RNA samples with a 260 of 280 ratio from 1.8 to 2.0 and a 260 of 230 ratio from 2.0 to 2.5 were used for the next cDNA library preparation.

Library construction and high-throughput sequencing

We combined equivalent amounts of total RNA from each tissue mixture (≥ 10μg total RNA) and delivered it to Shanghai Majorbio Bio-pharm Biotechnology Co., Ltd. (Shanghai, China) for the next steps. Briefly, first, the library was constructed by the Truseq^{TM} RNA sample prep kit (Illumina, USA) following the manufacturer's

[Downloaded from jifro.ir on 2026-06-25]

[DOR: 20.1001.1.15622916.2016.15.1.38.2]

protocol step by step after the fragmentation of total RNA. Then, Adaptor-ligated fragments were separated by size on a 2% TAE-agarose gel (Certified Low-Range Ultra Agarose, Bio-Rad), followed by the excising of desired range of cDNA fragments (200 ± 25 bp) from the gel. Next, the concentration of obtained cDNA was examined by TBS 380 Fluorometer (Invitrogen, USA). After enriching and amplifying the selectively cDNA fragments using the cBot Truseq PE Cluster Kit v3-cBot-HS (Illumina, USA), the cDNA library was sequenced on a PE flow cell using the Illumina HiSeq^{TM} 2000.

Data filtering and de novo assembly

After transformation the image data using base calling, raw reads data were obtained. Next, clean reads were available by cleaning and quality checks to the former raw reads. De novo assembly of the clean reads was performed using Trinity (http://trinityrnaseq.sourceforge.net/) with default K-mers = 25. Contigs without ambiguous bases were obtained by conjoining the K-mers in an unambiguous path. Then, the reads were mapped back to contigs using Trinity to construct unigenes with the paired-end information. After then, the contigs were connected with Trinity, and sequences that could not be extended on either end were obtained. Such sequences are defined as unigenes. Finally, the overlapping unigenes from the libraries were assembled into a continuous sequence using the overlapping ends of different sequences, and redundant sequences were removed to acquire non-redundant unigenes as long as possible using the TIGR Gene Indices Clustering (TGICL) tools (Pertea *et al*., 2003). The parameters were set at a similarity of 94 %.

Gene annotation

Unigenes were aligned with sequences in the National Center for Biotechnology Information (NCBI) non-redundant protein (Nr) database (http://www.ncbi.nlm.nih.gov/), the String database (http://string-db.org/), the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database (http://www.genome.jp/kegg) using BLASTx with an E value threshold of $<$ le^{-5}, through which the possible functional classifications and molecular pathways could also be predicted. The Blast2GO program (Conesa *et al*., 2005) was used to obtain GO annotation of the unigenes based on BLASTx hits according to the GO database (http://www.geneontology.org/) with an E value threshold of $<le^{-5}$. In addition, unigenes were aligned with the NCBI nucleotide (Nt) databases (http://blast.ncbi.nlm.nih.gov) using

[Downloaded from jifro.ir on 2026-06-25]

[DOR: 20.1001.1.15622916.2016.15.1.38.2]

BLASTn with an E value of $<le^{-5}$.

Molecular markers detection and primer design

Potential SSR markers were detected among the 92,347 unigenes using the Msatcommander software (http://code.google.com/p/msatcommander/). The parameters were adjusted for identification of perfect mono-, di-, tri-, tetra-, penta-, and hexa-nucleotide motifs with a minimum of 10, 6, 4, 4, 4 and 4 repeats, respectively. Primer pairs were designed and tagged using primer 3 (Rozen and Skaletsky, 2000). Meanwhile, candidate SNP markers were detected using the samtools software (http://samtools.sourceforge.net/) and VarScan software (http://varscan.sourceforge.net/).

Results

General characteristics

Using an Illumina paired-end sequencing platform, a total of 61.92 million sequences with 94.20 % Q20 bases (base quality more than 20) were yielded. After eliminating adapter sequences and filtering out the low-quality reads, a total of 58.09 million clean reads with 98.48 % Q20 bases were generated from the cDNA libraries. The total length of the clean reads was about 5,679.46 million bp. The GC content (ratio of guanine and cytosine) of clean reads was also determined, giving rise to an overall GC content of 41.64 %, indicating a low GC content in the cDNA of *H. cumingii*. Based on the clean reads, a total of 92,347 non-redundant unigenes with average lengths of 1,150.61 bp, were assembled from the library. The unigenes size distributions for the library were consistent, which implied that the Illumina sequencing solution was reproducible and reliable. The unigenes fell between 351 and 28,861 bp in length. Among these unigenes, the length of 26,387 (28.57 %) ranged from 401 to 600 bp, 21,915 (23.73 %) ranged from 601 to 1,000 bp, and 32,806 (35.53 %) were more than 1,000 bp in length (Fig. 1). Open reading frames (ORFs) were extracted using the Trinity software. As a result, 36,848 unigenes were extracted ORFs, while 55,499 were not.

Comparative analysis of unigenes

For the annotation of non-redundant unigenes, a sequence similarity search was conducted against the NCBI non-redundant (Nr), String and nucleotide (Nt) databases using the BLASTx, BLASTx and BLASTn software with a cutoff E value of le^{-5}, respectively. As a result, 32.53 % (30,043) and 8.50 % (7,850) of the consensus sequences showed homology with proteins in the Nr and String databases, respectively, and 0.35 % (321) of the unigenes showed similarity to Nucleotide sequences in the Nt database.

[Downloaded from jifro.ir on 2026-06-25]

[DOR: 20.1001.1.15622916.2016.15.1.38.2]

In addition, 53.56 % (16,090) of the mapped sequences of the Nr database showed highest match with *Crassostrea gigas*, followed by *Capitella teleta* with 7.69 % (2,311), while only 0.35 % (104) unique sequences matched the registered sequences of *H. cumingii* (Table 1). This is possibly due to the relative lack of registered information in *H. cumingii*.

Functional annotation

Gene Ontology (GO) could provide a structured and controlled vocabulary concept for describing genes and their products in three categories: cellular component, molecular function and biological process (Ashburner *et al*., 2000). Based on Nr databases, 11,174 unigenes were assigned to one or more ontologies. Under the cellular component category (Fig. 2), the two majority (22.47 %) of classifications were involved in cell and cell part with the same number of unigenes (5,181), followed by organelle (3,670; 15.92 %) and organelle part (2,168; 9.40 %). Under the molecular function category (Fig. 2), 5,772 (45.27 %) unigenes were assigned to binding, followed by catalytic activity (5,262; 41.27 %), and transporter activity (494; 3.87 %). For the biological process category (Fig. 2), 17.78 % was categorized as 'cellular process', 14.82 % as 'metabolic process', 12.07 % as 'single-organism process', and 7.87 % as 'biological regulation'.

As we all known, COG analysis and KEGG pathway analysis are helpful for predicting potential genes and their functions at a whole transcriptome level. The predicted metabolic pathways, together with the COG analysis, are useful for further investigations of gene function in future studies. Based on sequence homology, 6,340 unigenes had a COG classification. These unigenes were classified into 25 COG categories (Fig. 3). The most common category was 'general function prediction only' with 1,422 (22.43 %) unigenes, followed by transcription (536; 8.45 %), signal transduction mechanisms (517; 8.15 %), replication, recombination and repair (497; 7.84 %), posttranslational modification, protein turnover and chaperones (491; 7.74 %), translation, ribosomal structure and biogenesis (341; 5.38 %) and Cytoskeleton (313; 4.94 %) (Fig. 3). According to the KEGG results, 11,317 unigenes were mapped onto 320 predicted KEGG metabolic pathways. The numbers of unique sequences mapped to various pathways ranged from 1 to 2003. These pathways were divided into six groups (Fig. 4), of which metabolism was the largest group (8,569), followed by human diseases (8,181), organism systems (5,867), cellular processes (3,351), environmental information processing (3,004) and genetic information processing (2,312).

[Downloaded from jifro.ir on 2026-06-25]

[DOR: 20.1001.1.15622916.2016.15.1.38.2]

Table 1. Species distribution of the BLASTX results against nr database (the former sixteen).

Species	Isogenes hit number	Percentage（%）
Strongylocentrotus purpuratus	740	2.46
Nematostella vectensis	327	1.09
Xenopus (Silurana) tropicalis	244	0.81
Danio rerio	219	0.73
Ixodes scapularis	211	0.70
Tribolium castaneum	186	0.62
Daphnia pulex	178	0.59
Oryzias latipes	170	0.57
Anolis carolinensis	164	0.55
Takifugu rubripes	137	0.46
Amphimedon queenslandica	131	0.44
Maylandia zebra	126	0.42
Hydra vulgaris	124	0.41
Xenopus laevis	120	0.40
Gallus gallus	114	0.38
Hyriopsis cumingii	104	0.35

Table 2: Selected genes of interest for sex determination/differentiation in the *H. cumingii* transcriptome.

Gene name	Hits	Similarity (%)	Value
DMRT	2	61-100	170-229
SOX2	1	100	313
SOX9	3	44-66	87-465
SOX14	1	77	361
P450	59	45-100	120-558
ZFY1	2	62-83	192-727
Testis-specific	19	60-975	63-94
sex-regulated protein janus-A	1	80	136

DMRT: mab-3-related transcription factor, SOX : SRY-related HMG-domain containing transcription factor, p450: cytochrome P450 aromatase, ZFY : Zinc finger Y-chromosomal protein. 1

[Downloaded from jifro.ir on 2026-06-25]

[DOR: 20.1001.1.15622916.2016.15.1.38.2]

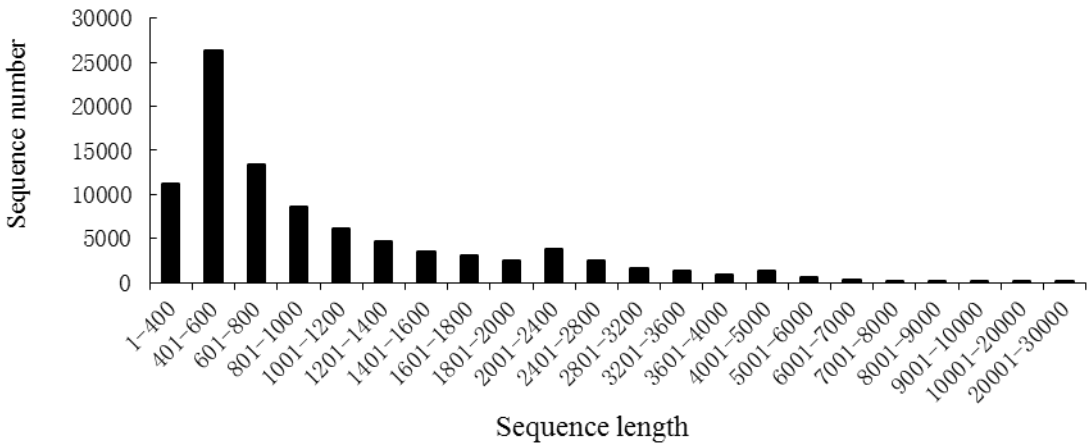

Figure 1: Sequence length distributions of unigenes. In total, 92,347 non-redundant unigenes were assembled, and fell between 351 to 28,861 bp in length.

In the metabolism pathway, 8,569 unigenes were classified into twelve subgroups, of which most were mapped to Global and overview maps (3,165), Carbohydrate metabolism (991), amino acid metabolism (983), lipid metabolism (779), Glycan biosynthesis and metabolism (555) and Xenobiotics biodegradation and metabolism (500). In addition, the human diseases, organismal systems, genetic information processing, cellular processes and environmental information processing pathways were classified into 11, 9, 4, 4 and 3 subgroups, respectively.

Putative molecular markers

A total of 8,014 SSRs were identified from the transcriptomic dataset. Of these, the most abundant type of repeat motif was mononucleotide (37.02 %), followed by tri-nucleotide (29.35 %), di-nucleotide (28.45 %), tetra-nucleotide (4.99 %), penta-nucleotide (0.12 %) and hexa-nucleotide (0.06 %) repeat units. The frequencies of EST-SSRs with different numbers of repeat units were calculated. SSRs with 10 repeat motifs (25.37 %) were the most common, followed by 4 repeat motifs (21.54 %), 6 repeat motifs (11.59 %), 11 repeat motifs (8.90 %), 5 repeat motifs (7.79 %) and 7 repeat motifs (7.44 %). Among these, A/T (33.53 %) represented the dominant type, followed by AC/GT (14.89 %), AG/CT (8.88 %), AAC/GTT (7.99 %), AAT/ATT (5.99 %), ATC/GAT (5.17 %) and AT/CG (4.60 %). The frequency of the remaining 47 types of motifs accounted for 18.95 %. Additionally, 3,790 primer pairs were obtained.

[Downloaded from jifro.ir on 2026-06-25]

[DOR: 20.1001.1.15622916.2016.15.1.38.2]

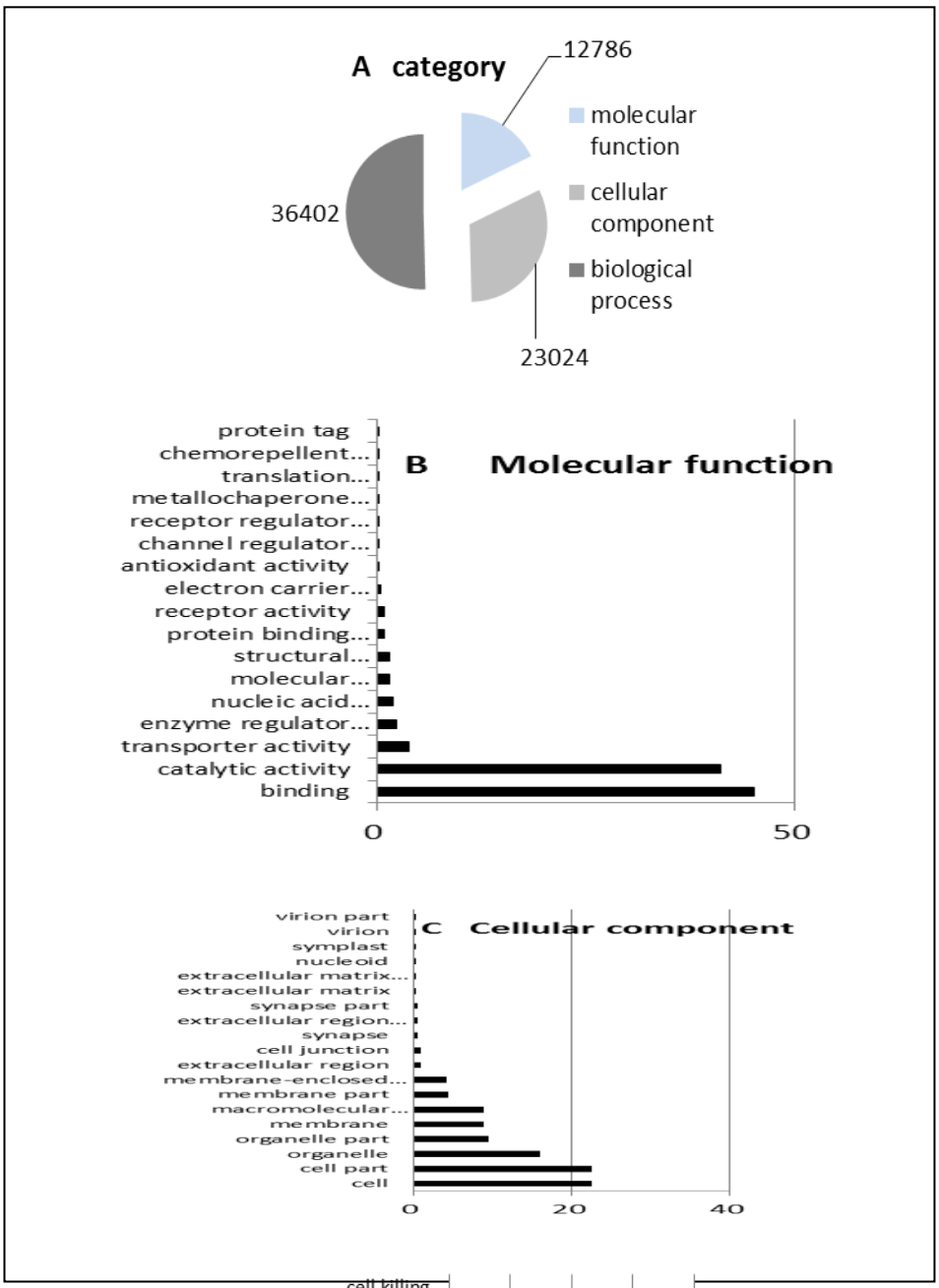

Figure 2: Gene Ontology classifications s in *H. cumingii*. The results are summarized in three ical process, cellular component and molecular function 5.27 %) (B), cell (5,181, 22.47 %) and cell part (5,181, 22.47 ess (6471, 17.78 %) (D) were the majority classifications in t individually.

[Downloaded from jifro.ir on 2026-06-25]

[DOR: 20.1001.1.15622916.2016.15.1.38.2]

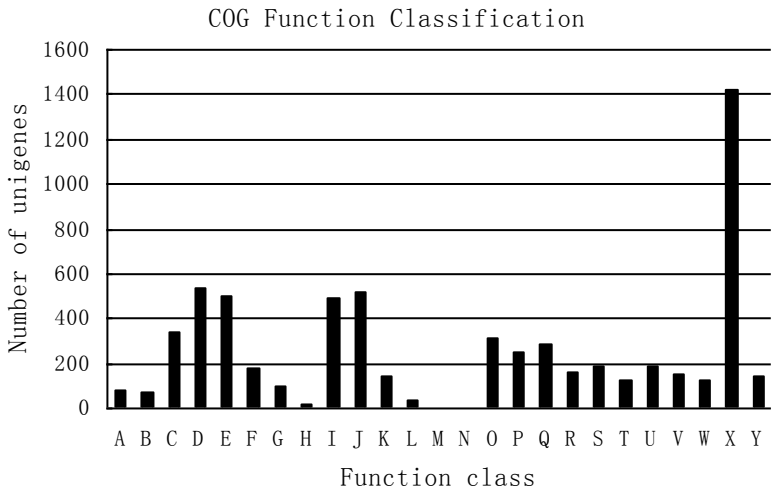

Figure 3: Clusters of orthologous groups (COG) classifications of *H. cumingii* unigenes A: RNA processing and modification; B: Chromatin structure and dynamics; C: Translation, ribosomal structure and biogenesis; D: Transcription; E: Replication, recombination and repair; F: Cell cycle control, cell division, chromosome partitioning; G: Cell wall/membrane/envelope biogenesis; H: Cell motility; I: Posttranslational modification, protein turnover, chaperones; J: Signal transduction mechanisms; K: Intracellular trafficking, secretion, and vesicular transport; L: Defense mechanisms; M: Extracellular structures; N: Nuclear structure; O: Cytoskeleton; P: Energy production and conversion; Q: Amino acid transport and metabolism; R: Nucleotide transport and metabolism; S: Carbohydrate transport and metabolism; T: Coenzyme transport and metabolism; U: Lipid transport and metabolism; V: Inorganic ion transport and metabolism; W: Secondary metabolites biosynthesis, transport and catabolism; X: General function prediction only; Y: Function unknown.

SNPs were identified from alignments of unigene assembly. In total, 38,288 SNPs were identified, of which 24,546 were putative transitions (Ts) and 13,742 were putative transversions (Tv), giving a mean In: Ts:Tv ratio of 1.79:1 across the transcriptome of *H. cumingii*. On average, one SNP was found every 2.78 kb in the unigenes. The A/G, C/T SNP types were the most common. In contrast, C/G types were the smallest SNP types.

Discussion

General characteristics and functional annotation

Recently, high-throughput techniques have been widely used to examine physiological processes and applications in diverse fields of biology (Wang *et al*., 2012; Liao *et al*., 2013). Roche-454 massive parallel pyrosequencing has been reported in *H. cumingii* for the purpose of identifying genes involved in the nacre coloring (Bai *et al*., 2013).

[Downloaded from jifro.ir on 2026-06-25]

[DOR: 20.1001.1.15622916.2016.15.1.38.2]

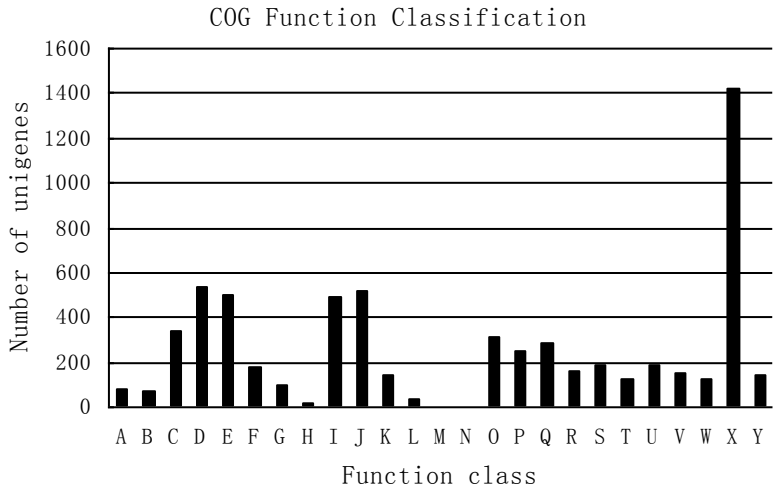

Figure 3: Clusters of orthologous groups (COG) classifications of *H. cumingii* unigenes A: RNA processing and modification; B: Chromatin structure and dynamics; C: Translation, ribosomal structure and biogenesis; D: Transcription; E: Replication, recombination and repair; F: Cell cycle control, cell division, chromosome partitioning; G: Cell wall/membrane/envelope biogenesis; H: Cell motility; I: Posttranslational modification, protein turnover, chaperones; J: Signal transduction mechanisms; K: Intracellular trafficking, secretion, and vesicular transport; L: Defense mechanisms; M: Extracellular structures; N: Nuclear structure; O: Cytoskeleton; P: Energy production and conversion; Q: Amino acid transport and metabolism; R: Nucleotide transport and metabolism; S: Carbohydrate transport and metabolism; T: Coenzyme transport and metabolism; U: Lipid transport and metabolism; V: Inorganic ion transport and metabolism; W: Secondary metabolites biosynthesis, transport and catabolism; X: General function prediction only; Y: Function unknown.

In our research, a relatively complete transcriptome of *H. cumingii* using de novo transcriptome sequencing was performed, to identify sex determination/differentiation genes. As a result, a total of 92,347 non-redundant unigenes with average lengths of 1,150.61 bp were assembled, and a significant number of putative functions and metabolic pathways associated with the unique sequences were identified. The GO terms (59), COG items (25) and KEGG subgroups (6) were a little more in our transcriptome compared with the related classifications reported by Bai *et al*. (2013) (58, 24, 3). There were two probable reasons for these results; firstly, the different sequencing platform, Illumina paired-end and FLX 454, and secondly, different sampling individuals and tissues.

[Downloaded from jifro.ir on 2026-06-25]

[DOR: 20.1001.1.15622916.2016.15.1.38.2]

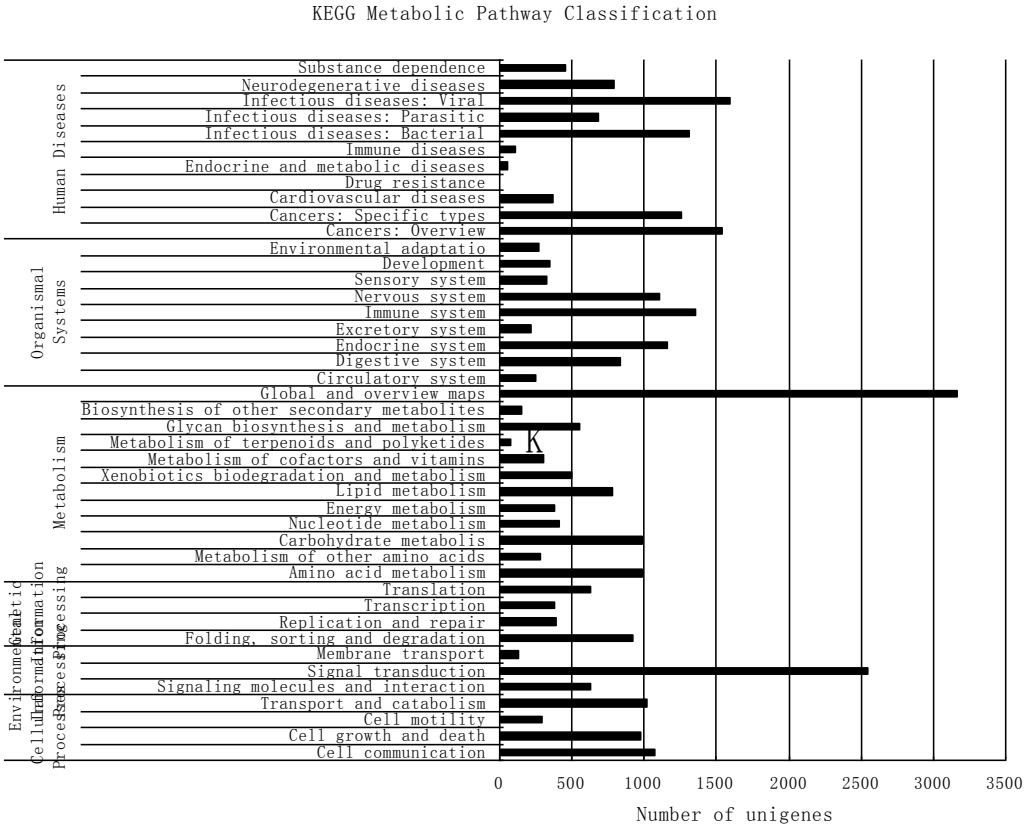

Figure 4 : KEGG metabolic pathway classification of *H. cumingii* based on six categories: Cellular processes, Environmental information processing, Genetic information processing, Metabolism, Organismal systems and Human Diseases.

Here, the samples were collected from six tissues including gonads, hepatopancreas, foot, mantel, gills and adductor muscles from three selected populations instead of mantel and pearl sac from the P- and W-line host mussels according to Bai *et al*. (2013).

Genes of interest involved in sex determination

It is believed that graft tissue piece implantation during gonad maturation period is not conducive to pearl quality of the mussel (Deng and Cai, 2007) and research also showed that total weights, grain weights and grain sizes of the pearls in 3- and 4-year-old males of *H. cumingii* were all significantly greater than that in female individuals (Zhao *et al*., 2013). So it's very important to study the sex determination mechanism. Although the reproductive biology of freshwater mussels has been well researched (Pan *et al*., 2010), limited available and valuable sex determination and differentiation information is exists (Guan, 2005). Previous studies showed that the majority of the researched 15 species of Neritidae (Archaeogastropoda, Gastroroda) have XO sex chromosomes,

[Downloaded from jifro.ir on 2026-06-25]

[DOR: 20.1001.1.15622916.2016.15.1.38.2]

while most species of Viviparidae and Potamididae (Mesogastropoda, Gastroroda) were XO or XO type. Whereas, the sex chromosomes about Lamellibranchia were little researched, existence of sex chromosomes in *H. cumingii* (Unionidae, Lamellibranchia) is not convincing (Yang *et al*., 2008), so it is uncertain whether *H. cumingii* possesses the same sex determination mechanism as Gastroroda species.

Sex determination and sex differentiation are very complicated processes. Up to now, many studies have mainly focused on dealing with the human and biomodel organism, and a number of critical genes involved in sex differentiation and reproduction have been identified, such as Sex-determining region Y protein (*SRY*) (Sinclair *et al*., 1990), Doublesex- and mab-3-related transcription factor 1 (*DMRT1*) (Raymond *et al*., 2000; Carlsson and mahlapuu, 2002; Uhlenhaut *et al*., 2009), *SRY*-related HMG-domain containing transcription factor (*SOX*) (Ner, 1992; Vilain and McCabe, 1998), Transcription factor GATA-4 **(***GATA-4*) (Hang *et al*., 1995), Nuclear receptor subfamily 0 group B member 1 (*DAX-1*) (Guo *et al*., 1996), Wilms tumor protein (*WT1*) (Buaas *et al*., 2009), Steroid genic factor 1 (*SF-1*) (Oba *et al*., 1996), Muellerian-inhibiting factor (*AMH*) (Muensterberg and Lovell-Badge, 1991), etc. In this study, majority of putative genes related to sex determination/differentiation found here were identified for the first time, such as *Dmrt, SOX*, Zinc finger Y-chromosomal protein 1 (*ZFY1*) (Connor and Ashworth, 1992; Nagamine and Carlisle, 1997). Only cytochrome *P450* aromatase (*P450*) (White *et al*., 2000) had been more studied in *H. cumingii*. The research described herein provides no evidence for a master gene whose presence or absence determines sex or sex differentiation of *H. cumingii*. Further studies are needed to understand the molecular functions of these putative genes as well as the molecular mechanism of sex determination and sex differentiation.

Molecular markers

Because of the high variability, abundance, neutrality and co-dominance of microsatellite DNA (Liu and Cordes, 2004), much research on genomic simple sequence repeats (SSRs) in *H. cumingii* have been done so far for the purpose of constructing genetic linkage, performing QTL analysis and evaluating the level of genetic variation (Wang *et al*., 2006; Li *et al*., 2009; Bai *et al*., 2011), and the identification of SSRs from expressed sequences has been proved to be a useful way for providing a rich source of valuable molecular markers (Bai *et al*., 2011). EST-SSRs associated with known function genes can be more useful for

[Downloaded from jifro.ir on 2026-06-25]

[DOR: 20.1001.1.15622916.2016.15.1.38.2]

comparative gene mapping (Liu *et al*., 1999) and also facilitate physical mapping. RNA-Seq is particularly useful as a shotgun method for generating EST data and identifying candidate genes, and tends to be more widely used for finding putative molecular markers (Blanca *et al*., 2011; Ma *et al*., 2012). The number of SSRs (8,014) found in this study was a little fewer, while SNPs (38,288) were more compared with the number of SSRs (9,474) and SNPs (27,303) reported before using the 454 technology (Bai *et al*., 2013). Although the SSRs and SNPs identified herein should be next verified before further use, they will be necessary for accelerating breeding programs to improve pearl quality in the future.

Acknowledgments

This work was supported by National Natural Science Foundation of China (NSFC) (no. 31302192); Specialized Research Fund for the Doctoral Program of Higher Education (SRFDP) (no. 20123305110001); Zhejiang Science and Technology Project (2012C12907-5); Huzhou Municipal Natural Science Foundation(2014YZ06).

.

References

Ashburner, M., Ball, C.A., Blake, J. A., Botstein, D., Butler, H., Cherry, J.M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M. and Sherlock, G., 2000. Gene ontology: tool for the unification of biology. *Nature Genetics*, 25, 25–29.

Bai, Z. Y., Niu, D. H. and Li, J. L., 2011. Development and characterization of EST-SSR markers in the freshwater pearl mussel (*Hyriopsis cumingii*). *Conservation Genetics Resources*, 3, 765–767.

Bai, Z. Y., Zheng, H. F., Lin, J. Y., Wang, G. L. and Li, J. L., 2013. Comparative analysis of the transcriptome in tissues secreting purple and white nacre in the pearl mussel *Hyriopsis cumingii*. *PLOS One*, 8 (**1**), e53617.

Blanca, J., Cañizares, J., Roig, C., Ziarsolo, P., Nuez, F. and Picó, B., 2011. Transcriptome characterization and high throughput SSRs and SNPs discovery in *Cucurbita pepo* (Cucurbitaceae). BMC Genomics, 12, 104.

Buaas, F. W., Val, P. and Swain, A., 2009. The transcription co-factor CITED2 functions during sex determination and early gonad development. *Human Molecular Genetics*, 18, 2989–3001.

Carlsson, P. and Mahlapuu, M., 2002. Forkhead transcription factors: key players in development and metabolism. *Developmental Biology*,

[Downloaded from jifro.ir on 2026-06-25]

[DOR: 20.1001.1.15622916.2016.15.1.38.2]

250, 1–23.

Collins, L. J., Biggs, P. J., Voelckel, C. and Joly, S., 2008. An approach to transcriptome analysis of non-model organisms using short-read sequences. *Genome informatics*, 21, 3–14.

Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M. and Robles, M., 2005. Blast2GO : a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21(**18**), 3674–3676.

Connor, F. A. and Ashworth, A., 1992. Sequence of cDNA for Xenopus XZFY-1. *Nucleic Acids Res*earch, 20, 5845–5845.

Deng, C. M. and Cai, Y. Y., 2007. Marine commercial shellfish and farming. China: China Agriculture Press. 349P.

Dong, Q. A. and Shi, A. J., 1999. The secreting activities of the pearl sac cells of *Hytiopsis cumingii*. *Journal of Fisheries of China*, 2, 001.

Fei, Z., Pan, J., Xu, Z., Ding, J. and Zhang, J., 2004. Study of the elimination of suspended substances and chlorophyll in water by *Hytiopsis cumingii* (Lea). *Transactions of Oceanology and Limnology*, 2, 40–45.

Gao, X. G., Han, J., Lu, Z. C., Li, Y. F. and He, C. B., 2013. De novo assembly and characterization of spotted seal *Phoca largha* transcriptome using Illumina paired-end sequencing. *Comparative Biochemistry and Physiology Part D: Genomics Proteomics*, 8(**2**), 103–110.

Guan, Y. Y., 2005. The study of Karyotype and sex-determining gene in imposex *Thais clavigera*. China: Shantou University. 32P.

Guo, W., Burris, T. P., Zhang, Y. H., Huang, B. L., Mason, J., Copeland, K. C., Kupfer, S. R., Pagon, R. A. and McCabe, E. R., 1996. Genomic sequence of the DAX1 gene: an orphan nuclear receptor responsible for X-linked adrenal hypoplasia congenita and hypogonadotropic hypogonadism. *The Journal of Clinical Endocrinology and Metabolism*, 81(**7**), 2481–2486.

Hale, M. C., McCormick, C. R., Jackson, J. R. and Dewoody, J. A., 2009. Nextgeneration pyrosequencing of gonad transcriptomes in the polyploid lake sturgeon (*Acipenser fulvescens*): the relative merits of normalization and rarefaction in gene discovery. *BMC Genomics*, 10, 203.

Hong, X. T., Xiang, L. X. and Shao, J. Z., 2006. The immunostimulating effect of bacterial genomic DNA on the innate immune responses of bivalve mussel, *Hyriopsis cumingii* (Lea). *Fish & shellfish immunology*, 21(**4**), 357–364.

Joubert, C., Piquemal, D., Marie, B., Manchon, L., Pierrat, F.,

[Downloaded from jifro.ir on 2026-06-25]

[DOR: 20.1001.1.15622916.2016.15.1.38.2]

Zanella-Cléon, I., Cochennec-Laureau, N., Gueguen, Y. and Montagnani, C., 2010. Transcriptome and proteome analysis of *Pinctada margaritifera* calcifying mantel and shell: focus on biomineralization. *BMC Genomics*, 11, 613.

Liao, X. L., Cheng, L., Xu, P., Lu, G. Q., Wachholtz, M., Sun, X. W. and Chen, S. L., 2013. Transcriptome analysis of crucian carp (*Carassius auratus*), an important aquaculture and hypoxia-tolerant species. *PLOS One*, 8(**4**), e62308.

Li, J., Qian, R., Bao, B., Wang, G. and Qi, N., 2004. RAPD analysis on genetic diversity among the stocks of *Hyriopsis cumingii* from the five large lakes of China. *Journal of Shanghai Fisheries University*, 14(**1**), 1-5.

Li, J. L., Wang, G. L. and Bai, Z. Y., 2009. Genetic variability in four wild and two farmed stocks of the Chinese freshwater pearl mussel (*Hyriopsis cumingii*) estimated by microsatellite DNA markers. *Aquaculture*, 287 (**3-4**), 286–291.

Liu, Z., Tan, G., Li, P. and Dunham, R. A., 1999. Transcribed dinucleotide microsatellites and their associated genes from channel catfish *Ictalurus punctatus*. *Biochemical and Biophysical Research Communications*, 259, 190–194.

Liu, Z. J. and Cordes, J. F., 2004. DNA marker technologies and their applications in aquaculture genetics. *Aquaculture*, 238, 1–37.

Luo, Y. M., Wei, K. J. and Hu, L., 2007. Advances in pearl rearing with *Hyriopsis cumingii*. *Reservoir Fisheries*, 27, 33–35.

Ma, K. Y., Qiu, G. F., Feng, J. B. and Li, J. L., 2012. Transcriptome analysis of the oriental river prawn, *Macrobrachium nipponense* using 454 pyrosequencing for discovery of genes and markers. *PLOS One*, 7 (**6**), e39727.

McGinty, E. L., Zenger, K. R., Jones, D. B. and Jerry, D. R., 2012. Transcriptome analysis of biomineralisation-related genes within the pearl sac: Host and donor oyster contribution. *Marine Genomics*, 5, 27–33.

Meyer, E., Aglyamova, G. V., Wang, S., Buchanan-Carter, J., Abrego, D., Colbourne, J. K., Wilis, B. L. and Matz, M. V., 2009. Sequencing and de novo analysis of a coral larval transcriptome using 454 GSFlx. *BMC Genomics*, 10, 219.

Muensterberg, A. and Lovell-Badge, R., 1991. Expression of the mouse anti-Mullerian hormone gene suggests a role in both male and female sexual differentiation. *Development*, 113, 613–624.

Nagamine, C. M. and Carlisle, C., 1997. Zfy2/1 fusion gene fails to replicate

[Downloaded from jifro.ir on 2026-06-25]

[DOR: 20.1001.1.15622916.2016.15.1.38.2]

Zfy1 expression pattern in fetal gonads. *Genomics*, 43, 397–398.

Ner, S. S., 1992. HMGs everywhere. *Current Biology*, 2, 208–210.

Oba, K., Yanase, T., Nomura, M., Morohashi, K., Takayanagi, R. and Nawata, H., 1996. Structural characterization of human Ad4bp (SF-1) gene. *Biochemical and Biophysical Research Communications*, 226, 261–267.

Pan, B. B., Li, J. L. and Bai, Y. Z., 2010. Histological study on ovarian development and oogenesis of *Hyriopsis cumingii* cultured in the pon d. *Journal of Shanghai Ocean University*, 19(**4**), 452–456.

Parchman, T. L., Geist, K. S., Grahnen, J. A., Benkman, C. W. and Buerkle, C. A., 2010. Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. *BMC Genomics*, 11, 180.

Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., Lee, Y., White, J., Cheung, F., Parvizi, B., Tsai, J. and Quackenbush, J., 2003. TIGR gene indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics*, 19(**5**), 651–652.

Qiu, A. D. and Shi, A. J., 1998. Effects of pH on nacre secretion of freshwater pearl mussel *(Hyriopsis cumingii*). *Acta Zoologica Sinica*, 45(**4**), 361–370.

Raymond, C. S., Murphy, M. W., O'Sullivan, M. G., Bardwell, V. J. and Zarkower, D., 2000. Dmrt1, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation. *Genes & Development*, 14, 2587–2595.

Rozen, S. and Skaletsky, H., 2000. Primer3 on the WWW for general users and for biologist programmers. In: Bioinformatics methods and protocols: Methods in molecular biology. USA: Humana Press, 365-386.

Sinclair, A. H., Berta, P., Palmer, M. S., Hawkins, J. R., Griffiths, B. L. Smith, M. J., Foster, J. W., Frischauf, A., Lovell-Badge, R. and Goodfellow, P. N., 1990. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature*, 346, 240–244.

Uhlenhaut, N. H., Jakob, S., Anlag, K., Eisenberger, T., Sekido, R. Kress, J., Treier, A. C, Klugmann, C., Klasen, C., Holter, N. I., Riethmacher, D., Schütz, G., Cooney, A. J., Lovell-Badge, R. and Treier, M., 2009. Somatic sex reprogramming of adult ovaries to testes by FOXL2 ablation. *Cell*, 139, 1130–1142.

Vilain, E. and McCabe, E. R., 1998. Mammalian sex determination: from

[Downloaded from jifro.ir on 2026-06-25]

[DOR: 20.1001.1.15622916.2016.15.1.38.2]

gonads to brain. *Molecular Genetics and Metabolism,* 65, 74–84.

Wang, G. L., Wang, J. J. and Li, J. L., 2006. Preliminary study on applicability of microsatellite primers developed from *Crassostrea gigas* to genomic analysis of *Hyriopsis cumingii*. *Journal of Fisheries of China,* 1, 30–38.

Wang, X. W., Luan, J. B., Li, J. M., Bao, Y. Y., Zhang, C. X. and Liu, S. S., 2010. De novo characterization of a whitefly transcriptome and analysis of its gene expression during development. *BMC Genomics,* 11, 400.

Wang, S. F., Wang, X. F., He, Q. W., Liu, X. X., Xu, W. L., Li, L. B., Gao, J. W. and Wang, F. D., 2012. Transcriptome analysis of the roots at early and late seedling stages using Illumina paired-end sequencing and development of EST-SSR markers in radish. *Plant Cell Reports*, 31, 1437–1447.

White, J. A., Ramshaw, H., Taimi, M., Stangle, W., Zhang, A. Everingham, S., Creighton, S., Tam, S., Jones, G. and Petkovich, M., 2000. Identification of the human cytochrome P450, P450RAI-2, which is predominantly expressed in the adult cerebellum and is responsible for all-trans-retinoic acid metabolism. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 6403-6408.

Xu, Q. Q., Guo, L. G., Xie, J. and Zhao, C. Y., 2011. Relationship between quality of pearl cultured in the triangle mussel *Hyriopsis cumingii* of different ages and its immune mechanism. *Aquaculture*, 315, 196–200.

Yang, P. H., Yang, W. D. and Wang, X. Y., 2008. Research on chromosomes of *hyriopsis cumingii* in Dongting Lake area. *Journal of Hunan University of Arts and Science (Natural Science Edition)*, 20(**1**), 64–68.

Zhao, Y. C., Bai, Z. Y., Fu, L. L., Liu, Y., Wang, G. L. and Li, J. L., 2013. Comparison of growth and pearl production in males and females of the freshwater mussel, *Hyriopsis cumingii*, in China. *Aquacult*ure *International*, 21, 1301–1310.

[Downloaded from jifro.ir on 2026-06-25]

[DOR: 20.1001.1.15622916.2016.15.1.38.2]