

De Novo Characterization of Japanese Scallop *Mizuhopecten yessoensis* Transcriptome and Analysis of Its Gene Expression following Cadmium Exposure

Xiao-lin Meng^{1,2,3}, Mei Liu¹, Ke-yong Jiang¹, Bao-jie Wang¹, Xue Tian³, Shu-juan Sun¹, Zuo-yong Luo^{1,2}, Chu-wen Qiu^{1,2}, Lei Wang^{1*}

1 R&D Center of Marine, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China, **2** University of Chinese Academy of Sciences, Beijing, China, **3** College of Animal Science and Veterinary Medicine, Shanxi Agriculture University, Taigu, China

Abstract

Background: Japanese scallop has been cultured on a large-scale in China for many years. However, serious marine pollution in recent years has resulted in considerable loss to this industry. Moreover, due to the lack of genomic resources, limited research has been carried out on this species. To facilitate the understanding at molecular level immune and stress response mechanism, an extensive transcriptomic profiling and digital gene expression (DGE) database of Japanese scallop upon cadmium exposure was carried out using the Illumina sequencing platform.

Results: RNA-seq produced about 112 million sequencing reads from the tissues of adult Japanese scallops. These reads were assembled into 194,839 non-redundant sequences with open reading frame (ORF), of which 14,240 putative amino acid sequences were assigned biological function annotation and were annotated with gene ontology and eukaryotic orthologous group terms. In addition, we identified 720 genes involved in response to stimulus and 302 genes involved in immune-response pathways. Furthermore, we investigated the transcriptomic changes in the gill and digestive gland of Japanese scallops following cadmium exposure using a tag-based DGE system. A total of 7,556 and 3,002 differentially expressed genes were detected, respectively, and functionally annotated with KEGG pathway annotations.

Conclusion: This study provides a comprehensive transcripts sequence resource for the Japanese scallop and presents a survey of gene expression in response to heavy metal exposure in a non-model marine invertebrate via the Illumina sequencing platform. These results may contribute to the in-depth elucidation of the molecular mechanisms involved in bivalve responses to marine pollutants.

Citation: Meng X-l, Liu M, Jiang K-y, Wang B-j, Tian X, et al. (2013) De Novo Characterization of Japanese Scallop *Mizuhopecten yessoensis* Transcriptome and Analysis of Its Gene Expression following Cadmium Exposure. PLoS ONE 8(5): e64485. doi:10.1371/journal.pone.0064485

Editor: Shu-Biao Wu, University of New England, Australia

Received: November 13, 2012; **Accepted:** April 15, 2013; **Published:** May 31, 2013

Copyright: © 2013 Meng et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Financial support for this study was provided by the Main Direction Program of Knowledge Innovation of Chinese Academy of Sciences (KZCX2-EW-Q215). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: wanglei@ms.qdio.ac.cn

Introduction

The Japanese scallop, *Mizuhopecten yessoensis*, is a cold water scallop species and is distributed in northern Japan, Sakhalin, the Kuril Islands and northern Korea [1,2]. Due to its great commercial value, this scallop was introduced to China from Japan in 1982 and expanded rapidly along the northern coastline. By 2005, the annual production of Japanese scallop had reached 150,000 tonnes in China [3].

Despite its economic importance, in recent years, *M. yessoensis* aquaculture has been hampered due to a variety of reasons such as population degradation, high temperature in summer, and other environment factors which were drastically influenced by climatic conditions and anthropogenic activities in coastal water [4,5]. Many chemical contaminants, including organochlorine compounds, herbicides, domestic and municipal wastes, petroleum products and heavy metals are now recognized to have serious adverse effects on aquaculture environments, even when released

at low levels [6,7]. However, the impact of these pollutants on *M. yessoensis* is unclear, especially at the molecular level. Previous related studies on *M. yessoensis* have mainly focused on testing bioaccumulation in tissues [8,9], effects on antioxidant enzyme activity [4,10], lipid peroxidation [11,12], DNA strand break [13,14], the discovery of genetic markers [15–18] and the construction of gene maps [19]. Nevertheless, due to the lack of genomic resources such as genome and transcriptome sequences, these studies were limited and an overall understanding of the mechanism of action of these pollutants in *M. yessoensis* is a high priority. To achieve this, a global understanding of the transcriptome profiling of *M. yessoensis* is the first and necessary step.

The massive parallel next generation sequencing technologies have facilitated the production of high coverage sequence data, enabling genome wide assays of transcriptional activities and have been applied in many types of aquatic animals such as zebrafish [20], carp [21], sea bass [22], and clam [23]. Although the 454

and SOLID platforms have their own superiority, recent algorithmic and experimental advances have greatly increased the applicability of Illumina sequencing and *de novo* assembly, which has been successfully and increasingly used for non-model species [22,24–26]. In addition, Illumina sequencing technology has been shown to be highly replicable and was proven to be a superior method to study mRNA expression levels due to its ability to identify differentially expressed genes [27]. To our knowledge, this is the first report on the transcriptome profiling of adult *M. yessoensis* using the Illumina facility with the aim of constructing a database on this species.

Cadmium pollution is a serious problem in Bohai Bay, China, according to an ecological assessment conducted from 2001–2005 [28]. Therefore, understanding its toxicological mechanism and effective early warning of aquaculture risk in *M. yessoensis*, according to the biomarker monitoring system, is now necessary. In this study, digital gene expression (DGE) technology was applied to analyze the differential expression of genes in the gill and digestive gland of *M. yessoensis* following cadmium (Cd) exposure, and a sensitive biomarker database was constructed. Our results could provide valuable and reliable data for bivalve aquaculture and are expected to improve our understanding of the toxicological mechanism of marine pollutants.

Results and Discussion

Illumina Paired-end Sequencing and Reads Assembly

To obtain a global overview of *M. yessoensis* transcriptome and gene activity at nucleotide resolution, a cDNA pool was prepared from seven organs including the adductor muscle, digestive gland, gill, gonad (male and female), kidney, visceral mass and mantle and sequenced using the Illumina sequencing platform. Approximately 112.26 million 93 bp reads were generated. The raw reads produced in this study have been deposited in the NCBI SRA database (accession number: SRR653778). After removing adaptors and low quality reads, reliable reads were *de novo* assembled with Velvet and Oases software. A total of 217,190 contigs were assembled ranging from 100 to 29,088 bp, with an average size of 436 bp. Among these contigs, 184,390 (~84.89%) of the clean contigs were smaller than 500 bp, 12,703 (5.85%) were between 500 and 1,000 bp, 13,847 (~6.38%) were between 1,000–3,000 bp in length, and 6,250 (~2.88%) were longer than 3,000 bp. The length distribution of assembled contigs was shown in Figure 1. The previously reported *M. yessoensis* transcriptome by 454 sequencing generated 32,590 contigs with a mean size of 618 bp [29], which is 180 bp longer than the length in our current study, yet we obtained more detailed data in terms of quantity of the database. After the gene prediction program, a total of 194,839 contigs (with an average length of 461 bp) from these non-redundant sequences were found including putative open reading frames (ORF) which encoded proteins using the GetORF program of the EMBOSS suite. A summary of the data was shown in Table 1.

To demonstrate the quality of the sequencing data, 6 contigs were selected and 6 pairs of primers were designed for RT-PCR amplification. Agarose gel electrophoresis results showed that 5 of the 6 primer pairs obtained a band of the expected size, and the five identified PCR products were further confirmed by Sanger sequencing (data not shown).

Annotation of Predicted Proteins

All the predicted protein sequences (translation by the Transeq program of the EMBOSS suite) were compared with the NCBI non-redundant (nr) protein database and the Swiss-Prot database

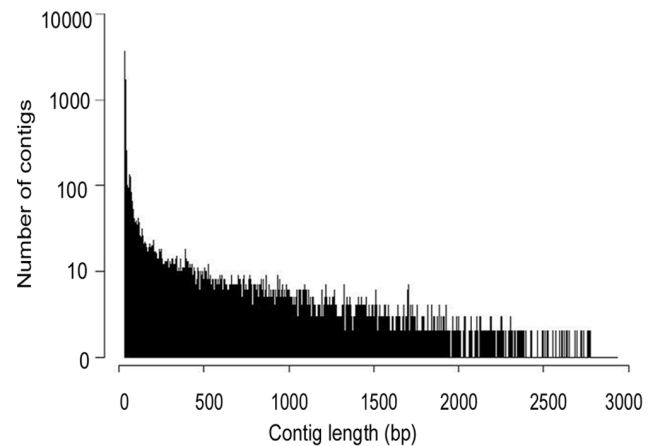


Figure 1. Assembled contig length distribution of the *M. yessoensis* transcriptome. The x-axis indicates contig size and the y-axis indicates the number of contigs of each size. doi:10.1371/journal.pone.0064485.g001

for gene annotation using BLASTx with an E-value cut-off of $1e^{-3}$. A total of 14,240 predicted protein sequences obtained biological function annotation, while the remainder would require more genetic data for annotation in future. Gene ontology (GO) assignments were used to classify the functions of the predicted proteins. Based on GoPipe software, a total of 5,245 predicted proteins were assigned at least one GO term for describing cellular components, molecular functions and biological process (Figure 2). Briefly, in the cellular components category, cell (30.3%) was the dominant group followed by intracellular (25.8%). For molecular functions, binding (27.5%) was the most representative GO term, followed by catalytic activity (16.2%). With regard to biological process, the most represented categories were cellular process (19.5%) and metabolic process (16.2%) (Figure 2). These results corresponded with the previously sequenced *M. yessoensis* via 454 sequencing [29]. To further evaluate the completeness of the transcriptome library and the effectiveness of the annotation process, eukaryotic orthologous groups (KOG) were used. In total, 34,989 annotated putative proteins were classified functionally into 25 molecular families, including signal transduction mechanism, cellular structure, biochemical metabolism, defence system, gene expression and so on (Figure 3).

Besides GO annotation and KOG analysis, we mapped all the predicted proteins to the reference canonical pathways in KEGG (Kyoto Encyclopedia of Genes and Genomes) [30] for functional categorization and annotation. In total, we assigned 2,702 proteins into different KEGG pathways and the most representative pathways were infectious diseases (678 members), cancers (482 members), signal transduction (438 members) and carbohydrate metabolism (352 members) (Table 2). These well-categorized and annotated resources could provide an important and valuable database for investigating specific bioprocesses, identification of functional genes and analyzing specific traits during *M. yessoensis* research.

Functional Genes Involved in Environmental Stress and Immunity

In many bivalve species, responses to environmental stress and changes in immunity are used as biomarkers to evaluate the biological effects of different types of pollutants in aquatic animals, including metallothionein (MT), lysosomal membrane integrity, acetylcholinesterase (AChE), the cytochrome P450 (CYP) family,

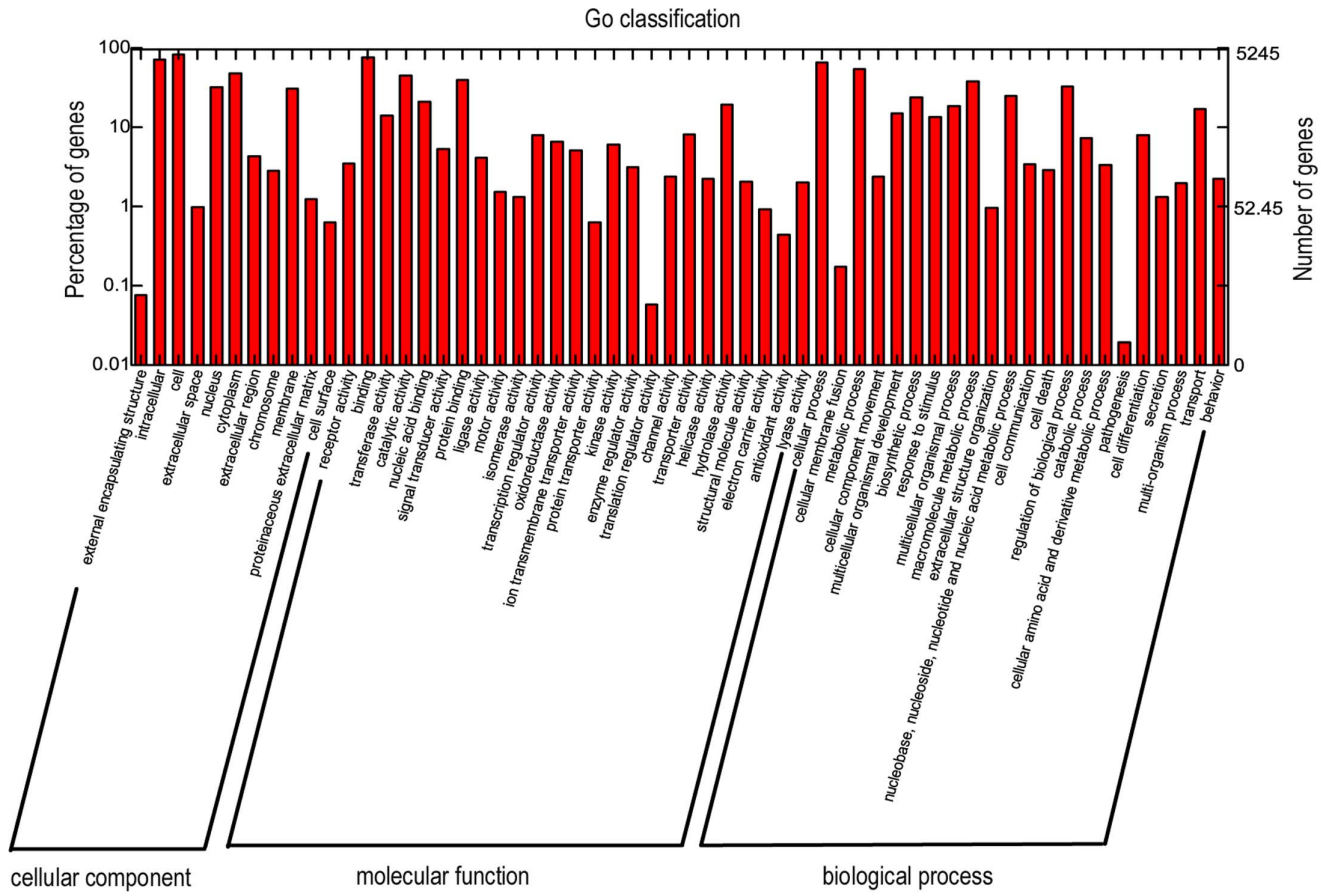


Figure 2. GO annotations of *M. yessoensis* transcriptome data. A total of 5,249 contigs are assigned to at least one GO term including biological process, cellular component and molecular function. doi:10.1371/journal.pone.0064485.g002

glutathione-dependant oxidoreductases and heat shock proteins (HSP) [31–36]. In addition, due to a lack of immunoglobulins, invertebrates have developed unique systems of biological host defense, the so called innate immunity [37]. However, the exact molecular and cellular basis of the immune system remains poorly understood. In this study, the sequence and annotation information from BLAST, GO and KEGG annotations provided valuable gene resources for research on the molecular biology of these traits in *M. yessoensis*.

The GO annotation identified 720 sequences that potentially respond to stimulus (GO: 0050896) (Table S1). DNA damage in marine organisms, one of the major effects following the pollutants stimulus, may be relevant to DNA strands broken which can be induced by free radicals, organic and inorganic contaminants, heavy metals, etc. Therefore, it is essential for cells to efficiently respond to DNA damage through coordinated and integrated DNA-damage checkpoints and repair pathways. However, the mechanism of the DNA damage repair in bivalves is not very clear. The DNA repair protein family, DNA mismatch repair protein family and DNA excision repair protein family [38–40] which play an important role in DNA damage repair after challenge were presented abundant in this category. These data could be valuable for future research. HSP have a long history of use in studies of organism stress response. In this term of GO annotation, the HSP families including HSP70, HSP90, HSP110 and HSP transcription factor were detected which play essential roles as molecular chaperones by assisting the correct folding of

nascent and stress-accumulated misfolded proteins and by preventing their aggregation [36], but others, such as HSP27, HSP40, HSP10, HSP22 were not detected. In addition, other genes which have been used as biomarkers in environmental monitoring were detected, such as glutathione S-transferases (GST), Glutathione peroxidase (GPx), CYP family, SOD, and CAT. It is worth noting that the HSP families were the most abundant transcripts in this category in Hou’s report [29] and the stress-associated endoplasmic reticulum protein (SERP2) was also highly expressed, which is different from our data, due possibly to the different samples used in the two experiments.

Although knowledge of bivalve immune-related genes has increased in the last few years with the development of biotechnology, the available information is still scarce and fragmentary [41]. With regard to the immune system, KEGG analysis resulted in 251 genes involved in 13 immune-response pathways (Table 2, Table S1) such as Fc gamma R-mediated phagocytosis signaling pathway, chemokine signaling pathway, complement and coagulation cascades, Toll-like receptor signaling pathway and NOD-like receptor signaling pathway. The complement system consists of a tightly regulated network of proteins that play crucial role in host defense and inflammation. C3 and C1 q are the central components in this system and were detected in this database. Although the complement pathway has not been extensively described in bivalves, our results support the presence of this defense mechanism. Toll-like receptors (TLRs) are an ancient family of pattern recognition receptors that play key roles

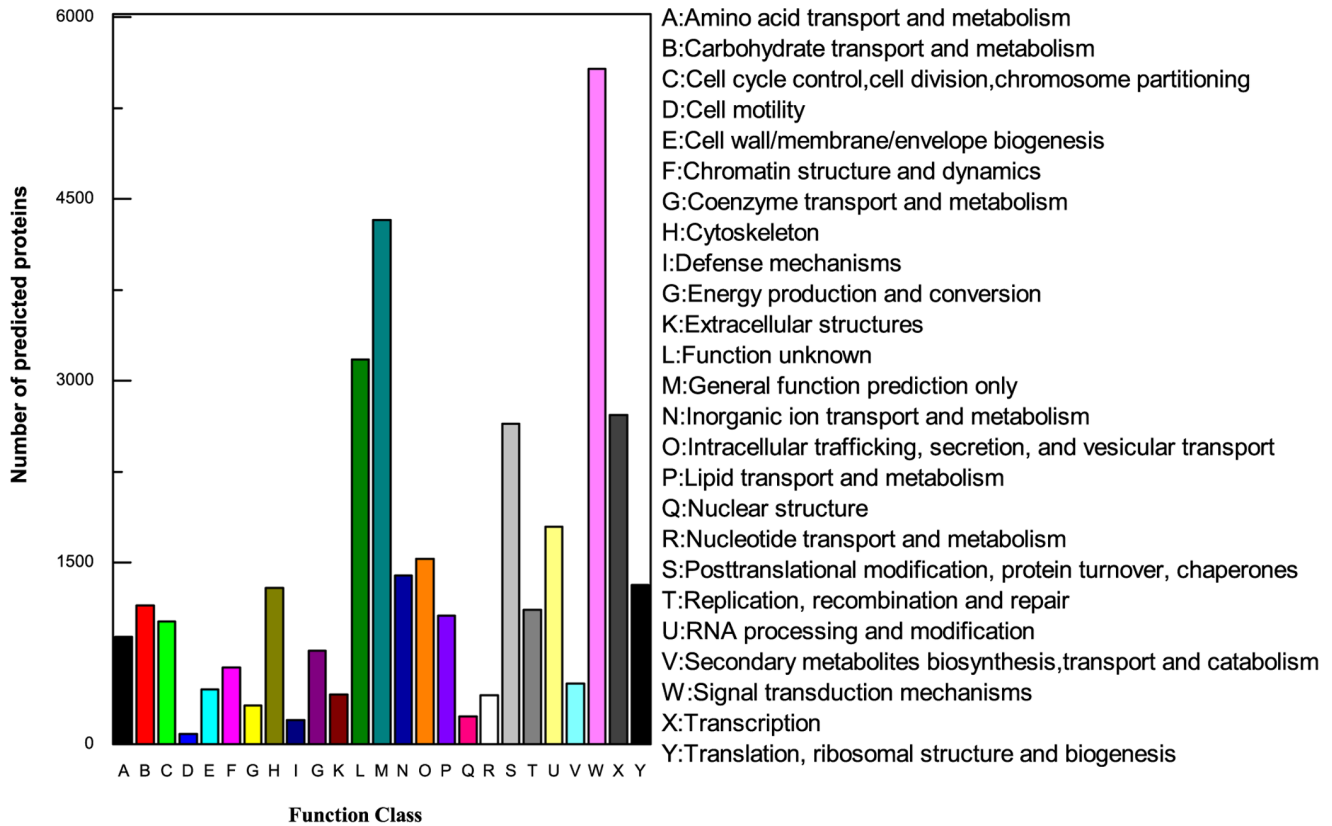


Figure 3. KOG annotations of predicted proteins. A total of 34,839 predicted proteins have a KOG classification among the 25 categories. doi:10.1371/journal.pone.0064485.g003

in detecting non-self substances and activating the immune system. TLR1, TLR 4, MYD88, TRAF and IRAK, were identified in the database, which have been established to play an essential role in the activation of innate immunity by recognizing specific patterns of microbial components [42]. Tumor necrosis factor receptor-associated factor 3 (TRAF3), a highly versatile regulator, played important roles beyond the TNFR-superfamily (SF), mediating certain innate immune receptor and cytokine receptor signals [43]. ESTs with homology to the TRAF3 have been detected in the pearl oyster [44] and were also presented in the RIG-I-like receptors (RLRs) and TLRs signaling pathways. PI3 K [45], MAPK [46], and p38 [47], which are also crucial in cellular responses to external stress signals, were found in corresponding pathways. Therefore, the large set of immune-relevant genes and

their roles in responses to environmental stress may significantly improve our understanding of the immune mechanism in *M. yessoensis*.

Digital Gene Expression Profile Analysis

DGE analysis based on the Solexa/Illumina system, which avoids many of the inherent limitations of microarray analysis, was performed to identify the genes involved in the response of *M. yessoensis* to cadmium (Cd) exposure. We sequenced four DGE libraries: gill in the control (CG), digestive gland in the control (CD), gill in the treated group (TG) and digestive gland in the treated group (TD), and generated 5.75, 9.21, 8.96 and 6.95 million raw reads, respectively (Table 3). After removing the low quality reads, the number of clean reads ranged from 5.67 to 9.12 million (98.75–99.0%). The data sets are available at the NCBI SRA with the accession number: SRR654701. Gene annotation was performed by tag mapping analysis using the 217,190 non-redundant sequences from RNA-seq based transcriptome analysis as the reference transcript database. Among the clean reads generated by Illumina sequencing of the four libraries, about 1.86 million to 2.95 million sequences were mapped to the entire reference database (Table 3). The level of gene expression was analyzed by calculating the number of mapped reads for each gene locus and then normalizing it to the number of Reads per Kilobase per Million (RPKM) [48]. As shown in Figure 4, the results indicated that the major types of expressed genes were represented in fewer than 18 copies and a small percentage of the genes were highly expressed.

Table 1. Summary of Illumina transcriptome sequencing and assembly for *M. yessoensis*.

Total number of reads	112,265,296
Total base pairs (bp)	10,440,672,528
Average read length (bp)	93
Total number of non-redundant contigs	217,190
Mean length of non-redundant contigs (bp)	436
Contigs with coding protein sequences	194,839
Mean length of contigs with coding protein sequences (bp)	461

doi:10.1371/journal.pone.0064485.t001

Table 2. KEGG biochemical mappings for *M. yessoensis*.

KEGG categories represented	Unique sequences (The number of enzymes)
Metabolism	
Carbohydrate Metabolism	352 (172)
Energy Metabolism	117 (96)
Lipid Metabolism	308 (166)
Nucleotide Metabolism	184 (112)
Amino Acid Metabolism	301 (169)
Metabolism of Other Amino Acids	102 (61)
Glycan Biosynthesis and Metabolism	226 (124)
Metabolism of Cofactors and Vitamins	133 (110)
Metabolism of Terpenoids and Polyketides	32 (24)
Biosynthesis of Other Secondary Metabolites	43 (23)
Xenobiotic Biodegradation and Metabolism	142 (58)
Genetic Information Processing	
Transcription	144 (135)
Translation	278 (227)
Folding, Sorting and Degradation	321 (285)
Replication and Repair	202 (122)
Membrane Transport	
ABC transporters	27 (24)
Signal Transduction	438 (261)
Signal Molecules and Interaction	86 (73)
Cellular Process	
Transport and Catabolism	293 (241)
Cell Motility	59 (56)
Cell Growth and Death	317 (154)
Cell Communication	156 (120)
Organismal systems	
Immune system	302 (155)
Endocrine System	242 (152)
Circulatory System	47 (36)
Digestive System	194 (97)
Excretory System	70 (58)
Nervous System	282 (131)
Sensory System	32 (18)
Development	76 (68)
Environmental Adaptation	25 (19)
Human Diseases	
Cancers	482 (166)
Immune Diseases	38 (31)
Neurodegenerative Diseases	201 (119)
Cardiovascular Diseases	52 (28)
Endocrine and Metabolism Diseases	33 (28)
Infectious Diseases	678 (304)

doi:10.1371/journal.pone.0064485.t002

Changes in Gene Expression in the Gill and Digestive Gland Following Cd Exposure

To calculate the expression abundance of each gene between CG and TG or CD and TD, the MARS model [49] in the DEGseq package was used. A total of 7,556 significantly changed

gene entities were observed between CG and TG corresponding to 3,673 genes up-regulated and 3,883 genes down-regulated (Table S2). Between CD and TD, a total of 3,002 differentially expressed genes were detected with 1,627 up-regulated genes and 1,375 down-regulated genes (Table S3). This suggested that the number

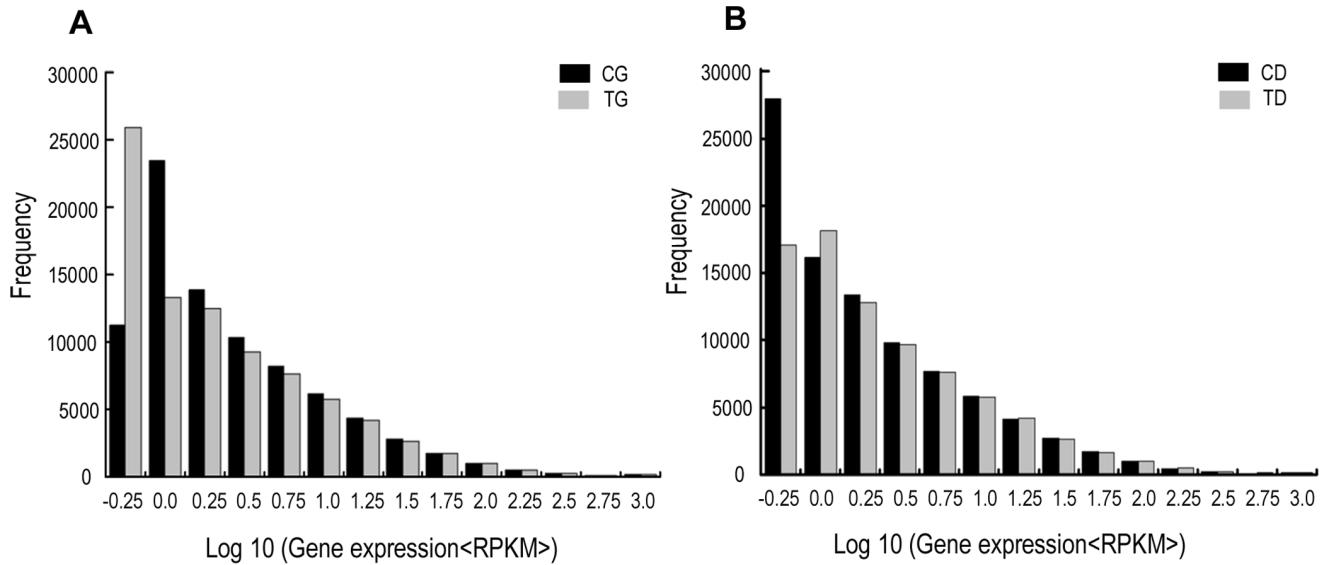


Figure 4. The level of gene expression for each gene. Gene expression level was determined by calculating the number of reads for each gene and then normalized to RPKM (Reads per kilobase per Million). A indicates the distribution of genes in CG and TG, B indicates the distribution of genes in CD and TD.

doi:10.1371/journal.pone.0064485.g004

of differentially expressed genes between CG and TG was more than those between CD and TD. The top 20 most differentially expressed genes between samples (TG vs CG; TD vs CD) were then analyzed (Table S4). We observed that most of the highly expressed genes in both of the compared groups were orphan sequences which meant that no homologues were found in the Nr database. This might be due to a lack of clear annotation information on molluscs in the database and the function of these unknown genes needs to be identified in the future. To identify more realistic different expressed genes, negative binomial distribution was introduced [50] and we obtained 70 up- and 47 down-regulated genes in gill and 81 up- and 43 down-regulated genes in digestive gland (Table S5).

Function Annotation of Differentially Expressed Genes

To understand the functions of differentially expressed genes (DEGs), we mapped all the genes with expression abundance more than 2-fold to terms in the KEGG database and compared this with the whole transcriptome background. Among all the genes with KEGG annotation, 479 differentially expressed genes were observed between CG and TG (Table S6). Most of them were involved in proteasome, ribosome biogenesis in eukaryotes, ATP-binding cassette (ABC) transporters, MAPK signaling pathway

and oxidative phosphorylation. Between CD and TD, 154 differentially expressed genes were annotated with KEGG pathway annotation (Table S6). The main enriched metabolism pathways concentrated on ABC transporters, glycine, serine and threonine metabolism, steroid hormone biosynthesis and glutathione metabolism. Cellular process including lysosome and peroxisome were also enriched.

Global cellular stress response has been defined as all proteins over-produced due to environmental stress. Following Cd exposure, several protein families including ABC, HSP and CYP were observed over transcribed in gill or digestive gland of *M. yessoensis*. Previous work has indicated that ABCB- and ABCC-type transporters confer multixenobiotic resistance and form an environment-tissue barrier in bivalve gills [51]. This protective mechanism is referred to as multixenobiotic resistance (MXR) which is supposedly ubiquitous in aquatic invertebrates. Also, other research suggests that the ABC multidrug transporters could be act as a detoxifying mechanism of various toxic agents in aquatic organism [52]. This might be the reason that ABC families also found significantly up-regulated in digestive gland in our results. In addition to ABCB and ABCC, ABCA, ABCD and ABCG were detected simultaneously up-regulated. As we discussed previously, HSP families are critical factors during the process of environmental stress. The present study found that HSP70, HSP110 in gill and HSP70 in digestive gland were over transcribed.

Comparing the enriched different expressed genes between gill and digestive gland following Cd exposure, several interesting pathways and genes were explored. NADH dehydrogenase, ATPase and cytochrome c oxidase of oxidative phosphorylation were only induced in gill of *M. yessoensis* by Cd, suggesting a compensation of partial uncoupling of oxidative phosphorylation [53] and the key role of the gill in energy metabolism. DNA damage and DNA damage repair were the major effects of pollutants stimulus. Related DNA-damage-inducible protein and DNA mismatch repair protein were also detected up-regulated only in gill. This probably related with the function of gill which

Table 3. Statistics of DGE sequencing.

Summary	CG	TG	CD	TD
Raw reads	5,745,741	8,958,064	9,210,131	6,946,099
Clean reads	5,674,105	8,864,949	9,118,668	6,873,165
Clean reads (%)	98.75%	98.96%	99.00%	98.95%
Low quality (%)	1.25%	1.04%	1.00%	1.05%
Reads mapped to genes	1,856,721	2,865,585	2,951,986	2,198,404
Mapped rate (%)	32.72%	32.32%	32.37%	31.99%

doi:10.1371/journal.pone.0064485.t003

acts as the main tissue for filtration of suspended matter and forms the first defense barrier.

Genes encoding detoxification enzymes played important role in bivalves when exposed to chemical pollutants and related evidences in previous study have indicated that digestive gland was responsible for detoxification of environmental pollutants and suggested it as a target organ for detection/identification of biomarkers of pollutants [23]. These detoxification enzymes include CYP families, GST, AChE and so on. CYP are one of the major phase I-type classes of detoxification enzymes found in terrestrial and aquatic organisms ranging from bacteria to vertebrates [54]. In our pathway enrichment analysis to digestive gland, totally 9 transcripts of cytochrome P450 families were found and eight of them presented up-regulate including CYP17, CYP1A1, CYP3A, CYP5A, CYP4F, CYP26A, CYP98A. Although CYP also be detected in gill, the highest expression was observed in digestive gland. GSTs constitute a large protein family, with a pivotal role in detoxification of xeno-compounds [23]. An increased GST activity has been observed after exposure to a broad set of pollutants. Following Cd exposure, two transcripts encoding GST were detected over-expressed in digestive gland in pathway of glutathione metabolism. However, the same tendency was not observed in gill. It could be due to the tissue-specific expression and sensitivity to dose/type of chemicals of GST in bivalves. At the cellular level, loss of lysosomal membrane integrity has been observed as a consequence of oxidative stress induced by several classes of chemicals. Reduced lysosomal membrane stability is also linked to increased autophagy [23,55]. Therefore, Cd exposure may induce lysosomal damage in the digestive gland of *M. yessoensis*.

Experimental Validation of Differentially Expressed Genes

To further evaluate the DGE library, nine genes expressed in the gill or digestive gland with clear annotation (2 up-regulated, 2 down-regulated and 2 controls for each tissue) were randomly selected for analysis by qRT-PCR. The qRT-PCR results displayed the same expression tendency as the DGE library (Table 4). Interestingly, yellow *-x1c* (*Y-x1c*) which was significantly down-regulated in the gill was highly expressed in the digestive gland. However, yellow-like genes have only been identified in insect species and in a number of bacteria [56]. The function of yellow-like genes is largely unknown and requires further research. GPx, often used as a biomarker of environmental stimulus, was significantly down-regulated in the gill. This finding is consistent with several previous studies [57,58] which directly tested the enzyme activity using a biochemical test. In contrast, CYP, another biomarker used in monitoring environmental contamination [34], was down-regulated in the digestive gland, but displayed different expression in other studies [59,60], probably due to different species, tissues or exposure time and dosage. GTPase and IMAP family 1 (*GIMAP1*), believed to be a regulator of cell death [61], were down-regulated in the digestive gland, which may mean immunologic injury and needs to be identified by other ways. *SPF27* and *scramblase* were used as controls and did not show a difference in expression in gill and digestive gland tissues. These findings provide us with more valuable molecular targets for the in-depth study of oxidative damage and immunologic injury in *M. yessoensis*.

Conclusions

In this study, we generated a comprehensive transcriptome of adult *M. yessoensis* using the Illumina sequencing platform. A single run produced 194,839 distinct sequences with 14,240 sequences having explicit annotation. A significant number of genes involved

Table 4. Validation of DGE library by qRT-PCR.

Gene	Gene ID	DGE library	fold by DGE	fold by qPCR
<i>Hzf3</i>	locus_34761	TG vs. CG	3.48	4.24
<i>Nephrin</i>	locus_134559	TG vs. CG	7.25	10.16
<i>Y-x1c</i>	locus_22146	TG vs. CG	-10.52	-18.26
	locus_4770	TD vs. CD	7.37	6.37
<i>Gpx</i>	locus_16910	TG vs. CG	-4.14	-2.18
<i>Hsp70</i>	locus_20269	TD vs. CD	2.27	2.63
<i>CYP</i>	locus_14839	TD vs. CD	-5.77	-2.85
<i>GIMAP1</i>	locus_20540	TD vs. CD	-5.40	-6.33
	locus_100	TG vs. CG	-0.86	/
<i>SPF27</i>	locus_100	TD vs. CD	-0.49	/
	locus_10003	TG vs. CG	0.23	/
<i>Scramblase</i>	locus_10003	TD vs. CD	0.89	/

doi:10.1371/journal.pone.0064485.t004

in immunity and response to stimulus were found in this transcriptome. These findings substantially supplement existing sequence resources for *M. yessoensis*. To our knowledge, this is the first publication using Illumina sequencing technology in this species, which lack a reference genome. In addition, genes differentially expressed in the gill and digestive gland following cadmium exposure were identified and functionally annotated with the KEGG database. These data would provide potential molecular targets in shellfish for functional studies of genes responding to marine pollutants and would serve as a valuable reference for identifying biomarkers in marine environmental pollution monitoring.

Materials and Methods

Experimental Scallop

Adult Japanese scallops (*Mizuhopecten yessoensis*) of both sexes were collected from Dalian Zhangzidao (Dalian, Liaoning Province, China) in 2011. Housing and care of scallops and collection of tissue samples for use in described experiments were conducted in accordance with the International Guiding Principles for Biomedical Research Involving Animals (<http://www.cioms.ch/frame1985texts/guidelines.html>). Only healthy scallops of a homogeneous size (8.730 ± 0.167 cm (mean \pm SD, $n = 30$), shell height) were used. They were kept in running aerated sea water (salinity 30‰) at $16 \pm 1^\circ\text{C}$ which was pumped from the first bathing beach of Qingdao and filtered prior to the experiments. Half the seawater was changed daily and the scallops were fed with the microalgae *Spirulina*.

Cadmium Exposure and RNA Preparation

The scallops in the experimental groups were exposed to $\text{CdCl}_2 \cdot 2.5 \text{H}_2\text{O}$ (Kanto Chemical Co., Tokyo, Japan) and the final Cd^{2+} concentration in the seawater was 0.1 mg/L which was twenty times that of the standard according to the 'Water quality standard for fisheries of China' ($\text{Cd}^{2+} \leq 0.005$ mg/L). Three replicates were prepared for both the control and Cd treated group. They were all kept under the same conditions described above and the changed seawater was resupplied with the corresponding concentration of Cd. No mortality was observed during the experimental period. After 14 days exposure, tissues including the adductor muscle, digestive gland, gill, gonad (male

and female), kidney, visceral mass and mantle in the control were collected for RNA-seq. In addition, the gill and digestive gland in the control or treated groups were collected, respectively, for DGE analysis. At least three independent biological replicates for each sample were harvested. Total RNA was isolated using TRIzol reagent (Gibco BRL) following the manufacturer's protocol. The quantity and quality of total RNA was confirmed using a NanoDrop spectrophotometer (Thermo Scientific, USA) and an Agilent 2100 Bioanalyzer (Agilent Technologies). It was then treated with DNase I (Ambion, USA) for 1 h at 37°C to remove residual DNA.

cDNA Library Preparation and Illumina Sequencing for RNA-seq

One paired-end (PE) cDNA library was generated from the pooled total RNA of the adductor muscle, digestive gland, gill, gonad (male and female), kidney, visceral mass and mantle in equal quantity. In brief, oligo (dT) beads were used to isolate poly (A) mRNA (mixture). First-strand cDNA was synthesized using a random hexamer-primer and reverse transcriptase (Invitrogen). Second-strand cDNA was synthesized using RNase H (Invitrogen) and DNA polymerase I (TaKaRa, Japan). The cDNA was sonicated for 5 min using the cup horn Sonic Dismembrator 550 (Fisher Scientific). The fragmented library was fractionated on 8% polyacrylamide gels, and 300–500 bp base pair fractions were excised. The gel purified cDNA product was modified for sequencing using the TruSeq™ DNA Sample Prep Kit- Set A (Illumina, USA) and then PCR-amplified with TruSeq PE Cluster Kit (Illumina, USA).

Analysis of Illumina Sequencing Results

The cDNA library was sequenced on the Illumina sequencing platform (GA II) and raw PE reads with an average length of 93 bp were generated. The raw reads were then assembled into the non-redundant consensus using Velvet (version 0.7.62; <http://www.ebi.ac.uk/zerbino/velvet/>), which was developed for assembly of short reads using the de Bruijn graph algorithm [62]. In addition, another publicly available program, Oases (version 0.1.8; <http://www.ebi.ac.uk/zerbino/oases/>) was used, which was also developed for *de novo* assembly of short reads in order to obtain best assembly results [63]. Adapter sequences were then trimmed and low quality reads were removed. Short sequences (<100 bp in length) were also removed using the custom Perl Program [64]. Putative open reading frames (ORF) within contigs were then obtained which encoded proteins using the GetORF program of the EMBOSS suite [65]. The ORF of each predicted protein was then used for BLASTx searches against the NCBI non-redundant (nr) protein database and Swiss-Prot protein database with an E-value threshold to 10^{-3} . Functional annotation using gene ontology terms (GO; <http://www.geneontology.org>) was analyzed using the BLASTp algorithm against the Swiss Prot and TrEMBL databases by the GoPipe program according to gene2go software [66] at <http://gopipe.fishgenome.org>. The COG and KEGG pathways annotation was performed using Blastall software against the Cluster of Orthologous Groups database and Kyoto Encyclopedia of Genes and Genomes database [67].

Digital Gene Expression (DGE) Library Preparation and Sequencing

Four tag libraries from the control gill (CG), the control digestive gland (CD), the treated group gill (TG) and the treated group digestive gland (TD) were prepared for DGE which was performed in parallel using the Digital Gene Expression Tag

Profile Kit (Illumina). Briefly, total RNA from the four samples was enriched using oligo-dT magnetic beads. The first and second cDNA strands were synthesized and bead bound cDNA was subsequently digested with *Nla*III. The 3'-cDNA fragments were then purified with oligo(dT) beads and 5' ends were added with Illumina adapter 1, which contained a recognition site for the endonuclease *Mme*I for cutting 17 bp downstream of the recognition site (CATG) to produce tags with adapter 1. After removing 3' fragments with magnetic bead precipitation, an Illumina GEX adapter 2 was introduced at 3' ends of the tags. The resulting cDNA tags with adapter-ligated were amplified using PCR amplification for about 15 cycles. The four enriched cDNA tag libraries were then gel-purified and checked using the Agilent 2100 Bioanalyzer. Each cDNA tag library was digested and the single-chain molecules were fixed onto an Illumina proprietary sequencing chip (flow cell) and underwent deep sequencing using the Illumina Genome Analyzer.

Aligning DGE Tags to the Reference Transcriptome

Raw sequencing reads were filtered by the Illumina pipeline. Prior to mapping reads to the reference transcriptome generated by RNA-seq, all low quality sequences such as short sequences (<21 nt), empty reads, and singletons (tags that occurred only once) were filtered. To monitor mapping events on both strands, both sense and antisense sequences were included in the data process. Only perfect matches of 17-bp tags plus the CATG *Nla*III recognition site were allowed.

Analysis of Differentially Expressed Genes

For gene expression analysis, the number of expression tags from the four samples was calculated and normalized to Reads per Kilobase per Million (RPKM) [48], respectively. Then the expression abundance of each gene between the control and treated group (gill or digestive gland) was counted based on the MARS model (MA-plot-based method with Random Sampling model) [49] using the DEGseq package. In this study, we used stringent value FDR (false discovery rate) <0.001 and the absolute value of $|\log_2^{\text{Ratio}}| \leq 1$ as the threshold to judge the significant difference in gene expression. In order to get more realistic data, the algorithm based on negative binomial distribution was implemented [50]. For pathway enrichment analysis, all differentially expressed genes were mapped to terms of the KEGG database and significantly enriched KEGG terms were identified.

Quantitative Real-time PCR (qRT-PCR) Verification

Nine genes were selected for confirmation of DGE data by qRT-PCR using the SYBR Premix Ex Taq kit (Takara, Japan) according to the manufacturer's instructions and performed in triplicate on the stratagene Mx3005P system. Briefly, total RNA was isolated from the gill and digestive gland of *M. yessoensis* in the control and Cd treated group (at least 4 scallops for each sample), using TRIzol reagent (Invitrogen, USA). First strand cDNA was obtained from 2 µg of total RNA using the PrimeScript II 1st strand cDNA synthesis kit (Takara, Japan). The specific primers used for qRT-PCR are listed in Table S7 and β-actin was used as an endogenous control. The relative gene expression was analyzed using the comparative threshold cycle (CT) method established by Livak et al. [68].

Supporting Information

Table S1 Genes involved in environmental stress and immunity. A list of the potential genes involved in environmental stress and immunity based on BLAST, GO and KEGG annotation.

(XLSX)

Table S2 Different expressed genes between CG and TG. A list of the different expressed genes in the gill of *M. yessoensis* after Cd exposure. FDR (<0.001) and P-value (<0.001) were used as the standard.

(XLSX)

Table S3 Different expressed genes between CD and TD. A list of the different expressed genes in the digestive gland of *M. yessoensis* after Cd exposure. FDR (<0.001) and P-value (<0.001) were used as the standard.

(XLSX)

Table S4 The top 20 differentially expressed genes between CG and TG or CD and TD. The top 20 differentially expressed genes after Cd exposure in the gill and digestive gland were analyzed.

(XLSX)

Table S5 Different expressed genes between CG and TG or CD and TD. A list of different expressed genes based on negative binomial distribution.

(XLSX)

Table S6 Pathway enrichment analysis of different expressed genes. A list of pathway enrichment analysis of different expressed genes between TG and CG or TD and CD.

(XLSX)

Table S7 Primers for qRT-PCR validation.

(XLSX)

Acknowledgments

We thank Dr. Yuanhuai Han (College of Agriculture, Shanxi Agriculture University) and Dr. Xingchun Wang (College of life science, Shanxi Agriculture University) for useful advice and assistance.

Author Contributions

Conceived and designed the experiments: XLM LW. Performed the experiments: XLM ML KYJ BJW XT SJS ZYL CWQ. Analyzed the data: XLM XT. Wrote the paper: XLM XT.

References

- Sakurai I, Seto M (2000) Movement and orientation of the Japanese scallop *Patinopecten yessoensis* (Jay) in response to water flow. *Aquaculture* 181: 269–279.
- Sato M, Nagashima K (2001) Molecular characterization of a mitochondrial DNA segment from the Japanese scallop (*Patinopecten yessoensis*): demonstration of a region showing sequence polymorphism in the population. *Mar biotechnol* 3: 370–379.
- Li Q, Xu K, Yu R (2007) Genetic variation in Chinese hatchery populations of the Japanese scallop (*Patinopecten yessoensis*) inferred from microsatellite data. *Aquaculture* 269: 211–219.
- Liu W, He C, Li W, Zhou Z, Gao X, et al. (2010) Discovery of host defence genes in the Japanese scallop *Mizuhopecten yessoensis* Jay by expressed sequence tag analysis of kidney tissue. *Aquac Res* 41: 1602–1613.
- Nagashima K, Sato M, Kawamata K, Nakamura A, Ohta T (2005) Genetic structure of Japanese scallop population in Hokkaido, analyzed by mitochondrial haplotype distribution. *Mar Biotechnol* 7: 1–10.
- Torres MA, Barros MP, Campos SCG, Pinto E, Rajamani S, et al. (2008) Biochemical biomarkers in algae and marine pollution: a review. *Ecotox Environ safe* 71: 1–15.
- Haynes D, Johnson JE (2000) Organochlorine, heavy metal and polyaromatic hydrocarbon pollutant concentrations in the Great Barrier Reef (Australia) environment: a review. *Mar Pollut Bull* 41: 267–278.
- Chelomin V, Bobkova E, Lukyanova O, Chekmasova N (1995) Cadmium-induced alterations in essential trace element homeostasis in the tissues of scallop *Mizuhopecten yessoensis*. *Comp Biochem Phys C* 110: 329–335.
- Evtushenko Z, Lukyanova O, Belcheva N (1990) Cadmium bioaccumulation in organs of the scallop *Mizuhopecten yessoensis*. *Mar Biol* 104: 247–250.
- Shan Z, Li H, Bao X, He C, Yu H, et al. (2011) A selenium-dependent glutathione peroxidase in the Japanese scallop, *Mizuhopecten yessoensis*: cDNA cloning, promoter sequence analysis and mRNA expression. *Comp Biochem Phys B* 159:1–9.
- Chelomin V, Belcheva N (1992) The effect of heavy metals on processes of lipid peroxidation in microsomal membranes from the hepatopancreas of the bivalve mollusc *Mizuhopecten yessoensis*. *Comp Biochem Phys C* 103: 419–422.
- Lukyanova ON, Khotimchenko YS (1995) Lipid peroxidation in organs of the scallop *Mizuhopecten yessoensis* and sea-urchin *Strongylocentrotus intermedius* during the reproductive cycle. *Comp Biochem Phys B* 110: 371–377.
- Slobodskova VV, Zhukovskaya AF, Chelomin VP (2012) DNA damage in the gill cells of the marine scallop *Mizuhopecten yessoensis* during anoxic stress and aerobic recovery. *Ocean Sci J* 47: 95–100.
- Sasaki YF, Izumiya F, Nishidate E, Ishibashi S, Tsuda S, et al. (1997) Detection of genotoxicity of polluted sea water using shellfish and the alkaline single-cell gel electrophoresis (SCE) assay: a preliminary study. *Mutat Res-gen Tox En* 393: 133–139.
- Sato M, Kawamata K, Zaslavskaya N, Nakamura A, Ohta T, et al. (2005) Development of microsatellite markers for Japanese scallop (*Mizuhopecten yessoensis*) and their application to a population genetic study. *Mar Biotechnol* 7: 713–728.
- Wang X, Song B, Qiu X, Meng X (2009) Development of EST-SSRs in scallop (*Patinopecten yessoensis*) from sequence database. *Conserv Genet* 10: 1129–1131.
- Liu W, Li H, Bao X, He C, Li W, et al. (2011) The First Set of EST-derived Single Nucleotide Polymorphism Markers for Japanese Scallop, *Patinopecten yessoensis*. *J World Aquacult Soc* 42: 456–461.
- Wang X, Hu X, Li J, Li L, Hou R, et al. (2012) Characterization of 38 EST-derived SNP markers in Zhikong scallop (*Chlamys farreri*) and their cross-species utility in Yesso scallop (*Patinopecten yessoensis*). *Conservation Genet Resour*: 1–7.
- Xu K, Li Q, Kong L, Yu R (2008) A first-generation genetic map of the Japanese scallop *Patinopecten yessoensis*-based AFLP and microsatellite markers. *Aquac Res* 40: 35–43.
- Hegeudus Z, Zakrzewska A, Ágoston V, Ordas A, Rác P, et al. (2009) Deep sequencing of the zebrafish transcriptome response to mycobacterium infection. *Mol Immunol* 46: 2918–2930.
- Ji P, Liu G, Xu J, Wang X, Li J, et al. (2012) Characterization of common carp transcriptome: Sequencing, de novo assembly, annotation and comparative genomics. *PLoS one* 7: e35152. Doi:10.1371/journal.pone.0035152.
- Xiang L, He D, Dong W, Zhang Y, Shao J (2010) Deep sequencing-based transcriptome profiling analysis of bacteria-challenged *Lateolabrax japonicus* reveals insight into the immune-relevant genes in marine fish. *BMC Genomics* 11: 472.
- Milan M, Coppe A, Reinhardt R, Cancela LM, Leite RB, et al. (2011) Transcriptome sequencing and microarray development for the Manila clam, *Ruditapes philippinarum*: genomic tools for environmental monitoring. *BMC Genomics* 12: 234.
- Lesley JC, Patrick JB, Voelckel C (2008) An approach to transcriptome analysis of non-model organisms using short-read sequences. *Genome Informatics* 21: 3–14.
- Li R, Fan W, Tian G, Zhu H, He L, et al. (2009) The sequence and de novo assembly of the giant panda genome. *Nature* 463: 311–317.
- Wang XW, Luan JB, Li JM, Bao YY, Zhang CX, et al. (2010) De novo characterization of a whitefly transcriptome and analysis of its gene expression during development. *BMC Genomics* 11: 400.
- Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* 18: 1509–1517.
- Peng S, Hu Y, Bai Z (2009) Pollution assessment and ecological risk evaluation for heavy metals in the sediments of Bohai Bay. *Journal of Waterway and Harbor* 30: 57–60. (in Chinese with English abstract).
- Hou R, Bao Z, Wang S, Su H, Li Y, et al. (2011) Transcriptome Sequencing and De Novo Analysis for Yesso Scallop (*Patinopecten yessoensis*) Using 454 GS FLX. *PLoS one* 6: e21560. doi:10.1371/journal.pone.0021560.
- Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, et al. (1999) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 27: 29–34.
- Escobedo-Fregoso C, Mendez-Rodríguez L, Monsalvo-Spencer P, Llera-Herrera R, Zenteno-Savín T, et al. (2010) Assessment of Metallothioneins in Tissues of the Clam *Megapitaria squalida* as Biomarkers for Environmental Cadmium Pollution From Areas Enriched in Phosphorite. *Arch Environ Con Tox* 59: 255–263.
- Zorita I, Apraiz I, Ortiz-Zarragoitia M, Orbea A, Cancio I, et al. (2007) Assessment of biological effects of environmental pollution along the NW Mediterranean Sea using mussels as sentinel organisms. *Environ Pollut* 148: 236–250.
- Wepener V, van Vuren JHJ, Chatiza FP, Mbizi Z, Slabbert L, et al. (2005) Active biomonitoring in freshwater environments: early warning signals from biomarkers in assessing biological effects of diffuse sources of pollutants. *Phys Chem Earth* 30: 751–761.
- Bucheli TD, Fent K (1995) Induction of cytochrome P450 as a biomarker for environmental contamination in aquatic ecosystems. *Crit Rev Env Sci Tec* 25: 201–268.
- Regoli F, Principato G (1995) Glutathione, glutathione-dependent and antioxidant enzymes in mussel, *Mytilus galloprovincialis*, exposed to metals under field and laboratory conditions: implications for the use of biochemical biomarkers. *Aquat Toxicol* 31: 143–164.

36. Gupta SC, Sharma A, Mishra M, Mishra RK, Chowdhuri DK (2010) Heat shock proteins in toxicology: How close and how far? *Life Sci* 86: 377–384.
37. Iwanaga S, Lee BL (2005) Recent advances in the innate immunity of invertebrate animals. *J Biochem Mol Biol* 38: 128–150.
38. Tainer JA, Thayer MM, Cunningham RP (1995) DNA repair proteins. *Curr Opin Struc Biol* 5: 20–26.
39. Li GM (2007) Mechanisms and functions of DNA mismatch repair. *Cell Res* 18: 85–98.
40. Sancar A (1996) DNA excision repair. *Annu Rev Biochem* 65: 43–81.
41. Moreira R, Balseiro P, Planas JV, Fuste B, Beltran S, et al. (2012) Transcriptomics of In Vitro Immune-Stimulated Hemocytes from the Manila Clam *Ruditapes philippinarum* Using High-Throughput Sequencing. *PLoS one* 7(4): e35009. doi:10.1371/journal.pone.0035009.
42. Takeda K, Akira S (2004) TLR signaling pathways. *Semin Immunol* 16: 3–9.
43. Hildebrand JM, Yi Z, Buchta CM, Poovassery J, Stunz LL, et al. (2011) Roles of tumor necrosis factor receptor associated factor 3 (TRAF3) and TRAF5 in immune cell functions. *Immunol Rev* 244: 55–74.
44. Huang XD, Liu WG, Guan YY, Shi Y, Wang Q, et al. (2012) Molecular cloning, characterization and expression analysis of tumor necrosis factor receptor-associated factor 3 (TRAF3) from pearl oyster *Pinctada fucata*. *Fish Shellfish Immunol* 33: 652–658.
45. Wymann MP, Marone R (2005) Phosphoinositide 3-kinase in disease: timing, location, and scaffolding. *Current Opin Cell Biol* 17: 141–149.
46. Kaminska B (2005) MAPK signalling pathways as molecular targets for anti-inflammatory therapy—from molecular mechanisms to therapeutic benefits. *Biochim Biophys Acta* 1754: 253–262.
47. Kumar S, Boehm J, Lee JC (2003) p38 MAP kinases: key signalling molecules as therapeutic targets for inflammatory diseases. *Nat Rev Drug Discov* 2: 717–726.
48. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat methods* 5: 621–628.
49. Wang L, Feng Z, Wang X, Zhang X (2010) DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26: 136–138.
50. Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Bio* 11: R106.
51. Luckenbach T, Epel D (2008) ABCB-and ABCC-type transporters confer multixenobiotic resistance and form an environment-tissue barrier in bivalve gills. *Am J Physiol-Reg I* 294: R1919–R1929.
52. Kingtong S, Chitramvong Y, Janvilisri T (2007) ATP-binding cassette multidrug transporters in Indian-rock oyster *Saccostrea forskali* and their role in the export of an environmental organic pollutant tributyltin. *Aquat Toxicol* 85: 124–132.
53. Owen J, Hedley BA, Svendsen C, Wren J, Jonker MJ, et al. (2008) Transcriptome profiling of developmental and xenobiotic responses in a keystone soil animal, the oligochaete annelid *Lumbricus rubellus*. *BMC Genomics* 9: 266.
54. Snyder MJ (2000) Cytochrome P450 enzymes in aquatic invertebrates: recent advances and future directions. *Aquat Toxicol* 48: 529–547.
55. Moore MN, Allen JL, McVeigh A, Shaw J (2006) Lysosomal and autophagic reactions as predictive indicators of environmental impact in aquatic animals. *Autophagy* 2: 217–220.
56. Ferguson LC, Green J, SurrIDGE A, Jiggins CD (2011) Evolution of the insect yellow gene family. *Mol Biol Evol* 28: 257–272.
57. Company R, Serafim A, Bebianno M, Cosson R, Shillito B, et al. (2004) Effect of cadmium, copper and mercury on antioxidant enzyme activities and lipid peroxidation in the gills of the hydrothermal vent mussel *Bathymodiolus azoricus*. *Mar Environ Res* 58: 377–381.
58. Maria V, Bebianno M (2011) Antioxidant and lipid peroxidation responses in *Mytilus galloprovincialis* exposed to mixtures of benzo (a) pyrene and copper. *Comp Biochem Phys C* 154: 56–63.
59. Wong C, Yeung H, Cheung R, Yung K, Wong M (2000) Ecotoxicological assessment of persistent organic and heavy metal contamination in Hong Kong coastal sediment. *Arch Environ Con Tox* 38: 486–493.
60. Ueng YF, Liu C, Lai CF, Meng LM, Hung YY, et al. (1996) Effects of cadmium and environmental pollution on metallothionein and cytochrome P450 in tilapia. *B Environ Contam Tox* 57: 125–131.
61. Amy S, Tracey L, John P, Carine D, Christine C, et al. (2009) Expression of GIMAP1, a GTPase of the immunity-associated protein family, is not up-regulated in malaria. *Malaria J* 8: 53.
62. Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18: 821–829.
63. Schulz MH, Zerbino DR, Vingron M, Birney E (2012) Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28: 1086–1092.
64. Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, et al. (2002) The Bioperl toolkit: Perl modules for the life sciences. *Genome Res* 12: 1611–1618.
65. Rice P, Longden I, Bleasby A (2000) EMBOSS: the European molecular biology open software suite. *Trends Genet* 16: 276–277.
66. Cheng ZZ, Xue CH, Zhu S, Zhou FF, Xuefeng BL, et al. (2005) GoPipe: streamlined gene ontology annotation for batch anonymous sequences with statistics. *Prog Biochem Biophys* 32: 187–191.
67. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 38: D355–D360.
68. Livak KJ, Schmittgen TD (2001) Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods* 25: 402–408.