

RESEARCH ARTICLE

# Transcriptome Sequencing and Characterization of Japanese Scallop *Patinopecten yessoensis* from Different Shell Color Lines

Jun Ding<sup>1‡</sup>, Le Zhao<sup>1‡</sup>, Yaqing Chang<sup>1\*</sup>, Wenming Zhao<sup>2</sup>, Zhenlin Du<sup>2</sup>, Zhenlin Hao<sup>1</sup>

**1** Key Laboratory of Mariculture & Stock Enhancement in North China's Sea, Ministry of Agriculture, Dalian Ocean University, Dalian, China, **2** Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

‡ These authors contributed equally to this work.

\* [yaqingchang@hotmail.com](mailto:yaqingchang@hotmail.com)



OPEN ACCESS

**Citation:** Ding J, Zhao L, Chang Y, Zhao W, Du Z, Hao Z (2015) Transcriptome Sequencing and Characterization of Japanese Scallop *Patinopecten yessoensis* from Different Shell Color Lines. PLoS ONE 10(2): e0116406. doi:10.1371/journal.pone.0116406

**Academic Editor:** Cynthia Gibas, University of North Carolina at Charlotte, UNITED STATES

**Received:** March 16, 2014

**Accepted:** December 9, 2014

**Published:** February 13, 2015

**Copyright:** © 2015 Ding et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The raw reads produced in P1, P2, P3, and P4 have been submitted to the NCBI SRA database (accession numbers: SRR1185949, SRR1185962, SRR1185963, and SRR1185966).

**Funding:** This research was supported by the National High-Tech Research and Development Program of China (No 2012AA10A410) and the Introduction of Advanced Agricultural Science and Technology Key Project ("948" Project:2011-G12). The funders had no role in study design, data

## Abstract

Shell color is an important trait that is used in breeding the Japanese scallop *Patinopecten yessoensis*, the most economically important scallop species in China. We constructed four transcriptome libraries from different shell color lines of *P. yessoensis*: the left and right shell mantles of ordinary strains of *P. yessoensis* and the left shell mantles of the 'Ivory' and 'Maple' strains. These four libraries were paired-end sequenced using the Illumina HiSeq 2000 platform and contained 54,802,692 sequences, 40,798,962 sequences, 74,019,262 sequences, and 44,466,166 sequences, respectively. A total of 214,087,082 expressed sequence tags were assembled into 73,522 unigenes with an average size of 1,163 bp. When the data were compared against the public Nr and Swiss-Prot databases using BlastX, nearly 30.55% (22,458) of the unigenes were significantly matched to known unique proteins. Gene Ontology annotation and pathway mapping analysis using the Kyoto Encyclopedia of Genes and Genomes categorized unigenes according to their diverse biological functions and processes and identified candidate genes that were potentially involved in growth, pigmentation, metal transcription, and immunity. Expression profile analysis was performed on all four libraries and many differentially expressed genes were identified. In addition, 5,772 simple sequence repeats were obtained from the *P. yessoensis* transcriptomes, and 464,197, 395,646, and 310,649 single nucleotide polymorphisms were revealed in the ordinary strains, the 'Ivory' strain, and the 'Maple' strain, respectively. These results provide valuable information for future genomic studies on *P. yessoensis* and improve our understanding of the molecular mechanisms involved in the growth, immunity, shell coloring, and shell biomineralization of this species. These resources also may be used in a variety of applications, such as trait mapping, marker-assisted breeding, studies of population genetics and genomics, and work on functional genomics.

collection and analysis, decision to publish, or preparation of the manuscript.

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

## Introduction

The Japanese scallop *Patinopecten yessoensis* is a bivalve that is naturally distributed along the coastline of northern Japan, the northern Korean Peninsula, and the Far East region of Russia. Because of its great commercial value, *P. yessoensis* has become one of the most important marine aquacultured shellfish in the north region of China since it was introduced in 1982 [1]. The annual production of scallops reached 1,200,000 tons and 2,000,000 hectares by 2007, accounting for 58 billion yuan of output value in China. More than 50% of this value (200,000 tons and 30 billion yuan of output value) came from *P. yessoensis* [2].

Consumers evaluate the quality and value of seafood according to its color, and certain colors often represent high quality. In *P. yessoensis*, the colors of the left and right shells are typically reddish-brown and white, respectively, and the left shell is typically smaller than the right shell. A small percentage of *P. yessoensis* individuals possess two white shells and display better growth traits than wild-type individuals, despite being raised in similar culture conditions [3]. This fact suggests that the color patterns of scallop shells could be a useful marker for selective breeding. Therefore, our research group's long-term selective breeding program focused on developing an improved 'Ivory' strain of *P. yessoensis* in which both shells were completely white, growth rates were rapid, and scallops were tolerant to high temperatures. As an additional point of comparison, our group created another strain, 'Maple,' a hybrid of ordinary strain and 'Ivory' individuals. This strain has a red and white pattern on the left shell.

The mantle is a unique organ tissue in bivalve mollusks, located midway between the visceral mass and shell. The mantle is in direct contact with the environment and has a sensory function. It can adjust the valves in response to unfavorable environmental conditions [4], and it controls inflow of water into the shell's internal chamber, which is responsible for respiratory and filter feeding purposes. Another main function of the mantle is to secrete biomineralization proteins to form the shell [5]. The mantle pallial and mantle edge are the main secretory tissues. Previous studies on the different shell colors of *P. yessoensis* mainly focused on the scallops' physiological properties [6], genetic structure, and microsatellite markers [7–8]. However, the underlying molecular differences and functional consequences of those differences with respect to shell color and performance among strains are unknown, due to the lack of genomic resources available for *P. yessoensis*. As expressed sequenced tags (ESTs) provide comprehensive information with respect to the dynamics of the scallops' transcriptome, it presents a valuable resource for breeding and research.

Due to cost and time limitations, it is infeasible to consider whole genome sequencing for this species. Fortunately, RNA-Seq based on next-generation sequencing (NGS) is an option. This is a high-throughput technology that offers great advantages for the examination of the fine structure of a transcriptome [9]. Especially when no genome sequence is available, transcriptome sequencing provides an effective way to obtain large amounts of sequence data [10]. In addition, transcriptome sequences exclude non-coding DNA, so the sequences that are obtained contain a high percentage of functional information, helping to reveal the molecular mechanism of functional genes [11–13]. Moreover, the availability of a large number of genetic markers developed using NGS technologies is facilitating trait mapping and marker-assisted breeding [14].

In this study, we sampled the mantles from different shell color lines of *P. yessoensis* and used Illumina paired-end sequencing technology to generate a large EST dataset. Many simple sequence repeats (SSRs) and single-nucleotide polymorphisms (SNPs) were identified. This study is the first characterization of *P. yessoensis* from different shell color lines by analyzing large-scale transcriptome sequences. These sequences will serve as a valuable resource for the development of molecular markers, as well as research on gene mapping, comparative genomics, and functional gene discovery.

**Table 1. Summary statistics of *Patinopecten yessoensis* mantle transcriptome assembly using Trinity software.**

Unigene Number	Max. Length (bp)	Min. Length (bp)	Mean Length (bp)	Median Length (bp)	N50	N90	Total Length (bp)
	33,371	300	1,163	714	1,768	484	85,514,539

doi:10.1371/journal.pone.0116406.t001

## Results and Discussion

### Paired-end sequencing and assembly

Four *P. yessoensis* cDNA libraries representing different shell color lines were constructed and used for Illumina paired-end sequencing to generate representative transcripts of a wide range of biological processes. The P1 library represents the left shell mantle of the ordinary strain of *P. yessoensis*, the P2 library represents the right shell mantle of the ordinary strain, the P3 library represents the left shell mantle of the ‘Ivory’ strain, and the P4 library represents the left shell mantle of the ‘Maple’ strain. A total of 214,105,082 raw reads with an average length of 101 bp were acquired with four libraries. The raw reads produced in P1, P2, P3, and P4 have been submitted to the NCBI SRA database (accession numbers: SRR1185949, SRR1185962, SRR1185963, and SRR1185966). After the low-quality reads were filtered out of the sequence data, 166,521,376 (77.78%) high-quality reads remained and were used for the *de novo* assembly. An overview of the sequencing procedure is presented in [S1 Table](#). Because no reference genome exists for *P. yessoensis*, the high-quality reads from all four libraries were combined and assembled into a reference transcriptome using Trinity software[15]. This assembly yielded a total of 73,522 unigenes with an average length of 1163 bp, a minimum length of 300 bp, and a maximum length of 33,371 bp. An overview of the sequencing and assembly process is presented in [Table 1](#).

### Functional annotation of the *P. yessoensis* transcriptome

Several complementary methods were used to annotate the *P. yessoensis* transcriptome assembly. First, the unigenes were compared against the public Nr and Swiss-Prot databases using BlastX (E-value < 1e<sup>-5</sup>). A total of 22,458 (30.55%) unigenes were assigned to biological functions, leaving more than half of the unigenes (69.45%) not matched to known genes. The inability to annotate a large percentage of unigenes is likely a consequence of the paucity of sequences available in public databases from phylogenetically closely related species. Similar situations exist in the cases of other marine animals, including *Arctica islandica* (32.84% annotation rate after transcriptome analysis)[16] and *Laternula elliptica* (16.93% annotation rate after transcriptome analysis)[17].

Moreover, the length of query sequences always influences the results of Blast comparison, so short reads obtained from sequencing are seldom matched to known genes [18]. In our study, sequences shorter than 300 bp were eliminated from assembly results in order to ensure that gene matches were meaningful. Among sequences shorter than 999 bp, only 14.54% were annotated, and among sequences longer than 1000 bp, the rate increased to 58.19% ([Table 2](#)). Annotation efficiency increased with the length of the sequence.

As for the species distribution of these successfully annotated unigenes, 62.14% of the hits matched to the Bivalvia class in general, including 13,577 (60.45%) sequences that were annotated to proteins from *Crassostrea gigas*; 137 (0.61%) matched to *Azumapekten farreri*; 69 (0.31%) matched to *Mytilus galloprovincialis*; 45 (0.20%) matched to *Argopecten irradians*; 42 (0.18%) matched to *Pinctada fucata*; and 66 (0.30%) matched to another Bivalvia species. In previous work, Hou et al. [19] matched only 4.10% of sequences from the

**Table 2. Summary statistics of functional annotation of *Patinopectin yessoensis* transcriptome.**

Category	All sequences	300–999bp	≥ 1000 bp
Total number of unigenes	73,522	46,670	26,852
Unigene matches against Nr and Swiss-Prot	22,410	6,787	15,623
Unigene matches against GO	13,575	3,226	10,349
Unigene matches against KEGG	15,179	3,762	11,417

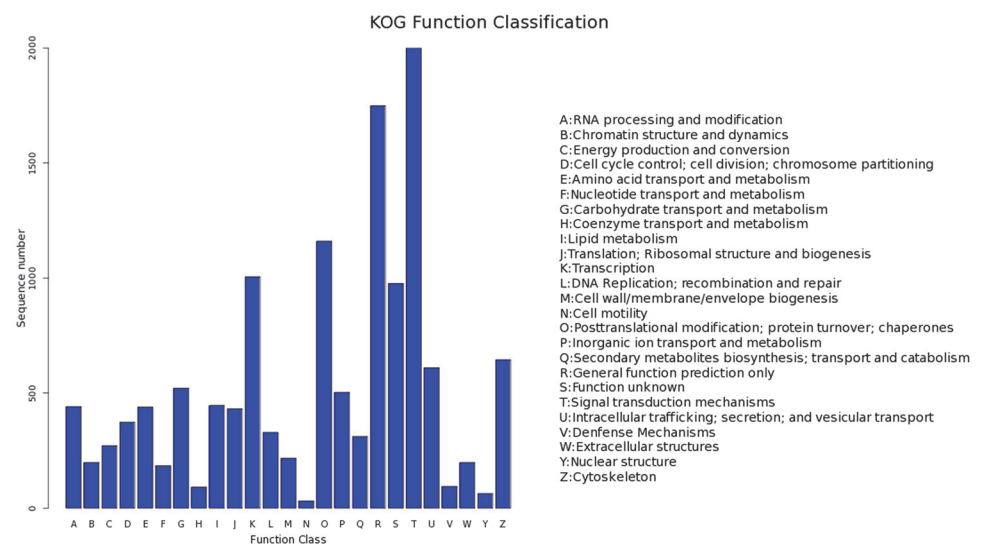
doi:10.1371/journal.pone.0116406.t002

*P. yessoensis* transcriptome to the Bivalvia class. This result suggests that as large-scale sequencing of marine animals continues to expand, annotation efficiency will increase. A total of 599 (2.67%) annotated unigenes in this study matched to prokaryotes and protozoans, apparently due to contamination from seawater, and these sequences have been removed. Therefore, a total of 21,859 annotated unigenes (S2 Table) were applied to the subsequent analysis, including the mapping of functional genes, analysis of differential expression, and SSR and SNP mining.

### EuKaryotic Orthologous Groups (KOG) classification

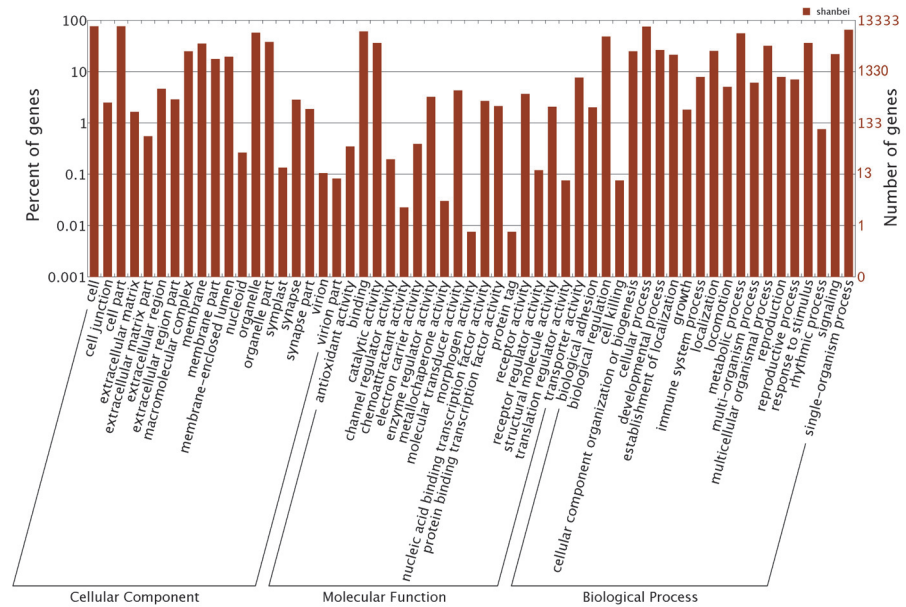
KOG analysis was carried out to provide a deeper understanding of the functions of the unigenes. About 13,905 unigenes were classified into 25 functional categories. The category of ‘signal transduction mechanisms’ contained the largest number of unigenes (2,619, 18.83%) (Fig. 1), followed by the ‘general function prediction only’ cluster (1,749, 12.58%) and the ‘post-translational modification; protein’ cluster (1160, 8.34%).

The categories of greatest interest in the present study were inorganic ion transport and metabolism (503, 3.61%), defense mechanisms (94, 0.68%) and signal transduction mechanisms (2,619, 18.83%). Because the genes in these categories were likely related to immune function, biomineralization, and shell coloring, these categories should be considered for the development of molecular markers in the *P. yessoensis* breeding programs.



**Fig 1. EuKaryotic Orthologous Groups (KOG) function classifications of *Patinopectin yessoensis*.**

doi:10.1371/journal.pone.0116406.g001



**Fig 2. Gene Ontology classifications of assembled sequences in *Patinopecten yessoensis* transcriptome.**

doi:10.1371/journal.pone.0116406.g002

### Gene Ontology (GO) classification

In addition to KOG analysis, we conducted GO analysis on the basis of sequence similarities to known proteins in the Nr databases. This analysis provided hierarchical relationships that represent information on molecular functions, cellular components, and biological processes. A total of 13,333 unigenes were annotated by GO analysis with one or more GO term (Fig. 2), for a total of 1,037 GO assignments. Among these, 10,457 unigenes were annotated to the ontology of molecular functions, 11,963 to biological processes, and 10,997 to cellular components. For cellular components, the major represented categories were cell (GO: 0005623) and cell part (GO:0044464). For biological processes, cellular processes (GO: 0009987) was the most represented GO term, followed by single-organism processes (GO: 0044699). Genes involved in other important biological processes, such as growth, immune system processes, and biological regulation were also identified. Furthermore, we also found a number of unigenes that were involved in interesting categories, such as biomineralization and pigmentation, which may play a role in the shell and shell color formation (S3 Table). Regarding molecular functions, genes involved in binding (GO: 0005488) and metabolic process (GO: 0008152) were highly represented.

### Kyoto Encyclopedia of Genes and Genomes (KEGG) classification

KEGG pathway analysis was carried out on the assembled unigenes to reveal the biochemical pathways operating in *P. yessoensis*. The results annotated 8,500 unigenes into 292 different pathways (Table 3). Among these, metabolic pathways contained the largest number of unigenes, and there were three major subgroups involved in carbohydrate metabolism, amino acid metabolism, and lipid metabolism, respectively (Fig. 3). Another pathway of interest included the subgroups of betalain biosynthesis, flavonoid biosynthesis, indole alkaloid biosynthesis, anthocyanin biosynthesis, calcium signaling, and carotenoid biosynthesis. These pathways are all related to the synthesis of biological pigments, so we inferred that the pathways probably play significant roles in biomineralization and shell coloring.

**Table 3. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway mapping for *Patinopecten yessoensis*.**

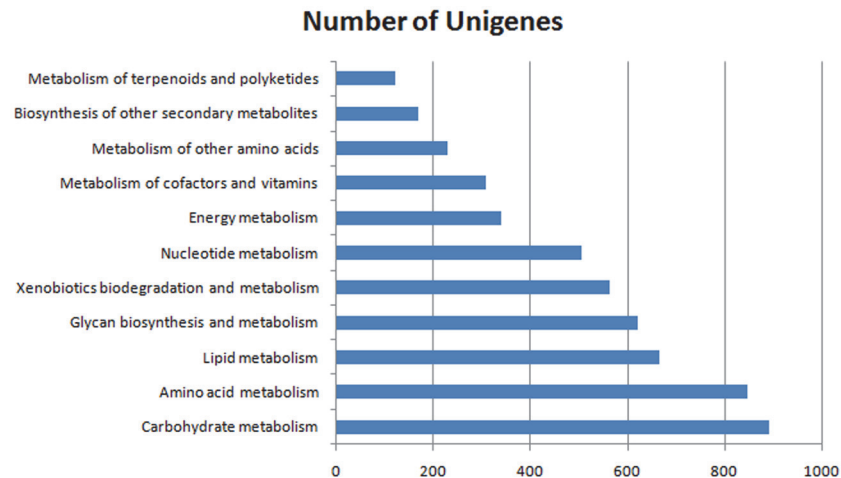
KEGG Pathways	Sub-Pathways	Number of Unigenes
<b>Metabolism</b>	Carbohydrate metabolism	893
	Energy metabolism	341
	Lipid metabolism	667
	Amino acid metabolism	848
	Metabolism of other amino acids	231
	Glycan biosynthesis and metabolism	622
	Metabolism of cofactors and vitamins	311
	Metabolism of terpenoids and polyketides	123
	Biosynthesis of other secondary metabolites	171
	Xenobiotics biodegradation and metabolism	564
	Nucleotide metabolism	506
<b>Genetic Information Processing</b>	Transcription	406
	Translation	608
	Folding, sorting, and degradation	909
	Replication and repair	279
<b>Environmental Information Processing</b>	Membrane transport	84
	Signal transduction	1,665
	Signaling molecules and interaction	769
<b>Cellular Processes</b>	Transport and catabolism	971
	Cell motility	253
	Cell growth and death	944
	Cell communication	945
<b>Organismal Systems</b>	Immune system	1,332
	Endocrine system	891
	Circulatory system	277
	Digestive system	851
	Excretory system	257
	Nervous system	623
	Sensory system	332
	Development	350
Environmental adaptation	192	

doi:10.1371/journal.pone.0116406.t003

### Functional genes involved in growth, immunity, and biomineralization

The important economic traits of growth and immunity have been the focus of much work on economically valuable marine shellfish [19–21]. The unigene and annotation information from Nr, Swiss-Prot, GO, and KEGG all provide valuable genomic resources for *P. yessoensis* and will enable further study of the molecular basis of growth, immune function, and shell biomineralization.

Many environmental factors, including water temperature and salinity, hormone effects, and molecular factors, affect the growth and development of *P. yessoensis* [22–23]. In the present study, we identified a series of genes related to growth, such as transforming growth factor (TGF)-beta regulator 1, TGF-beta receptor-associated protein 1, vascular endothelial growth factor and receptor, acidic fibroblast growth factor intracellular-binding protein, bone morphogenetic protein, connective tissue growth factors, epidermal growth factor and receptors, fibroblast growth factor and receptors, and mitogen-activated protein kinase.



**Fig 3. Categories classified by Kyoto Encyclopedia of Genes and Genomes (KEGG).**

doi:10.1371/journal.pone.0116406.g003

The living environment of *P. yessoensis* is filled with various kinds of parasites and pathogens, but scallops have a fairly complete immune system that protects them from infections and other damage [24]. Unlike vertebrates, shellfish have no specific immune lymphoid cells or antibodies, and their cellular immunity relies on phagocytosis [25]. We classified genes involved with immune function, including those that produce lysozyme, acid phosphatase, alkaline phosphatase, beta-glucuronidase, lectin, superoxide dismutase, catalase, toll-like receptor, heat shock protein, stress-associated endoplasmic reticulum protein, and stress-activated protein kinase JNK.

Among these genes, the ones that produce lysozyme include the lysosomal enzymes acid phosphatase, alkaline phosphatase, and  $\beta$ -glucuronidase. During phagocytosis, lysosomal enzyme is released into the blood serum [26]. Lysozyme is a ubiquitous antibacterial enzyme that lyses bacterial cell walls and occurs in many tissues and secretions. Extensive studies have suggested that it plays an important role in the body's defenses against infection [27]. Acid phosphatase and alkaline phosphatase have direct antibacterial activities and could function as regulatory factors to influence the phagocytic process. Lectins have been reported in many marine bivalves, such as clams, oysters, and mussels. They are important pattern recognition receptors that recognize pathogen-associated molecular patterns and deliver appropriate signals to the cell to alert the innate immune system [28].

For KEGG analysis, a total of 1,332 sequences were classified into 15 immune-response pathways, including the chemokine signaling pathway, complement and coagulation cascades, antigen processing and presentation, the toll-like receptor signaling pathway, the NOD-like receptor signaling pathway, the RIG-I-like receptor signaling pathway, Fc gamma R-mediated phagocytosis, the cytosolic DNA-sensing pathway, and leukocyte transendothelial migration. These pathways may play important roles in the body immunity of *P. yessoensis*.

Shells, one of the most interesting features of bivalves, are secreted by the mantle in a process called biomineralization. Chitin is an important component in bivalve nacre formation that promotes the biomineralization process by forming the framework for other macromolecular components [29]. We identified four genes related to chitin biosynthesis: chitin synthase, chitin deacetylase isoform A, chitinase A1, and chitinase domain-containing protein 1. Calcium metabolism also plays an interesting role in shell formation, as more than 95% of the nacre's weight is calcium carbonate [30]. We identified several genes related to calcium

**Table 4. Statistics regarding alignment of sequences.**

	P1	P2	P3	P4
Number of reads	41,865,770	30,897,906	57,955,052	35,802,648
Number of reads mapped	39,189,962	28,733,671	54,285,922	33,500,633
% of reads mapped	93.61%	93.00%	93.67%	93.57%
Total number of unique reads	36,699,002	26,871,380	50,633,697	31,232,858
Total % of unique reads	87.66%	86.97%	87.37%	87.49%

doi:10.1371/journal.pone.0116406.t004

metabolism, such as calreticulin, calmodulin, calmodulin-binding protein, calcitonin receptor, calcium-binding protein, calcium and integrin-binding protein, and calcyclin-binding protein.

Functional analysis using the Illumina sequencing database identified genes that are potentially related to growth, immunity, and biomineralization. Further experiments should demonstrate the functions and expression patterns of these candidate genes and should analyze their potential roles in growth and immune function of *P. yessoensis* of different shell color lines.

### Differential gene expression among the four shell color strains

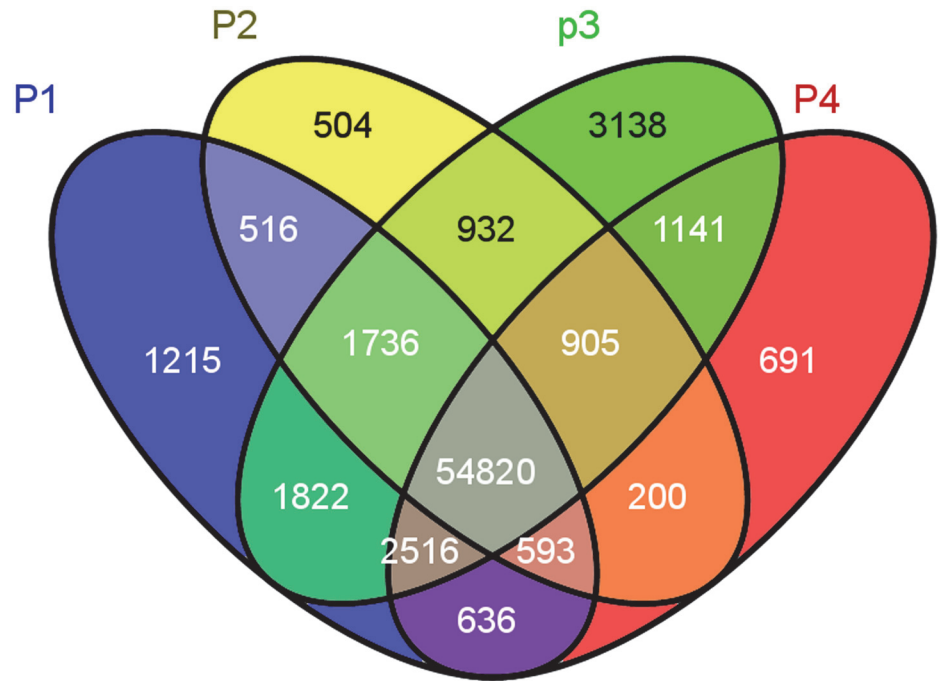
A total of 21,859 annotated unigenes which did not match prokaryotic and protozoan sequences were used as the reference transcriptome for mapping each set of reads from the different shell morphs. The reads of the four *P. yessoensis* transcriptome libraries generated by Illumina paired-end sequencing were mapped to assembly reference sequences. The mapping rates exceeded 93% for total reads mapped and 86% for total unique reads mapped (Table 4).

### Variation of gene expression among different shell color lines of *P. yessoensis*

To characterize the transcriptome of *P. yessoensis* strains with different shell colors, we collected samples from ordinary strains of *P. yessoensis* (left and right shell mantle), the ‘Ivory’ strain, and the ‘Maple’ strain (left shell mantle). These four samples represented three distinct strains of *P. yessoensis*, and we compared the gene expression of these four transcriptomes. A total of 63,854, 60,206, 67,010, and 61,502 genes were detected in the P1, P2, P3, and P4 libraries, respectively, and 54,820 genes were co-expressed among all four transcriptome libraries (Fig. 4). Of these, 3,138 genes were exclusively expressed in P3, which is more than in any other library, and 1,215, 504, and 691 genes were detected only in the P1, P2, and P4 transcriptomes, respectively (Fig. 4). Transcriptomes of P1 and P2, which come from the “original” strain of *P. yessoensis*, exhibited high similarity and were grouped first (Pearson’s correlation  $r = 0.843$ ); next, P4 joined in them as the first group, and finally, P3 formed the second group (Fig. 5). The gene expression profiles among the mantle of the P2 library, which produced a white shell (the right shell mantle of the ordinary strain) were similar to those of P3 (shell mantle of the ‘Ivory’ strain) (Pearson’s correlation  $r = 0.761$ ), as expected. The variation observed in the transcriptomes of different *P. yessoensis* strains reflects the overall effects of all non-genetic and genetic factors, and it is clear that more changes in gene expression occur in the ‘Ivory’ strain of *P. yessoensis*.

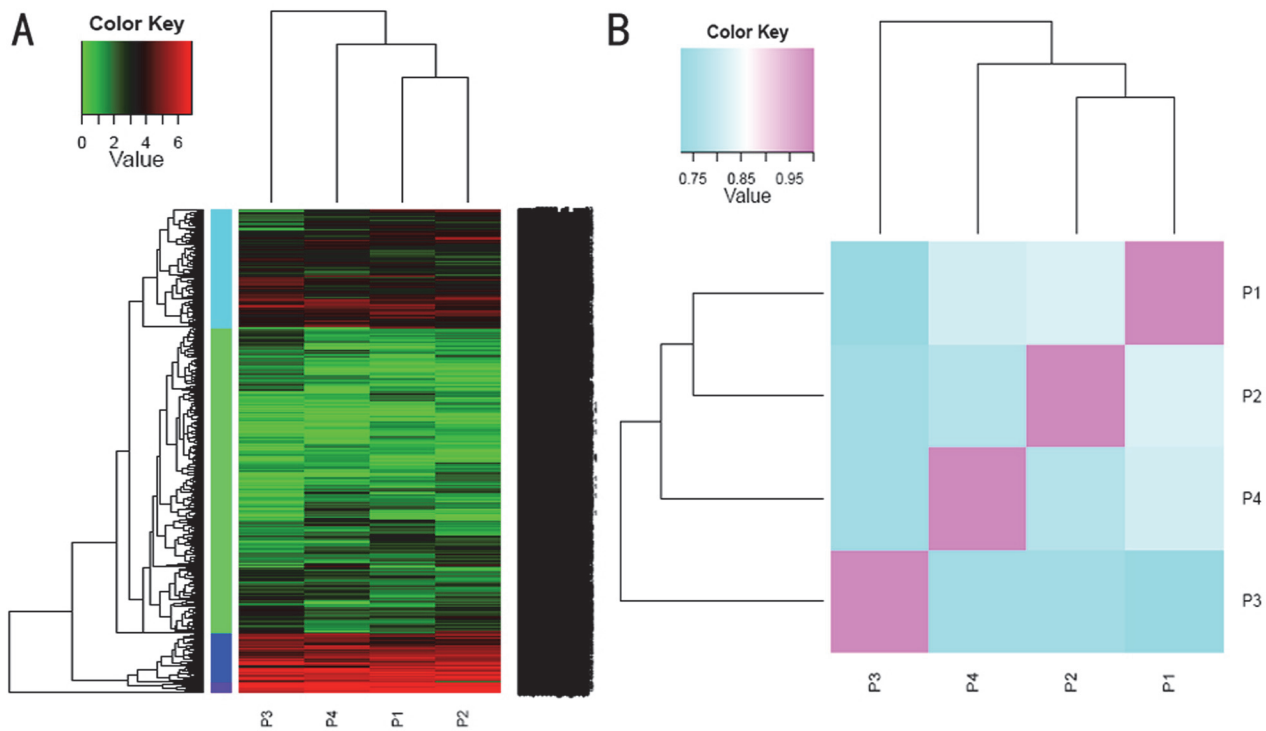
In order to determine the differences in *P. yessoensis* from different shell color lines more accurately, we used DEGseq [31] to screen the differentially expressed genes (DEG) ( $p < 0.05$ ) with a relation model that chose P3 as the control group and compared it with the other three transcriptome libraries. We found 2,338 differentially expressed genes with 1,097 up-regulated genes and 1,241 down-regulated genes in the P1 library, compared with the control group





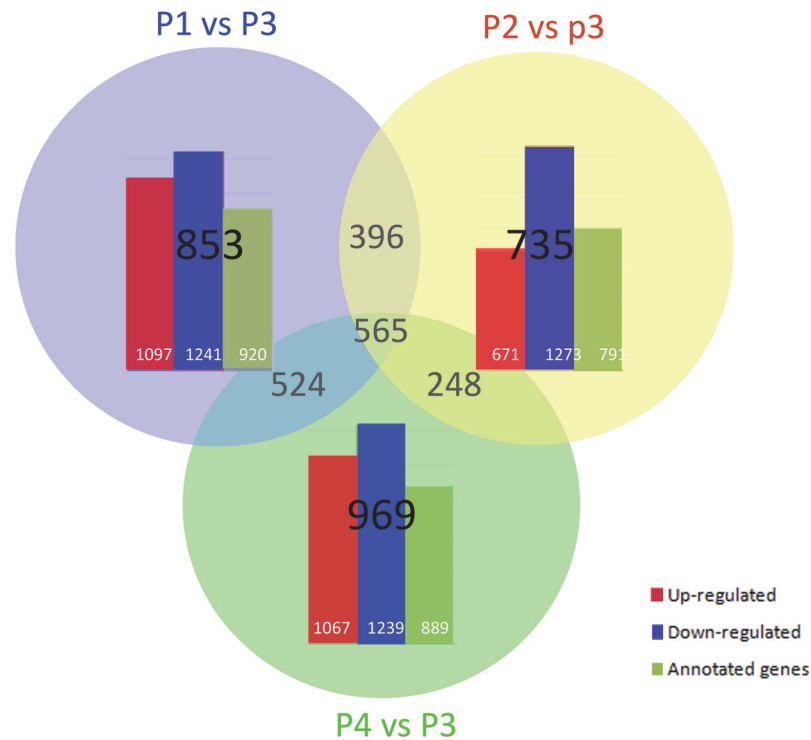
**Fig 4. Venn diagram of the expressed genes along the four different *Patinopecten yessoensis* transcriptomes.** A total of 54,820 genes were co-expressed among the four samples. The Venn diagram serves as an interactive tool for comparing the list, which was plotted using Venn.

doi:10.1371/journal.pone.0116406.g004



**Fig 5. Comparisons of transcriptional profiles across samples.** (a) Hierarchical clustering of transcripts and samples. (b) The hierarchically clustered Spearman correlation matrix resulting from comparing the transcript expression values for each pair of samples.

doi:10.1371/journal.pone.0116406.g005



**Fig 6. Changes in gene expression profiling among the different treatment.** Up-regulated (red), down-regulated (blue), and annotation (green) unigenes were quantified and presented by histogram. Comparisons of DEG in P1/P3, P2/P3, and P4/P3 are presented by Venn chart.

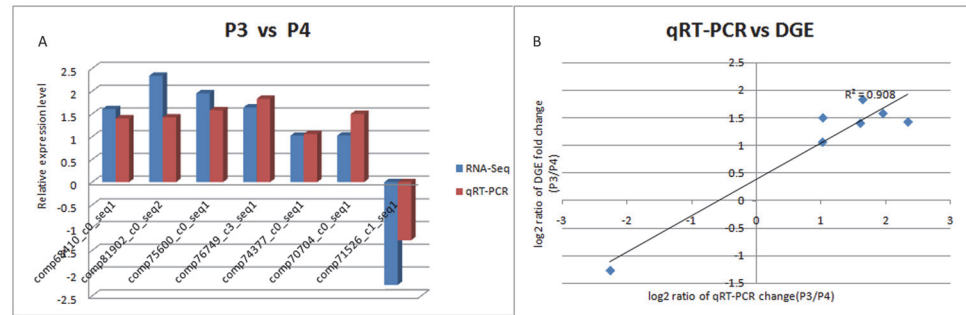
doi:10.1371/journal.pone.0116406.g006

library (P3). A total of 1,944 and 2,306 DEGs were also identified in the P2 and P4 libraries, respectively, compared with P3 (Fig. 6). To confirm whether our sequencing and analysis were reliable and valid, we randomly selected seven DEGs and measured their expression in the same RNA sample of P3 and P4 by real-time reverse transcription (RT)-PCR. All seven genes showed uniformly consistent results in RT-PCR and transcriptome sequencing (Fig. 7), which indicates that transcriptome sequencing was reliable and we can make reasonable deductions from the functional enrichment analysis of the DEGs.

### Functional analysis of differentially expressed genes

In a comparison of the DEGs with a reference gene database (21,859 annotated unigenes), 920, 791 and 889 DEGs were annotated, in P1/P3, P2/P3, and P4/P3, respectively (Fig. 6). However, a large number of DEGs could not be annotated, including some highly expressed genes. In order to better understand the functions of DEGs, GO and KEGG pathway enrichment analysis was carried out on the DEGs.

Betalain biosynthesis (ko00965,  $p < 0.01$ ) and tyrosine metabolism (ko00350,  $p < 0.01$ ) were significantly enriched in the DEGs of P1/P3, P2/P3, and P4/P3; Moreover, across all three comparisons (P1/P3, P2/P3, and P4/P3) 565 DEGs (248 annotated genes) were consistently differentially expressed. As betalain biosynthesis and tyrosine metabolism were also significantly enriched among these genes, we hypothesize these two pathways were involved in coloration. Betalains are natural pigments, nitrogen-containing water-soluble compounds derived from tyrosine, that confer yellow/red colors [32]. The intermediate products derived from the oxidation of tyrosine and indoles constitute melanin, which is an irregular light-absorbing



**Fig 7. Quantitative reverse-transcription-PCR validation of differentially expressed gene (DEG) results.** (a) Fold changes measured by mRNA-Seq and real-time RT-PCR. (b) Correlation between the expression fold change level of DEG between P3 and P4.

doi:10.1371/journal.pone.0116406.g007

polymer [33]. Furthermore, we found that tyrosinase-like protein existed in both the betalain biosynthesis and tyrosine metabolism pathways, and this gene was expressed to a greater extent in the mantles with red shells. Whether tyrosinase-like protein is related to the formation of red shells is an interesting question that will be explored in future study.

The shell is formed by the mantle in a process that is similar to pearl formation [34]. Although there are different views on the relationship between the amount of trace metals and the color of the shell and pearl, researchers generally believe that trace metals are linked closely with shell and pearl coloring [34]. Some amount of the trace metals is stored in the aragonite in its ionic state, and the other trace metals unite with porphyrins to form metal-porphyrins, resulting in different colors in shells and pearls [35–36]. In our study, the pathways of porphyrin and chlorophyll metabolism (ko00860) were also enriched in the DEGs from the P3 library over those in the P1, P2, and P4 libraries. Some ferritin, which has a lower expression in the mantle of ‘Ivory’ strains of *P. yessoensis*, was also found in these pathways.

In order to determine which metal ions mainly affected the color of the shell in *P. yessoensis*, we compared the amount of metal elements in red shells and white shells. The amounts of Fe and Zn in red shells were significantly higher, 3.04 and 2.41 times these amounts of white shells, respectively (Table 5). Through analysis of these DEGs, we obtained several genes that are involved in metal transport. It has been shown that ferritin controls the concentration and distribution of iron in mollusk shells, sculpting shell morphology and coloration [37–38]. Metalloreductase STEAP2, divalent metal transporter 1, and zinc transporter ZIP12 also play important roles in Fe<sup>2+</sup> and Zn<sup>2+</sup> transport [39–42]. In addition, these genes in the P1 library (representing the reddish-brown shell) existed at higher levels than in the other libraries (representing the white and red-and-white shells). The measured results of trace elements were confirmed by the transcriptome analysis result, and both of these analyses suggest that formation of the reddish-brown shell requires more iron and zinc ions. Moreover, iron-porphyrins and zinc-porphyrins also produce red and pink colors, respectively [43].

The DEG identification can be improved by searching for SNPs that are associated with these genes. Further identification of SNPs within these genes could explain shell color

**Table 5. Amounts of trace elements measured in *Patinopecten yessoensis* (mg/kg).**

Sample	Mn	Cu	Zn	Fe	Mg
Left shell of ‘Ivory’ strain	17.10	> 0.10	4.02	18.40	938.00
Left shell of ordinary strain	19.20	> 0.10	9.70	55.6	964.00

doi:10.1371/journal.pone.0116406.t005

variation between different shell color lines and could provide a foundation for further studies into the molecular mechanisms controlling mollusk shell coloration.

## Detection of SSRs and SNPs

Transcriptome is a valuable resource for the effective and convenient development of genetic markers. Both SSRs and SNPs are valuable molecular markers for the *P. yessoensis* breeding program and offer the greatest potential for identifying functional genes of economically important traits. Approximately 3 to 7% of expressed genes contain putative SSR motifs [44], and SSR markers have been widely used to construct genetic maps in marine animals [45]. In our study, we performed a general screen for SSRs on the sequences that were integrated from the four *P. yessoensis* transcriptomes. The screening produced a total of 5,772 SSRs with 142 motifs. Dinucleotides (2,639) and trinucleotides (2,989) were major types of SSRs, followed by tetranucleotides (131) and pentanucleotides (13). The most common SSR type, considering sequence complementarity, was AAC/GTT (1,191), followed by AG/CT (1,090), AT/AT (901), AAG/CTT (506), AAT/ATT (317), and ATC/ATG (274).

The ESTs generated in our study through Illumina sequencing originated in three different strains of *P. yessoensis* samples. We screened for SNPs that were present in all three strains. In ordinary strains of *P. yessoensis*, a total of 464,197 high-quality SNPs were generated, among which 228,136 were transitions, 183,247 were transversions, and 52,814 were in indels. In 'Ivory,' there were 395,646 high-quality SNPs, among which 192,782 were transitions, 156,693 were transversions, and 46,171 were indels. In 'Maple,' there were 310,649 high-quality SNPs, among which 153,280 were transitions, 122,883 were transversions, and 34,486 were indels (Fig. 8).

The SSRs and SNPs generated from this study constitute an important and valuable resource for further studies on the analysis of marker development, genetic linkage mapping, and interesting traits in *P. yessoensis*.

## Conclusions

We performed *de novo* transcriptome sequencing for *P. yessoensis* from different shell color lines using the Illumina platform. A large number of candidate genes potentially involved in growth, immunity, and shell biomineralization were identified and are worthy of further investigation. In addition, genes that are differentially expressed in different shell color lines of *P. yessoensis* were identified and functionally annotated using the GO and KEGG databases. This information provides potential molecular targets in shell coloring and metal transcription. A large number of SNPs and SSRs identified in this study will provide a valuable material foundation for future genetic linkage analysis and will help to further aquaculture breeding programs for this species.

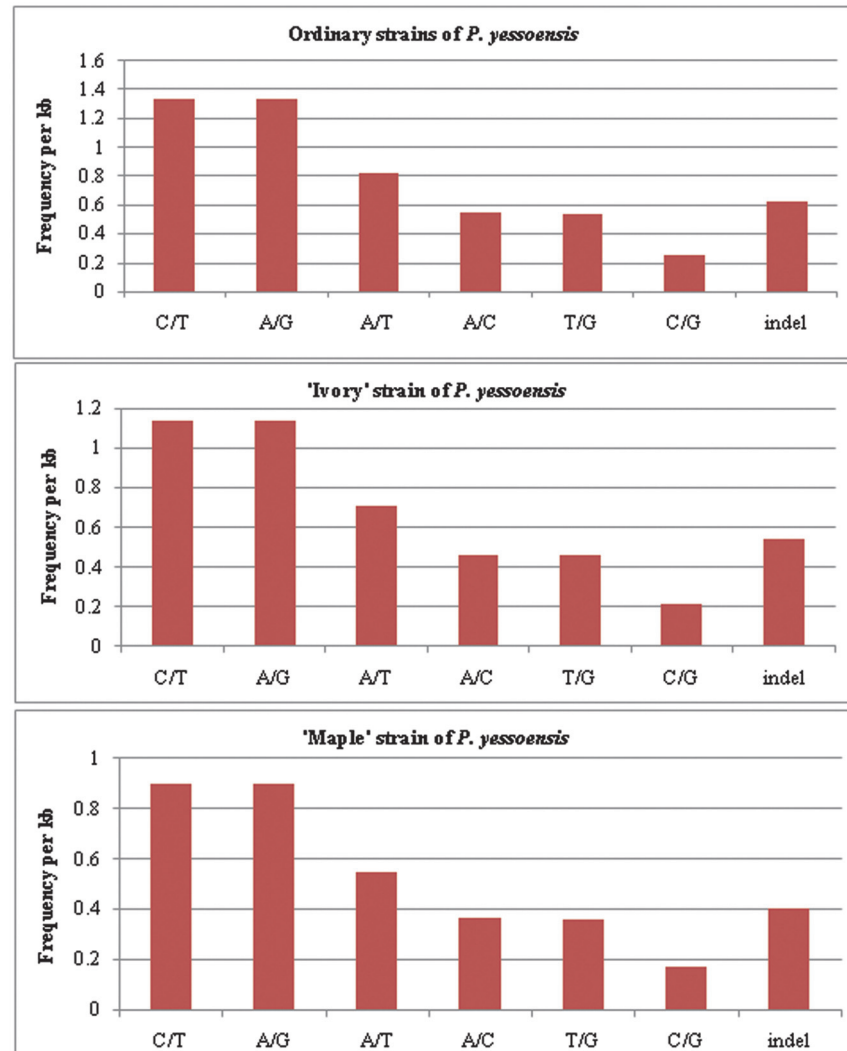
## Materials and Methods

### Ethics statement

Not applicable. Our research did not involve human participants or samples. Housing and care of *P. yessoensis* and collection of tissue samples for use in the experiments described were conducted in accordance with the International Guiding Principles for Biomedical Research Involving Animals (<http://www.cioms.ch/frame> 1985 texts of guidelines.html).

### Sampling of and RNA extraction from mantle tissues of *P. yessoensis*

In order to obtain more and better information from the transcriptomes of different color lines of *P. yessoensis* by Illumina paired-end sequencing, we collected 2-year-old live individuals of ordinary, 'Ivory,' and 'Maple' strains of *P. yessoensis*. We maintained all of these scallops in the same



**Fig 8. Frequency distribution of single nucleotide polymorphisms identified in *Patinopectin yessoensis* transcriptsomes from different shell color lines.**

doi:10.1371/journal.pone.0116406.g008

cultivation environment, at the Key Laboratory of Mariculture, Ministry of Agriculture, Dalian, Liaoning Province, China, in 2013. The scallops were healthy and homogeneous in size, with shell weights of  $12.64 \pm 0.32$  g and shell lengths of  $4.82 \pm 0.23$  cm. The mantles of the left and right shells were randomly collected from at least three independent ordinary strains of *P. yessoensis* as samples (Sample P1 and Sample P2) and were used for RNA extraction, as were the mantles of the left shell of 'Ivory' and 'Maple' (Sample P3 and Sample P4). Total RNA was isolated from each sample using an RNAprep pure tissue kit (TIANGEN, CHN), following the manufacturer's protocol. The quantity and quality of total RNA were confirmed using the NanoDrop2000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and gel electrophoresis.

### cDNA library preparation for Illumina sequencing

Each paired-end cDNA library was generated using the Genomic Sample Prep Kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. The quality of the library

was assessed using Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), and cluster amplification was performed using the TruSeq PE Cluster Kit and a cBot (Illumina). The four cDNA libraries were sequenced with a paired-end module at Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China.

### *De novo* assembly of Illumina sequencing reads

Before assembly, the 101-bp raw paired-end reads were filtered to obtain high-quality clean reads by removing adaptors, PCR duplicates, and low-quality sequences (reads with a base quality below 20). Due to the lack of genetic information about *P. yessoensis*, the clean reads from four libraries were mixed together as a reference database, and *de novo* assembly of these clean reads was performed using the short reads assembling program Trinity (default parameters were used) [15]. All assembled sequences were compared with the NCBI non-redundant protein database, Swiss-Prot database, the KEGG pathway, and the KOG database using BlastX with an E-value of less than  $10^{-5}$ , and the best aligned results were used to decide the sequence direction of unigenes. If the results of different databases conflicted, a priority order of Nr, Swiss-Port, KEGG, and KOG was followed. When a unigene did not align to any of the databases, ESTScan was used to predict the sequence direction [46].

### Gene annotation and classification

Functional annotations were performed by sequence comparison with public databases. The unigenes were compared against the Nr, Swiss-Prot, KEGG, and KOG databases by BlastX (E-value  $< 10^{-5}$ ), and the highest sequence similarity with the given unigenes was defined as the unigene functional annotation and used in the following analysis. For Nr annotation, we used the Blast2GO program to obtain the GO annotations of the unigenes (E-value  $< 10^{-5}$ ) [47]. We also used WEGO software to conduct GO functional classification for all unigenes [48]. Unigenes were aligned to the KOG database to predict and classify possible functions, and pathway annotation was performed using Blastall software against the KEGG database [49]. In addition, any elements of the unigenes that matched to prokaryote and protozoan databases were considered contaminants and were removed from the analysis.

### Comparison of the transcriptomes of four cDNA libraries

The clean reads from four cDNA libraries were mapped on the reference database by BWA software with the default parameter value [15], and the mapping rates were all around 80%. The RPKM (reads per kilobase per million reads) [50] was applied to measure the gene expression levels. We used the calculated gene expression to compare the differences in gene expression between four samples. We performed pair-wise comparisons of the gene expression of these four transcriptomes using the DEGseq [31] package, and  $p < 0.05$  was used as the threshold to screen the differentially expressed genes. Then, the hierarchical clustering of transcripts and samples was used to show the relative expression levels of each transcript in each sample, and a comparison of the transcript expression values for each pair of samples was used to show the hierarchically clustered Spearman correlation matrix [51]. For pathway and GO enrichment analysis, we selected  $p < 0.05$  as a threshold to identify the significantly enriched KEGG and GO terms.

To assess the reliability of our sequencing and analysis by real-time RT-PCR, we used the same RNA samples for sequencing from P3 and P4, and we synthesized cDNA using the PrimeScript RT reagent Kit (TaKaRa, DL, CHN). Real-time PCR was performed in triplicate according to the manufacturer's instructions by using the SYBR Premix Ex Taq (Tli RNaseH Plus) Kit (TaKaRa, DL, CHN) on an ABI 7500 real-time PCR system and analyzed by the  $2^{-\Delta\Delta Ct}$  method using  $\beta$ -actin as the reference control [52]. All the primer sequences used for RT-PCR are listed in S4 Table.

## Trace metal composition in the reddish-brown shell and white shell

Samples of red and white shell were collected from the left shells of live individuals of ordinary and 'Ivory' strains of *P. yessoensis* from the Key Laboratory of Mariculture, Ministry of Agriculture, Dalian, Liaoning Province, China, in 2013. The sampled *P. yessoensis* individuals were healthy and homogeneous in size, with a shell weight of  $4.64 \pm 0.32$  g and a shell length of  $2.82 \pm 0.23$  cm. We opened the scallops with a scalpel, ensured the integrity of the shell, and washed the sample with deionized water. Then, these shells were oven-dried at 65°C for more than 24 h, the reddish-brown shell and the white shell were transferred to separate mortars, and both were ground to a fine powder. Finally, the amounts of the trace metals Mn, Cu, Zn, Fe, and Mg from both samples were evaluated using atomic absorption spectrometry.

## The identification of SNPs and SSRs

All types of SSRs were identified and localized microsatellite motifs were identified using MICroSATellite (MISA, <http://pgrc.ipk-gatersleben.de/misa/>) software. The criteria were that sequences had at least six repeats of dinucleotide and five repeats of all other motifs (from trinucleotide to hexanucleotide).

As the four transcriptomes of differently colored shell lines belong to three strains of *P. yessoensis*, we screened for the presence of SNPs in all three strains. We identified potential SNPs using the GATK program [53] and filtered the results using two parameters (SNP quality  $\geq 30$  and reads depth  $\geq 5$ ), producing 467,098 putative SNPs.

## Supporting Information

**S1 Table. Summary statistics of *Patinopecten yessoensis* sequencing raw reads.**  
(XLSX)

**S2 Table. Sequences with significant BLAST matches against biological functions.**  
(XLSX)

**S3 Table. Statistics of the Gene Ontology (GO) term involved in immune, biomineralization, and pigmentation.**  
(XLSX)

**S4 Table. Statistics on primer sequences.**  
(XLSX)

## Acknowledgments

We thank Mr. Junwei Zhu (Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China) and Mr. Lianjiang Chi (Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China) for their useful advice and assistance.

## Author Contributions

Conceived and designed the experiments: LZ JD. Performed the experiments: LZ JD YQC ZLH. Analyzed the data: LZ ZLD WMZ. Contributed reagents/materials/analysis tools: JD. Wrote the paper: LZ.

## References

1. Wang QC (1984) Introduction of Japanese scallop and prospect of culture it in northern China. *Fish Sci* 3:24–27.

2. Li N (2009) What accounting for orange color of Yesso scallop muscle (*Patinopecten yessoensis*) and its application in breeding. Thesis for Ph.D., Ocean University of China.
3. Chang YQ, Chen XX, Ding J, Cao XB, Li RL, et al. (2007) Genetic diversity in five scallop populations of the Japanese scallop (*Patinopecten yessoensis*). *Acta Ecologica Sinica* 27:1145–1152.
4. Charles GH (1966) Sense organs (less cephalopods). In *Physiology of Mollusca*. Volume II. Edited by: Wilbur KM, Yonge CM. New York: Academic Press. p. 455–511.
5. Morse MP, Zardus JD (1997) Bivalvia. In *Microscopic Anatomy of Invertebrates*. Volume 6A. Edited by: Harrison FW, Kohn AJ. Mollusca II. New York: Wiley-Liss, Inc. p. 7–118.
6. Ding J, Chang YQ, Zhang J, Hao ZL, Zhao P, et al. (2011) Analysis and evaluation of nutritive composition in 'Ivory' strains and ordinary strains of *Mizuhopecten yessoensis*. *Journal of Agricultural Science and Technology* 13: 121–128.
7. Zhang J, Chang YQ, Ding J, Wang XY, Xu MJ (2011) Genetic structure of *Patinopecten yessoensis* with different shell colours and correlation of microsatellite markers with growth traits. *Chinese Agricultural Science Bulletin* 27:83–91.
8. Wang T, Ding J, Yu JP, Chang YQ (2009) ISSR analysis on genetic diversity and differentiation of five different geographic populations of *Patinopecten yessoensis*. *Journal of Yantai University (Natural Science and Engineering Edition)* 22: 35–41.
9. Feng C, Chen M, Xu Cj, Bai L, Yin XR, et al. (2012) Transcriptomic analysis of Chinese bayberry (*Myrica rubra*) fruit development and ripening using RNA-Seq. *BMC Genomics* 13: 19. doi: [10.1186/1471-2164-13-19](https://doi.org/10.1186/1471-2164-13-19) PMID: [22244270](https://pubmed.ncbi.nlm.nih.gov/22244270/)
10. Fu BD, He SP (2012) Transcriptome analysis of silver carp (*Hypophthalmichthys molitrix*) by paired-end RNA sequencing. *DNA Res* 19: 1–12. doi: [10.1093/dnares/dsr037](https://doi.org/10.1093/dnares/dsr037) PMID: [22039173](https://pubmed.ncbi.nlm.nih.gov/22039173/)
11. Andersen JR, Lubberstedt T (2003) Functional markers in plants. *Trends in Plant Science* 8: 554–560. PMID: [14607101](https://pubmed.ncbi.nlm.nih.gov/14607101/)
12. Morozova O, Marra MA (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92: 255–264. doi: [10.1016/j.ygeno.2008.07.001](https://doi.org/10.1016/j.ygeno.2008.07.001) PMID: [18703132](https://pubmed.ncbi.nlm.nih.gov/18703132/)
13. Guo S, Zheng Y, Joung JG, Liu S, Zhang Z, et al. (2010) Transcriptome sequencing and comparative analysis of cucumber flowers with different sex types. *BMC Genomics* 11: 384. doi: [10.1186/1471-2164-11-384](https://doi.org/10.1186/1471-2164-11-384) PMID: [20565788](https://pubmed.ncbi.nlm.nih.gov/20565788/)
14. Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotechnol* 27: 522–530. doi: [10.1016/j.tibtech.2009.05.006](https://doi.org/10.1016/j.tibtech.2009.05.006) PMID: [19679362](https://pubmed.ncbi.nlm.nih.gov/19679362/)
15. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760. doi: [10.1093/bioinformatics/btp324](https://doi.org/10.1093/bioinformatics/btp324) PMID: [19451168](https://pubmed.ncbi.nlm.nih.gov/19451168/)
16. Philipp EER, Wessels W, Gruber H, Strahl J, Wagner AE, et al. (2012) Gene expression and physiological changes of different populations of the long-lived bivalve *Arctica islandica* under low oxygen conditions. *PLoS ONE* 7(9): e44621. doi: [10.1371/journal.pone.0044621](https://doi.org/10.1371/journal.pone.0044621) PMID: [23028566](https://pubmed.ncbi.nlm.nih.gov/23028566/)
17. Clark MS, Thorne MA, Vieira FA, Cardoso JC, Power DM, et al. (2010) Insights into shell deposition in the Antarctic bivalve *Laternula elliptica*: gene discovery in the mantle transcriptome using 454 pyrosequencing. *BMC Genomics* 11:362.
18. Novaes E, Drost DR, Farmerie WG, Pappas GJ Jr., Grattapaglia D, et al. (2008) High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. *BMC Genomics* 9: 312. doi: [10.1186/1471-2164-9-312](https://doi.org/10.1186/1471-2164-9-312) PMID: [18590545](https://pubmed.ncbi.nlm.nih.gov/18590545/)
19. Zhang GF, Fang XD, Guo XM, Li L, Luo RB, et al. (2012) The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 490:49–54. doi: [10.1038/nature11413](https://doi.org/10.1038/nature11413) PMID: [22992520](https://pubmed.ncbi.nlm.nih.gov/22992520/)
20. Shi MJ, Lin Y, Xu GR, Xie LP, Hu XL, et al. (2013) Characterization of the Zhikong scallop (*Chlamys farreri*) mantle transcriptome and identification of biomineralization-related genes. *Marine Biotechnology* 15:706–715. doi: [10.1007/s10126-013-9517-0](https://doi.org/10.1007/s10126-013-9517-0) PMID: [23860577](https://pubmed.ncbi.nlm.nih.gov/23860577/)
21. Zhang LL, Li L, Zhu YB, Zhang GF, Guo XM (2014) Transcriptome analysis reveals a rich gene set related to innate immunity in the eastern oyster (*Crassostrea virginica*). *Marine Biotechnology* 16:17–33. doi: [10.1007/s10126-013-9526-z](https://doi.org/10.1007/s10126-013-9526-z) PMID: [23907648](https://pubmed.ncbi.nlm.nih.gov/23907648/)
22. Chen S, Xiao YP, Wu PY (2007) Temperature tolerance research of scallop (*Patinopecten yessoensis*) in Nanji Island of Zhejiang Province in China. *Journal of Zhejiang Ocean University(Natural Science)* 2:160–164.
23. Liu MJ, Chang YQ, Yang YF, Liu YP, Zhang J, et al. (2011) Effects of gradual salinity changes on immune parameters of scallop (*Patinopecten yessoensis*). *Journal of Agricultural Science and Technology* 13:129–135.
24. Ke JY, Chen YS, Dai CJ (2009) Review of researches done on the malacozoan immune mechanism. *Journal of Ningde Teachers College (Natural Science)* 21:113–117.



25. Anderson RS (1988) Effects of anthropogenic agents on bivalve cellular and humeral defense mechanism[M]. *Am Fish Spec Public* 18:238–242.
26. Canesi L, Gallo G, Gavioli M, Pruzzo C (2002) Bacteria—hemocyte interactions and phagocytosis in marine bivalves. *Microsc Res Tech* 57:469–476. PMID: [12112429](#)
27. Xue QG, Hellberg ME, Schey KL, Itoh N, Eytan RI, et al. (2010) A new lysozyme from the eastern oyster, *Crassostrea virginica*, and a possible evolutionary pathway for i-type lysozymes in bivalves from host defense to digestion. *BMC Evol Biol* 10:213. doi: [10.1186/1471-2148-10-213](#) PMID: [20633278](#)
28. Medzhitov R, Janeway C Jr. (2000) Innate immune recognition: mechanisms and pathways. *Immunol Rev* 173:89–97. PMID: [10719670](#)
29. Sonnhammer ELL, von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* 6:175–182. PMID: [9783223](#)
30. Li JZ (2001) Comprehensive utilization of shells. *Journal of Huaihai Institute of Technology* 10:22–23.
31. Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11: R106, doi: [10.1186/gb-2010-11-10-r106](#) PMID: [20979621](#)
32. Tanaka Y, Sasaki N, Ohmiya A (2008) Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *The Plant Journal* 54:733–749, doi: [10.1111/j.1365-313X.2008.03447.x](#) PMID: [18476875](#)
33. Riley PA (1997) Melanin. *International Journal of Biochemistry & Cell Biology* 29:1235–1239. doi: [10.1016/j.biocel.2015.01.003](#) PMID: [25595463](#)
34. Lu YZ (2012) Effect of metal ions on the nacre color and metallothionein expression in *Pinctada martensii*. Thesis for M.S., Guangdong Ocean University.
35. Geoffrey AS, Jack RC, Gilda HL (1992) Magnesium porphyrin radical cations: a theoretical study of substituent effects on the ground state. *Journal of the American Chemical Society* 114:9538–9544.
36. Duan CH, Ding YL, Mou ZG, Xu D, Wei Q (2005) Studies and applications of porphyrin color reagents. *Journal of Jilin University (Information Science Edition)* 19:32–37.
37. Jackson DJ, Wörheide G, Degnan BM (2007) Dynamic expression of ancient and novel molluscan shell genes during ecological transitions. *BMC Evol Biol* 7:160. PMID: [17845714](#)
38. Yano M, Nagai K, Morimoto K, Miyamoto H (2006) Shematrin: a family of glycine-rich structural proteins in the shell of the pearl oyster *Pinctada fucata*. *CompBiochemPhysiol B BiochemMolBiol* 144:254–262.
39. Ohgami RS, Campagna DR, McDonald A, Fleming MD (2006) The Steap proteins are metallo-reductases. *Blood* 108:1388–94. PMID: [16609065](#)
40. Collins JF, Prohaska JR, Knutson MD (2010) Metabolic crossroads of iron and copper. *Nutr Rev* 68:133–147. doi: [10.1111/j.1753-4887.2010.00271.x](#) PMID: [20384844](#)
41. Au C, Benedetto A, Aschner M (2008) Manganese transport in eukaryotes: the role of DMT1. *Neurotoxicology* 29(4):569–76. doi: [10.1016/j.neuro.2008.04.022](#) PMID: [18565586](#)
42. Gunshin H, Mackenzie B, Berger UV, Gunshin Y, Romero MF, et al. (1997) Cloning and characterization of a mammalian proton-coupled metal-ion transporter. *Nature* 388: 482–8. PMID: [9242408](#)
43. Zhang HF, Pan JH (2005) Applications of porphyrin and metalloporphyrin. *Journal of Chemical Education* 26:3–8.
44. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29:644–652, doi: [10.1038/nbt.1883](#) PMID: [21572440](#)
45. Wang LL, Zhang H, Song LS, Guo XS (2007) Loss of allele diversity in introduced populations of the hermaphroditic bay scallop *Argopecten irradians*. *Aquaculture* 271(1–4):252–259.
46. Iseli C, Jongeneel CV, Bucher P (1999) ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proc Int Conf Intell Syst Mol Biol* pp. 138–48. PMID: [10786296](#)
47. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, et al. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21(18): 3674–6. PMID: [16081474](#)
48. Ye J, Fang L, Zheng HK, Zhang HY, Chen J, et al. (2006) WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res* 34: W293–7. PMID: [16845012](#)
49. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Research* 32: D277–D280. PMID: [14681412](#)
50. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5(7):621–628. doi: [10.1038/nmeth.1226](#) PMID: [18516045](#)

51. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, et al. (2013) *De novo* transcript sequence reconstruction from RNA-Seq using the Trinity platform for reference generation and analysis. *Nature Protocols* 8: 1494–512. doi: [10.1038/nprot.2013.084](https://doi.org/10.1038/nprot.2013.084) PMID: [23845962](https://pubmed.ncbi.nlm.nih.gov/23845962/)
52. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-</sup>(Delta Delta C(T)) Method. *Methods* 25: 402–408. PMID: [11846609](https://pubmed.ncbi.nlm.nih.gov/11846609/)
53. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20 (9):1297–1303. doi: [10.1101/gr.107524.110](https://doi.org/10.1101/gr.107524.110) PMID: [20644199](https://pubmed.ncbi.nlm.nih.gov/20644199/)