Molecular dominance investigation for large-sized parents of Chinese Mitten Crab (*Eriocheir sinensis*) based on ovarian transcriptome

Fujun Xuan (swimming_crab@126.com)
Yancheng Teachers University
Longlong Fu
Jiangsu Freshwater Fisheries Research institute
Kun Wang
Yancheng Teachers University
Zonglin Hua
Yancheng Teachers University
Xinxin Xiong
Yancheng Teachers University
Jianguang Zhang
Jiangsu Da-ren aquatic products Co., Ltd
Jianlin Pan
Jiangsu Freshwater Fisheries Research Institute
Boping Tang
Yancheng Teachers University
Weibing Guan
Shanghai Ocean University
Yongxu Cheng
Shanghai Ocean University

Research article

Keywords: Eriocheir sinensis breeding, ovarian transcriptome, differentially expressed genes, pathways, bioinformatics analysis

Posted Date: November 16th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-108246/v1

Abstract

Background: Germplasm degradation is one of the important issues for the healthy development of Chinese Mitten Crab (*Eriocheir sinensis*) industry. Nowadays, nursery units and crab farmers have realized the seedling superiority of large-sized parents (≥200g male, ≥150g female), but the underlying molecular mechanisms remain unclear. Here we investigated the advantage of breeding with large-sized parents based on ovarian transcriptome.

Results: RNA isolation and transcriptome sequencing were performed using ovarian tissue obtained from three large-sized female *E. sinensis* and three medium-sized female *E. sinensis*. The differentially expressed genes (DEGs) were investigated between two groups, followed by the exploration of common DEGs (co-DEGs) among the DEGs in the current study and published genes for growth, reproduction, and immune response pathways. Quantitative real-time polymerase chain reaction (qRT-PCR) was performed to validate all co-DEGs. Finally, a pathway enrichment analysis was performed to investigate the potential pathways associated with the validated co-DEGs. A total of 5,307 up-regulated genes and 3,465 down-regulated genes were investigated between the two groups, of which 43 co-DEGs were explored when all the DEGs in the current study were integrated with 338 genes in 15 reproductive, immune, and growth pathways. A total of nine co-DEGs, including *TRINITY_DN13931_c0_g1*, *TRINITY_DN1908_c0_g2*, and *TRINITY_DN6686_c0_g1*, were enrolled for further study after qRT-PCR validation. These validated co-DEGs were mainly enriched in pathways such as apoptosis, insulin signaling, and mTOR signaling.

Conclusion: Our original work might lay an important molecular foundation for further exploring the advantages of using large-sized parents for *E. sinensis* breeding in the future. Based on the comparative analysis of gene expression, the advantages of breeding large-sized female *E. sinensis* parents at the molecular level, such as growth, reproduction, and immunity, were analyzed for the first time. The apoptosis pathways, insulin signaling pathways, and mTOR signaling pathway might be involved in the advantage of breeding large-sized female *E. sinensis* parents.

Highlights

1. Molecular advantage of large-sized female *sinensis* parent investigated
2. Apoptosis, insulin, and mTOR pathways related to large-sized advantage
3. Genes like *TRINITY_DN13931_c0_g1* were vital for large-sized *sinensis* breeding
4. 8,772 differentially expressed genes between large-sized and medium-sized crabs

Background

The Chinese mitten crab *Eriocheir sinensis* is a traditional aquatic treasure in China [1] . It is considered a luxury aquatic food for Chinese people due to its delicate flavor [2]. The artificial culture of *E. sinensis* has become both a leading and pillar industry in the cultivation of new freshwater varieties in China, and plays an important role in promoting the sustainable and healthy development of freshwater aquaculture

in China [3]. High quality germplasm resources are beneficial for improving the breeding and economic effects of *E. sinensis*[4], but the degradation of germplasm resources in recent years is one of the most critical factors restricting its production. For example, to reduce broodstock costs, nearly all commercial hatcheries in China select small or medium-sized *E. sinensis* (females: 60–100 g/ind.; males: 80–150 g/ind.) as parents [5].

Parents with excellent characteristics contribute to solving the problem of germplasm degradation in *E. sinensis* [6]. At present, aquaculture and scientific research have increasingly recognized the influence of parental size on crab production [7, 8]. The selection of large *E. sinensis* （female: ≥ 150 g/ind.; male: ≥ 200 g/ind.）as parents for group selection is a key measure for purification and rejuvenation [9]. It was proven that the fecundity, egg holding capacity, and spawning capacity of offspring increased with the increase in parents' size [7, 8]. Meanwhile, the offspring of large-sized parents result in a higher proportion of large-sized market crabs and higher economic benefits [10]. Moreover, a previous study showed that the growth characters and nutritional status in the juvenile phase (coin size) of the offspring of large-sized parents were better than those from medium-sized parents [11]. Although genome sequencing, assembly, and annotation provide a valuable resource for understanding the molecular expression of animal growth and development processes, obtaining the genome sequence of any crab species is very difficult [12].

Transcriptome analysis is an effective method for evaluating gene expression and its biological function [13]. A previous study showed that transcriptome profiling of the testes during the reproduction in *E. sinensis* can be realized using Illumina sequencing [14]. The molecular mechanism of the process of growth, immunity, and reproduction in *E. sinensis* has become one of the main research directions in this field [15, 16]. By comparing the transcripts from the ovarian library, a previous transcriptome analysis successfully identified gonadal-associated genes, which was helpful in understanding the regulatory mechanism related to the reproductive pathway [17]. The immune response during the development of *E. sinensis* was also explored at the RNA expression level [18]. However, knowledge of the underlying molecular mechanisms associated with growth, reproduction, and immune responses in the regulation of large-sized *E. sinensis* is still very limited. To understand the advantage of using large-sized *E. sinensis* as parents at the molecular level in more detail, the regulatory pathways associated with the growth, reproduction, and immune response of *E. sinensis* need to be studied.

In this study, tissues were obtained from the ovaries of large *E. sinensis* and medium-sized *E. sinensis*, followed by RNA isolation and transcriptome sequencing. The differentially expressed genes (DEGs) were investigated based on the sequencing data between the two groups. Then, the common DEGs (co-DEGs) were explored among the DEGs identified in the current study and the genes in growth, reproduction, and immune response pathways that have been previously published. All co-DEGs were validated by quantitative real-time PCR (qRT-PCR). Finally, a pathway enrichment analysis was performed to investigate the potential pathways associated with the validated co-DEGs in the large *E. sinensis.* This study hopes to reveal the advantage of using large-sized *E. sinensis* parents in breeding from a molecular perspective.

Results

Correlation analysis between samples

After quality control and data pre-processing, the sequencing data were enrolled for the correlation analysis between the F-O-D group and F-O-Z group to test the reliability of the experiment and reasonable sample selection. The PCA analysis showed that the samples between the two groups were distributed in a decentralized way, while the samples within groups were clustered (Fig. 1). The correlation clustering heatmap showed that the correlation of samples enrolled in this study was high (Fig. 2).

Differentially expressed genes analysis based on sequencing data

Based on the sequencing data, a total of 5,307 up-regulated genes and 3,465 down-regulated genes were revealed between the F-O-D group and F-O-Z group. The volcano plot for the result of the DEGs analysis is shown in Fig. 3A, which directly indicates the relative expression values of genes among samples in different groups. The scatter plot for all these DEGs is shown in Fig. 3B. Furthermore, the results of the cluster analysis for gene expression patterns showed that the expression of genes in the same sample was aggregated, and the difference was obvious among the samples (Fig. 4).

Co-differentially expressed genes investigation and quantitative real-time polymerase chain reaction verification

Based on the reference retrieval, there were seven reproduction-associated pathways (retinol metabolism, cell cycle, mTOR signaling pathway, Wnt signaling pathway, insulin signaling pathway, ovarian steroidogenesis, and progesterone-mediated oocyte maturation), four growth-associated pathways (mTOR signaling pathway, PI3K-Akt signaling pathway, Wnt signaling pathway, and TGF-beta signaling pathway), and seven immune-associated pathways (cell cycle, lysosome, endocytosis, phagosome, apoptosis, cell adhesion molecules (CAMs), and natural killer cell mediated cytotoxicity) screened out. A total of 338 genes enriched in these 15 pathways (after removing overlapping pathways) were explored. Combined with the DEGs investigated in the current study, a total of 43 DEGs related to reproductive, immune, and growth processes were further revealed. Finally, the qRT-PCR analysis showed that a total of nine co-DEGs (*TRINITY_DN6686_c0_g1*, *TRINITY_DN6296_c2_g1*, *TRINITY_DN2026_c1_g1*, *TRINITY_DN1908_c0_g2*, *TRINITY_DN605_c0_g2*, *TRINITY_DN2717_c0_g1*, *TRINITY_DN608_c5_g1*, *TRINITY_DN13931_c0_g1*, and *TRINITY_DN35647_c0_g1*) showed the same trend as the sequencing results.

Enrichment analysis of verified co-differentially expressed genes

The KEGG enrichment analysis was performed on the PCR-verified co-DEGs between the two groups (Table 1). The results showed that these DEGs were mainly enriched in pathways such as apoptosis (ko04210, genes: *TRINITY_DN13931_c0_g1*, *TRINITY_DN1908_c0_g2*, and *TRINITY_DN6686_c0_g1*), insulin signaling pathway (ko04910, genes: *TRINITY_DN13931_c0_g1*, *TRINITY_DN2717_c0_g1*, and

TRINITY_DN608_c5_g1), and mTOR signaling pathway (ko04150, genes: *TRINITY_DN13931_c0_g1* and *TRINITY_DN2717_c0_g1*) (Fig. 5).

Table 1 The result of pathway enrichment analysis based on co-differentially expressed genes

TermID	Description	P value	GeneCount	GeneIDs
ɔ04960	Aldosterone-regulated sodium reabsorption	0.000351	2	TRINITY_DN13931_c0_g1, TRINITY_DN2717_c0_g1
ɔ04210	Apoptosis	0.000449	3	TRINITY_DN13931_c0_g1, TRINITY_DN1908_c0_g2, TRINITY_DN6686_c0_g1
ɔ04910	Insulin signaling pathway	0.000606	3	TRINITY_DN13931_c0_g1, TRINITY_DN2717_c0_g1, TRINITY_DN608_c5_g1
ɔ04930	Type II diabetes mellitus	0.001092	2	TRINITY_DN13931_c0_g1, TRINITY_DN2717_c0_g1
ɔ04923	Regulation of lipolysis in adipocytes	0.001726	2	TRINITY_DN13931_c0_g1, TRINITY_DN2717_c0_g1
ɔ04213	Longevity regulating pathway - multiple species	0.004083	2	TRINITY_DN13931_c0_g1, TRINITY_DN2717_c0_g1
ɔ04070	Phosphatidylinositol signaling system	0.004444	2	TRINITY_DN13931_c0_g1, TRINITY_DN2026_c1_g1
ɔ04066	HIF-1 signaling pathway	0.005613	2	TRINITY_DN13931_c0_g1, TRINITY_DN2717_c0_g1
ɔ04211	Longevity regulating pathway	0.006909	2	TRINITY_DN13931_c0_g1, TRINITY_DN2717_c0_g1
ɔ04072	Phospholipase D signaling pathway	0.007843	2	TRINITY_DN13931_c0_g1, TRINITY_DN2717_c0_g1
ɔ04914	Progesterone-mediated oocyte maturation	0.00833	2	TRINITY_DN13931_c0_g1, TRINITY_DN605_c0_g2
ɔ04931	Insulin resistance	0.00833	2	TRINITY_DN13931_c0_g1, TRINITY_DN2717_c0_g1
ɔ04071	Sphingolipid signaling pathway	0.009344	2	TRINITY_DN13931_c0_g1, TRINITY_DN6686_c0_g1
ɔ04145	Phagosome	0.009871	2	TRINITY_DN1908_c0_g2, TRINITY_DN2026_c1_g1
ɔ04068	FoxO signaling pathway	0.011532	2	TRINITY_DN13931_c0_g1, TRINITY_DN2717_c0_g1
ɔ04152	AMPK signaling pathway	0.012703	2	TRINITY_DN13931_c0_g1, TRINITY_DN2717_c0_g1
ɔ04810	Regulation of actin cytoskeleton	0.01789	2	TRINITY_DN13931_c0_g1, TRINITY_DN2026_c1_g1
ɔ04014	Ras signaling pathway	0.019308	2	TRINITY_DN13931_c0_g1, TRINITY_DN2717_c0_g1
ɔ04140	Autophagy - animal	0.020773	2	TRINITY_DN13931_c0_g1, TRINITY_DN6686_c0_g1
ɔ04015	Rap1 signaling pathway	0.022284	2	TRINITY_DN13931_c0_g1, TRINITY_DN2717_c0_g1
ɔ04150	mTOR signaling pathway	0.023841	2	TRINITY_DN13931_c0_g1, TRINITY_DN2717_c0_g1
ɔ04151	PI3K-Akt signaling pathway	0.028778	2	TRINITY_DN13931_c0_g1, TRINITY_DN2717_c0_g1
ɔ04060	Cytokine-cytokine receptor interaction	0.03443	1	TRINITY_DN35647_c0_g1
ɔ04932	Non-alcoholic fatty liver disease (NAFLD)	0.038816	2	TRINITY_DN13931_c0_g1, TRINITY_DN2717_c0_g1
ɔ04973	Carbohydrate digestion and absorption	0.047067	1	TRINITY_DN13931_c0_g1

Discussion

Although the advantage of using large-sized *E. sinensis* in actual production has been proven, the detailed molecular mechanism for this is still unclear. In this study, based on the *E. sinensis* ovarian transcriptome data analysis, a total of 5,307 up-regulated DEGs and 3,465 down-regulated DEGs were revealed between the large and medium-size groups. A total of 43 co-DEGs were explored when all the DEGs in the current study were integrated with 338 genes from 15 reproductive, immune, and growth-associated pathways reported in the literature. After qRT-PCR validation, a total of nine co-DEGs, including *TRINITY_DN13931_c0_g1*, *TRINITY_DN1908_c0_g2*, and *TRINITY_DN6686_c0_g1,* were finally investigated. The pathway enrichment analysis showed that these validated co-DEGs were mainly enriched in pathways such as the insulin signaling pathway (a reproduction-associated pathway), mTOR signaling pathway (a growth-associated pathway), and apoptosis (an immune-associated pathway).

The close relationship between insulin signaling and reproduction has already been revealed in the body development of helminths and insects [19]. Insulin signaling is an important pathway that regulates developmental and differentiation functions such as reproduction in animals [20]. Also, the control of body size involves growth regulating pathways such as insulin [21]. A previous study showed that the insulin-like androgenic gland hormone gene extensively participates in the reproduction of the female mud crab [22]. With the help of transcriptome sequencing, the multiple insulin-like peptides that participate in the insulin signaling pathway have been identified in crustacean species [23]. Importantly, female crabs are considered an important platform for the investigation of the biological function of insulin-like peptides in crabs [24]. A previous study indicated that the inhibitory role of insulin in vitellogenesis and oocyte maturation in the female mud crab is visualized via certain signaling pathways, such as autocrine and paracrine pathways [20]. Interestingly, the stimulation of vitellogenin expression is realized via insulin in the crab [23] . These insulin-associated proteins, including insulin-like peptides, are not only involved in ovarian development [24], but also negatively regulate glucose metabolism and participate in the immune response of *E. sinensis* against pathogen infection [25] .

A previous study showed that the mTOR pathway is associated with the insulin signaling pathway during the regulation of animal development, life span, and the length of larval development [26]. mTOR is a member of the phosphatidylinositol 3-kinase-related kinase family, which regulates various biological processes including growth control [27]. Since crustaceans must shed their exoskeleton periodically to develop and grow, the mTOR signaling pathway, which regulates growth and molting, has been widely studied by researchers [28]. It was proved that mTOR signaling gene expression contributes to the growth associated function in *Gecarcinus lateralis* [29]. The mTOR signaling genes are involved in molt regulation during the thermal acclimation of juvenile Dungeness crabs [30]. A previous study showed that mTOR integrates intrinsic signals (molting hormones) and extrinsic signals (thermal stress) to regulate molting and growth in decapod crustaceans. The molting gland is the source of steroid hormone production and the consequent regulation of the molt cycle in decapod crustaceans, and is responsive to both external environmental and internal physiological signals in the crab [28]. In crustaceans such as

crabs, mTOR activity either directly or indirectly controls the transcription of genes that drive the activation of the molting gland [31].

The apoptosis pathway has been shown to participate in the immune process [32]. Recently, the effector caspases such as Sp-caspase, which exhibit an immune response and cell apoptosis, have been identified in the mud crab [33]. In crabs, apoptosis can induce the activity of hemocytes, which play a vital role in defending against pathogen invasion after pathogen stimulation [34]. For *E. sinensis*, a previous study showed that the reduction of apoptosis via certain proteins (such as Es IAP1) can regulate the activity of cells during the growth process [35]. A recent transcriptomic analysis showed that the growth inhibition and survival rate reduction induced by various factors (such as hepatopancreatic necrosis disease) were significantly correlated with apoptosis in the development of *E. sinensis*[36]. Importantly, genes such as *nm23* and caspase in the apoptosis pathway participate in resistance to adverse environments during the growth of *E. sinensis* individuals[37, 38]. Thus, large-sized parents of *E. sinensis* might display a strong adverse environment resistance via gene expression in apoptosis, which further benefits the improvement of individual survival rates and economic value. In this study, the sequencing and differential expression analysis revealed nine co-DEGs, including *TRINITY_DN13931_c0_g1*, *TRINITY_DN1908_c0_g2*, and *TRINITY_DN6686_c0_g1*, between large-sized *E. sinensis* and medium-sized *E. sinensis*. Importantly, these co-DEGs were enriched in pathways including apoptosis, insulin signaling, and mTOR signaling. Thus, we speculated that apoptosis, insulin signaling, and mTOR signaling pathways enriched by *TRINITY_DN13931_c0_g1*, *TRINITY_DN1908_c0_g2*, and *TRINITY_DN6686_c0_g1* might be associated with the advantage of using large-sized parents in *E. sinensis* breeding.

Conclusions

Our original work might lay an important molecular foundation for further exploring the advantages of using large-sized parents for *E. sinensis* breeding in the future. Based on the comparative analysis of gene expression, the molecular-level advantages of using large-size female parents when breeding *E. sinensis*, such as growth, reproduction, and immunity, were analyzed for this study. Apoptosis pathways, insulin signaling pathways, and the mTOR signaling pathway might play important roles in the advantage of large-size parents in *E. sinensis* breeding.

Methods

Animal and sample collection

A total of three large female *E. sinensis* (weight: 248.60 ± 3.12 g) (F-O-D group) and three medium-sized female *E. sinensis* (weight: 102.87 ± 1.52 g) (F-O-Z group) were provided by a local crab breeding unit in the mating season (November 30, 2019). All crabs were washed and the ovarian parts of their reproductive system were dissected for the comparative analysis of the transcriptome. Transcriptome sequencing was provided by Shanghai Origingene Biomedical Technology Co., Ltd.

RNA isolation and sequencing

Briefly, using the TRIzol kit (Takara, Dalian, Chian), the total RNA in different samples was extracted, followed by purification via NanoDrop 2000 (Thermo Fisher, Massachusetts, USA). The reverse transcription cDNA library of the total RNA (5 μg/sample) was implemented using the TruseqTM RNA sample prep kit for Illumina® (New England BioLabs, Massachusetts, USA). Then, after treatment with repaired end modification, the cDNAs were amplified via 15 cycles of the repair chain reaction. Finally, the Illumina Hiseq 4000 platform (Illumina) with paired-end method [39] was used for cDNA cluster sequencing.

Pretreatment of RNA-seq dat*a*

Quality control of the raw reads was implemented using FastqStat.jar tools (version: 1.0) as follows: i) removal of barcode sequences and bases containing non-AGCT at the 5' end; ii) elimination of the paired reads with N > 10%; iii) exclusion of low-quality reads in which bases with Q < 20; and iv) exclusion of reads with a length < 25 bp. The quality control of clean reads was performed using fastqc (version 0.11.4) [40].

Transcriptome assembly and gene expression analysis

Trinity is a software suitable for *de novo* assembly without a reference genome based on the Illumina short fragment sequence (clean data) [41]. In this study, the *de novo* transcriptome assembly with sequence reads obtained from the Illumina platform was performed by Trinity Assembler (version: 2.6.6), followed by gene prediction using the ORF method in Trinity. Then, based on the RNA-seq data quantification software Kallisto (version: 0.43.1), the quantitative analysis of the gene expression levels was investigated among different samples [42]. Transcripts per million reads (TPM) were used as quantitative indicators in the current analysis.

Relation analysis among samples

A correlation analysis was performed to investigate the correlation among samples, especially among biological replicates. Furthermore, a principal component analysis (PCA) analysis was performed to identify the impact of a large sample based on the expression of the sample clustering.

Differentially expressed genes analysis

The differential expression analysis of genes between the two groups was performed using edgeR (version: 3.24)[43]. A false discovery rate (FDR) < 0.05, and |log2 fold change (FC)| > 1 were selected as cutoff thresholds for the DEGs investigation. The results were visualized using a volcano plot and scatter plot. Then, the clustering analysis of DEGs was performed with a distance algorithm (between samples: Spearman correlation coefficient; between genes: Pearson correlation coefficient) to reveal genes with the same or similar expression patterns using plot_cluster_exp (version: 1.1.0). The results of clustering were visualized by a heatmap.

Co-differentially expressed genes exploration based on the literature search

The pathways and enriched genes related to reproduction, growth, and immunity were screened for by reference retrieval. Electronic literature databases including Embase, PubMed, and the Cochrane Library were searched. In addition, a literature review was used to select relevant studies. Then, the genes enriched in these pathways were intersected with DEGs in the current study to obtain the co-DEGs associated with reproduction, growth, and immune responses.

Quantitative real-time polymerase chain reaction assay

In order to screen the DEGs with the same trend as the sequencing results, quantitative verification based on a qRT-PCR was performed on 12 co-DEGs that had higher FC. Briefly, according to the manufacturer's instructions, total RNA (10 μl) was extracted and quantified from each sample in each group using TRIzol reagent (Invitrogen, California, USA), and reverse-transcribed using SYBR Green PCR Mastermix (Applied Biosystems, Woolston, Warrington, UK). Then, the qRT-PCR was performed with the help of the ABi-PrisMR 7500 sequence detection system (Applied Biosystem, Foster City, California, USA). The gene *ACTB* was used as a reference. The amplification transcription reaction conditions were as follows: 95°C for 30 s, 40 cycles at 95°C for 5 s, and 60°C for 34 s. At the end point of each extended period, the signal was obtained, and then the amplification curve was studied. The primer sequences are listed in Table 2. The relative expression of the target genes was calculated using the $2^{-\Delta\Delta CT}$ method [44].

Table 2 The primers used in this study.

Gene	Forward sequence (5'-3')	Reverse sequence (5'-3')	PCR Products(bp)
ACTB	ACGGAGCGAGGCTACACC	CGGCAGTGGTCATTTCCTG	112
6686	TGGGCACGGTGTTCGCTAT	GCAACGCCTTCCTTCTGGT	195
6296	GCTCCCAGGACACCGAACA	CCACCGTGCAGGACAACC	120
2026	TTCACTATCTGTATCGCCCTCA	CCCCGCTATGCTCAGTTGT	115
1908_c0_g2	AGTACGGCAAGAGGAGCAAG	GTCGGAGTGTTCCAGGGTG	119
605	CTGCTGGGAAAATGAACCTTA	TGAGTGGGGAAGTGATTATGAA	110
2717	TGAAGGAAATGGTAAGCAAAGG	AATTATGTCTGGGAAGGACGG	134
608	TCATCCACGCCAAATCCG	ATCACGTAGAAGTTGTCCCCAT	184
13931	AGAGCAGCACAGCCAAACC	GCGGCAGACCTGAGAAAACA	170
35647	TGTCCCTACAGAACTTTCTCCC	CACCCGCAGCCTTCCACT	102

Pathway enrichment analysis

The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of the co-DEGs verified by qRT-PCR were performed using go_enrichment (version: 2.1.0) software [45, 46]. A P-value < 0.05 was considered a cutoff value for significant KEGG enrichment.

Declarations

Ethics approval and consent to participate

Not applicable. The sampling location was not privately-owned or protected, and field sampling did not involve protected species.

Consent for publication

Not applicable.

Competing Interest

All authors declare no conflicts of interest.

Funding

This work was supported by the earmarked fund for Jiangsu Agricultural Industry Technology System (NO. JATS [2020] 362); Jiangsu Agricultural Science and Technology Innovation Fund [NO. CX(19)2031]; Freshwater Fisheries Research Institute of Jiangsu Province Independent Research Project [NO. SZL201802]; Ministry of Agriculture modern Agricultural Industrial Technology System Special Project (2017-2020) freshwater crab ecosystem culture (NO. CARS-48) and Key projects of Science and Technology Department of Ningxia Autonomous Region (2020-2021).

Authors' contributions

Fujun Xuan and Longlong Fu drafted the manuscript and involved in the experimental work. Kun Wang, Zonglin Hua and Xinxin Xiong involved in the experimental work and helped to analyzed data. Jianguang Zhang performed sample preparation. Jianlin Pan and Boping Tang involved in the experimental work. Fujun Xuan, Weibing Guan and Yongxu Cheng conceived and designed the study. All authors read and gave final approval of the final manuscript.

Availability of data and materials

Sequencing data presented in this study has been deposited at NCBI under the accession PRJNA669606.

Acknowledgements

The authors thank the Shanghai Major Biotechnology Corporation for the assistance with data processing and bioinformatic analysis. We are indebted to Ming Lei and Lingjiang He for support and discussions.

References

[1]Li X, Li Z, Liu J, De Silva, SS. Advances in precocity research of the Chinese mitten crab *Eriocheir sinensis*. Aquacult Int. 2011; 19(2): 251-267.

[2]Chen DW, Zhang M, Shrestha S. Compositional characteristics and nutritional quality of Chinese mitten crab (*Eriocheir sinensis*). Food Chem. 2007; 103(4): 1343-1349.

[3]Wang J, Xu P, Zhou G, Li X, Lu Q, Liu X, Zhou J, Wang C. Genetic improvement and breeding practices for Chinese mitten crab, *Eriocheir sinensis*. J World Aquacult Soc. 2018; 49(2): 292-301.

[4]Wang S, He Y, Wang Y, Tao N, Wu X, Wang X, Qiu W, Ma M. Comparison of flavour qualities of three sourced *Eriocheir sinensis*. Food Chem. 2016; 200: 24-31.

[5]Sui L, Wille M, Cheng Y, Wu X, Sorgeloos P. Larviculture techniques of Chinese mitten crab *Eriocheir sinensis*. Aquaculture. 2011; 315(1-2): 16-19.

[6]Bai A, Zhou X. The Research Progress of Polyploid Induction to *Eriocheir sinensis*. Jiangsu Agri Sci. 2012; 20(11): 248-251.

[7]Mao HC, Wang GL, Yang YC, Tao C, Ma XZ, Wang CH. Productive evaluation on the effect of ecological nursery of *Eriocheir sinensis* in different sizes of parent crab ponds. Fish Sci Tech Info. 2014; 41(5): 233-236.

[8]Wang SB, Liu NG, Jiang XD, Deng D, Zhang JB, Cheng YX, Wu XG. Experiment summary of three parent specifications of Chinese mitten crab breeding in soil pond. Sci Fish Farming. 2019; 35: 12-13.

[9]Sui L, Zhang F, Wang X, Bossier P, Sorgeloos P, Hänfling B. Genetic diversity and population structure of the Chinese mitten crab *Eriocheir sinensis* in its native range. Mar Biol. 2009; 156(8): 1573-1583.

[10]Zhang HS. Experimental study on the culture of Chinese mitten crab seedling from large size parents. Rural Sci Tech. 2018; 9: 110-111.

[11]Chen JW, Ma XZ, Wang W, Yang YC, Tao C. The comparative study on growth characteristics of offspring produced by female parents with different weight of the Chinese mitten crab (*Eriocheir Sinensis*). Chin J Zool. 2016; 51(5): 895-906.

[12]Song L, Bian C, Luo Y, Wang L, You X, Li J, Qiu Y, Ma X, Zhu Z, Ma L. Draft genome of the Chinese mitten crab, *Eriocheir sinensis*. GigaScience. 2016; 5(1): s13742-13016-10112-y.

[13]Li X, Chen J, Hu X, Huang Y, Li Z, Zhou L, Tian Z, Ma H, Wu Z, Chen M. Comparative mRNA and microRNA expression profiling of three genitourinary cancers reveals common hallmarks and cancer-specific molecular events. PloS one. 2011; 6(7): e22570.

[14]He L, Wang Q, Jin X, Wang Y, Chen L, Liu L, Wang Y. Transcriptome profiling of testis during sexual maturation stages in *Eriocheir sinensis* using Illumina sequencing. PloS one. 2012; 7(3):

[15]Li X, Cui Z, Liu Y, Song C, Shi G. Transcriptome analysis and discovery of genes involved in immune pathways from hepatopancreas of microbial challenged mitten crab *Eriocheir sinensis*. PloS one. 2013; 8(7): e68233.

[16]Liu Z, Yu P, Cai M, Wu D, Zhang M, Chen M, Zhao Y. Effects of microplastics on the innate immunity and intestinal microflora of juvenile *Eriocheir sinensis*. Sci Total Environ. 2019; 685: 836-846.

[17]Gao J, Wang X, Zou Z, Jia X, Wang Y, Zhang Z. Transcriptome analysis of the differences in gene expression between testis and ovary in green mud crab (*Scylla paramamosain*). BMC genomics. 2014; 15: 585.

[18]Dong C, Zhao J, Song L, Wang L, Qiu L, Zheng P, Li L, Gai Y, Yang G. The immune responses in Chinese mitten crab *Eriocheir sinensis* challenged with double-stranded RNA. Fish shellfish immunol. 2009; 26(3): 438-442.

[19]Dissous C. Venus kinase receptors at the crossroads of insulin signaling: their role in reproduction for helminths and insects. Front Endocrinol. 2015; 6: 118.

[20]Bao C, Yang Y, Huang H, Ye H. Inhibitory role of the mud crab short neuropeptide F in vitellogenesis and oocyte maturation via autocrine/paracrine signaling. Front Endocrinol. 2018; 9: 390.

[21]Nijhout HF, Riddiford LM, Mirth C, Shingleton AW, Suzuki Y, Callier V. The developmental control of size in insects. Wiley Interdisciplinary Reviews: Developmental Biology. 2014; 3: 113-134.

[22]Huang X, Ye H, Huang H, Yang Y, Gong J. An insulin-like androgenic gland hormone gene in the mud crab, *Scylla paramamosain*, extensively expressed and involved in the processes of growth and female reproduction. Gen Comp Endocr. 2014; 204: 229-238.

[23]Huang X, Feng B, Huang H, Ye H. 2017. In vitro stimulation of vitellogenin expression by insulin in the mud crab, *Scylla paramamosain*, mediated through PI3K/Akt/TOR pathway. Gen Comp Endocr. 2017; 250: 175-180.

[24]Huang X, Ye H, Feng B, Huang H. Insights into insulin-like peptide system in invertebrates from studies on IGF binding domain-containing proteins in the female mud crab, *Scylla paramamosain*. Mol Cell Endocrinol. 2015; 416: 36-45.

[25]Wang L, Chen H, Wang L, Song L. An insulin-like peptide serves as a regulator of glucose metabolism in the immune response of Chinese mitten crab *Eriocheir sinensis*. Dev Comp Immunol. 2020; 108: 103686.

[26]Jia K, Chen D, Riddle DL. The TOR pathway interacts with the insulin signaling pathway to regulate *C. elegans* larval development, metabolism and life span. Development. 2004; 131(16): 3897-3906.

[27]Sarbassov DD, Ali SM, Sabatini DM. Growing roles for the mTOR pathway. Curr Opin Cell Biol. 2005; 17(6): 596-603.

[28]Das S, Pitts NL, Mudron MR, Durica DS, Mykles DL. Transcriptome analysis of the molting gland (Y-organ) from the blackback land crab, *Gecarcinus lateralis*. Comp Biochem Phys D. 2016; 17: 26-40.

[29]Cosenza KS. Role of ecdysteroids on Myostatin and mTOR signaling gene expression in molt-dependent growth and atrophy of skeletal muscle in *Gecarcinus lateralis* and *Carcinus maenas*, The Colorado State University Libraries. 2016;

[30]Wittmann A, Chang ES, Mykles DL. mTOR signaling genes are involved in molt regulation during thermal acclimation of juvenile Dungeness crabs (*Metacarcinus magister*).2015;

[31]Shyamal S, Das S. Guruacharya A, Mykles D, Durica D. Transcriptomic analysis of crustacean molting gland (Y-organ) regulation via the mTOR signaling pathway. Sci Rep. 2018; 8(1): 1-17.

[32]Lum JJ, Bauer DE, Kong M, Harris MH, Li C, Lindsten T, Thompson CB. Growth factor regulation of autophagy and cell survival in the absence of apoptosis. Cell. 2005; 120(2): 237-248.

[33]Li J, Dong L, Zhu D, Zhang M, Wang K, Chen F. An effector caspase Sp-caspase first identified in mud crab *Scylla paramamosain* exhibiting immune response and cell apoptosis. Fish shellfish immunol. 2020; 103, 442-453.

[34]Johansson MW, Keyser P, Sritunyalucksana K, Söderhäll K. Crustacean haemocytes and haematopoiesis. Aquaculture. 2000; 191: 45-52.

[35]Qu C, Sun J, Xu Q, Lv X, Yang W, Wang F, Wang Y, Yi Q, Jia Z, Wang L. An inhibitor of apoptosis protein (Es IAP1) from Chinese mitten crab *Eriocheir sinensis* regulates apoptosis through inhibiting the activity of Es Caspase-3/7-1. Sci Rep. 2019; 9(1): 1-10.

[36]Yan B, Liu X, Zhou Y, Zhang M, Fang P, Jiang M, Yuan R, Hu X, Cao G, Xue R. Transcriptomic analysis reveals that hepatopancreatic necrosis disease in *Eriocheir sinensis* (Chinese mitten crabs) may be the result of autophagy and apoptosis. Aquaculture. 2020; 515: 734579.

[37]Jin XK, Li WW, Wang Q. Caspase and nm23: Apoptosis genes linked to the antibacterial response of the Chinese mitten crab. Bioengineered bugs. 2011; 2(3): 174-177.

[38]Jin XK, Li WW, He L, Lu W, Chen LL, Wang Y, Jiang H, Wang Q. Molecular cloning, characterization and expression analysis of two apoptosis genes, caspase and nm23, involved in the antibacterial response in Chinese mitten crab, *Eriocheir sinensis*. Fish shellfish immunol. 2011; 30: 263-272.

[39]Ran M, Chen B, Li Z, Wu M, Liu X, He C, Zhang S, Li Z. 2016. Systematic Identification of Long Non-Coding RNAs in Immature and Mature Porcine Testes. Biol Reprod. 2016; 77(4): 1-9.

[40]Andrews S. Babraham bioinformatics-FastQC a quality control tool for highthroughput sequence data. URL: https://www. bioinformatics. babraham. ac. uk/projects/fastqc/(accessed 06.12. 2018). 2015;

[41]Bankar KG, Todur VN, Shukla RN, Vasudevan M. Ameliorated de novo transcriptome assembly using Illumina paired end sequence data with Trinity Assembler. Genomics data. 2015; 5: 352-359.

[42]Du Y, Huang Q, Arisdakessian C, Garmire LX. Evaluation of STAR and Kallisto on Single Cell RNA-Seq Data Alignment. G3: Genes, Genomes, Genetics. 2020; 10: 1775-1783.

[43]Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26: 139-140.

[44]Livak KJ. 2001. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2(-Delta Delta C(T))Method. Methods. 2001; 25(4): 402-408.

[45]Consortium GO. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. 2004; 32: 258-261.

[46]Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000; 28: 27-30.

Figures

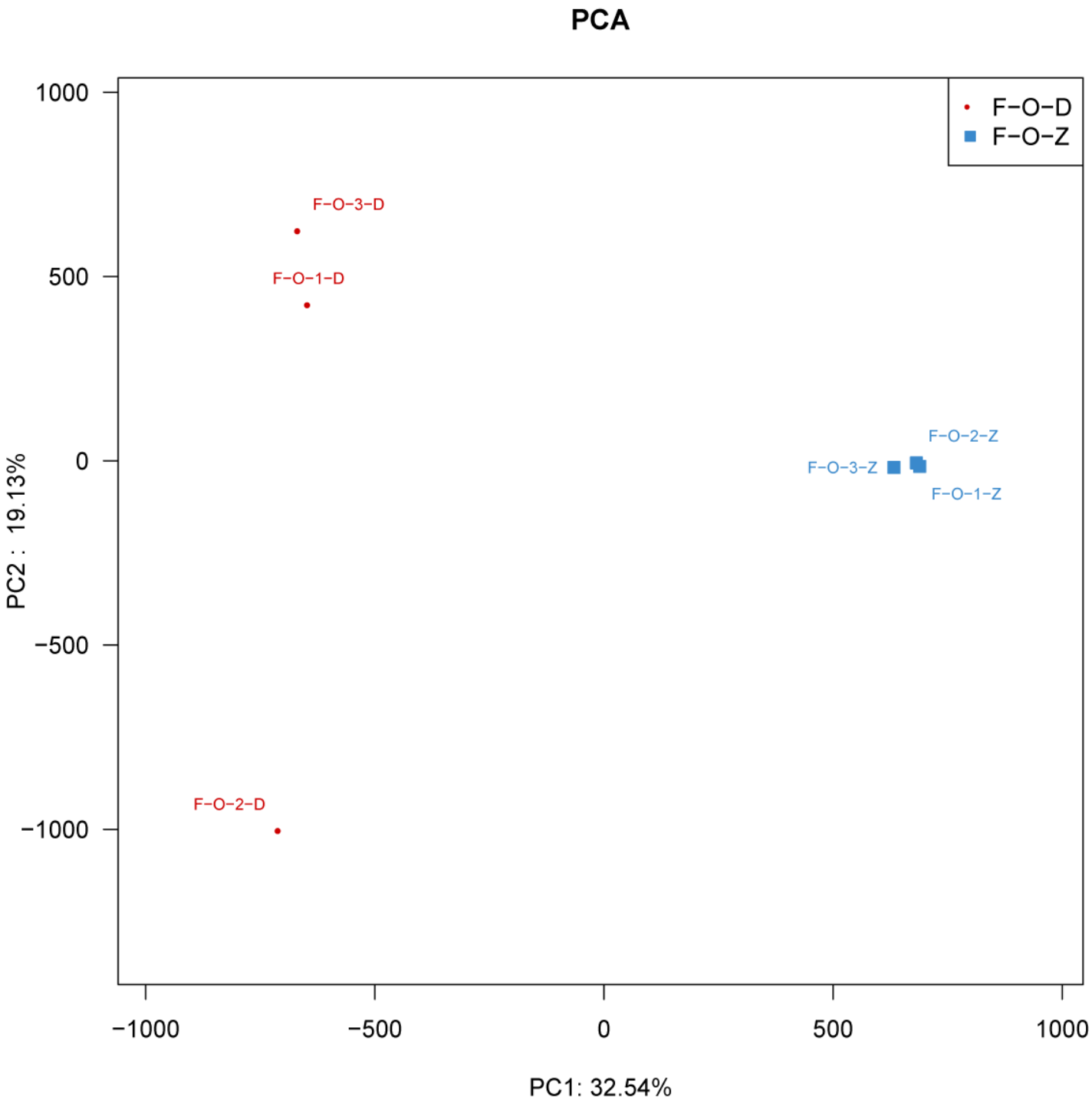

Figure 1

Principal component analysis of the sequenced data. The x-axis and y-axis represent the new data set corresponding to the principal components (PCs) after dimensionality reduction, which was used to represent the difference between samples; the value in the coordinate axis label represents the percentage of variance of the corresponding PC interpretation of the population. The dots represent samples, with the same color/shape representing the same sample group.

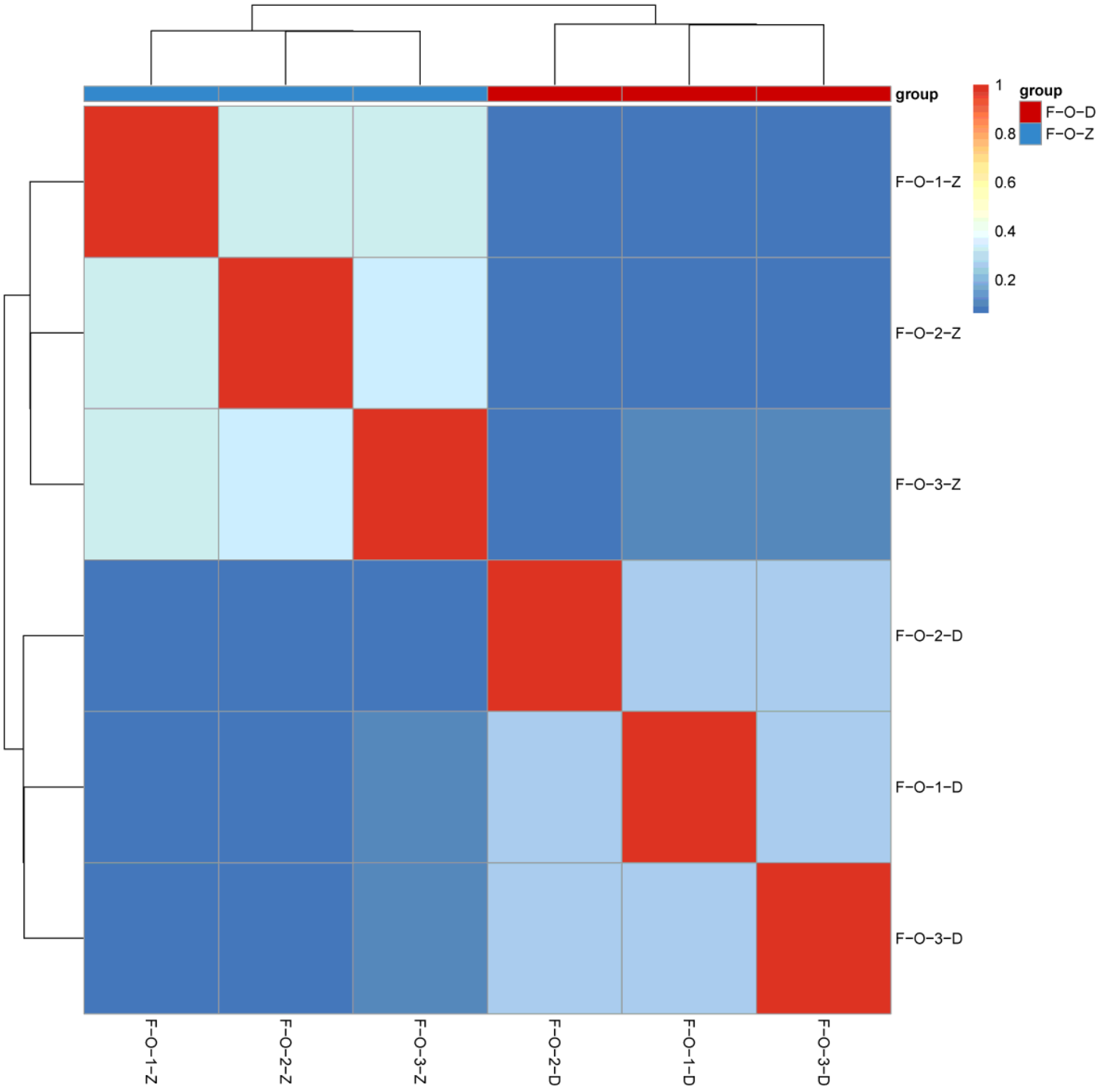

Figure 2

Clustering of samples based on correlation coefficients. The colors relate to the correlation coefficient: red for a high expression value, and blue for a low expression value. From blue to red, the higher the value, the higher the correlation.

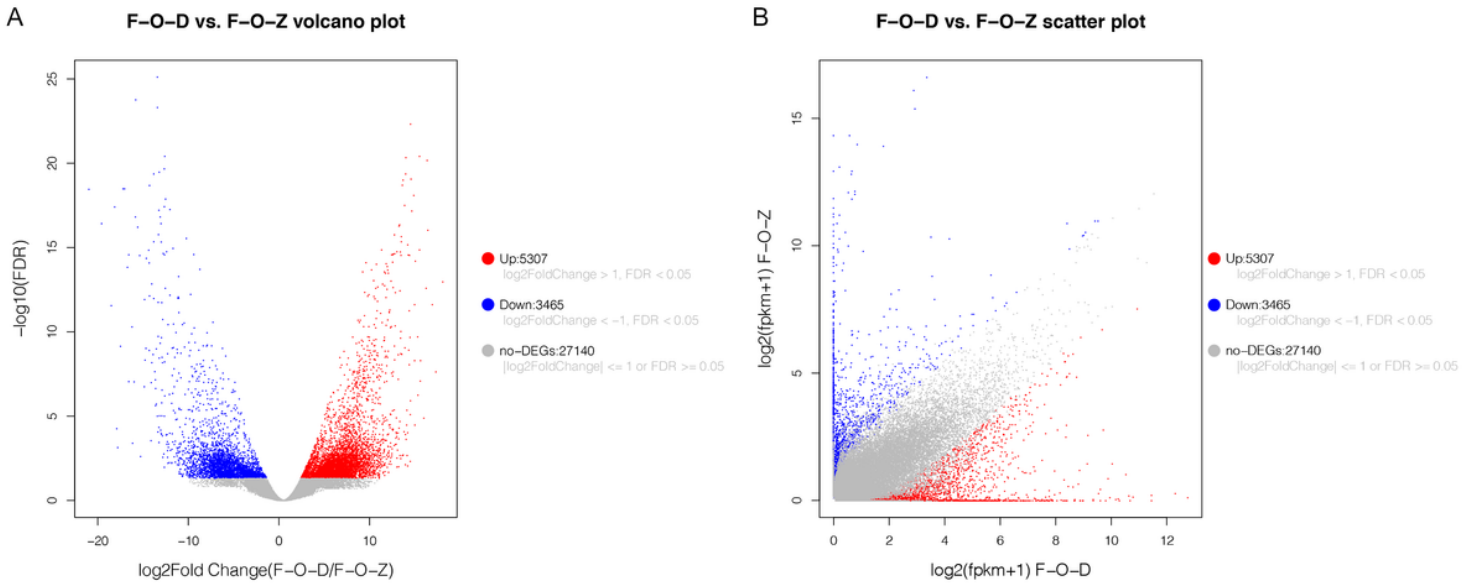

Figure 3

Comparison of differentially expressed genes (DEGs) between F-O-D group (large-sized female crabs) and F-O-Z group (medium-sized female crabs). A) The volcano plot for DEGs between the two groups; the x-axis represents the multiple change value of the expression difference for genes between two samples; the y-axis represents the statistical test value of the difference for gene expression; the red nodes represent up-regulated DEGs, while the blue nodes represent down-regulated DEGs. B) the scatter plot for DEGs between the two groups; the x-axis represents the expression levels (transcript per million reads or TPM value) of genes in the F-O-D group; the y-axis represents the expression levels (TPM value) of genes in the F-O-Z group; the red nodes represent up-regulated DEGs, while the blue nodes represent down-regulated DEGs.

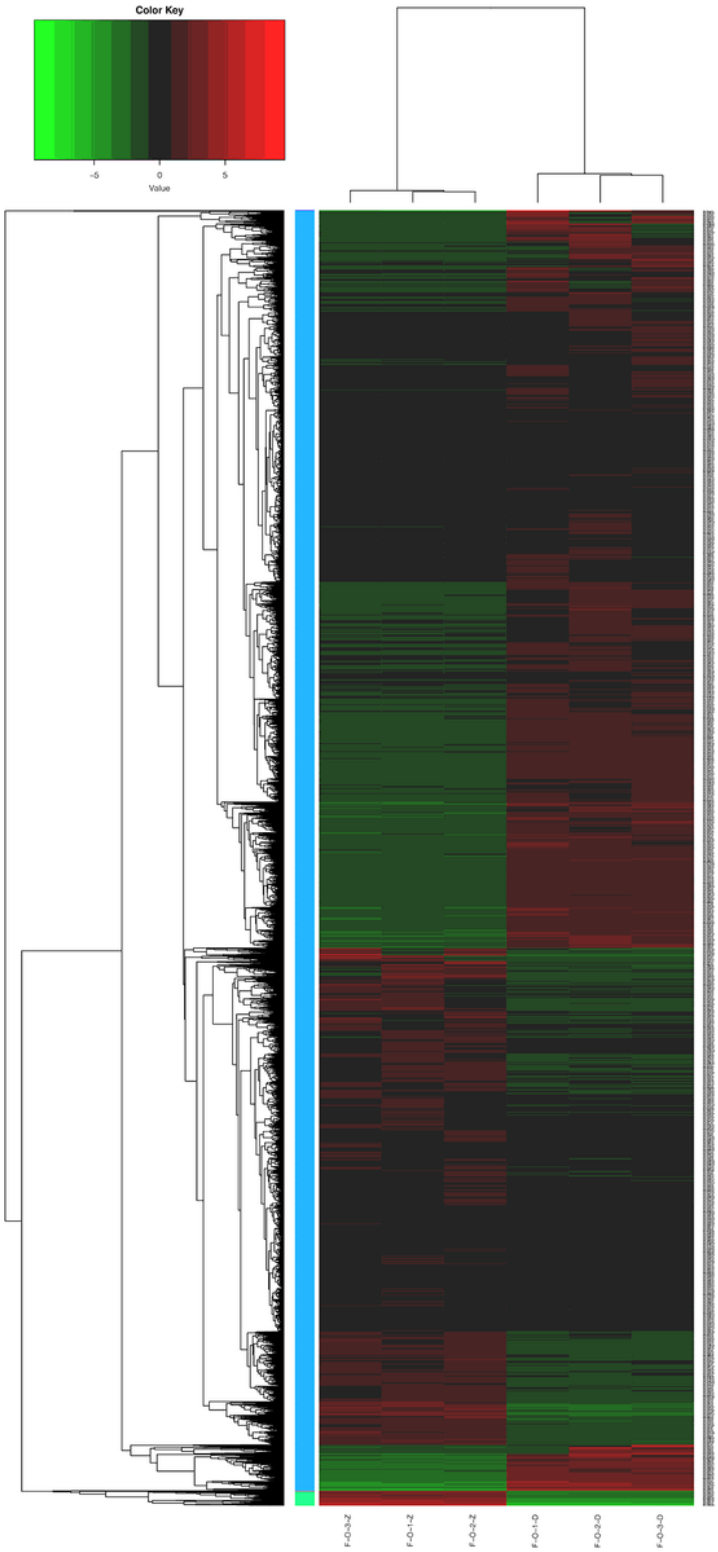

Figure 4

Cluster diagram of patterns for differentially expressed genes (DEGs) among samples. Each column represents a sample and each row represents a gene. The color in the graph represents the gene expression level in the group of samples (log10 FPKM). Red represents a high expression level of the gene in the sample, and green represents a low expression level.

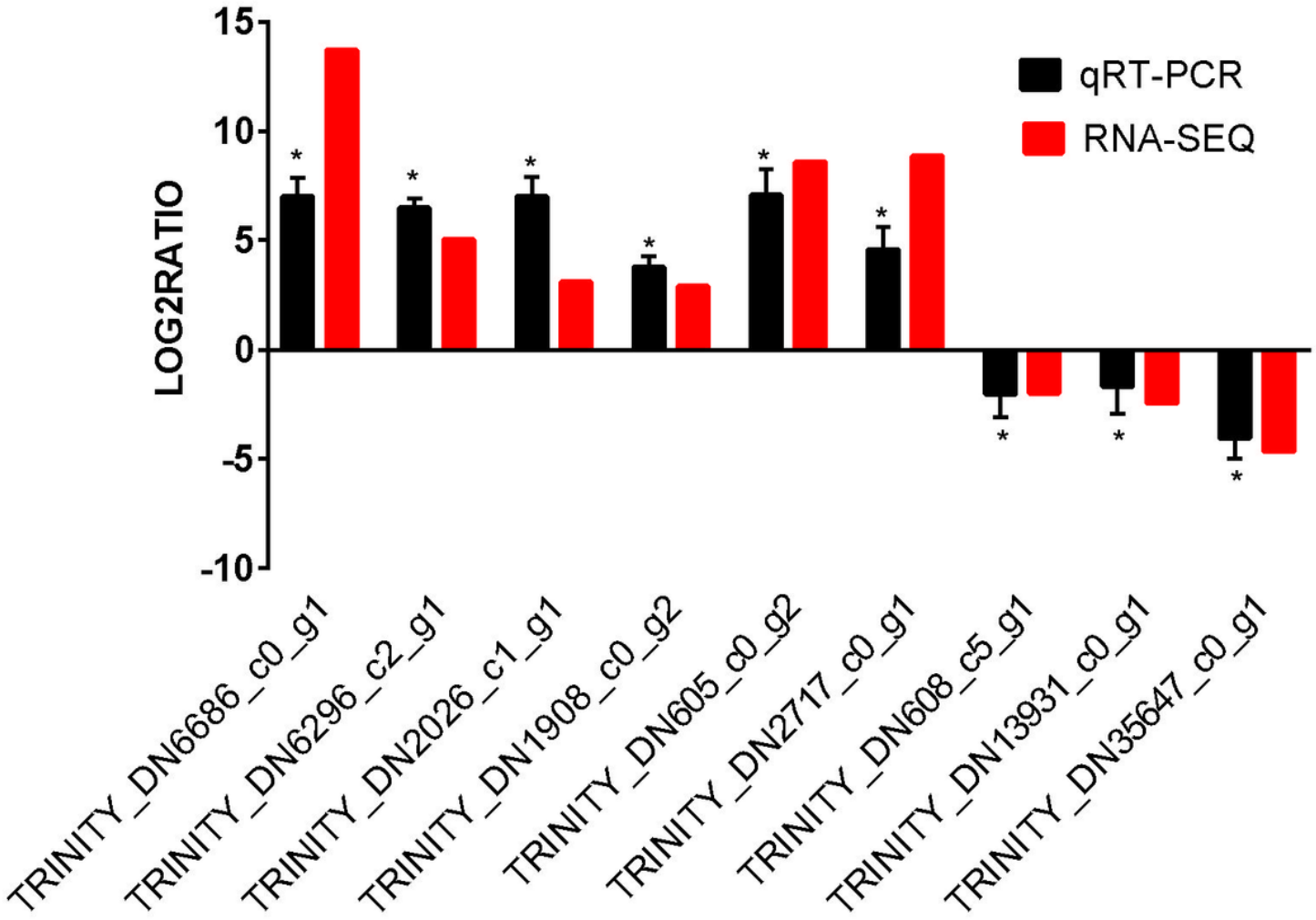

Figure 5

The relative expression levels of 11 genes had the same expression trend as that in the sequencing results. Data are expressed as mean ± standard deviation. *p < 0.05 compared with control.