Pakistan J. Zool., vol. 49(2), pp 535-542, 2017. DOI: http://dx.doi.org/10.17582/journal.pjz/2017.49.2.535.542

Development of Simple Sequence Repeats (SSR) by Transcriptome in Chinese Mitten Crab (*Eriocheir sinensis* H. Milne Edwards)

Fukuan Du[1,2], Yan Li[2], Zhixin Wen[1] , Ruguo Chen[1] and Pao Xu[2]*

[1]*Jiangsu Red Grease Crab Co. Ltd, Xinghua, Jiangsu, 225766, China*

[2]*Wuxi Fisheries College, Nanjing Agricultural University, Wuxi, Jiangsu, 214081, China*

ABSTRACT

The Chinese mitten crab (*Eriocheir sinensis*) is one of the most studied and economically important crustaceans in China. However, the production scale does not meet the market demand. This resulted in excessive fishing, which almost caused extinction of the species in the the Yangtze River. Therefore, it is very urgent to carry out genetic breeding to improve yield. In this study, we developed SSR markers by transcriptome, which could be used to genetic research. In this study, we performed transcriptome sequencing of six libraries from hepatopancreas samples in *E. sinensis*. After removing low-quality reads, short reads and reads belonging to mitochondria, a total of 278,260,960 clean reads corresponding to mRNAs were obtained, these reads covered a total of 34,780,598,273 bases. A total of 48,657 SSR loci were recognized, of which 10,310 unigene sequences contain more than one SSR. These genetic markers are ideal molecular markers for the following investigating genetic diversity, constructing genetic maps, and marker assisted selection.

Article Information
Received 27 August 2016
Revised 08 October 2016
Accepted 13 October 2016
Available online 08 February 2017

Authors' Contributions
PX and RC designed the study. YL and ZW performed the experimental work. FD performed bioinformatics analysis and wrote the article.

Key words
Eriocheir sinensis, Transcriptome, SSR

INTRODUCTION

The Chinese mitten crab *Eriocheir sinensis*, native to East Asia, is the most important aquatic economic animals with a high commercial value as a food source throughout the northern and central coastal regions of China (Wang *et al*., 2008). *E. sinensis* has received increasing attention over the years and has been cultured by fishermen because of its wonderful flavor and high commercial value. The aquaculture production in China reached 600,000 tons per year (Cui *et al*., 2015). However, the production scale does not meet the market demand. This resulted in excessive fishing, which almost caused extinction of the species in the Yangtze River. Therefore, it is very urgent to perform genetic research on *E. sinensis* to protect the wild population and carry out genetic breeding to improve yield.

Currently, transcriptome has been applied in the *E. sinensis*. Sex-biased genes are considered to account for most of phenotypic differences between males and females. In order to explore the sex-biased gene expression in crab, several studies have been performed the whole-body (Liu *et al*., 2015), testis (Zhang *et al*., 2011; Jiang *et al*., 2009) and accessory sex glands (He *et al*., 2013) transcriptome analysis in male and female juveniles of *E. sinensis* using next-generation sequencing technology. Hui *et al*. (2014) detailed the molecular basis of osmoregulation and the stress adaption mechanisms of larvae at key developmental stages with a comparative transcriptomic analysis of *E. sinensis* megalopae before and after desalination. In the study of Li *et al*. (2015b), comparative transcriptomic analysis between the fifth zoeae and megalopae of *E. sinensis* was performed. Patterns of differential gene expression at the two developmental stages were screened in order to reveal the molecular basis of the principal change in the larvae at the two stages. These studies provided the different tissues RNA sequencing for *E. sinensis* and will facilitate further studies on molecular mechanisms of sex control, stress adaption and development in crab.

DNA markers are useful for the genetic research and the genetic breeding of aquacultured organisms (Wei *et al*., 2016). Several types of genetic markers, including microsatellites (Liu *et al*., 2011), SNP (Ma *et al*., 2011b), complete mitochondrial DNA (Qi *et al*., 2007) and AFLP (Wei, 2010) have been developed to assist in the improvement and enhancement of the economically important traits of aquatic animals. However, the SSR markers development in *E. sinensis* is limited.

Microsatellites are nuclear genetic markers that are considered an ideal molecular marker system for investigating genetic diversity (Dudaniec *et al*., 2010), constructing genetic maps (Ma *et al*., 2011a; Song *et al*.,

* Corresponding author: xup13806190669@163.com
0030-9923/2017/0002-0535 $ 9.00/0

2012), and marker assisted selection (MAS) (Ma *et al*., 2014; Fuji *et al*., 2007). So in this study, we will develop SSR markers by transcriptome, which could be used to genetic research.

MATERIALS AND METHODS

Experimental animals

E. sinensis (average weight, 12.4 ± 2.04 g) were adapted to the conditions in a 7.0 × 5.0 × 1.0 m^3 aquarium with a water temperature of 28.5 ± 1.0 °C, pH 7.2, and dissolved oxygen concentration of 9.2 ± 0.5 mg O_2/L dechlorinated and aerated water. The crabs were fed twice daily, at 7:00 AM and 5:00 PM. At the onset of the experiments, all crabs appeared healthy.

Tissue sampling

During the experiments, euthanized crabs were submerged immediately in crushed ice to retard degradation of RNA. Blood was collected into ammonium-heparinized capillary tubes at the arthrodial membrane of the last walking leg. All crabs appeared healthy during dissection and their hepatopancreas were removed and placed in liquid nitrogen. Plasma was separated by centrifugation. Plasma and hepatopancreas samples were stored at −80°C until later analysis.

Animal welfare and experimental procedures were carried out in accordance with the Guide for the Care and Use of Laboratory Animals (Ministry of Science and Technology of China, 2006), and were approved by the animal ethics committee of Chinese Academy of Fishery Sciences.

RNA sequencing and assembly

The samples were used to construct six separate, normalized cDNA libraries. Transcriptome sequencing was carried out on an Illumina HiSeq 2500 platform that generated approximately 278,260,960-bp paired-end (PE) raw reads (OE Biotech Co., Shanghai, China, Table I). Raw data were deposited in the NCBI Sequence Read Archive under the accession number (PRJNA318957). After removing adaptor sequences, ambiguous 'N' nucleotides (with the ratio of 'N' greater than 5%) and low quality sequences (with quality score less than 10), the remaining clean reads were assembled using trinity software (Grabherr *et al*., 2011) as described for *de novo* transcriptome assembly without a reference genome.

Unigene SSR analysis

SSR detection is done with software Microsatellite (MISA, http://pgrc.ipk-gatersleben.de/misa/misa.html) using unigenes as reference. Keep the SSRs which the length of the both ends on the Unigene are more than 150 bp. We use these sequences to design primers using software primer 3-2.3.4. Then, filtering primers by follows: 1) No SSRs in the primer. 2) Align the primers to unigenes sequence with 5' site allowed 3' mismatch and 3' site allowed 1 mismatch. 3) Remove the primers which aligned to more than one unigene. 4) Find SSRs on the product sequences by SSR finder. Keep the product which SSR finder's result is same to the MISA's result.

RESULTS

Generation of Coilia nasus *transcriptome data*

In this study, we performed transcriptome sequencing of six libraries from hepatopancreas samples in *E. sinensis* via an Illumina HiSeq 2500 platform sequencer: 46.4, 46.4, 46.3, 46.3, 46.5 and 46.3 million reads were obtained from the six libraries. After removing low-quality reads, short reads and reads belonging to mitochondria, a total of 278,260,960 clean reads corresponding to mRNAs were obtained, these reads covered a total of 34,780,598,273 bases (Table I).

De novo *assembly of* Coilia nasus *transcriptome data*

Using the Trinity assembly program, we generated a total of 99,983 unigenes (Table II). The length distribution of unigenes larger than 200 bp is shown in Figure 1. The mean unigene size and N50 were 814.8 bp and 1,074 bp, respectively. About half of the unigenes (48,157; 48.2%) were ≥ 500 bp. The largest unigene was 21,266 bp in length (Table II). The GC content frequency distribution was showed in Figure 2.

Table I.- Summary of sequence data generated for the *Coilia nasus* transcriptome and quality filtering.

Sample	Raw reads	Raw bases	Clean reads	Clean Nucleotides (nt)	Q30 percentage (%)
1	46,355,800	5,794,475,000	46,355,800	5,794,137,861	94.14
2	46424000	5,803,000,000	46,424,000	5,802,660,977	94.54
3	46328880	5,791,110,000	46,328,880	5,790,773,653	94.47
4	46294340	5,786,792,500	46,294,340	5,786,453,822	94.54
5	46544800	5,818,100,000	46,544,800	5,817,766,889	94.32
6	46313140	5,789,142,500	46,313,140	5,788,805,071	94.67

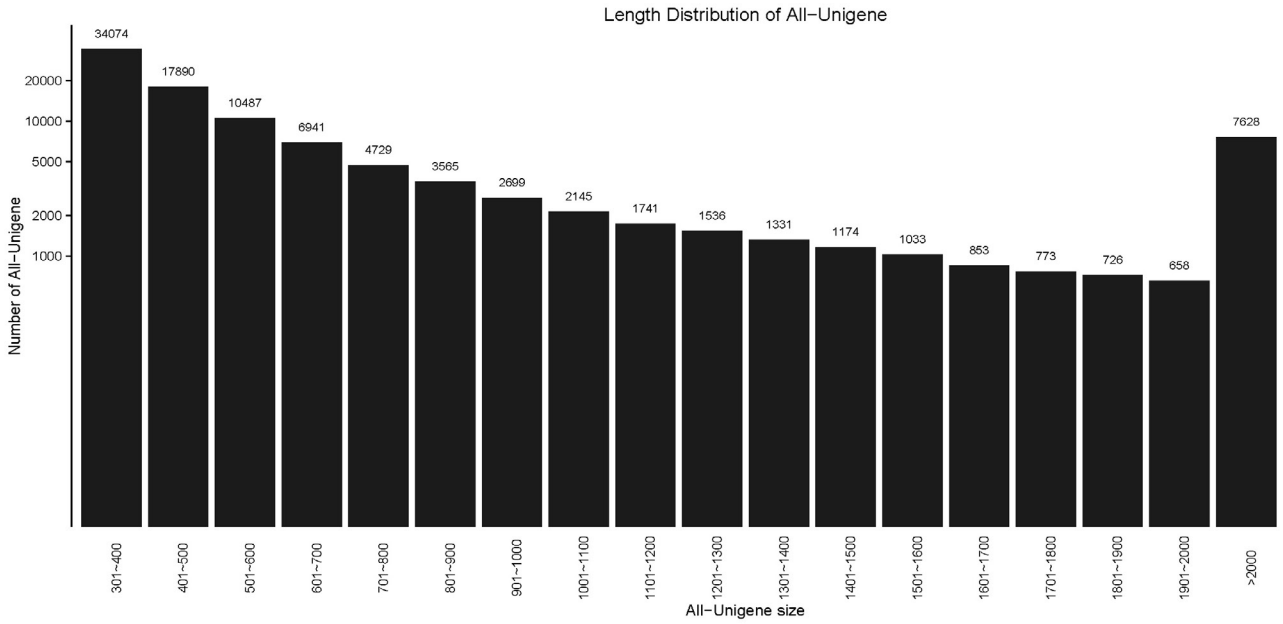

Fig. 1. Length distribution of unigenes.

Table II.- Assembly statistics of reads.

Parameter	Numbers
Number of Unigene	99,983
Total bases of Unigene (bp)	81,466,380
Unigene mean lengths (bp)	814.8
Number of Unigene ≥500 bp	48,157
N50	1,074
Max length (bp)	21,266

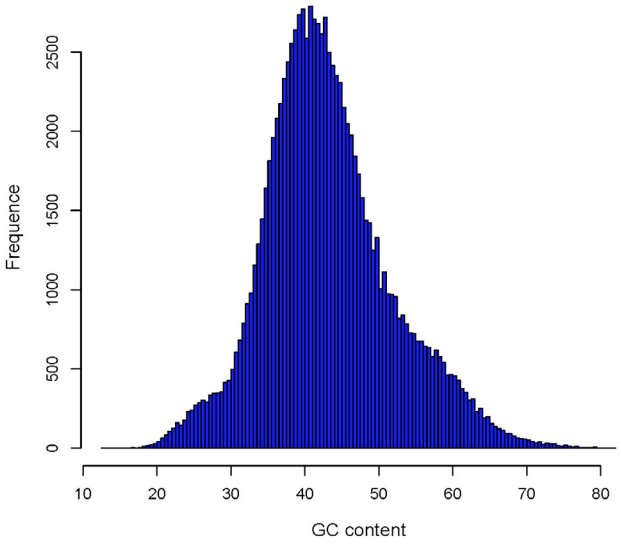

Fig. 2. GC content frequency distribution.

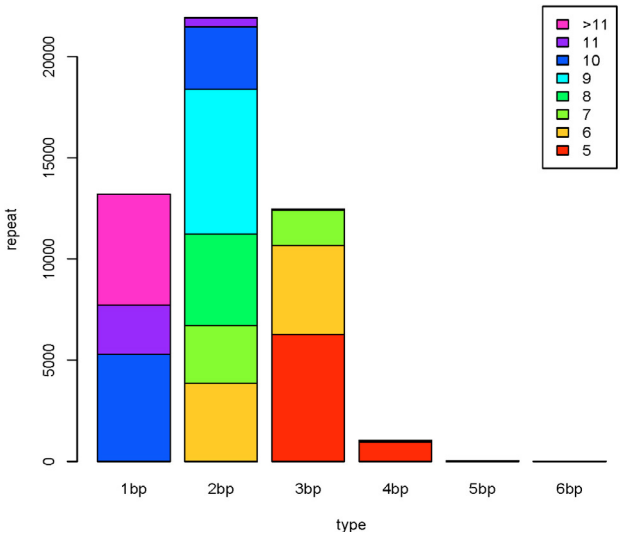

Fig. 3. SSR type statistics.

Unigene SSR filter and analysis

SSR detection is done with software MicroSAtellite (MISA) using unigenes as reference. In this study, a total of 48,657 SSR loci were recognized, of which 10,310 unigene sequences contain more than one SSR (Table III). The most common type was di-nucleotide, followed by mono-nucleotide, tri-nucleotide, hexa-nucleotide, penta-nucleotide and tetra-nucleotide (Fig. 3). The number of repeat motifs per locus ranged from 5 to 12, in which SSRs with 10 repeats were the most abundant, followed

Table III.- Statistics of SSR types.

Repeats	5	6	7	8	9	10	11	>11	Total	%
1bp	0	0	0	0	0	5295	2420	5489	13204	27.14
2bp	0	3845	2866	4513	7151	3078	449	14	21916	45.04
3bp	6267	4398	1740	62	2	0	1	3	12473	25.63
4bp	945	72	5	2	2	0	0	3	1029	2.11
5bp	20	3	1	0	1	0	0	2	27	0.06
6bp	4	3	0	0	0	0	1	0	8	0.02
Total	7236	8321	4612	4577	7156	8373	2871	5511	48657	100
%	14.87	17.10	9.48	9.41	14.71	17.21	5.90	11.33	100	

by loci with six, five, and nine repeats. The number of SSR repeats of > 11 was rare (Table III). The most abundant di-nucleotide repeat motif was AC/GT, followed by AG/CT, AT/AT; CG/GG was the least abundant. Among the tri-nucleotide repeat units, the dominant motif was AGG/CCT, followed by ACC/GCT and AAT/ATT (Table IV, placed at the end).

DISCUSSION

Several studies have been performed focusing on the *E. sinensis* transcriptome of eyestalk, testis, and the Y-organ. These studies secreted and produced a variety of neuropeptide hormones which regulates important physiological activities of the crustacean animals, such as growth, metabolism and reproduction (Wang *et al*., 2016). Huang *et al*. (2015) used RNA-Seq to investigate transcriptomic profiles of the hepatopancreas and identified differentially expressed genes at four molting stages of *E. sinensis*. This study provided novel insights into the functions of the hepatopancreas in energy metabolism and biological processes pertaining to molting in crustaceans. However, these studies did not develop SSR markers in the transcriptome. In our study, similar numbers of unigene sequences were generated after assembly compared to the previous study, which was partly due to the fact the same tissues were collected for sequencing. In parallel, the length distribution of contigs and unigenes in these studies were also coincidental, which indicated that the Illumina-based sequencing technology was successful and the assembly was relatively reliable. Therefore, the large number of unigenes obtained from this work could substantially increase the genomic information of *E. sinensis*.

Many studies have demonstrated that transcriptome sequencing was a powerful tool for identifying SSR markers (Luo *et al*., 2016; Fang *et al*., 2015; Li *et al*., 2015a; Shabbir, 2014). Various SSRs derived from transcriptome sequencing have been extensively used in aquatic animals genetic diversity analyses (Mu *et al*., 2015; Chen *et al*., 2011; Long *et al*., 2015). However, until now no SSRs were identified and used to evaluate genetic diversity for *E. sinensis*. In this study, a lot of unigene sequences that contained microsatellite loci were detected. The percentage of SSRs contained sequences was higher than *Macrobrachium rosenbergii* (Jung *et al*., 2011) and *Portunus trituberculatus* (Lv *et al*., 2014), but was lower than *Macrobrachium nipponense* (Jin *et al*., 2013). The discrepant SSRs frequency might be mainly caused by the parameters of tools for searching microsatellite loci. The mean density of SSR distribution was one microsatellite locus per 1.6 kb. This density was higher than the density in *M. nipponense* (1/7.8 kb) (Jin *et al*., 2013). The different distribution frequency of microsatellite loci may be partially related to genome composition, data size, and microsatellite screening criteria. Di-nucleotide repeats were the most abundant repeat type in our study. This finding was in accordance with previous studies using *M. rosenbergii*, *M. nipponense* and *P. trituberculatus* (Jung *et al*., 2011; Lv *et al*., 2014; Jin *et al*., 2013). This difference may also be caused by different parameters for detecting microsatellite and different genome composition of each species. The dominant di-nucleotide and tri-nucleotide repeat motif of *E. sinensis* was AC/GT and AGG/CCT. The lowest frequent motif was CG/CG which was also rare in studies using the other crustacean animals (Zhou *et al*., 2016).

ACKNOWLEDGMENTS

This work was supported by a grant from the China Postdoctoral Science Foundation (2016M590427), the National Key Technology R&D Program (2012BAD25B07); The Jiangsu Postdoctoral Sustentation Fund, China (1501145B).

Availability of supporting data

Raw sequencing data is available through the

NCBI Sequence Read Archive under Project Accession SRP034828 (http://trace.ncbi.nlm.nih.gov/Traces/sra/). All samples were sequenced as 90 bp paired-end reads on an Illumina HiSeq2500 sequencer.

Conflict of interest statement

We declare that we have no conflict of interest.

REFERENCES

Chen, L., Xu, H., Li, H., Wu, J., Ding, H. and Liu, Y., 2011. Isolation and characterization of sixteen polymorphic microsatellite loci in the golden apple snail *Pomacea canaliculata*. *Int. J. mol. Sci.,* **12**: 5993-5998. https://doi.org/10.3390/ijms12095993

Cui, Z., Hui, M., Liu, Y., Song, C., Li, X., Li, Y., Liu, L., Shi, G., Wang, S., Li, F., Zhang, X., Liu, C., Xiang, J. and Chu, K.H., 2015. High-density linkage mapping aided by transcriptomics documents ZW sex determination system in the Chinese mitten crab *Eriocheir sinensis*. *Heredity (Edinb),* **115**: 206-215. https://doi.org/10.1038/hdy.2015.26

Dudaniec, R.Y., Storfer, A., Spear, S.F. and Richardson, J.S., 2010. New microsatellite markers for examining genetic variation in peripheral and core populations of the coastal giant salamander (*Dicamptodon tenebrosus*). *PLoS One,* **5**: e14333. https://doi.org/10.1371/journal.pone.0014333

Fang, D.A., Zhou, Y.F., Duan, J.R., Zhang, M.Y., Xu, D.P., Liu, K., Xu, P. and Wei, Q., 2015. Screening potential SSR markers of the anadromous fish *Coilia nasus* by de novo transcriptome analysis using Illumina sequencing. *Genet. mol. Res.,* **14**: 14181-14188. https://doi.org/10.4238/2015.November.13.1

Fuji, K., Hasegawa, O., Honda, K., Kumasaka, K., Sakamoto, T. and Okamoto, N., 2007. Marker-assisted breeding of a lymphocystis disease-resistant Japanese flounder (*Paralichthys olivaceus*). *Aquaculture,* **272**: 291-295.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. and Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.,* **29**: 644-652. https://doi.org/10.1038/nbt.1883

He, L., Jiang, H., Cao, D., Liu, L., Hu, S. and Wang, Q., 2013. Comparative transcriptome analysis of the accessory sex gland and testis from the Chinese mitten crab (*Eriocheir sinensis*). *PLoS One,* **8**: e53915. https://doi.org/10.1371/journal.pone.0053915

Huang, S., Wang, J., Yue, W., Chen, J., Gaughan, S., Lu, W., Lu, G. and Wang, C., 2015. Transcriptomic variation of hepatopancreas reveals the energy metabolism and biological processes associated with molting in Chinese mitten crab, *Eriocheir sinensis*. *Sci. Rep.,* **5**: 14015. https://doi.org/10.1038/srep14015

Hui, M., Liu, Y., Song, C., Li, Y., Shi, G. and Cui, Z., 2014. Transcriptome changes in *Eriocheir sinensis* megalopae after desalination provide insights into osmoregulation and stress adaption in larvae. *PLoS One,* **9**: e114187. https://doi.org/10.1371/journal.pone.0114187

Jiang, H., Yin, Y., Zhang, X., Hu, S. and Wang, Q., 2009. Chasing relationships between nutrition and reproduction: A comparative transcriptome analysis of hepatopancreas and testis from *Eriocheir sinensis*. *Comp. Biochem. Physiol.,* Part D: *Genom. Proteom.,* **4**: 227-234.

Jin, S., Fu, H., Zhou, Q., Sun, S., Jiang, S., Xiong, Y., Gong, Y., Qiao, H. and Zhang, W., 2013. Transcriptome analysis of androgenic gland for discovery of novel genes from the oriental river prawn, *Macrobrachium nipponense*, using Illumina Hiseq 2000. *PLoS One,* **8**: e76840. https://doi.org/10.1371/journal.pone.0076840

Jung, H., Lyons, R.E., Dinh, H., Hurwood, D.A., Mcwilliam, S. and Mather, P.B., 2011. Transcriptomics of a giant freshwater prawn (*Macrobrachium rosenbergii*): *de novo* assembly, annotation and marker discovery. *PLoS One,* **6**: e27938. https://doi.org/10.1371/journal.pone.0027938

Li, C., Ling, Q., Ge, C., Ye, Z. and Han, X., 2015a. Transcriptome characterization and SSR discovery in large-scale loach *Paramisgurnus dabryanus* (Cobitidae, Cypriniformes). *Gene,* **557**: 201-208. https://doi.org/10.1016/j.gene.2014.12.034

Li, Y., Hui, M., Cui, Z., Liu, Y., Song, C. and Shi, G., 2015b. Comparative transcriptomic analysis provides insights into the molecular basis of the metamorphosis and nutrition metabolism change from zoeae to megalopae in *Eriocheir sinensis*. *Comp. Biochem. Physiol.,* Part D: *Genom. Proteom.,* **13**: 1-9.

Liu, B., Teng, S.S., Shao, Y.Q., Chai, X.L. and Xiao, G.Q., 2011. Isolation and characterization of 39 novel polymorphic EST-SSR loci for the blood

clam, *Tegillarca granosa*. *Conserv. Genet. Resour.,* **4**: 375-378. https://doi.org/10.1007/s12686-011-9552-y

Liu, Y., Hui, M., Cui, Z., Luo, D., Song, C., Li, Y. and Liu, L., 2015. comparative transcriptome analysis reveals sex-biased gene expression in juvenile Chinese mitten crab *Eriocheir sinensis*. *PLoS One,* **10**: e0133068. https://doi.org/10.1371/journal.pone.0133068

Long, Y., Wang, Y., Wu, S., Wang, J., Tian, X. and Pei, X., 2015. *De novo* assembly of transcriptome sequencing in *Caragana korshinskii Kom*. and characterization of EST-SSR markers. *PLoS One,* **10**: e0115805. https://doi.org/10.1371/journal.pone.0115805

Luo, H., Xiao, S., Ye, H., Zhang, Z., Lv, C., Zheng, S., Wang, Z. and Wang, X., 2016. Identification of immune-related genes and development of SSR/SNP markers from the spleen transcriptome of *Schizothorax prenanti*. *PLoS One,* **11**: e0152572. https://doi.org/10.1371/journal.pone.0152572

Lv, J., Liu, P., Gao, B., Wang, Y., Wang, Z., Chen, P. and Li, J., 2014. Transcriptome analysis of the *Portunus trituberculatus*: *de novo* assembly, growth-related gene identification and marker discovery. *PLoS One,* **9**: e94055. https://doi.org/10.1371/journal.pone.0094055

Ma, H., Chen, S., Yang, J., Chen, S. and Liu, H., 2011a. Genetic linkage maps of barfin flounder (*Verasper moseri*) and spotted halibut (*Verasper variegatus*) based on AFLP and microsatellite markers. *Mol. Biol. Rep.,* **38**: 4749-4764. https://doi.org/10.1007/s11033-010-0585-1

Ma, H., Jiang, W., Liu, P., Feng, N., Ma, Q., Ma, C., Li, S., Liu, Y., Qiao, Z. and Ma, L., 2014. Identification of transcriptome-derived microsatellite markers and their association with the growth performance of the mud crab (*Scylla paramamosain*). *PLoS One,* **9**: e89134. https://doi.org/10.1371/journal.pone.0089134

Ma, H., Ma, Q., Ma, C. and Ma, L., 2011b. Isolation and characterization of gene-derived single nucleotide polymorphism (SNP) markers in *Scylla paramamosain*. *Biochem. Syst. Ecol.,* **39**: 419-424. https://doi.org/10.1016/j.bse.2011.05.024

Mu, X., Hou, G., Song, H., Xu, P., Luo, D., Gu, D., Xu, M., Luo, J., Zhang, J. and Hu, Y., 2015. Transcriptome analysis between invasive *Pomacea canaliculata* and indigenous *Cipangopaludina cahayensis* reveals genomic divergence and diagnostic microsatellite/SSR markers. *BMC Genet.,* **16**: 12. https://doi.org/10.1186/s12863-015-0175-2

Qi, D., Guo, S., Tang, J., Zhao, X. and Liu, J., 2007. Mitochondrial DNA phylogeny of two morphologically enigmatic fishes in the subfamily Schizothoracinae (*Teleostei: Cyprinidae*) in the Qinghai-Tibetan Plateau. *J. Fish Biol.,* **70**: 60-74. https://doi.org/10.1111/j.1095-8649.2007.01366.x

Shabbir, M.I., 2014. A simple method of bead counting for next generation DNA sequencing. *Pakistan J. Zool.,* **46**: 1385-1389.

Song, W., Li, Y., Zhao, Y., Liu, Y., Niu, Y., Pang, R., Miao, G., Liao, X., Shao, C., Gao, F. and Chen, S., 2012. Construction of a high-density microsatellite genetic linkage map and mapping of sexual and growth-related traits in half-smooth tongue sole (*Cynoglossus semilaevis*). *PLoS One,* **7**: e52097. https://doi.org/10.1371/journal.pone.0052097

Wang, B., Ning, Q., Hao, T., Yu, A. and Sun, J., 2016. Reconstruction and analysis of a genome-scale metabolic model for *Eriocheir sinensis* eyestalks. *Mol. Biosyst.,* **12**: 246-252. https://doi.org/10.1039/C5MB00571J

Wang, C., Li, C. and Li, S., 2008. Mitochondrial DNA-inferred population structure and demographic history of the mitten crab (*Eriocheir sensu stricto*) found along the coast of mainland China. *Mol. Ecol.,* **17**: 3515-3527. https://doi.org/10.1111/j.1365-294x.2008.03850.x

Wei, D., 2010. RFLP analysis of the ribosomal DNA ITS region of three geographical populations of *Tegillarca granosa*. *J. Anhui agric. Sci.,* **10**: 1-7.

Wei, Z., Chen, R.Y., Laghari, M.Y., Xu, D.D., Mao, G.M., Shi, H.L. and Bao, L., 2016. Characterization of *Nibea albiflora* transcriptome: sequencing, *de novo* assembly, annotation and comparative genomics. *Pakistan J. Zool.,* **48**: 427-434.

Zhang, W., Wan, H., Jiang, H., Zhao, Y., Zhang, X., Hu, S. and Wang, Q., 2011. A transcriptome analysis of mitten crab testes (*Eriocheir sinensis*). *Genet. Mol. Biol.,* **34**: 136-141. https://doi.org/10.1590/S1415-47572010005000099

Zhou, T., Li, Z.H., Bai, G.Q., Feng, L., Chen, C., Wei, Y., Chang, Y.X. and Zhao, G.F., 2016. Transcriptome sequencing and development of genic SSR markers of an endangered Chinese endemic genus *Dipteronia* Oliver (Aceraceae). *Molecules,* **21**: 166. https://doi.org/10.3390/molecules21030166

Table IV.- Frequency of classified repeat types (considering sequence complementary).

Repeats	5	6	7	8	9	10	11	12	13	14	15	16	Total
AC/GT		2271	1775	2899	4406	1623	194	4					13172
AG/CT		1132	868	1429	2426	1060	154	6					7075
AT/AT		427	220	182	315	394	101	2					1641
CG/CG		15	3	3	4	1		1	1				28
AAC/GTT	402	260	131	4									797
AAG/CTT	420	246	137	8								1	812
AAT/ATT	632	615	339	9									1595
ACC/GGT	1172	788	245	8									2213
ACG/CGT	28	10	6	4			1						49
ACT/AGT	495	587	212	3							1		1298
AGC/CTG	503	266	154	4									927
AGG/CCT	1931	1257	362	12	1								3563
ATC/ATG	394	247	116	7								1	765
CCG/CGG	290	122	38	3	1								454
AAAC/GTTT	70	6	1										77
AAAG/CTTT	52	4											56
AAAT/ATTT	124	7			1								132
AACC/GGTT	5												5
AACG/CGTT		1											1
AACT/AGTT	10	1								1			12
AAGC/CTTG	1												1
AAGG/CCTT	149	5									2		156
AAGT/ACTT	3	1											4
AATC/ATTG	20	3		1									24
AATG/ATTC	18	4	1										23
AATT/AATT	2												2
ACAG/CTGT	90	3											93
ACAT/ATGT	85	7	1		1								94
ACCG/CGGT	3												3
ACCT/AGGT	8	2		1									11
Repeats	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**Total**
ACGC/CGTG	16	4											20
ACGG/CCGT	3	1											4
ACTC/AGTG	63	2											65
ACTG/AGTC	87	9											96
AGAT/ATCT	102	7											109
AGCC/CTGG	10	3	1										14
AGCG/CGCT	1	1											2
AGGC/CCTG	6	1	1										8
AGGG/CCCT	10												10

ATCC/ATGG	7							7
AAAAC/GTTTT	2							2
AAAAG/CTTTT		1						1
AAAAT/ATTTT		1						1
AAATT/AATTT	1							1
AACCC/GGGTT	1							1
AAGAG/CTCTT						1		1
AAGCC/CTTGG	1							1
AAGGG/CCCTT	2							2
AATAC/ATTGT	1							1
AATGC/ATTGC	1							1
AATGT/ACATT	1							1
AATTC/AATTG	1							1
ACACC/GGTGT				1				1
ACAGC/CTGTG	1							1
ACCCC/GGGGT	1							1
ACCTC/AGGTG			1					1
ACCTG/AGGTC	1							1
ACGCG/CGCGT	1							1
ACGGC/CCGTG	1							1
ACGGG/CCCGT	1							1
AGAGG/CCTCT	1						1	2
AGCAT/ATGCT	1							1
AGCCC/CTGGG	1							1
AGGCG/CCTCG	1							1
AAAAAG/CTTTTT	1							1
AAAGAG/CTCTTT		1						1
AACTAT/AGTTAT		1						1
AAGGAC/CCTTGT	1							1
AATAGT/ACTATT					1			1
ACCATC/ATGGTG	1							1
AGATAT/ATATCT	1							1
AGCCCC/CTGGGG		1						1