J. Ocean Univ. China (Oceanic and Coastal Sea Research)
https://doi.org/10.1007/s11802-018-3525-7
ISSN 1672-5182, 2018 17 (3): 632-640
http://www.ouc.edu.cn/xbywb/
E-mail:xbywb@ouc.edu.cn

Characterization of Chiton *Ischnochiton hakodadensis* Foot Based on Transcriptome Sequencing

DOU Huaiqian[1), #], MIAO Yan[1), #], LI Yuli[1), *], LI Yangping[1)], DAI Xiaoting[1)], ZHANG Xiaokang[1)], LIANG Pengyu[1)], LIU Weizhi[1), 2)], WANG Shi[1), 2)], and BAO Zhenmin[1), 3)]

1) *Key Laboratory of Marine Genetics and Breeding of Ministry of Education*, *Ocean University of China*, *Qingdao* 266003, *China*
2) *Laboratory for Marine Biology and Biotechnology*, *Qingdao National Laboratory for Marine Science and Technology*, *Qingdao* 266071, *China*
3) *Laboratory for Marine Fisheries Science and Food Production Processes*, *Qingdao National Laboratory for Marine Science and Technology*, *Qingdao* 266071, *China*

(Received April 14, 2017; revised July 7, 2017; accepted March 20, 2018)

Abstract Chiton (*Ischnochiton hakodadensis*) is one of marine mollusks well known for its eight separate shell plates. *I. hakodadensis* is important, which plays a vital role in the ecosystems it inhabits. So far, the genetic studies on the chiton are scarce due in part to insufficient genomic resources available for this species. In this study, we investigated the transcriptome of the chiton foot using Illumina sequencing technology. The reads were assembled and clustered into 256461 unigenes, of which 42247 were divided into diverse functional categories by Gene Ontology (GO) annotation terms, and 17256 mapped onto 365 pathways by KEGG pathway mapping. Meanwhile, a set of differentially expressed genes (DEGs) between distal and proximal muscles were identified as the foot adhesive locomotion associated, thus were useful for our future studies. Moreover, up to 679384 high-quality single nucleotide polymorphisms (SNPs) and 19814 simple sequence repeats (SSRs) were identified in this study, which are valuable for subsequent studies on genetic diversity and variation. The transcriptomic resource obtained in this study should aid to future genetic and genomic studies of chiton.

Key words chiton; transcriptome sequencing; DEG; SNP; SSR

1 Introduction

Chitons in class Polyplacophora are marine mollusks, varying in size and being oval. Currently 940 extant species (Schwabe, 2005) and 430 fossil species (Puchalski *et al*., 2008) have been recognized. They widely distribute all over the world and live either on or under rocks. The chiton shell is composed of eight separate shell plates overlapping slightly at the front and back edges (Connors *et al*., 2012). Compared with other mollusks, such as those with a single- or two-piece shell type, the eight separate shell plates allow chitons to roll into a protective ball or cling tightly to irregular surfaces when they encounter dangers. The shell embedded in a tough muscular girdle that surrounds the chiton body (Vinther and Nielsen, 2005). In some species, the valves reduced or covered by the girdle tissue (Treves *et al*., 2003). Chitons look for food at night and are generally herbivorous though some of them are omnivorous and some are carnivorous (Kangas and Shepherd, 2013). Their muscular feet can protect them from predators and strong crashing waves, thus it is difficult for them to dislodge from surfaces. The chiton foot has considerable power of adhesion and can cling to rocks very tightly, similar to true limpets (Smith *et al*., 1999) and abalone (Byern and Grunwald, 2010). Although chitons have many interesting biological characteristics, genomic, transcriptomic and proteomic studies on them are relatively rare. Additional research at the genetic level is appreciated in order to understand them better.

Next-generation sequencing (NGS) technology has rapidly progressed, and bioinformatics tools have enabled us to analyze quickly massive amounts of sequence data at a low cost (Romkes and Buch, 2014), there is great potential to expand transcriptomic resources to additional non-model organisms. Many marine species have their transcriptomes sequenced. The sequenced included for example mussels (Bettencourt *et al*., 2010; Craft *et al*., 2010), clams (Huan *et al*., 2012; Milan *et al*., 2011), pearl oysters (Joubert *et al*., 2010; Kinoshita *et al*., 2011),

# These authors contributed equally to this work.
* Corresponding author. Tel: 0086-532-82031969
E-mail: lazqd2015@126.com

Patinopecten yessoensis (Hou *et al*., 2011), sea cucumbers (Du *et al*., 2012) and *Chlamys farreri* (Fu *et al*., 2014; Miao *et al*., 2015; Wang *et al*., 2013). The transcriptome of chitons has also been profiled previously (Riesgo *et al*., 2012) in a comparative study with other species. In addition to gene discovery, transcriptome sequencing serves also as an efficient way of discovering genetic variations such as single nucleotide polymorphisms (SNPs) and simple sequence repeats (SSRs).

In this study, we sequenced and analyzed the transcriptome of a chiton (*I. hakodadensis*) foot, and performed gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment for the genes differentially expressed between distal muscle and proximal muscles, aiming to identify genes involving in foot adhesion thus providing references for further studies.

2 Materials and Methods

2.1 Sample Collection and RNA Preparation

I. hakodadensis individuals were collected from Taipingjiao Cape, Qingdao (36°03′2.67′′N, 120°22′12.28′′E), China. After being washed three to five times with clean seawater, *I. hakodadensis* were transferred to an aquarium with circulating seawater maintained at approximately 20℃. To obtain samples for RNA extraction, *I. hakodadensis* were acclimated for 3□5 days with tissues collected from foot. The muscle tissue was cut into two parts, the distal (near the ground layer) and the proximal (away from the ground layer). A total of 6 samples were collected from 3 individuals, which were divided into the distal muscle (S1, S2, S3, hereafter referred to as 'group 1') and the proximal muscle (S1', S2', S3', hereafter referred to as 'group 2'). The samples were rinsed three to five times with sterile water, and frozen in liquid nitrogen.

2.2 Library Construction and Illumina Sequencing

Total RNA was extracted with the method described by Hu *et al*. (2006), and DNA contamination was removed by DNase I treatment. Equal amounts of RNA from each sample were prepared for cDNA synthesis. First-strand cDNA was synthesized using Moloney murine leukemia virus reverse transcriptase (Thermo, USA) following the manufacturer's protocol. PCR products were pooled and purified using the MinElute PCR Purification kit (Qiagen, Germany). The cDNA was sonicated to produce fragments approximately 100–400 bp in length and the majority 200 bp. The cDNA was then purified using QiaQuick PCR extraction kit (Qiagen, Germany). The 200 bp fragments were selected for PCR amplification. Libraries were then prepared using the fragmented cDNA following the user's guide of the NEBNext DNA Library Prep Master Mix Set for Illumina kit (New England Biolabs, USA). The libraries were paired-end (100 bp) sequenced on an Illumina HiSeq 2000 system (Illumina, USA).

2.3 Sequence Assembly and Annotation

The raw reads of all six samples were preprocessed by removing those with adaptors, unknown nucleotides larger than 5% and low-qualities. The clean ones were assembled into unigenes using Trinity software (Grabherr *et al*., 2011) set at the default parameters. Unigenes of six samples were then merged into a non-redundant unigene set using a customized Perl script. Functional annotation of these unigenes was performed by comparison against the Swiss-Prot database using BlastX with an E value cutoff of 1e-5. Then, the Blast results were imported into Blast2GO (Götz *et al*., 2008), and GO terms at level 2 were assigned to produce a broad overview of groups of genes for biological processes, molecular function, and cellular component categories. KEGG annotation was carried out using the KEGG Automatic Annotation Server (KAAS) (Moriya *et al*., 2007).

2.4 Analysis of Differentially Expressed Unigenes

The FPKM (Fragments Per Kilo bases per Million fragments) value of all-unigenes in each sample were obtained and used for comparing the expression difference between samples. Differential expression analysis was performed using the R package edgeR (Robinson *et al*., 2010). *P* values were adjusted using the Benjamini-Hochberg false discovery rate procedure (Benjamini and Hochberg, 1995). Genes with a *P* value ≤ 0.05 were considered to be differentially expressed genes (DEGs). For a general characterization of DEGs, GO and KEGG enrichment analyses were performed using OmicShare tools, a free online platform for data analysis (www.omicshare.com/tools).

2.5 SNP and SSR Discovery

Transcriptome analysis using next-generation sequencing is one of the most efficient methods for SNP identification (Liu *et al*., 2011). Reads from the *I. hakodadensis* transcriptome were first mapped to an assembled reference using the program BWA (Li *et al*., 2009) with the default settings. SNP identification was limited to the contigs containing at least eight reads each allele and the minor allele needs to have a frequency $\geq 20\%$. The alignment output from read mapping was then sorted, and duplicate reads were removed using Picard command line tools. The alignment files were then indexed using SAM tools (Li *et al*., 2009). SNP calling was performed using the VarScan2 software (Koboldt *et al*., 2012). A single locus was believed to be a SNP locus if two alleles were detected. MISA, a microsatellite identification tool, was used to identify and localize microsatellite motifs. All types of SSRs from di-nucleotide to hexa-nucleotide were searched using the default settings as shown below. The minimum length of SSR was considered to be six repeats for a di-nucleotide and four repeats for others.

3 Results and Discussion

3.1 Sequence Assembly and Functional Annotation

As listed in Table 1, 74190988 raw reads were obtained using the Illumina HiSeq 2000 system. After removing adaptors and low-quality reads, 65737373 clean reads were

obtained, and they were assembled into 256461 unigenes. Gene or isoform abundance was represented as the FPKM value, and those transcripts with FPKM values being equal to or larger than 0.5 were defined as being expressed.

Table 1 Summary statistics of the transcriptome assembly for *I. hakodadensis*

	S1	S1'	S2	S2'	S3	S3'
Raw data	1508257125	1518188625	1474395625	1516704375	1598014500	1658313250
Raw reads	12066057	12145509	11795165	12133635	12784116	13266506
Filtered reads	10739803	10573417	10519390	10873504	11296656	11734603.5
Reads length (bp)	100	100	100	100	100	100

The lengths of expressed transcripts ranged from 201 to 28155 bp with an average of 679 bp and an N50 length of 2229 bp. The size distribution of contigs is shown in Fig.1. A total of 256461 unigenes were used for annotation and expression analyes.

For a general characterization of transcriptomes, GO assignment at level 2 was carried out using all unigenes of *I. hakodadensis* transcriptome (Fig.2). Among 256461 unigenes, 24530 (9.56%) significantly matched with the deposits in Swiss-Prot database. A total of 42247 (16.47%) unigenes were classified into diverse functional categories by GO annotation. Among them, 38957 were mapped to biological processes, 39173 to cellular components, and 38587 to molecular functions. Furthermore, 'cellular process' (85.97%) was the highest represented group in the biological process category, followed by 'single-organism process' (75.84%). For the cellular component, 'cell' (89.17%) and 'cell part' (89.12%) were the significantly represented groups, and a high proportion of genes (80.83%) were involved in the binding category of the molecular function ontology. To identify the biological pathways involved in *I. hakodadensis* foot, the KEGG pathway annotations were obtained by aligning all unigenes with the KEGG database on the KEGG Automatic Annotation Server (KAAS). Among the 256461 unigenes, 17256 (6.73%) were associated with 365 pathways.

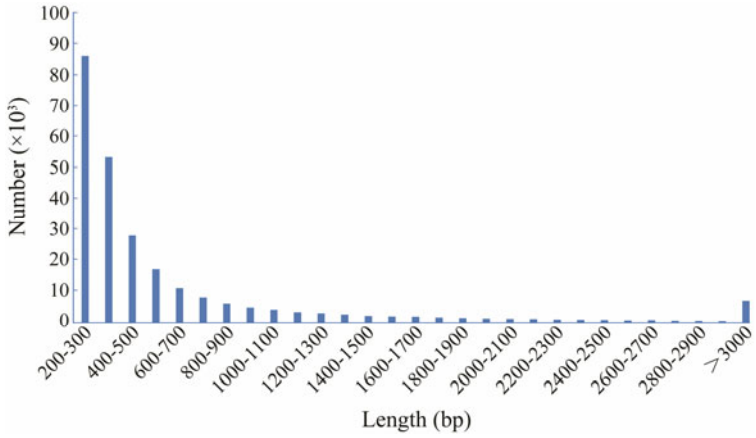

Fig.1 The length distribution of contigs of *I. hakodadensis*. Contigs were generated from *de novo* assembly of Illumina sequencing reads. *X*-axis represents sequence length intervals. *Y*-axis represents the number of unigenes in each interval.

3.2 Enrichment Analysis of DEGs in Chiton Foot

Gene differential expression analysis carried out using the R package edgeR revealed that 8650 out of 256461 unigenes (3.37%) were significantly differentially expressed between group 1 and group 2 with a P value < 0.05 (Fig.3). Among them, 4544 (52.53%) were significantly up regulated in group 1 (hereafter 'group 1 up-regulated unigenes') and 4107 (47.47%) in group 2 (hereafter group 2 up-regulated unigenes). To explore the possible molecular basis of chiton adhesion, we primarily analyzed the functions of the group 1 up-regulated unigenes.

GO enrichment analysis was conducted on 4544 group 1 up-regulated unigenes. Analyzing all unigenes of the top 10 GO terms showed that GO terms of 'single-multicellular organism process', 'anatomical structure development' and 'intracellular' are related to biological metabolic processes. Among them, the term 'proteinaceous extracellular matrix' was especially noted. A total of 67 unigenes (Table 2) in this term were mostly 'collagen' and 'fibrillin'.

Fig.4 shows the gene expression of collagen and fibrillin in group 1 and group 2 in chiton foot, and collagen and fibrillin show significantly higher expression levels in group 1 than in group 2. Collagen is a fundamental proteinaceous component in various connective tissues of multicellular animals and plays mechanically and physiologically important role in biological bodies. Previous studies have shown that in nematodes, collagen (in connective tissues) plays a role in movement. Their outer covering, called the cuticle, consists primarily of collagen, which helps the nematode move (Johnstone, 1994). Another study reported that the migrating epithelium contains collagen which must be liberated for continuous movements (Stenn *et al*., 1979). Some groups have isolated an adhesive protein with some features of collagen from the foot of the mussel *Geukensia demissa* (Waite

et al., 1989). Fibrillin is known as a glycoprotein which is essential for the formation of elastic fibers found in connective tissues (Kielty *et al*., 2002). Fibrillin-rich microfibrils are essential structural elements of all dynamic connective tissues (Kielty *et al*., 2003). It is a major component of the extracellular matrix and plays a role in cell adhesion function (D'Arrigo *et al*., 1998). These results implied that collagen and fibrillin may provide a basis for adhesive locomotion of the tissue.

KEGG enrichment analysis was then conducted in 4544 group1 up-regulated unigenes, and Table 3 shows all KEGG pathways with FDR < 0.05. All of the unigenes in the pathway 'Mucin type O-glycan biosynthesis' are N-acetylgalactosaminyl transferase and beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase 3, which are used for mucin synthesis. Mucins are major glycoprotein components of the mucus (Perezvilar and Hill, 1999) and are known to have adhesive properties (Beeley, 1985). Many animals use mucus for adhesion. Gastropod mollusks depend on mucus to hold themselves to the substratum during locomotion (Smith, 2002). Most commonly, mucus gels serve as a lubricating and protecting layer. Many mollusks, however, use similar gels to glue themselves to the substratum when they are inactive (Pawlicki *et al*., 2004). Thus, the mucin-related unigenes enriched in distal muscle (group 1) surely play an important role in chiton foot adhesion. These unigenes identified above may serve as potential candidate unigenes related to adhesion and could be used as references for future studies.

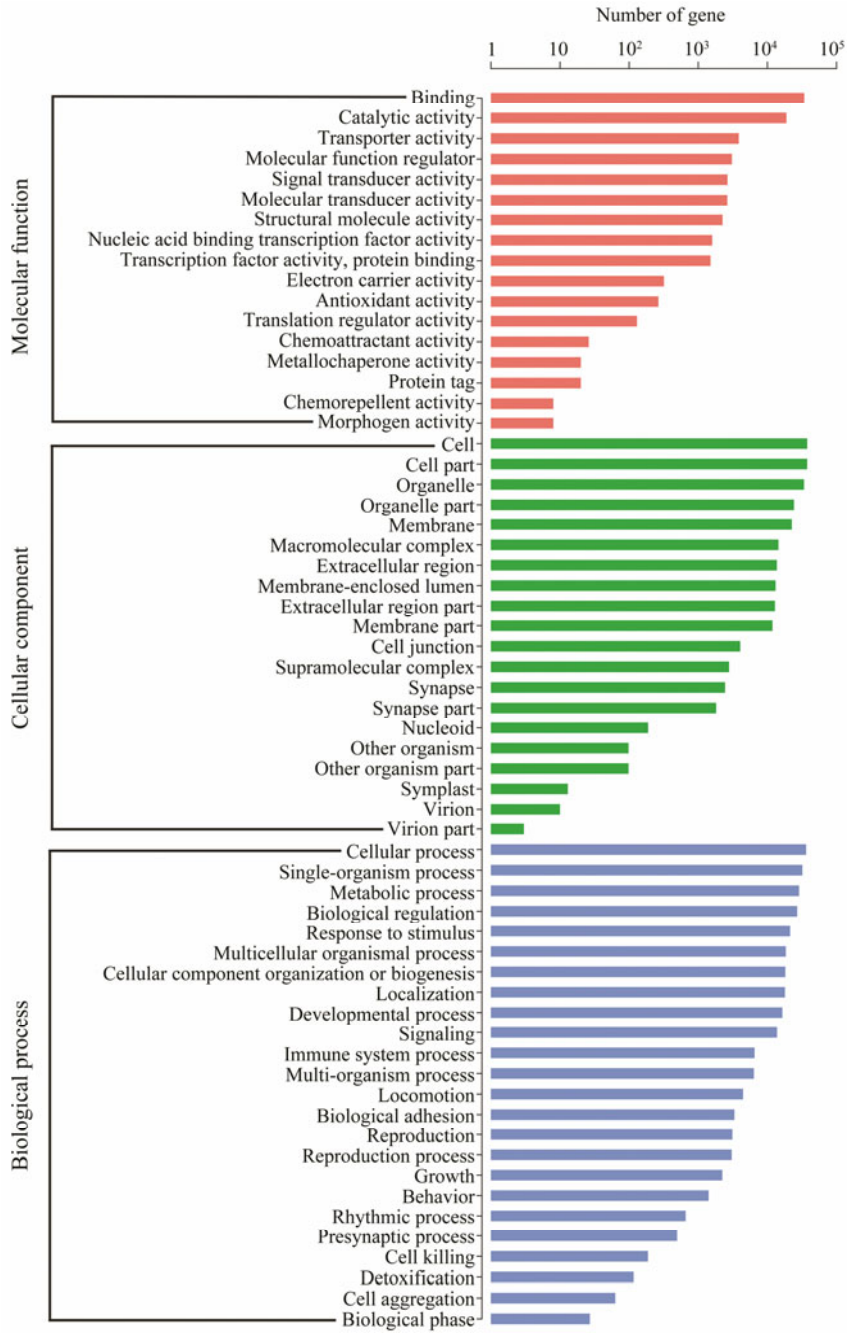

Fig.2 GO annotation of all unigenes. Unigenes with GO annotation were divided into three major categories: biological process, cellular component and molecular function.

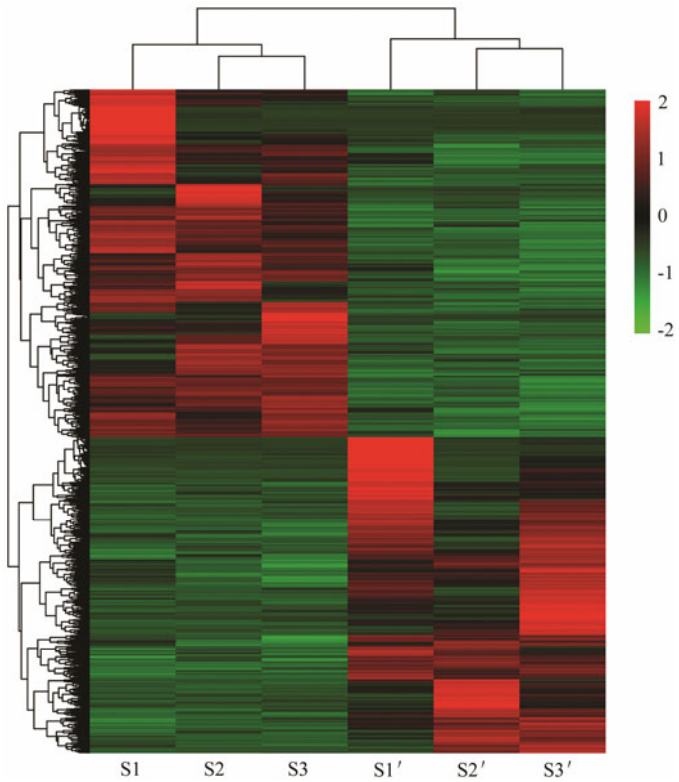

Fig.3 Heat map of differentially expressed unigenes. The significantly differentially expressed unigenes between group 1 and group 2.

Table 2 All unigenes in the GO term 'proteinaceous extracellular matrix' are enriched in group 1 biased unigenes of *I. hakodadensis* transcriptome

Gene ID	Group 1 average FPKM	Group 2 average FPKM	*P* value	SWISS
comp195817_c0_seq1	168.41	36.96	0.00233	COMA1_HUMAN Collagen alpha-1(XXII) chain
comp190425_c0_seq1	118.60	29.71	0.0369	COEA1_CHICK Collagen alpha-1(XIV) chain
comp188572_c1_seq1	116.22	19.38	0.000552	WFD18_RATWAP four-disulfide core domain 18
comp186371_c0_seq1	112.28	18.54	0.00095	WFD18_RATWAP four-disulfide core domain 18
comp189186_c1_seq1	44.66	8.01	0.000206	ITAX_MOUSE Integrin alpha-X
comp187327_c1_seq1	42.35	9.05	0.000664	CO6A3_HUMAN Collagen alpha-3(VI) chain
comp185102_c0_seq1	38.33	6.75	0.00433	MATN3_CHICK Matrilin-3
comp181636_c0_seq1	36.87	6.93	0.0107	MATN3_CHICK Matrilin-3
comp197247_c0_seq1	32.97	5.90	5.59E-05	MATN3_CHICK Matrilin-3
comp180141_c0_seq1	31.52	12.96	0.0154	MATN1_CHICK Cartilage matrix
comp187327_c0_seq1	31.38	6.26	0.000425	CO6A6_HUMAN Collagen alpha-6(VI) chain
comp183929_c0_seq1	26.38	10.76	0.0132	MATN1_CHICK Cartilage matrix
comp149214_c0_seq1	15.03	1.59	2.51E-05	FBN2_MOUSE Fibrillin-2
comp105094_c0_seq1	14.30	2.39	0.00211	FBN2_MOUSE Fibrillin-2
comp166584_c0_seq1	14.04	0.00	0.000379	AGRIN_HUMAN Agrin
comp170794_c0_seq1	11.85	0.00	0.000612	AGRIN_HUMAN Agrin
comp159455_c0_seq1	10.73	3.38	0.0237	CO6A4_MOUSE Collagen alpha-4(VI) chain
comp196367_c0_seq1	9.12	0.19	0.0104	FBN1_MOUSE Fibrillin-1
comp190654_c0_seq1	9.07	2.15	0.000456	WFD18_MOUSEWAP four-disulfide core domain 18
comp183591_c1_seq1	8.26	1.40	0.00465	MATN1_CHICK Cartilage matrix
comp183591_c0_seq1	8.13	1.45	0.00537	MATN1_CHICK Cartilage matrix
comp188437_c0_seq1	8.05	0.70	0.012	FBN1_HUMAN Fibrillin-1
comp195103_c0_seq1	7.92	1.88	0.00176	COMA1_HUMAN Collagen alpha-1(XXII) chain
comp153002_c0_seq1	7.78	0.49	1.79E-05	FBN2_MOUSE Fibrillin-2
comp196008_c2_seq1	7.63	1.99	0.000101	EMR1_RAT EGF-like module-containing mucin-like hormone receptor-like 1
comp154955_c0_seq1	7.56	0.81	0.000917	FBN2_MOUSE Fibrillin-2
comp186535_c0_seq1	7.16	0.50	0.0152	FBN1_MOUSE Fibrillin-1
comp167674_c1_seq1	6.71	0.92	0.000375	FBN1_HUMAN Fibrillin-1
comp143425_c0_seq1	5.15	1.11	0.0154	FBN1_HUMAN Fibrillin-1

(to be continued)

(continued)

Gene ID	Group 1 average FPKM	Group 2 average FPKM	*P* value	SWISS
comp161089_c0_seq1	4.18	0.59	0.00997	FBN2_MOUSE Fibrillin-2
comp195277_c1_seq1	4.11	0.21	0.00901	COCA1_HUMAN Collagen alpha-1(XII) chain
comp163936_c0_seq1	3.75	0.62	0.00528	FBN2_MOUSE Fibrillin-2
comp175122_c1_seq1	3.66	1.37	0.0153	WN10A_DANRE Wnt-10a
comp193211_c1_seq1	3.11	0.35	6.96E-05	FBN1_HUMAN Fibrillin-1
comp193211_c0_seq1	3.03	0.36	8.04E-05	FBN1_HUMAN Fibrillin-1
comp185245_c1_seq1	2.94	0.24	0.044	MFAP4_MOUSE Microfibril-associated glyco 4
comp186522_c0_seq1	2.72	1.06	0.0131	THSD1_PONAB Thrombospondin type-1 domain-containing 1
comp186522_c1_seq1	2.56	0.93	0.00504	THSD1_PONAB Thrombospondin type-1 domain-containing 1
comp165521_c0_seq1	2.48	0.13	2.62E-05	FBN1_HUMAN Fibrillin-1
comp163491_c0_seq1	2.44	0.27	0.000426	COCA1_HUMAN Collagen alpha-1(XII) chain
comp169688_c0_seq1	2.42	0.21	8.12E-05	COCA1_HUMAN Collagen alpha-1(XII) chain
comp197255_c0_seq1	2.33	1.16	0.0365	LAMA2_MOUSE Laminin subunit alpha-2
comp171741_c0_seq1	2.13	0.05	6.46E-06	LOXL2_XENLA Lysyl oxidase homolog 2
comp178893_c2_seq1	2.12	0.11	0.0147	COCA1_MOUSE Collagen alpha-1(XII) chain
comp187033_c1_seq1	1.99	0.83	0.0307	SLIT_DROME slit
comp185245_c0_seq1	1.76	0.26	0.0449	FCN2_RAT Ficolin-2
comp187033_c0_seq1	1.61	0.73	0.046	SLIT_DROME slit
comp171737_c0_seq1	1.37	0.02	0.00435	CL46_BOVIN Collectin-46
comp166750_c0_seq1	1.28	0.00	0.000901	CL46_BOVIN Collectin-46
comp124010_c0_seq1	1.23	0.00	0.0067	MFAP4_BOVIN Microfibril-associated glyco 4
comp164430_c0_seq1	0.98	0.05	0.0414	MFAP4_BOVIN Microfibril-associated glyco 4
comp169543_c1_seq1	0.96	0.00	0.0105	CALR_CRIGR Calreticulin
comp146296_c0_seq1	0.94	0.18	0.00567	USH2A_MOUSE Usherin
comp143390_c0_seq1	0.92	0.00	0.00977	ORF73_HHV8P ORF73
comp121879_c0_seq1	0.83	0.00	0.0479	NOTC1_MOUSE Neurogenic locus notch homolog 1
comp11478_c0_seq1	0.77	0.00	0.00318	LAMB1_MOUSE Laminin subunit beta-1
comp178893_c0_seq1	0.77	0.06	0.0195	COCA1_MOUSE Collagen alpha-1(XII) chain
comp157735_c0_seq1	0.58	0.16	0.00662	LAP2_HUMAN LAP2
comp154489_c0_seq1	0.57	0.11	0.0199	LAMB1_DROME Laminin subunit beta-1
comp150554_c0_seq1	0.50	0.11	0.0335	USH2A_HUMAN Usherin
comp49455_c0_seq1	0.49	0.08	0.0368	USH2A_HUMAN Usherin
comp20348_c0_seq1	0.46	0.00	0.0285	SLIT2_HUMAN slit homolog 2

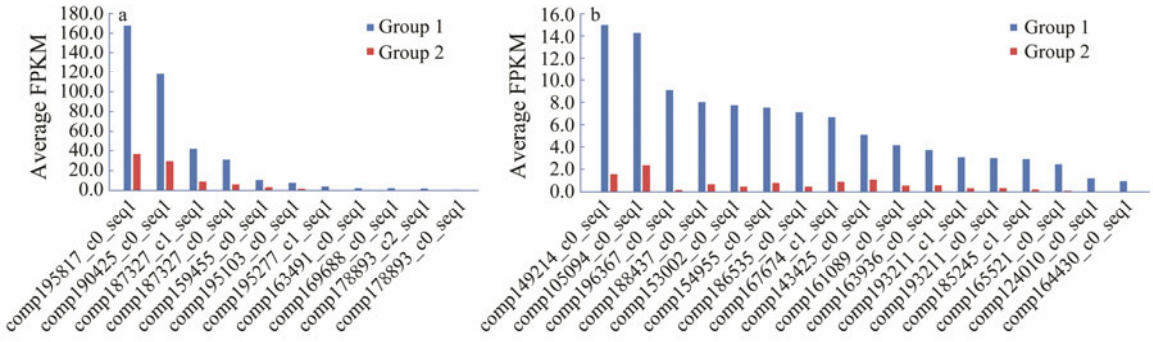

Fig.4 Expression profile of collagen (a) and fibrillin (b) genes enriched in the GO term 'proteinaceous extracellular matrix' in group 1 and group 2. Both collagen and fibrillin show significantly higher expression levels in group 1 than in group 2.

Table 3 KEGG pathways with FDR < 0.05 enriched in group1 biased unigenes

Pathway ID	Pathway	Gene no.	FDR
ko03030	DNA replication	22	3.42E-08
ko04110	Cell cycle	28	4.24E-05
ko00512	Mucin type O-glycan biosynthesis	11	6.67E-05
ko00533	Glycosaminoglycan biosynthesis - keratan sulfate	7	0.00161
ko04974	Protein digestion and absorption	16	0.0044
ko00520	Amino sugar and nucleotide sugar metabolism	16	0.0161
ko04914	Progesterone-mediated oocyte maturation	14	0.0176
ko04550	Signaling pathways regulating pluripotency of stem cells	16	0.0187
ko00240	Pyrimidine metabolism	23	0.0288

3.3 SNP and SSR Discoveries

A total of 679384 putative SNPs were identified from *I. hakodadensis*. For practical application in SNP genotyping assays, only biallelic SNPs were considered in this study. The overall frequency of all types of SNPs in *I. hakodadensis* transcriptome was one per 372 bp with transitions occurring more frequently than transversions (Fig.5). A/T was the most abundant transversion type, and C/G was the least abundant transversion type in *I. hakodadensis*. These results are similar to those in many other species (Liu *et al*., 2012; Zhou *et al*., 2014).

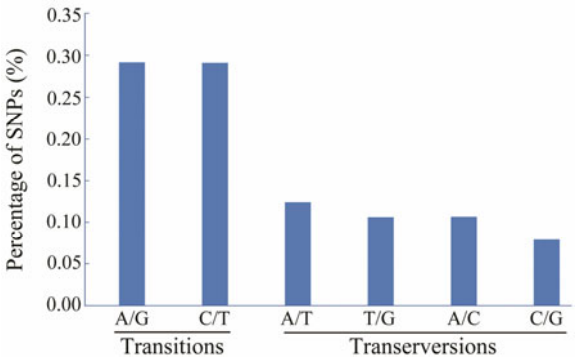

Fig.5 Classification of SNPs identified from *I. hakodadensis* transcriptome. In *I. hakodadensis*, transitions occurred more frequently than transversions. The overall frequency of all types of SNPs was one per 372 bp.

In addition, 19814 SSRs were identified from the transcriptome sequences of *I. hakodadensis*. The distribution pattern of SSR motifs is displayed in Table 4. Trinucleotide repeat type (11290) was the most frequent in *I. hakodadensis*, followed by di- (4639), tetra- (3675), penta- (182), and hexa- (28) nucleotide repeat types. The most prominent motif types in di-, tri-, tetra-, penta- and hexa-nucleotide repeats were AT, TGT, CAAT, AAAAC and CAGAAA, respectively. Thousands of high-quality SNPs and SSRs generated in this study can be used for further identification of genetic variation and diversity analysis of *I. hakodadensis*.

Table 4 Summary of simple sequence repeat (SSR) types in *I. hakodadensis* transcriptome

SSR type	No. of repeat	SSR length (bp)	No. of motif	Count	Major motif
Di-nucleotides	6-12	12-24	11	4639	AT
Tri-nucleotides	4-8	12-24	60	11290	TGT
Tetra-nucleotides	4-6	16-24	144	3675	CAAT
Penta-nucleotides	4-5	22-25	108	182	AAAAC
Hexa-nucleotides	4	24	22	28	CAGAAA

4 Conclusions

In this study, we conducted transcriptome sequencing of the chiton foot using the Illumina sequencing platform. Through transcriptome analysis, the differentially expressed genes (DEGs) between the distal muscle and the proximal muscle were identified and some potential genes might be related to foot adhesive locomotion at the transcriptome level. A large number of SNPs and SSRs were identified, which are expected to be used as markers for population genetic diversity analysis and variation identification in the future studies. This study provides transcriptomic resources for future genetic and genomic studies on chitons.

Acknowledgements

We acknowledge grant support from the National Natural Science Foundation of China (Nos. 31130054, 31472258), the AoShan Talents Program of Qingdao National Laboratory for Marine Science and Technology (No. 2015ASTP-ES02), and the Fundamental Research Funds for the Central Universities (No. 201564009).

References

Beeley, J. G., 1985. *Glycoprotein and Proteoglycan Techniques*. Elsevier, Amsterdam, 6pp.

Benjamini, Y., and Hochberg, Y., 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of The Royal Statistical Society. Series B (Methodological)*, **57** (1): 289-300.

Bettencourt, R., Pinheiro, M., Egas, C., Gomes, P., Afonso, M., Shank, T., and Santos, R. S., 2010. High-throughput sequencing and analysis of the gill tissue transcriptome from the deep-sea hydrothermal vent mussel *Bathymodiolus azoricus*. *BMC Genomics*, **11**: 559.

Byern, D. B. J. V., and Grunwald, D. B. I., 2010. *Biological Adhesive Systems: From Nature to Technical and Medical Application*. Springer, Vienna, 53pp.

Connors, M. J., Ehrlich, H., Hog, M., Godeffroy, C., Araya, S., Kallai, I., Gazit, D., Boyce, M., and Ortiz, C., 2012. Three-dimensional structure of the shell plate assembly of the chiton *Tonicella marmorea* and its biomechanical consequences. *Journal of Structural Biology*, **177** (2): 314-328.

Craft, J. A., Gilbert, J. A., Temperton, B., Dempsey, K. E., Ashelford, K., Tiwari, B., Hutchinson, T. H., and Chipman, J. K., 2010. Pyrosequencing of *Mytilus galloprovincialis* cDNAs: Tissue-specific expression patterns. *PLoS One*, **5** (1): e8875.

D'Arrigo, C., Burl, S., Withers, A. P., Dobson, H., Black, C., and Boxer, M., 1998. TGF-beta1 binding protein-like modules of fibrillin-1 and -2 mediate integrin-dependent cell adhesion. *Connective Tissue Research*, **37** (1-2): 29-51.

Du, H., Bao, Z., Hou, R., Wang, S., Su, H., Yan, J., Tian, M., Li, Y., Wei, W., and Lu, W., 2012. Transcriptome sequencing and characterization for the sea cucumber *Apostichopus japonicus* (Selenka, 1867). *PLoS One*, **7** (3): e33311.

Fu, X., Sun, Y., Wang, J., Xing, Q., Zou, J., Li, R., Wang, Z., Wang, S., Hu, X., and Zhang, L., 2014. Sequencing-based gene network analysis provides a core set of gene resource for understanding thermal adaptation in Zhikong scallop *Chlamys farreri*. *Molecular Ecology Resources*, **14** (1): 184-198.

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D.,

Nagaraj, S. H., Nueda, M. J., Robles, M., Talón, M., Dopazo, J., and Conesa, A., 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, **36** (10): 3420-3435.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., and Zeng, Q., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology*, **29** (7): 644-652.

Hou, R., Bao, Z., Wang, S., Su, H., Li, Y., Du, H., Hu, J., Wang, S., and Hu, X., 2011. Transcriptome sequencing and de novo analysis for Yesso scallop (*Patinopecten yessoensis*) using 454 GS FLX. *PLoS One*, **6** (6): e21560.

Hu, X., Bao, Z., Hu, J., Shao, M., Zhang, L., Bi, K., Zhan, A., and Huang, X., 2006. Cloning and characterization of tryptophan 2, 3-dioxygenase gene of Zhikong scallop *Chlamys farreri* (Jones and Preston 1904). *Aquaculture Research*, **37** (12): 1187-1194.

Huan, P., Wang, H., and Liu, B., 2012. Transcriptomic analysis of the clam *Meretrix meretrix* on different larval stages. *Marine Biotechnology*, **14** (1): 69-78.

Johnstone, I. L., 1994. The cuticle of the nematode *Caenorhabditis elegans*: A complex collagen structure. *Bioessays*, **16** (3): 171-178.

Joubert, C., Piquemal, D., Marie, B., Manchon, L., Pierrat, F., Zanellacléon, I., Cochenneclaureau, N., Gueguen, Y., and Montagnani, C., 2010. Transcriptome and proteome analysis of *Pinctada margaritifera* calcifying mantle and shell: Focus on biomineralization. *BMC Genomics*, **11**: 613.

Kangas, M., and Shepherd, S. A., 2013. Distribution and feeding of chitons in a boulder habitat at West Island, South Australia. *Journal of the Malacological Society of Australia*, **6** (3): 101-111.

Kielty, C. M., Baldock, C., Lee, D., Rock, M. J., Ashworth, J. L., and Shuttleworth, C. A., 2002. Fibrillin: From microfibril assembly to biomechanical function. *Philosophical Transactions of the Royal Society B Biological Sciences*, **357** (1418): 207-217.

Kielty, C. M., Wess, T. J., Haston, L., Ashworth, J. L., Sherratt, M. J., and Shuttleworth, C. A., 2003. *Fibrillin-rich Microfibrils: Elastic Biopolymers of the Extracellular Matrix*. Springer, Netherlands, 581-596.

Kinoshita, S., Wang, N., Inoue, H., Maeyama, K., Okamoto, K., Nagai, K., Kondo, H., Hirono, I., Asakawa, S., and Watabe, S., 2011. Deep sequencing of ESTs from nacreous and prismatic layer producing tissues and a screen for novel shell formation-related genes in the pearl oyster. *PLoS One*, **6** (6): e21238.

Koboldt, D. C., Zhang, Q., Larson, D. E., Dong, S., Mclellan, M. D., Ling, L., Miller, C. A., Mardis, E. R., Li, D., and Wilson, R. K., 2012. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research*, **22** (3): 568-576.

Li, H., Ding, G. L., and Li, Y., 2009. PAnnBuilder: An R package for assembling proteomic annotation data. *Bioinformatics*, **25** (8): 1094-1095.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R., 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*, **25** (16): 2078-2079.

Liu, S., Zhou, Z., Lu, J., Sun, F., Wang, S., Hong, L., Jiang, Y., Kucuktas, H., Kaltenboeck, L., and Peatman, E., 2011. Generation of genome-scale gene-associated SNPs in catfish for the construction of a high-density SNP array. *BMC Genomics*, **12**: 53.

Miao, Y., Zhang, L., Sun, Y., Jiao, W., Li, Y., Sun, J., Wang, Y., Wang, S., Bao, Z., and Liu, W., 2015. Integration of transcriptomic and proteomic approaches provides a core set of genes for understanding of scallop attachment. *Marine Biotechnology*, **17** (5): 523-532.

Milan, M., Coppe, A., Reinhardt, R., Cancela, L. M., Leite, R. B., Saavedra, C., Ciofi, C., Chelazzi, G., Patarnello, T., and Bortoluzzi, S., 2011. Transcriptome sequencing and microarray development for the Manila clam, *Ruditapes philippinarum*: Genomic tools for environmental monitoring. *BMC Genomics*, **12**: 234.

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., and Kanehisa, M., 2007. KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research*, **35** (2): 182-185.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, **5** (7): 621-628.

Pawlicki, J. M., Pease, L. B., Pierce, C. M., Startz, T. P., Zhang, Y., and Smith, A. M., 2004. The effect of molluscan glue proteins on gel mechanics. *Journal of Experimental Biology*, **207** (7): 1127-1135.

Perezvilar, J., and Hill, R. L., 1999. The structure and assembly of secreted mucins. *Journal of Biological Chemistry*, **274** (45): 31751-31754.

Puchalski, S. S., Eernisse, D. J., and Johnson, C. C., 2008. The effect of sampling bias on the fossil record of chitons (Mollusca, Polyplacophora). *American Malacological Bulletin*, **25** (1): 87-95.

Riesgo, A., Andrade, S. C. S., Sharma, P. P., Novo, M., Pérez-Porro, A. R., Vahtera, V., González, V. L., Kawauchi, G. Y., and Giribet, G., 2012. Comparative description of ten transcriptomes of newly sequenced invertebrates and efficiency estimation of genomic sampling in non-model taxa. *Frontiers in Zoology*, **9**: 33.

Robinson, M. D., McCarthy, D. J., and Smyth, G. K., 2010. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26** (1): 139-140.

Romkes, M., and Buch, S. C., 2014. Genotyping technologies: Application to biotransformation enzyme genetic polymorphism screening. *Methods in Molecular Biology*, **1105**: 99-115.

Schwabe, E., 2005. A catalogue of Recent and fossil chitons (Mollusca: Polyplacophora) Addenda. *Novapex*, **6** (4): 89-105.

Liu, S. K., Zhang, Y., Zhou, Z. C., Waldbieser, G., Sun, F., Lu, G. G., Zhang, J. R., Jiang, Y. L., Zhang, H., Wang, X. L., Rajendran, K. V., Khoo, L., Kucuktas, H., Peatman, E., and Liu, Z. J., 2012. Efficient assembly and annotation of the transcriptome of catfish by RNA-Seq analysis of a doubled haploid homozygote. *BMC Genomics*, **13**: 595.

Smith, A. M., 2002. The structure and function of adhesive gels from invertebrates. *Integrative and Comparative Biology*, **42** (6): 1164-1171.

Smith, A. M., Quick, T. J., and Peter, R. S., 1999. Differences in the composition of adhesive and non-adhesive mucus from the limpet *Lottia limatula*. *The Biological Bulletin*, **196** (1): 34-44.

Stenn, K. S., Madri, J. A., and Roll, F. J., 1979. Migrating epidermis produces AB2 collagen and requires continual collagen synthesis for movement. *Nature*, **277** (5693): 229-232.

Treves, K., Traub, W., Weiner, S., and Addadi, L., 2003. Aragonite formation in the chiton (Mollusca) girdle. *Helvetica Chimica Acta*, **86** (4): 1101-1112.

Vinther, J., and Nielsen, C., 2005. The early Cambrian Halkieria is a mollusc. *Zoologica Scripta*, **34** (1): 81-89.

Waite, J. H., Hansen, D. C., and Little, K. T., 1989. The glue protein of ribbed mussels (*Geukensia demissa*): A natural adhesive with some features of collagen. *Journal of Comparative Physiology B*, **159** (5): 517-525.

Wang, S., Hou, R., Bao, Z., Du, H., He, Y., Su, H., Zhang, Y., Fu, X., Jiao, W., and Li, Y., 2013. Transcriptome sequencing of Zhikong scallop (*Chlamys farreri*) and comparative transcriptomic analysis with Yesso scallop (*Patinopecten yessoensis*). *PLoS One*, **8** (5): e63927.

Zhou, Z. C., Dong, Y., Sun, H. J., Yang, A. F., Chen, Z., Gao, S., Jiang, J. W., Guan, X. Y., Jiang, B., and Wang, B., 2014. Transcriptome sequencing of sea cucumber *(Apostichopus japonicus*) and the identification of gene-associated markers. *Molecular Ecology Resources*, **14** (1): 127-138.

(Edited by Qiu Yantao)