Pakistan J. Zool., vol. 48(2), pp. 427-434, 2016.

Characterization of *Nibea albiflora* Transcriptome: Sequencing, *De Novo* Assembly, Annotation and Comparative Genomics

Wei Zhan,[1] Ruiyi Chen,[1] M.Y. Laghari,[2] Dongdong Xu,[1] Guomin Mao,[1] Huilai Shi[1] and Bao Lou[1]*

[1]*Marine Fisheries Research Institute of Zhejiang Province, Key Lab of Mariculture and Enhancement of Zhejiang Province, Zhoushan 316021, China*

[2]*Department of Fresh Water Biology and Fisheries, University of Sindh, Jamshoro, Pakistan*

ABSTRACT

Nibea albiflora is an economically and ecologically important fish species, which is widely distributed from South China Sea to the coastal water of Japan and Korea. Due to the economical and ecological importance of *N. albiflora*, genomic data are eagerly needed for genetic improvement. However, there is still no sufficient transcriptome data available for such valuable species. Therefore, presently transcriptome was sequenced deeply to provide well-assembled transcriptome sequences to *N. albiflora* research community. A total of 63,367,046 cleaned reads (78%) were obtained and were assembled, a total of 48,255 contigs ranging from 201 to 16,026 bp in length were generated. The average length was 725 bp, the N50 length was 1,279 bp. In result of assembled contigs, compared with the NCBI non-redundant (nr) protein database, a total of 19,027 contigs had a significant hit, corresponding to 13,743 unique protein accessions in the nr protein database. Top-Hit species distribution was then investigated, the *Maylandia zebra* is the species that returned the most BLAST hits with *N. albiflora* contigs, followed by *Oreochromis niloticus* and *Takifugu rubripes*. The contigs of the *N. albiflora* transcriptome had hits with 26.2% to 44.3% of the unique proteins of fugu, zebrafish, three-spined stickleback, medaka. In further, the cDNA SSR and SNP loci were identified for future marker development and genetic analysis. In the present investigation, the transcriptome of *N. albiflora* had been deeply sequenced, assembled and characterized, providing a valuable resource for a better understanding of the *N. albiflora* genome.

Article information

Received 28 August 2015
Revised 15 October 2015
Accepted 31 October 2015
Available online 1 March 2016

Authors' Contributions

WZ and BL designed the experiments. DX, GM and HS collected the samples, conducted the experiment and analyzed the data. RC and MYL wrote the article.

Key words

Comparative genomics, *de novo* assembly, *Nibea albiflora*, expressed sequence tag, cDNA contigs, micro-satellite identification, transcriptome.

INTRODUCTION

Nibea albiflora is a member of Sciaenidae and naturally distributes from the South China Sea to the coastal waters of Japan and Korea (Takita, 1974). As an economically important fishery and aquaculture species in China, *N. albiflora* was mainly obtained from wild populations in the early 1990s (Qiuttao *et al.*, 2012). However, overfishing pressure and water pollution in the past decade have led to a dramatic decline in a particular fish populations. The fishery ground and fishery season of *N. albiflora* have nearly disappeared in many regions of the China Sea (Han *et al.*, 2008).

Although it is a high-valued marine food fish species, little information is available at present to understand the genetic characteristics of *N. albiflora*. Only a few polymorphic microsatellite markers have been developed (Xing *et al.*, 2009; Ma *et al.*, 2011). The complete mitochondrial genome information of *N. albiflora* has been reported (Cheng *et al.*, 2011). The population genetic structure and genetic diversity have been investigated by using part of a mtDNA control region sequence (Han *et al.*, 2006), COI sequence (Xu *et al.*, 2012) and AFLP markers (Han *et al.*, 2006). In order to further explore its genetic characteristics and to protect germplasm resources of this valuable species, more genetic and genomic information is required.

Expressed sequence tag (EST) sequencing has been considered an efficient approach for genomic study and functional gene identification, especially for those species without a genome reference. In the past decade, hundreds of thousands ESTs have been sequenced for several aquaculture species using traditional Sanger's methods (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html), including Atlantic salmon (Wang *et al.*, 2010), catfish (Abernathy *et al.*, 2007) and rainbow trout (Wang *et al.*, 2010). These EST databases give ways for efficient gene discovery and transcriptome profiling in these species (Wang *et al.*, 2010; Abernathy *et al.*, 2007; Sha *et al.*, 2008), as well as genetic marker development (Xu *et al.*, 2012). In the past half-decade, high-throughput next-generation sequencing technologies had been developed and successfully used to obtain a large amount of transcriptome sequences at a lower cost, providing scientists the ability to collect sufficient genetic and genome resources for the many different species (Hampton *et al.*, 2011; Hou *et al.*, 2011; Jung *et al.*, 2011;

* Corresponding author: loubao6577@outlook.com
0030-9923/2016/0002-0427 $ 8.00/0

Yang *et al*., 2015).

In the present investigation, we performed *de novo* transcriptome sequencing of *N. albiflora* using Illumina Hiseq2000 platform. 63,367,046 cleaned reads were assembled into 48,255 cDNA contigs. Annotation and gene ontology analysis were then performed on these contigs and genomics research on *N. albiflora* and closely related species.

MATERIALS AND METHODS

Biological samples

Samples of *N. albiflora* were collected in coastal waters of Zhoushan, China. Twelve tissues including brain, muscle, liver, intestine, blood, head kidney, trunk kidney, skin, gill, spleen, gonad and heart were collected by dissecting individual. Tissue samples were stored in RNAlater (Qiagen, Hilden, Germany), transported to the laboratory at room temperature, and then stored at -20ºC prior to RNA extraction.

RNA extraction

Total RNA was extracted from 12 tissues using the TRIZOL Kit (Invitorgen, Carlsbad, CA, USA) following manufacturer's instructions. RNA samples were digested by DNase 1 to remove potential genomic DNA. Integrity and size distribution were checked with Bioanalyzer 2100 (Agilent technologies, Santa Clara, CA, USA). Equal amounts of the high quality RNA samples from each tissue were then pooled for cDNA synthesis and sequencing.

cDNA library construction, sequencing and assembly

RNA-seq library preparation and sequencing was carried out by Research Facility Center at Beijing Institutes of Life Science. cDNA libraries were prepared with ~2.5 μg of starting total RNA following the protocols of the IlluminaTruSeq RNA Sample Preparation Kit (Illumina). The final library had an average fragment size of 180 bp and final yields of ~400 ng. After KAPA quantitation and dilution, the library was sequenced on an IlluminaHiSeq 2000 with 101 bp paired-end reads. All sequenced reads in SRA format have been uploaded to the NCBI Short Read Archive with the accession number of SRR1016446. Adaptor sequences were trimmed and reads with low quality or length less than 70 were further removed by SolexaQA software (Cox *et al.*, 2010). Cleaned reads were assembled by TRINITY (Grabherr *et al*., 2011) with default parameters.

Functional annotation

The assembled transcriptome contigs were subjected to similarity search against NCBI non-redundant (nr) protein database using Blast2go (Conesa *et al*., 2005) with an e-value cutoff of 1e-6. Gene names and descriptions were assigned to each contig based on the BLASTx results. Gene ontology (GO) analysis was then conducted on the assembled transcriptome by using Blast2go. All assembled contigs were analyzed by transcripts_to_best_scoring_ORFs.pl from Trinity package (Grabherr *et al*., 2011) to search for Open Reading Frames (ORFs), which could distinguish between coding and non-coding sequences. KEGG pathways were assigned to assembled contigs using the online KEGG Automatic Annotation Server (KAAS) (Moriya *et al.*, 2007). The Bi-directional Best Hit (BBH) method was used to obtain KEGG Orthology (KO) assignment.

Assembly assessment

To compare the similarity to other teleost species, the transcriptome contigs were compared to Ensemble protein database of zebrafish (*Danio rerio*), fugu (*Takifugu rubripes*), medaka (*Oryzias latipes*), tetraodon (*Tetraodon nigroviridis*) and three-spined stickleback (*Gasterosteus aculeatus*) by using BLASTx program with e-value cutoff of 1e-5.

Repetitive element analysis and microsatellite identification

To identify all repetitive elements in the assembled contigs, RepeatMasker (Smit *et al*., 1996) was used with Repbase for all vertebrates. A perl-based script Msatfinder V 2.0.9 (Mi *et al*., 2005) was used for microsatellite identification. The mononucleotide repeats were ignored by modifying the configure file. The repeat thresholds for di-, tri-, tetra-, penta-, hexa-nucleotide motifs were set as 8, 5, 5, 5 and 5, respectively. Only microsatellite sequences with flanking sequences longer than 50 bp on both sides were identified for further marker development.

SNP identification

To identify putative single nucleotide polymorphism (SNP) loci in the transcriptome of *N. albiflora*, all RNA-seq reads were mapped onto the assembled contigs using BWA (Li and Durbin, 2009) and SAM tools (Li *et al*., 2009). The filtering threshold was set as bellowing, the read depth to no less than 10, and quality score to no less than 20. Then the output file was further extracted for useful information.

RESULTS AND DISCUSSION

Transcritptome sequencing and assembly of N. albiflora

To enable a comprehensive understanding and

profiling of the transcriptome of *N. albiflora*, mixed RNA originating from 12 tissues was sequenced by using Illumina Hiseq2000 sequencing technology. A total of 81,218,264 paired-end reads was generated with a read length of 101 bp. After the removal of ambiguous nucleotides, duplicates and low-quality sequences (Phred quality scores <20), a total of 63,367,046 cleaned reads (78%) were obtained. The raw transcriptome sequences in this study had been deposited in the NCBI SRA database (Accession number: SRR1016446). The cleaned reads were then assembled by Trinity. As a result, 48,255 contigs were generated, ranging from 201 to 16,026 bp in length. The average length was 725 bp, the N50 length was 1,279 bp. The contig length distribution is shown in Figure 1.

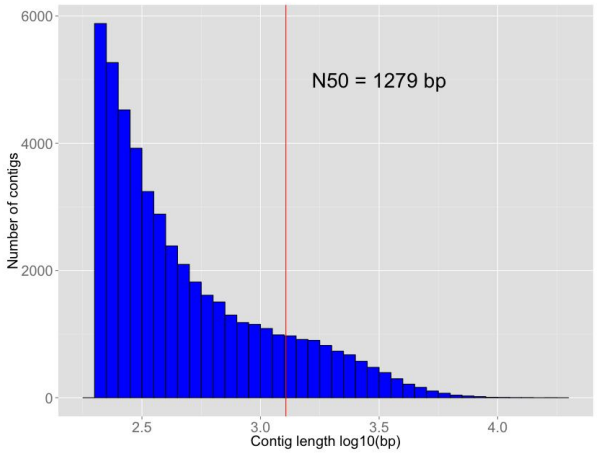

Fig. 1. Length distribution of assembled contigs of *N. albiflora*.

Functional annotation

All assembled contigs were first compared with the NCBI non-redundant (nr) protein database for functional annotation by using Blast2go software with an e-value cutoff of 1e-6. A total of 19,027 contigs has a significant hit, corresponding to 13,743 unique protein accessions in the nr protein database. Top-Hit species distribution was then investigated. As can be seen in Figure 2, the *Maylandia zebra* is the species that returned the most BLAST hits with *N. albiflora* contigs, followed by *Oreochromis niloticus* and *Takifugu rubripes*.

Gene ontology (GO) analysis was conducted on those 13,743 unique proteins by using blast2go. A total of 13,192 unique proteins were assigned at least one GO term for describing biological processes, molecular functions and cellular components. The blast2go output file was input into the BGI WEGO program and GO annotations were plotted (http://wego.genomics.org.cn) (Fig. 3). Of these, the molecular function ontology made up the majority (12246, 92.8%), followed by the biological process ontology (12035, 91.2%) and the cellular component ontology (12119, 91.9%). Briefly, for biological process, gene involved in cellular processes (GO:0009987) and metabolic processes (GO:0008152) were highly represented; for molecular functions, binding (GO:0005488) was the most represented GO term, followed by catalytic activity (GO:0003824); cells (GO:0005623) and organelles (GO:0043226) were the most represented categories for the cellular component. To assess the functional diversity of assembled transcriptome, GO annotation of zebrafish (Ensembl) were compared with those of *N. albiflora* transcriptome, reflecting a similar functional distribution on GO categories and indicating the sequence diversity of the transcriptome study.

In addition, KEGG pathway analysis was performed on all assembled contigs as alternative approach for functional categorization and annotation. Enzyme commission (EC) numbers were assigned to 2,830 unique sequences, which categorized them into different functional groups (Table I). Briefly, of these sequences with KEGG annotation, 942 (33.3%) were classified into the metabolism, including majority sub-groups of carbohydrate metabolism (191, 6.7%), amino acid metabolism (175, 6.2%) and energy metabolism (166, 4.9%). Sequences grouped into the Genetic information processing (GIP), accounted for 813 (2.8%), including translation (321, 11%), folding, sorting and degradation (284, 10%), transcription (135, 4.7%), replication and repair (111, 0.4%), etc. Organismal systems, cellular processes and environmental information processing (EIP) groups contained 854 (30.2%), 609 (21.5%) and 686 (24.2% KEGG annotated sequences, respectively. Well-categorized and annotated transcriptome could serve as important and valuable resources for gene identification analysis of specific traits in *N. albiflora* genetics and genomics. For instance, 321 transcript contigs associated with immune systems in KEGG analysis had been collected and 170 contigs associated with growth (Table I).

From all the assembled contigs of the *N. albiflora* transcriptome, 15178 ORFs were detected, with an average ORF length of 950 bp and with a range from 300 bp to 13,360 bp (Fig. 4). The remaining contigs contained no ORFs, indicating they are non-coding sequences and likely come from untranslated regions (UTR). The assembled transcriptomecontigs served as a reference for cSNPs identification from RNA-seq data. ORF analysis allows the discrimination of synonymous and non-synonymous SNPs and the identification of nonsense mutations in *N. albiflora.*

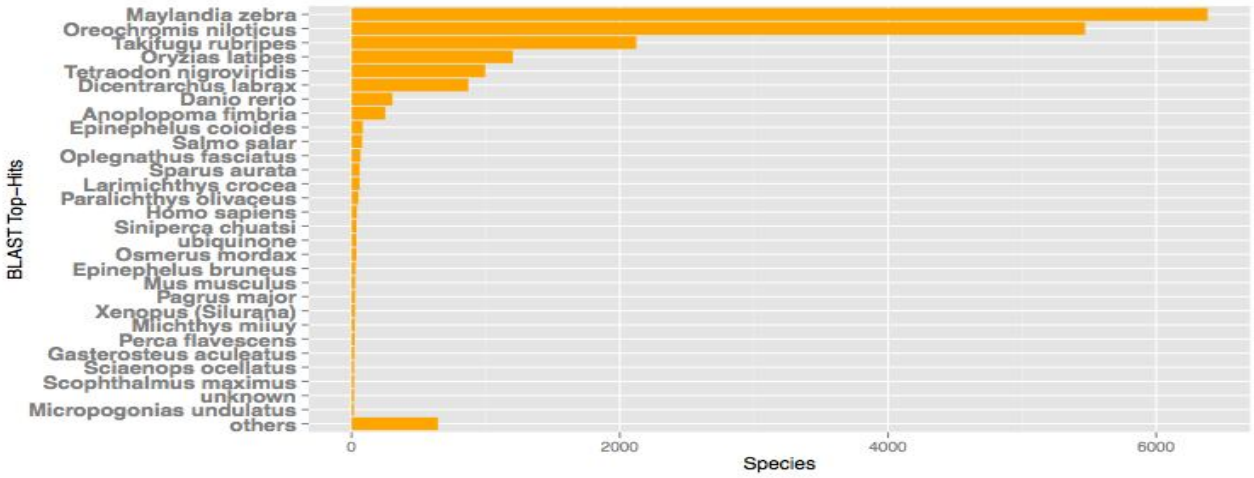

Fig. 2. Top-Hit species distribution of *N. albiflora* contigs.

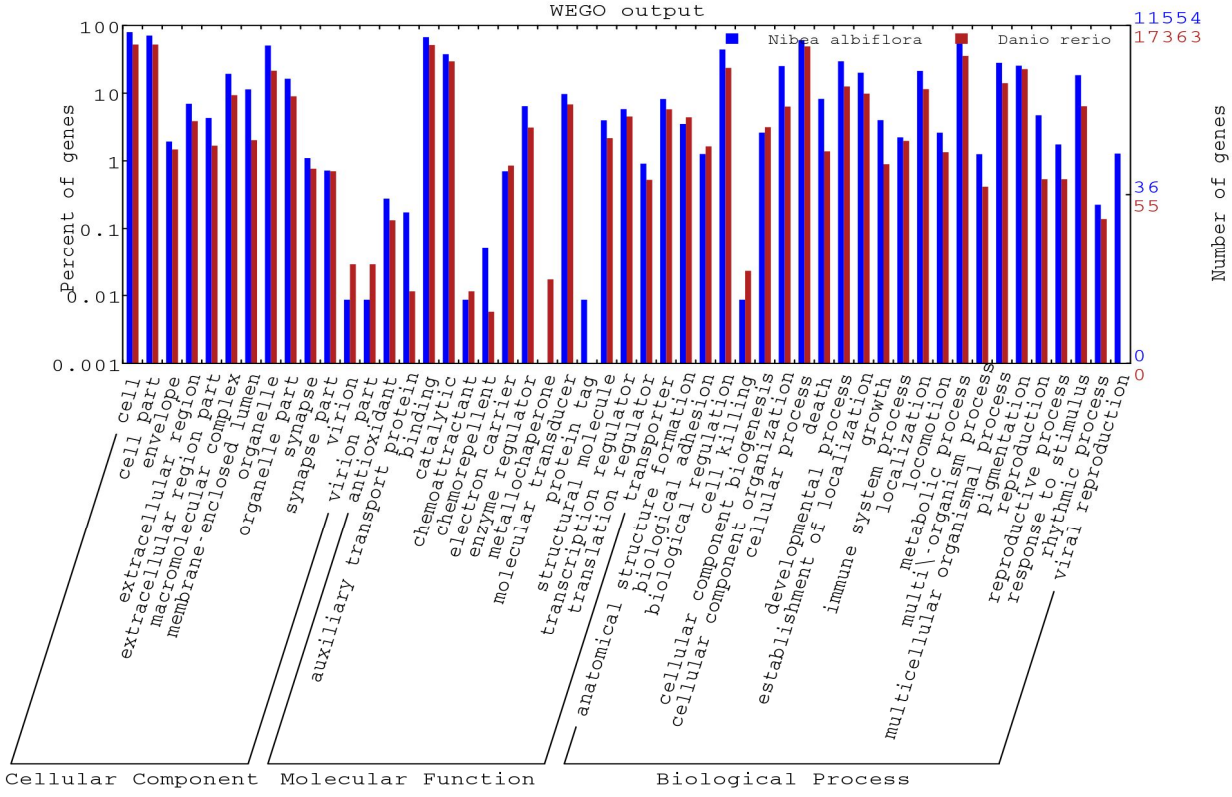

Fig. 3. Gene Ontology (GO) categories of the unigenes. Distribution of the GO categories assigned to the *N. albiflora* transcriptome. Unique transcripts (unigenes) were annotated in three categories: cellular components, molecular functions, and biological processes.

Table I.- KEGG biochemical mapping for *N. albiflora.*

KEGG categories represented	Unique sequences (Number of KO)
Metabolism	942 (792)
Carbohydrate metabolism	191 (152)
Amino acid metabolism	175 (145)
Energy metabolism	166 (141)
Nucleotide metabolism	121 (99)
Metabolism of cofactors and vitamins	105 (86)
Lipid metabolism	191 (160)
Glycan biosynthesis and metabolism	129 (117)
Metabolism of other amino acids	69 (53)
Xenobiotics biodegradation and metabolism	63 (47)
Biosynthesis of secondary metabolites	18 (15)
Metabolism of terpenoids and polyketides	24 (23)
Genetic information processing	813 (735)
Replication and repair	111 (106)
Folding, sorting and degradation	284 (251)
Transcription	135 (127)
Translation	321 (285)
Environmental information processing	686 (571)
Signal transduction	548 (449)
Signaling molecules and interaction	21 (185)
Membrane transport	20 (19)
Cellular processes	609 (504)
Cell motility	88 (68)
Cell growth and death	170 (148)
Transport and catabolism	284 (231)
Cell communication	180 (139)
Organismal systems	854 (722)
Immune system	321 (270)
Endocrine system	250 (202)
Circulatory system	90 (70)
Digestive system	156 (123)
Excretory system	73 (61)
Nervous system	197 (162)
Sensory system	20 (17)
Development	110 (92)
Environmental adaptation	52 (45)
Total	2,830 (2,463)

Assessment of transcriptome assembly

The assembled contigs of *N. albiflora* transcriptome were compared with protein database of zebra fish, fugu, medaka and three-spined stickleback by using BLASTx program with and e-value cutoff of 1e-5. There were 17,395 contigs (36.0%) with 11,698 unique protein hits, 16,851 contigs (34.9%) with 10,932 unique protein hits, 16,884 contigs (34.9%) with 12,549 unique protein hits, 17,217 contigs (35.7%) with 11,469 unique protein hits identified with significant hits on proteins of zebra fish, medaka, fugu and three-spined stickleback, respectively. The contigs of the *N. albiflora* transcriptome had hits with 26.2% to 44.3% of the unique proteins of fugu, zebrafish, three-spined stickleback, medaka (Table II). Obviously, the transcriptome similarity was relatively lower than expected. The current research however does not cover the whole transcriptome as the genetic material in this study was only collected from adult fish; transcripts from early development stages were therefore not included. In addition some rare transcripts may be missed or were only collected as singletons during the assembly, even though a high sequencing depth was applied. For better understanding and characterization of *N. albiflora* transcriptome, we would need a complete set of data from virtually every tissue across every lift stage and every circumstance, or the whole-genome sequencing and assembly.

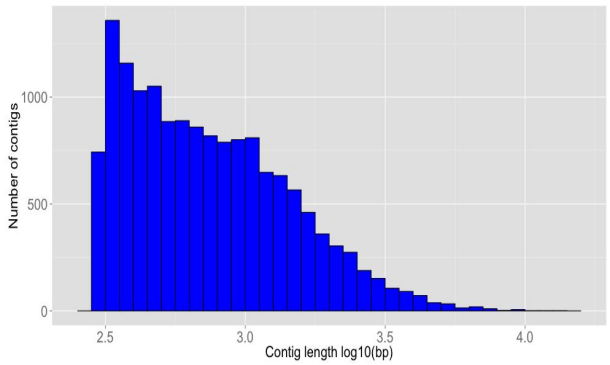

Fig. 4. Length distribution of identified ORF of *N. albiflora* transcriptome.

Table II.- Summary of BLASTx search results of *N. albiflora* transcriptome.

Database	*N. albiflora* hits	Unique protein	% of total unique proteins
NR	19,027	13,743	
Zebra fish	17,395	11,698	(27.5%) 42,555
Medaka	16,851	10,932	(44.3%) 24,674
Fugu	16,884	12,549	(26.2%) 47,841
Three-spined stickleback	17,217	11,469	(41.6%) 27,576

Repetitive element analysis and microsatellite identification

A total of 5,028 microsatellites were initially identified from 4,496 contigs, including di-, tri-, tetra-,

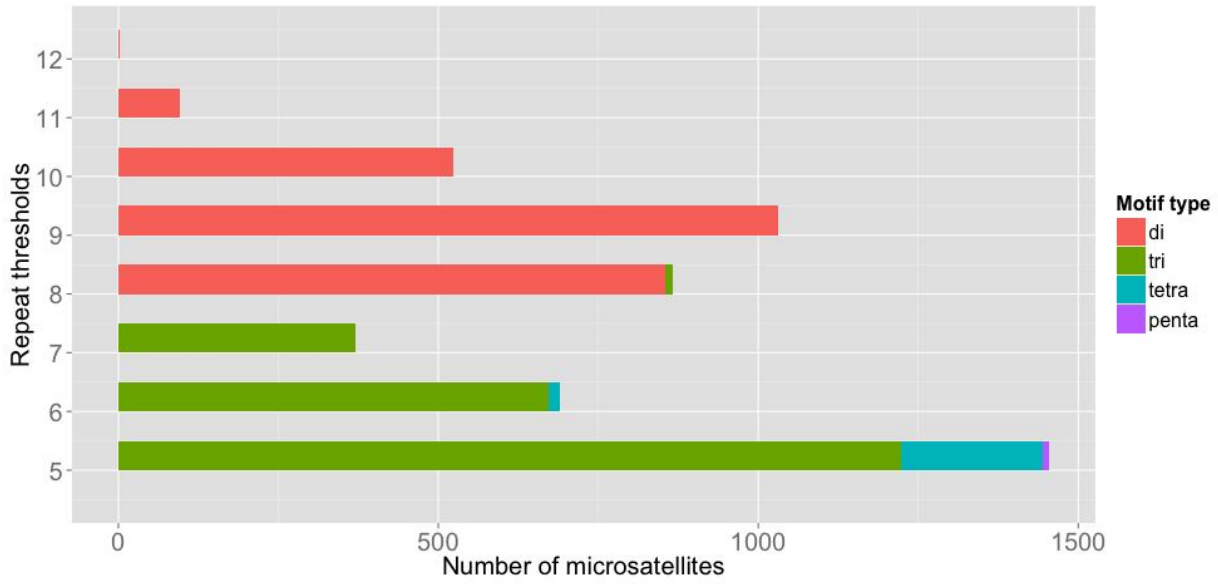

Fig. 5. The number distribution of microsatellites in *N. albiflora* transcriptome.

penta- and hexa-nucleotide repeats (Table III). The number distribution of microsatellites of every type of repeats is shown in Figure 5.

Table III.- Statistic of microsatellites identified from *N. albiflora* transcriptome.

Total number of contigs	**48,255**
Microsatellite identified	5,028
Di-nucleotide repeats	2,510
Tri--nucleotide repeats	2,279
Tetra--nucleotide repeats	230
Penta--nucleotide repeats	9
Number of contigs containing microsatellites	4,496

Table IV.- SNPs identified.

SNP classification	**Number of SNPs**
5' UTR	1,878
3' UTR	7,911
Coding region	12,691
synonymous	4,175
non-synonymous	8,504
Pre-terminated	12
Undefined	6,829
Total	29,309

The proportion of the repetitive elements in the *N. albiflora*transcriptome was assessed by using Repeatmasker. Repeat masking of the 35,007,163 bp of the *N. albiflora* contigs resulted in the detection of 646,243 (1.85%) of repeated sequences. The classification and respective proportion of the identified repetitive elements are shown in Table V. The most abundant type of repetitive elements in the sequences was Retroelements (0.64%), mostly LTR elements (0.32), followed by DNA transposon (0.26%). Various satellite sequences, low complexity and simple sequence repeats accounted for 0.02%, 0.50% and 0.39% of the base pairs, respectively.

SNP identification

For further application of the RNA-Seq data, SNPs were discovered using the assembled transcriptome. The short read of RNA-Seq data were aligned onto the reference transcriptome of *N. albiflora*, generating 29,309 SNPs after quality control and filtration (See Methods). The proportions of transition substitutions were 32.4% for A/G and 34.8% for C/T, compared with smaller proportions of transversion for A/C (8.8%), G/T (8.2%), A/T (8%) and C/G (7.8%). Among all SNPs detected, 14,605 were in the putative ORF region, of which 4,175 were synonymous and 8,504 were non-synonymous. The mean number of SNPs per kilobase in the ORF region was 12,691. Further analysis was done to classify identified SNPs (Table IV).

Table V.- The classification and respective proportion of the identified repetitive elements of *N. albiflora* transcriptome.

	Number of elements	Length occupied	Percentage of sequence
Retro elements	1514	224976 bp	0.64%
SINE:	244	23074 bp	0.07 %
Penelope	30	4365 bp	0.01%
LINEs:	617	89938 bp	0.26 %
CRE/SLACS	0	0	0
L2/CR1/Rex	348	46572 bp	0.13 %
R1/LOA/Jockey	3	267 bp	0.00 %
R2/R4/NeSL	7	817 bp	0.00 %
RTE/Bov-B	111	14291 bp	0.04 %
L1/CIN4	107	20188 bp	0.06 %
LTR elements:	653	111964 bp	0.32 %
BEL/Pao	98	25262 bp	0.07 %
Ty1/Copia	11	7252 bp	0.02 %
Gypsy/DIRS1	280	58596 bp	0.17 %
Retroviral	215	18045 bp	0.05 %
DNA transposons	837	90461 bp	0.26 %
hobo-Activator	301	24769 bp	0.07 %
Tc1-IS630-Pogo	248	41386 bp	0.12 %
En-Spm	0	0 bp	0.00 %
MuDR-IS905	0	0 bp	0.00 %
PiggyBac	12	1110 bp	0.00 %
Tourist/Harbinger	19	1987 bp	0.01 %
Other (Mirage, P-element, Transib)	1	43 bp	0.00 %
Rolling-circles	0	0	0
Unclassified:	55	3890 bp	0.01 %
Total interspersed repeats		319327 bp	0.91 %
Small RNA:	87	10636 bp	0.03%
Satellites：	87	8284 bp	0.02 %
Simple repeats:	3926	137546 bp	0.39 %
Low complexity:	3669	176617 bp	0.50 %

CONCLUSION

In this study, the transcriptome of *N. albiflora* was sequenced using the HiSeq2000 platform with high coverage, and then de novo assembled and functionally annotated. An ORF analysis had been identified. In addition, repetitive element analysis was conducted, and cDNA SSR and SNP loci were identified for future marker development and genetic analysis. Synonymous and non-synonymous sites were analyzed on unigenes.

ACKNOWLEDGEMENTS

This research was supported by grants from the National Science and Technology Project of China (NO.2012AA10A413-5), the National Natural Science Foundation of China (NO.41476127) and Project of Zhejiang Province of China (No.2012C12907-8, No.2013F20001, No.2014F10004 and No.2015F10001).

REFERENCES

Abernathy, J., Xu, P., Li, P., Xu, D. and Kucuktas, H., 2007. Generation and analysis of expressed sequence tags from the ciliate protozoan parasite *Ichthyophthirius multifiliis*. *BMC Genomics*, **8**:176.

Conesa, A., Götz, S., Garcia-Gomez, J. M., Terol, J., Talon, M. and Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**: 3674-3676.

Cox, M., Peterson, D. and Biggs, P., 2010. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinform.*, **11**: 485.

Cheng, Y., Xu T., Jin X. and Wang, R., 2011. Complete mitochondrial genome of the yellow drum *Nibea albiflora* (Perciformes, Sciaenidae). *Mitoch. DNA*, **22**: 80-82.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z. and Thompson, D.A., 2011. Full length transcriptome

assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.*, **29**: 644-652.

Hampton, M., Melvin, R.G., Kendall, A.H., Kirkpatrick, B.R., and Peterson, N., 2011. Deep sequencing the transcriptome reveals seasonal adaptive mechanisms in a hibernating mammal. *PLoS One*, **6**: e27021.

Han, Z., Gao, T., Wang, Z., Zhuang Z. and Su, T., 2006. Analysis of genetic diversity of *Nibea albiflora* by AFLP markers. *J. Fish. China*, **30**:640-646.

Han, Z., Gao, T., Yanagimoto, T. and Sakurai, Y., 2008. Genetic population structure of *Nibea albiflora*in Yellow Sea and East China Sea. *Fish. Sci.*, **74**: 544-552.

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

Jung, H., Lyons, R.E., Dinh, H., Hurwood, D.A. and McWilliam, S., 2011. Transcriptomics of a giant freshwater prawn (*Macrobrachium rosenbergii*): *De novo* assembly, annotation and marker discovery. *PLoS One*, **6**: e27938.

Li, H. and Durbin, R., 2009. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, **25**:1754-60.

Li, H., Handsaker, B., Wysoker, A. Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R., 2009. 1000 genome project data processing subgroup., 2009: the sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*, **25**:2078-2079.

Ma, C., Ma, H., Ma, L., Jiang, K., Cui, H. and Ma, Q., 2011. Isolation and characterization of 16 polymorphic microsatellite markers from *Nibea albiflora*. *Fish. Sci.*, **77**:707-711.

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C. and Kanehisa, M., 2007. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucl. Acids Res.*, **35**:182-185.

Qiuttao, H., Enhui, Z., Yonglin, L., Maocang, Y., Changjiang, H. and Qiaoxiang, D., 2012. Evaluation of activation and storage conditions for sperm of yellow drum *Nibea albiflora*. *Aquaculture*, **324**:319-322.

Sha, Z., Xu, P., Takano, T., Liu, H. and Terhune, J., 2008. The warm temperature acclimation protein Wap65 as an immune response gene: its duplicates are differentially regulated by temperature and bacterial infections. *Mol. Immunol*., **45**:1458-1469.

Smit, A.F.A., Hubley, R. and Green, P., 1996. *RepeatMasker Open-3.0*. http://www.repeatmasker.org.

Takita, T., 1974. Studies on the early life history of *Nibea albiflora* (Richardson) in Ariake Sound. *Bull. Fac. Fish. Nagasaki Univ.*, **38**:1-55.

Thurston, M.I. and Field, D., 2005. *Msatfinder: detection and characterization of microsatellites*. Available: http://www.genomics.ceh.ac.uk/msatfinder/. CEH Oxford, Mans- field Road, Oxford OX1 3SR (Accessed 4/26/2007).

Wang, S., Peatman, E., Abernathy, J., Waldbieser, G. and Lindquist, E., 2010. Assembly of 500,000 inter-specific catfish expressed sequence tags and large scale gene-associated marker development for whole genome association studies. *Genome Biol.*, **11**: 8.

Xing, S., Shao, C., Liao X., Tian, Y. and Chen, S., 2009. Isolation and characterization of polymorphic microsatellite loci from a dinucleotide-enriched genomic library of spotted maigre (*Nibea albiflora*). *Conserv. Genet.*, **10**:789-791.

Xu, D., Lou, B., Shi, H., Geng, Z., Li, S. and Zhang, Y., 2012. Genetic diversity and population structure of *Nibea albiflora* in the China Sea revealed by mitochondrial COI sequences. *Biochem. Syst. Ecol.*, 45:158-165.

Xu, J., Ji, P., Zhao, Z., Zhang, Y. and Feng, J., 2012. Genome-wide SNP discovery from transcriptome of four common carp strains. *PLoS One*, **7**: e48140.

Yang, X., Chen, Y., Liu, X., Liu, Q., Pi, X., Liu, Y. and Zheng, D., 2015. *De Novo* characterization of velvet skin transcriptome at the antlers tips of red deer (*Cervus elaphus*) and analysis of growth factors and their receptors related to regeneration. *Pakistan J. Zool.*, **48**: 151-157.