Turk. J. Fish.& Aquat. Sci. 19(12), 1049-1059
http://doi.org/10.4194/1303-2712-v19_12_07

RESEARCH PAPER

Turkish Journal of FISHERIES and AQUATIC SCIENCES

Sequencing and Transcriptome Analysis for Reproduction-Related Genes Identification and SSRs Discovery in Sequential Hermaphrodite *Amphiprion ocellaris*

Wei Yang[1,4],Baian Lin[3], Guangli Li[1], Huapu Chen[1,2,*], Min Liu[3,*]

[1]Zhanjiang City Key Laboratory for Marine Ecology and Environment, Guangdong Research Center for Reproductive Control and Breeding of Indigenous Valuable Fish Species, Fisheries College, Guangdong Ocean University, Zhanjiang 524088, China.
[2]Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266071, China.
[3]State Key Laboratory of Marine Environmental Science, College of Ocean and Earth Sciences, Xiamen University, Xiamen 361102, China.
[4]Food and Environmental Engineering Department, Yangjiang Polytechnic, Yangjiang 529566, China.

Article History
Received 29 March 2018
Accepted 09 December 2018
First Online 10 December 2018

Corresponding Author
Tel.: +86.18820706692
E-mail: chpsysu@hotmail.com

Keywords
False clown anemonefish
Amphiprion ocellaris
Sex change
Reproductive regulation
Candidate gene

Abstract

Amphiprion ocellaris is suggested to be an excellent research model for understanding socially controlled sex change of sequential hermaphrodite teleosts. However, there is a lack of knowledge regarding the genes involved in reproduction regulation in sex transition process. In order to enrich genetic data and sequence resources, Illumina HiSeq transcriptome sequencing was conducted for fully mature A. ocellaris using a multiple tissues mixing strategy. A total of 109,408,556 raw reads were obtained and 81,395 unigenes were generated by Trinity de novo assembly. Of which, 60,457 unigenes were annotated in at least one of the queried NR, COG, Swiss-Prot, KEGG and NT databases. After screening from reproduction-related Gene Ontology (GO) terms and pathways, plentiful candidate genes involved in reproductive process were identified, such as gnrh1, fsh, mtnr1a, avt, dmrt1, sox9a, amh, dax1, sf-1, foxl2 and cyp19a1. These candidates could make good starting points for profound mechanical analysis on the reproductive regulation during sex change. Moreover, 18,037 potential simple sequence repeats (SSRs) were detected in 14,212 unigenes for marker development. This transcript dataset will greatly increase available resources for gene mining and function analysis, and could contribute valuable information to in-depth studies on reproductive biology, genetics and genomics in the future.

Introduction

Teleosts exhibit the largest array of sexual patterns and sex-determining systems, such as gonochorism and hermaphroditism with diverse forms of protandry, protogyny and bi-directional sex change in sequential hermaphroditism. Sequential hermaphroditism is a unique reproductive strategy that is generally displayed in at least 27 families of teleosts, including Sparidae, Serranidae, Pomacentridae, Scaridae and Labridae (Sadovy de Mitcheson & Liu, 2008). Sequential hermaphrodites switch sex sometime during life history to function as a male in one phase and as a female in another. As thus, the reproductive biology and sex transition have become one of the most intriguing research hotspots, due to its significant value in academic research (e.g., sex-determination, evolution) and technological development (e.g., sex-change) for improving aquaculture efficiency (Frisch, 2004). Several species have been involved in gene expression analysis to uncover the regulation mechanism in natural sexual reversal. For example,

black porgy for (*Acanthopagrus schlegeli*) GnRH, GtH, *nr0b1* and *nr5a4* (An, Nelson, Habibi, & Choi, 2008; Wu, Tomy, & Chang, 2008), rice field eel (*Monopterus albus*) for *dmrt1* (Huang *et al*., 2005), *foxl2* (Hu, Guo, Gao, Tang, & Li, 2014), *amh* and *dax1* (Hu, Guo, Gao, Tang, & Li, 2015), cinnamon clownfish (*Amphiprion melanopus*) for GnRH (Kim, Shin, Habibi, Lee, & Choi, 2012), orange-spotted grouper (*Epinephelus coioides*) for *sox9* (Luo, Wei, Liu, Lin, & Zhu, 2010). However, comprehensive investigation at transcriptomic and genomic level is scarcely performed, and the complex pathways underlying sex change of sequential hermaphrodite teleosts remain unclear.

Species in the genus *Amphiprion* (Pomacentridae), also known as anemonefishes, are usually protandrous hermaphroditic, small in size (Sadovy de Mitcheson & Liu, 2008), and specifically, are socially controlled sex changers (Iwata, Nagai, & Sasaki, 2010). Within a social group, the female false clown anemonefish *Amphiprion ocellaris* is the largest in body size and dominant member; if the female dies or disappears from the social unit, the second-ranked male individual changes sex into female and the largest nonbreeder becomes a functional male (Iwata, Nagai, Hyoudou, & Sasaki, 2008). *A. ocellaris* is suggested to be an excellent model fish for socially controlled sex change mechanism research, due to its small size, easy cultivation in captivity and young age for sexual maturation. However, previous studies on *A. ocellaris* are mostly focused on histology (Jacob *et al*., 2016), social interaction (Iwata *et al*., 2008) and embryological development (Yasir & Qin, 2007). There is a lack of background knowledge and genetic resources regarding the reproduction-related genes. The mechanisms of reproductive regulation in sex reversal are rarely inspected and poorly understood.

As a version of Next-Generation Sequencing (NGS) technologies, the high throughput RNA sequencing has been utilized frequently in non-model species for rapid and cost-effective access to gene expression data and large-scale transcript sequences. Transcriptome analysis plays significant roles in gene discovery, gene expression and regulation analysis, molecular marker development (Lu *et al*., 2014; Xie *et al*., 2014; Ma *et al*., 2016; Yang *et al*., 2017). Illumina HiSeq RNA sequencing, Trinity based *de novo* assembly, annotation and differential expression analyses were performed in the present study. To the best of our knowledge, this is the first comprehensive transcriptome analysis on *A. ocellaris* base on the multiple tissue samples mixing strategy. Our study primarily aimed i) to markedly enrich genetic information and gene sequence dataset for functional gene mining, expression analysis and EST-SSR marker development, and ii) to identify candidate genes putatively involved in reproduction regulation for future research into molecular mechanism of sexual transition.

Materials and Methods

Animal Materials and Sample Preparation

Mature female and male *Amphiprion ocellaris* (body weight: females (n=5), 35.0±4.5 g; males (n=5), 19.3±3.2 g) were collected from an anemonefish hatchery center (Xiamen, China) on October 2016. Alive fish were sacrificed by decapitation following anesthetization with a MS-222 solution. Tissue samples such as brain, pituitary, head kidney, kidney, heart, muscle, spleen, liver, gut, gill and gonad (testis or ovary) were excised rapidly, then, quick-frozen in liquid-nitrogen immediately and stored in a -80°C freezer until RNA extraction.

RNA Extraction and Sequencing

Total RNA was isolated from each tissue using a Trizol reagent kit (Life Technologies, Carlsbad, CA, USA) following the instructions of manufacturer. The concentration of total RNA was quantified by Nanodrop 2000c (Thermo Scientific, Wilmington, DE, USA) using the absorbance value at 260 nm, and the purity was determined by $OD_{260/280}$ (accepted range: 1.8-2.0). The 18*S* and 28*S* ribosomal bands stained with ethidium bromide (EB) on a 0.8% agarose gel were utilized for RNA integrity assessment. Two pooled RNA samples (approximately 5µg each sex) were used for the construction of male and female cDNA libraries, respectively. In brief, mRNAs were extracted using an Oligo-dT Beads Kit (Qiagen, Hilden, Germany) and fragmented. A SuperScript double-stranded cDNA synthesis kit (Invitrogen, California, USA) was employed to synthesize double-strand cDNA, using 200 – 700 nt fragments as templates and random primers. Then, ligated fragments were obtained by a series of processes including purification of PCR products, end repair, dA-tailing, and ligation of Illumina adapters. Lastly, suitable fragments were isolated by agarose gel electrophoresis and purified for PCR amplification. The RNA-Seq was performed externally by Total Genomics Solution Co., Ltd. (Shenzhen, Guangdong, China). The cDNA libraries of the two sex groups were sequenced on Illumina HiSeq™ 2000 platform (Illumina, Inc., San Diego, USA) and paired-end (PE) reads with length of 125 bp were generated.

De Novo Assembly

The raw sequencing data was quality-controlled by SOAPnuke v1.5.0 with the parameters "-l 10 -q 0.5 -n 0.05 -p 1 -i". Specifically, the raw reads were filtered to produce high quality data via processes including the removal of adapter sequences, reads with ambiguous sequences (N) more than 10% and sequences with more than 20% low-quality bases (quality value < 20). After filtration, the clean reads with high quality were

de novo assembled by Trinity package (version: r20140717, http://trinityrnaseq.sourceforge.net/) with default parameters. The redundant final linear transcripts sequences were eliminated and the longest ones were defined as unigenes (Grabherr *et al*., 2011; Yang *et al*., 2017).

Functional Annotation and Classification

By means of BLAST 2.2.26+ software with an E-value cut-off of 1E-5, functional annotation was performed by sequence alignment against public databases. All assembled unigenes were submitted to search against protein databases including NCBI non-redundant protein (NR), Cluster of Orthologous Groups of proteins (COG), Swiss-Prot, Kyoto Encyclopedia of Genes and Genomes (KEGG) by BLASTx, and NCBI Nucleotide Sequences (NT) database by BLASTn. Sequences with the highest similarity scores from the databases were defined as the functional annotation for the related unigenes. Further analyses were carried out to explain the Gene Ontology (GO) function of the annotation results by Blast2GO (Conesa *et al*., 2005). Then GO term classification and visualization were finished by use of the WEGO statistical software (Ye *et al*., 2006). The KOBAS v2.0 was used to assay the KEGG annotation results for pathway categories (Xie *et al*., 2011).

SSR Loci Identification

Transcriptome SSR loci detection in the assembled sequences was performed using the Perl program MIcroSAtellite (MISA) software package (Thiel, Michalek, Varshney, & Graner, 2003). Because it is difficult to distinguish genuine mononucleotide repeats from polyadenylation products and single nucleotide stretch errors generated by sequencing, mononucleotide repeats were ignored in this study. A minimum of 6, 5, 5, 4 and 4 contiguous repeat units were used to identify perfect di-, tri-, tetra-, penta- and hexa-nucleotide SSR motifs, respectively. The Primer 5.0 software (Lalitha, 2000) was employed to design primer pairs flanking each SSR locus.

Results

Sequencing and Assembly

In totality, 109,408,556 raw PE reads (16,411,283,400 nt bases) were generated by Illumina sequencing. The quality control resulted in 105,603,122 clean PE reads (52,056,328 for female and 53,546,794 for male), which is equal to 15,840,468,300 nt bases sequencing data (Table 1). The GC ratio and Q20 value were 49.27% and 97.00%, respectively. The high-quality clean reads were *de novo* assembled and 81,395 unigenes with a mean sequence length of 1282.77 bp and an N50 value of 2549 bp were produced. The unigene lengths ranged from 201 bp to 60,846 bp and, 45,397 (55.77%) unigenes were ≥500 bp in length (Figure 1a). The original sequencing data are available from the NCBI Sequence Read Archive (SRA) database under the accessions SRX2558711 (male) and SRX2558712 (female).

Functional Annotation and Classification

A total of 60,457 (74.28%) unigenes were annotated in at least one of queried databases. Of these, 44,516 (54.69%), 58,890 (72.35%), 26,283 (32.29%), 26,375 (32.40%), 15,410 (18.93%) and 20,223 (24.84%) unigenes were matched to the NR, NT, Swiss-Prot, KEGG, COG and GO databases, respectively (Figure 1b). A total of 12,078 unigenes could be annotated in all five databases. The E-value distribution showed that 75.46% unigenes exhibited significant homology (below 1E-50) in Nr database (Figure 1c). Over 90% of the annotated unigenes displayed more than 70% similarities. The results confirmed that the transcriptome data were successfully annotated. In a further analysis of the matching sequences, all homologous genes came from 334 species, and most genes (29,820; 66.99%) were related to those of the bicolor damselfish *Stegastes partitus* (Figure 1d).

In order to classify the function of the unigenes, GO terms were assigned by Blast2GO analysis. All unigenes annotated in GO database with one or more GO terms were assigned into 61 functional groups

Table 1. Statistics for the sequencing and assembly of the *A. ocellaris*transcriptome

Sequencing		Assembly	
Total raw reads	109,408,556	Total length of sequences (bp)	104,411,064
Total raw bases(bp)	16,411,283,400	Number of unigenes	81,395
Total clean reads	105,603,122	Maximum sequence length(bp)	60,846
Total clean reads ratio(%)	96.53	Minimum sequence length(bp)	201
Total clean bases(bp)	15,840,468,300	GC content (%)	47.15
Clean reads GC ratio* (%)	49.27	N50 length (bp)	2549
Clean reads Q20* (%)	97.00	Mean length (bp)	1282.77

Note: *—"Q20" is Phred quality scores 20, which is used to measure the probability that a base is called incorrectly. Q20means a 1% chance of error. —"GC ratio" is the guanine-cytosine content of sequences.

(Figure 2a). Of which, 66,972 (50.68%) transcripts comprised the largest category of 'biological process', followed by 'cellular component' (39,672; 30.02%) and 'molecular function" (25,513; 19.31%). The level-two GO terms 'cellular process' (11,985; 9.07%), 'cell' (8342; 6.31%) and 'binding' (10,817; 8.18%) were the largest groups in the three main categories, respectively. After KEGG pathway analysis, 26,375 unigenes were assigned to six main categories, consisting of 307 different pathways (Figure 2b). Among these main categories, 'human diseases' was the most abundant group with 20,268 (29.06%) genes, followed by 'organismal systems' (19,156; 27.46%), 'environmental information processing' (12,619; 18.09%), 'metabolism' (9280; 13.30%), 'cellular processes' (5710; 8.19%) and 'genetic information processing' (2717; 3.90%).

Screening of Candidate Genes Involved in Reproduction

Based on the annotation and classification results, GO terms and KEGG pathways potentially associated with reproduction were further analyzed to screen out candidate gene sequences with reproductive function. Specifically, the target GO terms and KEGG pathways primarily included 'reproduction' (GO:0000003), 'reproductive process' (GO:0022414), 'hormone activity' (GO:0005179), 'receptor binding' (GO:0005102), 'receptor activity' (GO:0004872), 'steroid hormone mediated signaling pathway' (GO:0043401), 'steroid hormone receptor activity' (GO:0003707), 'steroid biosynthetic process' (GO:0006694), 'gonad development' (GO:0008406), 'GnRH signaling pathway' (ko04912), 'progesterone

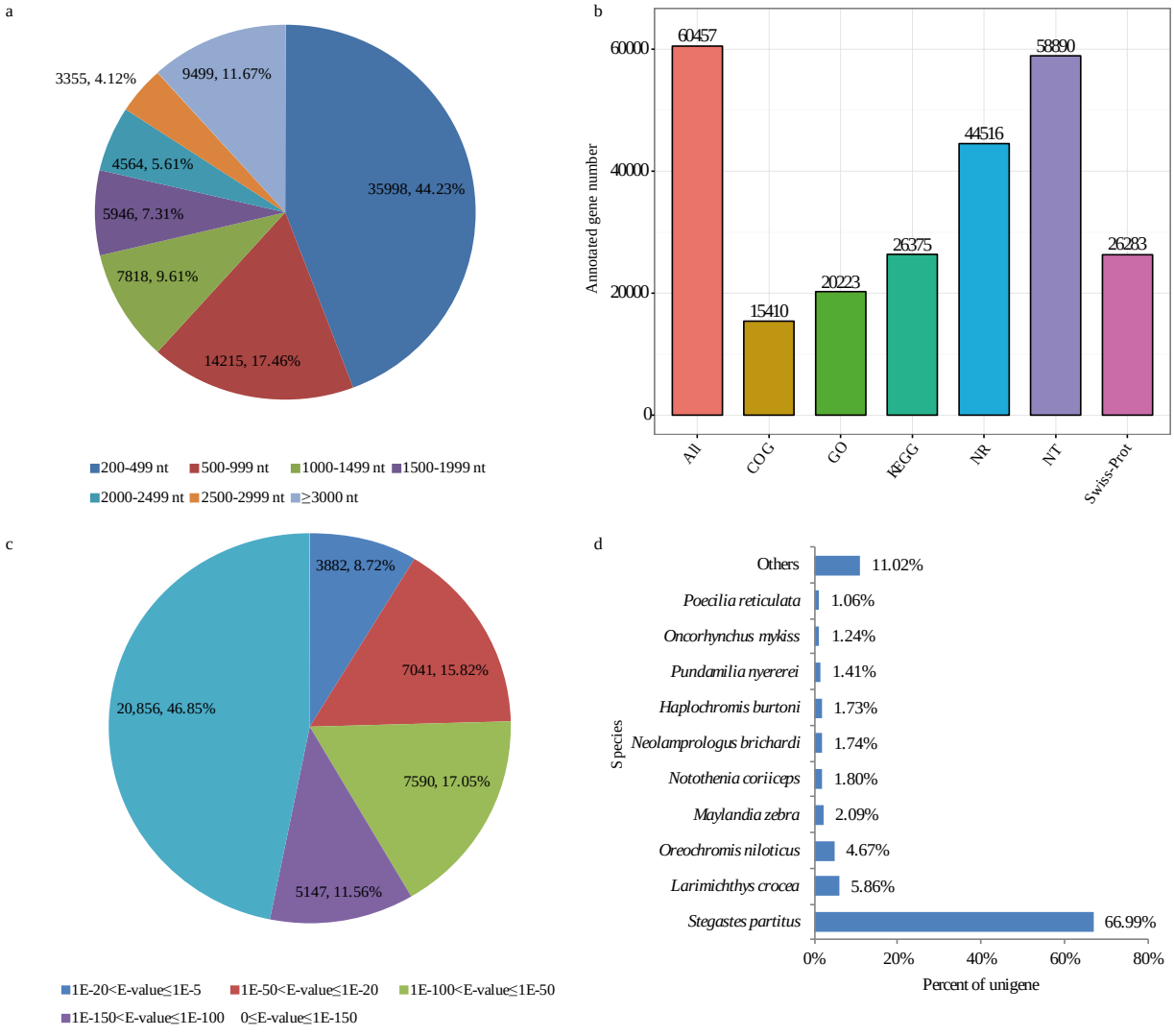

Figure 1. Statistical summary of assembly and annotation of unigenes
a. the sequence size distribution of assembled unigenes; **b.** a histogram of unigene annotation against the NR, NT, GO, Swiss-Prot, COG and KEGG databases, the number on the top of each bar indicatesthe unigenenumber; **c.**E-value distribution ofBLAST hits for unigene annotation against the NR database; **d.** the top ten matching species for annotated transcripts.

mediated oocyte maturation' (ko04914), 'estrogen signaling pathway' (ko04915), 'oocyte meiosis' (ko04114), 'steroid hormone biosynthesis' (ko00140), 'neurotrophin signaling pathway' (ko04722) and 'ovarian steroidogenesis' (ko04913).

As a result, lots of candidate genes with reproduction relevant roles were finally identified from the transcript data. Considering the different function, all candidate genes were generally classfied into three categories: gonadotropin-releasing hormone (GnRH)-gonadotropic hormone (GtH) axis and neuroendocrine regulatory pathway (Table 2), sex determination and differentiation (Table 3), and gonad development and gamete maturation (Table 4). The most represented gene products included GnRH1, arginine vasotocin (AVT), melatonin receptor type 1A(MTNR1A),

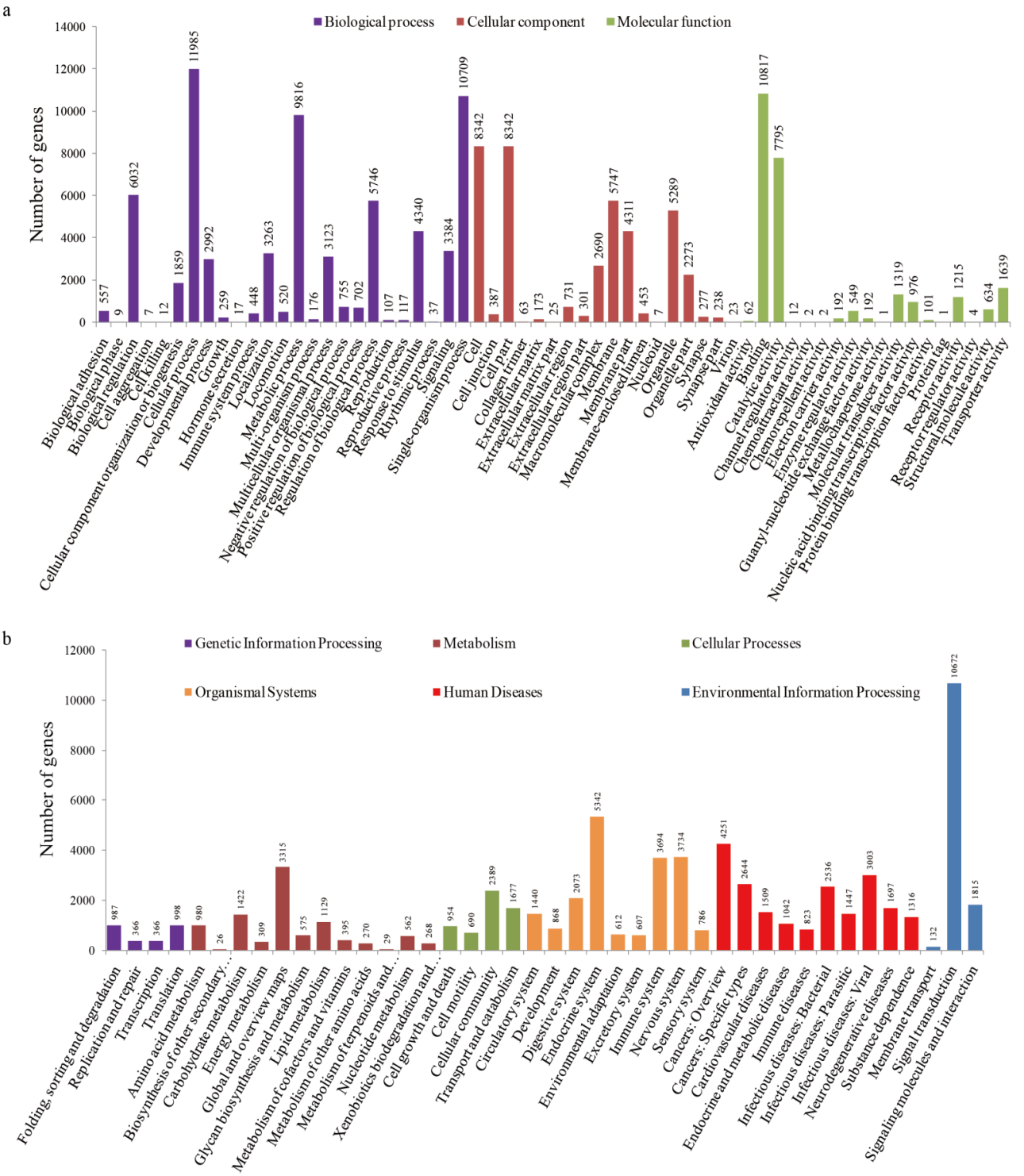

Figure 2. Gene Ontology (GO) and KEGG pathwaysfunctional classification of annotated unigenes.
a. a total of 20,223 unigenes showing significant similarity to homologous genes in GO databases wereassorted to three main categories: cellular components, molecular functions,and biological processes; b. KEEG pathway assignment based on six main categories: genetic information processing, metabolism, cellular processes, organismal systems, environmental information processing andhuman diseases.

doublesex and mab-3 related transcription factor 1 (DMRT1), anti-Müellerian hormone (AMH), nuclear receptor subfamily 0 group B member 1 (DAX1), steroidogenic factor 1 (SF-1), forkhead box l2 (FOXL2), wingless-type MMTV integration site family member-4 (WNT4) and cytochrome P450 family members (CYPs).

Discovery and Characterization of SSR Loci

Using the MISA software, a total of 14,212 (17.46%) unigenes were preliminarily identified to contain 18,037 SSR loci, with a distribution density of one SSR per 5.79 kb. Of which, 3083 (17.09%) unigenes contained more than one SSR. The SSR sequences and flanking primer pairs designed by Primer 5.0 are available from the corresponding author upon reasonable request. Among the identified SSRs, 3320 (18.41%) present in compound formation (Figure 3a). The most abundant repeat motif was di-nucleotide (6818; 46.33%), followed by tri-nucleotide (6785; 46.10%), tetra-nucleotide (713; 4.84%), penta-nucleotide (269; 1.83%) and hexa-nucleotide repeats (132; 0.90%) (Figure 3b). A frequency distribution analysis of the SSRs based on the motif sequence types was conducted (Figure 3c). The *A. ocellaris* transcriptome was rich in AC/GT (4533; 30.80%), AGG/CCT (2047; 13.91%) and AGC/GCT (1744; 11.85%) repeats.

Table 2. Reproduction-related genes involved in the GnRH-GtH aix and neuroendocrine pathway

Gene name	NR annotation	E-value	Typical transcript	Size (bp)
gnrh1	Gonadotropin-releasing hormone 1	1.6316E-51	Unigene36313	346
gnrhr2	Putative gonadotropin-releasing hormone II receptor	1.577E-177	Unigene30746	1175
fshβ	Follicle stimulating hormone beta subunit	4.2507E-66	Unigene14878	905
fshr	Follicle stimulating hormone receptor	1.425E-133	Unigene8771	633
lhb	Luteinizing hormone beta subunit	2.427E-74	Unigene39254	505
lhr	Luteinizing hormone receptor	1.3441E-86	Unigene38793	419
sl	Somatolactin	2.446E-143	Unigene35927	876
adcyap1	Pituitary adenylate cyclase-activating polypeptide isoform X1	1.743E-138	CL2426.Contig2*	1311
crf	Corticotropin-releasing factor	9.1985E-91	Unigene17419	1086
nsg2	Neuron-specific protein family member 2	8.699E-109	Unigene14435*	1826
mtnr1a	Melatonin receptor type 1A	0	Unigene10127	2077
nt5	5'-nucleotidase	0	Unigene9277	2681
avt	Arginine vasotocin	3.0453E-75	Unigene32991	1049
oxt	Isotocin	3.0163E-91	Unigene6587	777
oxtr2	Isotocin receptor 2	1.1498E-53	Unigene44138	303
sert	Sodium-dependent serotonin transporter-like	2.227E-25	CL6940.Contig2	923

Note: * — SSR locus containing unigene

Table 3. Reproduction-related genes involved in sex determination and differentiation

Gene name	NR annotation	E-value	Typical transcript	Size (bp)
dmrt1	Doublesex and mab-3 related transcription factor 1-like	6.8788E-55	Unigene51662	1266
foxl2	Forkhead box protein L2	0	Unigene10486	1968
fem1c	Protein fem-1 homolog C	0	Unigene991*	3749
zar1	Zygote arrest protein 1	0	Unigene12327*	1331
sf1	Steroidogenic factor 1	0	Unigene3392	648
gtf3a	Transcription factor IIIA	0	Unigene6402*	1216
amh	Muellerian-inhibiting factor	0	Unigene20226*	1789
amhr2	Anti-Muellerian hormone type-2 receptor-like	3.75E-176	Unigene871	965
dax1	Nuclear receptor subfamily 0 group B member 1-like	0	Unigene7318	1336
sox9a	Transcription factor Sox-9-A-like	0	Unigene13822*	1460
sox8	Transcription factor SOX-8	0	Unigene27545*	2657
wnt4	Protein Wnt-4-like	2.069E-69	Unigene54376	721
cyp19a1b	Brain cytochrome P450 aromatase	0	CL1310.Contig1	3363
cyp19a1a	Gonad cytochrome P450 aromatase	0	Unigene24667	1429
cyp17a1	Steroid 17-alpha-hydroxylase/17,20 lyase	0	Unigene26846	1850
star	Steroidogenic acute regulatory protein	5.717E-172	CL5812.Contig1	2390
er	Estrogen receptor	0	Unigene17217*	1843
erβ2	Estrogen receptor beta 2	0	CL200.Contig2	1820
gdf9	Growth/differentiation factor 9	0	Unigene19135	2250

Note: * — SSR locus containing unigene.

Table 4. Reproduction-related genes involved in gamete development and maturation

Gene name	NR annotation	E-value	Typical transcript	Size (bp)
col15a1	Collagen alpha-1(XV) chain-like	2.419E-154	Unigene10537	1714
col4a5	Collagen alpha-5(IV) chain	2.43E-155	Unigene29953	5385
pgrmc2	Membrane-associated progesterone receptor component 2	1.333E-105	Unigene8916	2085
ropn1l	Ropporin-1-like protein	9.167E-139	Unigene1259*	563
spin1	Spindlin-1-like	1.386E-149	CL4467.Contig1	2112
fstl1	Follistatin-related protein 1	0	Unigene9700*	1568
sox3	Transcription factor SOX-3	0	Unigene30401*	1957
sox17	Transcription factor SOX-17	0	Unigene2228	1707
sox6	Transcription factor SOX-6	2.1055E-60	Unigene10895*	1212
sox30	Transcription factor SOX-30-like	8.8215E-39	Unigene1522*	1325
zp2	Zona pellucida sperm-binding protein 2-like	0	Unigene15870	3020
thra	Thyroid hormone receptor alpha-like isoform X2	1.227E-162	CL6329.Contig2	1293
rassf1	Ras association domain-containing protein 1 isoform X1	0	CL1347.Contig1	2483
spef1	Sperm flagellar protein 1	2.807E-153	Unigene53307	975
spata16	Spermatogenesis-associated protein 16	0	Unigene51796	1487
spata4	Spermatogenesis-associated protein 4	6.068E-140	Unigene8946	2040
spata7	Spermatogenesis-associated protein 7	0	Unigene7461	2561
tctex1	Tctex1 domain-containing protein 1	1.8952E-91	Unigene51421	722
tdrp	Testis development-related protein	3.145E-128	Unigene3469	2079
tssk3	Testis-specific serine/threonine-protein kinase 3-like isoform X1	0	CL3264.Contig2	2664
spag6	Sperm-associated antigen 6-like	0	Unigene23482	2103
klhl10	Kelch-like protein 10	0	CL5280.Contig1	1934

Note: * – SSR locus containing unigene.

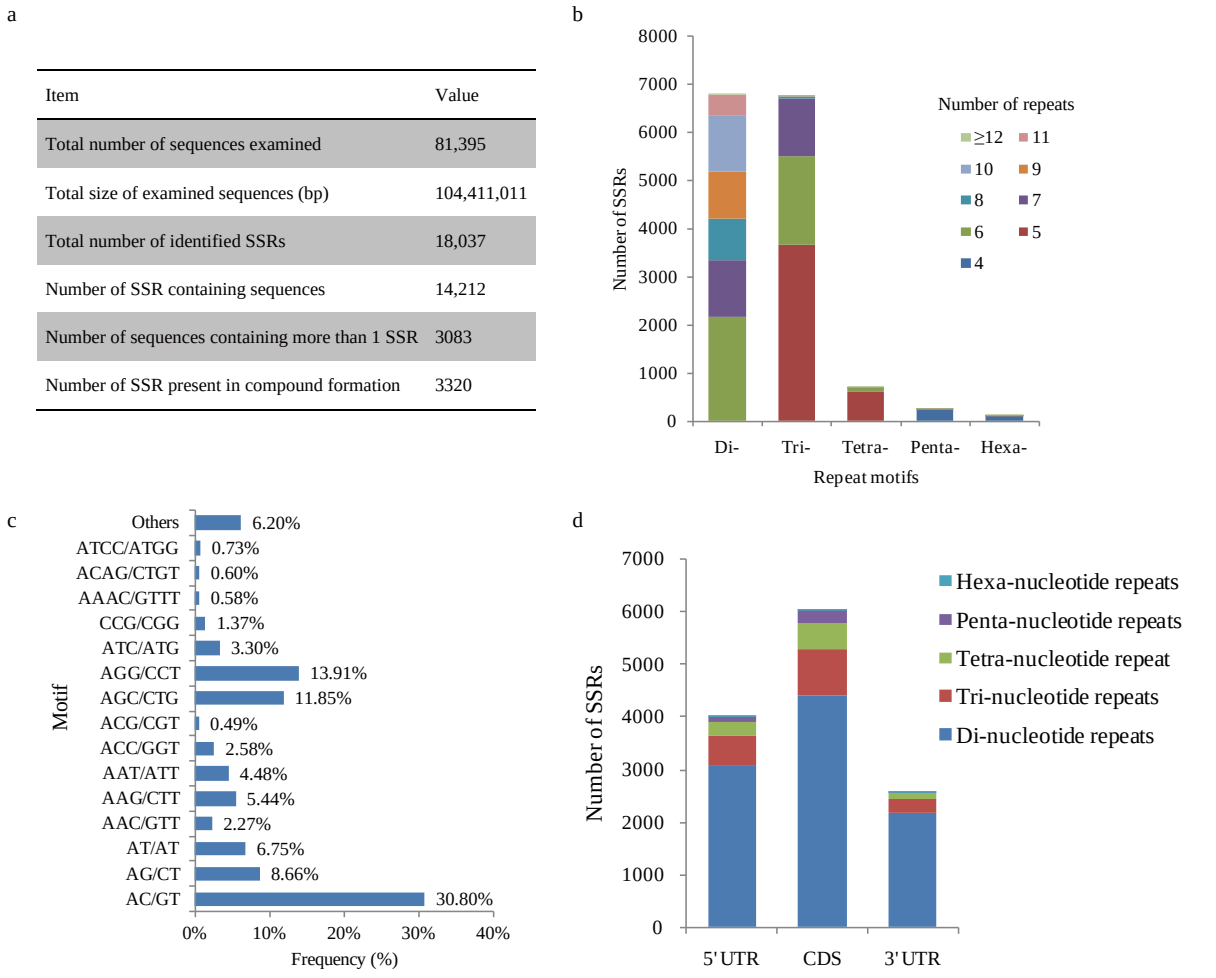

Figure 3. Characterization of SSR loci identified in the *A. ocellaris* transcriptome
a. statistics of SSR searching results; **b.** the distribution of repeat number for di-, tri-, tetra-, penta-, andhexa-nucleotide SSRrepeat motifs; **c.** the frequency distribution of classified types of repeat motifs; **d.** distribution of SSRs with different repeat motifs across the 5' UTR, CDS and 3' UTR.

The SSR location analysis showed that 6061 (33.6%) SSR loci were found in the coding region (CDS), 4040 SSRs (22.4%) in the 5' untranslated regions (5' UTR) and 2597 SSRs (14.4%) in the 3' UTR. Of the SSRs located in the 5' UTR, CDS or 3' UTR, the most abundant repeats were found to be di-nucleotide, accounting for 76.3% (3083), 72.8% (4412) and 84.1% (2184), respectively (Figure 3d). Fascinatingly, a total of 10 (17.2%) unigene sequences were identified as SSR-containing transcripts from the reproduction-related candidate genes (Table 2-4, Supplementary file). Of these reproduction-related EST-SSRs, almost all of the motif were di- or tri-nucleotides repeats (perfect motif type), except one SSR with compound motif type in Unigene30401 (*sox3*). Considering the SSR location, 3 EST-SSRs (CL2426.Contig2, Unigene14435, Unigene1259) were found in the 5' UTR, 3 (Unigene13822, Unigene17217, Unigene1522) in coding regions and 5 (CL2426.Contig2, Unigene991, Unigene20226, Unigene27545, Unigene30401) in the 3' UTR.

Discussion

Sequencing and Assembly

A multi-tissues strategy was utilized for cDNA libraries construction to generate representative transcriptomes of *A. ocellaris*, such strategy has been widely introduced in the RNA-Seq projects for teleosts (Lv *et al*., 2014; Ma *et al*., 2016; Yang *et al*., 2017). For high-throughput short read sequencing, high-quality assembly will provide benefit for post transcriptomic analysis like annotation, gene identification and comparative genomics. According to many past studies, the quality of assembly is evaluated mainly by the length distribution of contigs or transcripts (Jiang, Fan, & Xu, 2017). In this experiment, more than half of the *de novo* assembled unigenes were >500 bp in length and the mean length of unigenes reached 1282 bp (Table 1; Figure 1a). Similar results were given by Trinity transcriptome assemblies for other teleost fishes such as *Pelteobagrus fulvidraco* (1611 bp) (Lu *et al*., 2014), *Trachinotus ovatus* (1179 bp) (Xie *et al*., 2014) and *Scatophagus argus* (906 bp) (Yang *et al*., 2017), strongly indicating that our transcriptome data was assembled effectively. In contrast with other *de novo* assembly softwares e.g. CLC Genomics Workbench (Malachowicz, Wenne, & Burzynski, 2017), iAssembler (Chen *et al*., 2013) and Newbler (Wang *et al*., 2013), the most satisfying assembler Trinity could provide a unified solution for good-quality transcriptome reconstruction in species without a reference genome (Grabherr *et al*., 2011; Jiang *et al*., 2017). A longer assembled sequence can provide more adequate information for further gene investigation and identification of molecular markers. Here the BLAST match rates of CDS enhanced markedly from 200-300 bp (23.86%) to 1500-1600 bp (81.17%). On the other hand, the length of query sequences was crucial for determining the level of significance of the BLAST hits. In the present study, the ratio of unigenes with significant BLAST scores increased sharply from 200-500 bp to 500-1500 bp. These findings indicate that the proportion of sequences with matches in database is greater among the longer assembled sequences, which is consistent with other analytical results of the next-generation transcriptome sequencing (Zhu, Li, Yang, & Li, 2013; Shi *et al*., 2011; Yang *et al*., 2017).

Reproduction-Related Genes

GnRH-GtH axis has been demonstrated to take part in the reproductive regulation and sex reversal in socially controlled sex-changing fish (Ohta, Mine, Yamaguchi, & Matsuyama, 2008). In this study several candidate genes playing critical roles in GnRH-GtH axis, including *gnrh1*, *gnrhr*2, *fshβ*, *fshr*, *lhb* and *lhr*, were obtained by functional classification and screening (Table 2). The expression levels of GnRHs mRNA in mature cinnamon clownfish gonads were higher than the levels at earlier maturing stages (Kim *et al*., 2012), supporting the hypothesis that GnRHs engage in the modulation of gonadal development and sex change of hermaphrodite fish. Meantime, we identified a number of genes (e.g. *avt* and *mtnr1a*) involved in the neuroendocrine pathway which is widely believed to mediate the transmission of external signals from environment to GnRH-GtH axis (Table 2). The brain nanopeptide arginine vasotocin (AVT) is concerned with the highly plastic mechanism of sex determination in teleost fishes (Munday, Buston, & Warner, 2006). In *A. ocellaris* social rank formation can modulate the brain AVT production and may influence later sex differentiation (Iwata *et al*., 2010). Melatonin has also been implicated as a regulatory substance in seasonal reproduction and sexual maturation for fish species (Amano *et al*., 2000). Relatively higher serum melatonin levels after spawning permit initiation of the sex reversal in rice field eel (Shi, 2005), suggesting that melatonin is correlated with sex-change process. According to our latest experiments (unpublished data), brain *avt* and *mtnr1a* were significantly up-regulated during sex reversal in *A. ocellaris*. This finding strongly indicates that melatonin and AVT are both likely to play important roles in sex inversion. However, the exact functions of these genes are still unclear, and further research is urgently needed to elucidate their interaction patterns in neuroendocrine regulation network.

In addition, a few scarcely investigated genes involved in the neuroendocrine regulation of reproduction (e.g. *sl*, *oxt* and *oxtr2*) were identified for the first time in *A. ocellaris* (Table 2). Somatolactin is a pituitary hormone stimulating production of 11-ketotestosterone, testosterone by testicular fragments

and production of estradiol by ovarian follicles (Planas, Swanson, Rand-Weaver, & Dickhoff, 1992), and isotocin, another important brain nanopeptide, is considered to regulate female-specific behavior (Campbell, 2008). Although previous studies show that these genes may participate in reproductive process, the functions and mechanisms have not yet been established in sex-changing fish. Currently the upstream mechanisms controlling the sex transformation of anemonefish are still much less certain, as this species has a longer life span and takes a longer time to complete sex differentiation (Iwata *et al*., 2008, 2010). The identification of such functional genes would provide new insights into the roles of neuroendocrine pathway, and enable future investigations into the regulatory networks and mechanisms in socially controlled sex reversal.

Sexual differentiation during sex change is a perplexing and multi-level process requiring coordinate interaction of signals among the genetic, hormonal and environmental factors to control the proliferation, differentiation and development of reproductive tissues (Devlin & Nagahama, 2002; Wu *et al*., 2010). Numbers of genes with important roles in sex determination and differentiation were screened out, the transcript size of these genes ranged from 648 bp to 3749 bp (Table 3). Interestingly, a number of past studies have revealed that *dmrt1*, *sox9a*, *dax1*, *amh*, *sf-1*, *foxl2*, *cyp19a1a* and *wnt4* genes may perform critical roles in the sex differentiation and gametogenesis of sex-changing fish. *Sox9* exhibited an obviously phase-dependent expression pattern in the natural sex reversal of orange-spotted grouper (Luo *et al*., 2010), indicating its important functions for the initiation and maintenance of masculinization. *Dmrt1* is generally considered to take charge of testicular differentiation cascade and could down-regulate *cyp19a1a* in vertebrates (Guiguen, Fostier, Piferrer, & Chang, 2010). Zhu, Wang, Chen, and Guan (2016) found that *dmrt1* was absent in the ovary but present in Sertoli and gonia cells in testis and ovotestis of *Monopterus albus*, implying that it may be a essential regulatory factor for the transition of ovary to testis. The preliminary effects of *sf-1*, *dax1* and *amh* on sex-inversion regulation have been demonstrated as well. *Sf-1* expression increased significantly during and after testicular differentiation, while *dax1* was up-regulated markedly for the further development of differentiated gonad in black porgy (Wu *et al*., 2008). During the sex inversion in rice-field eel (Hu *et al*., 2015), up-regulation of *amh* and down-regulation of *dax1* are likely necessary for the activation of testis development, and the high *amh* expression level and a low *dax1* level facilitate the testis function maintenance. Hitherto most of these genes have been sparsely examined in anemonefish and the precise actions in sex transition are poorly understood. Hence more well-designed molecular biology and genetics research focused on steroidogenesis, gonad differentiation, gametogenesis and gamete maturation are sorely necessary and highly encouraged to unravel the in-depth regulatory mechanism underlying sex-switch of *A. ocellaris*. The available transcript resource derived from transcriptome data would provide good starting points for future studies.

Characterization of SSR Loci

Using the MISA software, about 17% of assembled unigenes were identified as SSR loci containing sequences in the *A. ocellaris* transcriptome. On average, one SSR could be found every 5.79 kb. The distribution density of microsatellites was similar to the reports for other teleosts such as *T. ovatus* (Xie *et al*., 2014) and *Salmo trutta* m. *trutta* (Malachowicz *et al*., 2017). Among these unigenes, 3083 contained more than one SSR and of which, 616 contained more than three SSRs (Figure 3a). The inequality of distribution of SSRs has been commonly detected in research of transcriptome-based SSR development for fishes, such as *S. argus* (Yang *et al*., 2017) and *Betta splendens* (unpublished data). In addition, di-nucleotide (AC/GT) and tri-nucleotide repeats (AGG/CCT, AGC/GCT) were found as the most abundant SSR motifs (Figure 3c). Similar results were also observed in the transcriptomic analysis of diverse species of fish including *T. ovatus* (Xie *et al*., 2014), *Scophthalmus maximus* (Ma *et al*., 2016), *S. argus* (Yang *et al*., 2017) and *S. trutta* m. *trutta* (Malachowicz *et al*., 2017). Such consistence implicates the conservativeness of the dominating microsatellite repeat types in teleosts.

Transcriptome- and EST-based SSRs mainly occur in the coding regions of annotated genes. In this study, 10 unigene sequences containing 11 SSR loci were identified from the reproduction-related transcripts. Of these detected EST-SSRs, three were located in the CDSs and eight in the 3' UTR or 5' UTR (Supplementary file). Previous studies concluded that different locations of SSRs could suggest their possible roles for gene expression and function (Zhang *et al*., 2014), and the SSR insertions in the promoter region may modulate gene expression (Fuganti *et al*., 2010). So that these SSR loci are likely to provide important marker resource for genetic analysis and determining functional genetic variation in above reproduction-related genes. We further noticed that all of the reproduction-related SSRs distributed in the CDSs were tri-nucleotides repeats. Such a propensity for tri-nucleotides may exist to reduce the incidence of frameshift mutations caused by nontriplet repeats by suppressing the other types of SSRs (Metzgar, Bytof, & Wills, 2000). These reproduction-related EST-SSRs could also be utilized as a helpful molecular tool for the genetic mechanism research on *A. ocellaris* reproduction regulation. Nowadays EST-SSR marker is an effective approach to identify phenotypic

associations and useful resource for genetic map construction, map-based gene clone, comparative genomics, genetic diversity analysis, functional genetic variation examination and molecular marker assisted selection (Zhang *et al*., 2014). To date, few polymorphic SSR has been developed in *A. ocellaris*. The newly identified EST-SSRs in our study will be a wealth of resource for developing polymorphic markers, which would be of good potential application in gene mapping, comparative genomics and a powerful tool for future genetic studies.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (Nos. 41706174 and 31702326), Marine Fishery Science and Technology Promotion Projects of Guandong Province (Nos. A201608B01 and 2017A0012), Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology (No. 2016LMFS-B12), Xiamen Southern Oceanographic Center of China (No. 13GZY001NF12) and Hainan Province Key Research and Development Program (No. ZDYF2018225).

References

Amano, M., Iigo, M., Ikuta, K., Kitamura, S., Yamada, H., & Yamamori, K. (2000). Roles of melatonin in gonadal maturation of underyearling precocious male masu salmon. *General and Comparative Endocrinology, 120*(2), 190-197. https://doi.org/10.1006/gcen.2000.7547

An, K.W., Nelson, E.R., Habibi, H.R., & Choi, C.Y. (2008). Molecular characterization and expression of three GnRH forms mRNA during gonad sex-change process, and effect of GnRHa on GtH subunits mRNA in the protandrous black porgy (*Acanthopagrus schlegeli*). *General and Comparative Endocrinology, 159*(1), 38-45. https://doi.org/10.1016/j.ygcen.2008.07.012

Campbell, A. (2008). Attachment, aggression and affiliation: the role of oxytocin in female social behavior. *Biological Psychology, 77*(1), 1-10. https://doi.org/10.1016/j.biopsycho.2007.09.001

Chen, X.H., Zeng, D.G., Chen, X.L., Xie, D.X., Zhao, Y.Z., Yang, C.L., ... Wang, H. (2013). Transcriptome analysis of *Litopenaeus vannamei* in response to white spot syndrome virus infection. *PLoS One, 8*(8), e73218. https://doi.org/10.1371/journal.pone.0073218

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics, 21*(18), 3674-3676. https://doi.org/10.1093/bioinformatics/bti610

Devlin, R.H., & Nagahama, Y. (2002). Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture, 208*(3-4), 191-364. https://doi.org/10.1016/S0044-8486(02)00057-1

Frisch, A. (2004). Sex-change and gonadal steroids in sequentially-hermaphroditic teleost fish. *Reviews in Fish Biology and Fisheries, 14*(4), 481-499. https://doi.org/10.1007/s11160-005-3586-8

Fuganti, R., Machado Mde, F., Lopes, V.S., Silva, J.F., Arias, C.A., Marin, S.R., ... Nepomuceno, A.L. (2010). Size of $AT_{(n)}$ insertions in promoter region modulates *Gmhsp17.6-L* mRNA transcript levels. *Journal of Biomedicine and Biotechnology*, *2010*, 847673. http://dx.doi.org/10.1155/2010/847673

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., ... Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology, 29*(7), 644-652. https://doi.org/10.1038/nbt.1883

Guiguen, Y., Fostier, A., Piferrer, F., & Chang, C.F. (2010). Ovarian aromatase and estrogens: a pivotal role for gonadal sex differentiation and sex change in fish. *General and Comparative Endocrinology, 165*(3), 352-366. https://doi.org/10.1016/j.ygcen.2009.03.002

Hu, Q., Guo, W., Gao, Y., Tang, R., & Li, D. (2014). Molecular cloning and analysis of gonadal expression of *foxl2* in the rice-field eel *Monopterus albus*. *Scientific Reports, 4*, 6884. https://doi.org/10.1038/srep06884

Hu, Q., Guo, W., Gao, Y., Tang, R., & Li, D. (2015). Molecular cloning and characterization of *amh* and *dax1* genes and their expression during sex inversion in rice-field eel*Monopterus albus*. *Scientific Reports, 5*, 16667. https://doi.org/10.1038/srep16667

Huang, X., Guo, Y., Shui, Y., Gao, S., Yu, H., Cheng, H., & Zhou, R. (2005). Multiple alternative splicing and differential expression of *dmrt1* during gonad transformation of the rice field eel. *Biology of Reproduction, 73*(5), 1017-1024. https://doi.org/10.1095/biolreprod.105.041871

Iwata, E., Nagai, Y., Hyoudou, M., & Sasaki, H. (2008). Social environment and sex differentiation in false clown anemonefish, *Amphiprion ocellaris*. *Zoological Science, 25*(2), 123-128. https://doi.org/10.2108/zsj.25.123

Iwata, E., Nagai, Y., & Sasaki, H. (2010). Social rank modulates brain arginine vasotocin immunoreactivity in false clown anemonefish (*Amphiprion ocellaris*). *Fish Physiology and Biochemistry, 36*(3), 337-345. https://doi.org/10.1007/s10695-008-9298-y

Jacob, H., Metian, M., Brooker, R.M., Duran, E., Nakamura, N., Roux, N., ... Lecchini, D. (2016). First description of the neuro-anatomy of a larval coral reef fish *Amphiprion ocellaris*. *Journal of Fish Biology, 89*(3), 1583-1591. https://doi.org/10.1111/jfb.13057

Jiang, Y., Fan, W.Q., & Xu, J.M. (2017). *De novo* transcriptome analysis and antimicrobial peptides screening in skin of *Paa boulengeri*. *Genes and Genomics, 39*(6), 653-665. https://doi.org/10.1007/s13258-017-0532-9

Kim, N.N., Shin, H.S., Habibi, H.R., Lee, J., & Choi, C.Y. (2012). Expression profiles of three types of GnRH during sex-change in the protandrous cinnamon clownfish, *Amphiprion melanopus*: Effects of exogenous GnRHs. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, 161*(2), 124-133. https://doi.org/10.1016/j.cbpb.2011.10.003

Lalitha, S. (2000). Primer premier 5. *BioTech Software and Internet Report, 1*(6), 270-272. https://doi.org/10.1089/152791600459894

Lu, J.G., Luan, P.X., Zhang, X.F., Xue, S.Q., Peng, L.N., Mahbooband, S., & Sun, X.W. (2014). Gonadal transcriptomic analysis of yellow catfish (*Pelteobagrus fulvidraco*): identification of sex-related genes and genetic markers. *Physiological Genomics, 46*(21), 798-

807. https://doi.org/10.1152/physiolgenomics.00088.2014

Luo, Y.S., Wei, H., Liu, X.C., Lin, H.R., & Zhu, Z.Y. (2010). Molecular cloning and mRNA expression pattern of *Sox9* during sex reversal in orange-spotted grouper (*Epinephelus coioides*). *Aquaculture, 306*(1-4), 322-328. https://doi.org/10.1016/j.aquaculture.2010.06.019

Lv, J.J., Liu, P., Gao, B.Q., Wang, Y., Wang, Z., Chen, P., & Li, J. (2014). Transcriptome analysis of the *Portunus trituberculatus*: *de novo* assembly, growth-related gene identification and marker discovery. *PLoS One, 9*(4), e94055. https://doi.org/10.1371/journal.pone.0094055

Ma, D.Y., Ma, A.J., Huang, Z.H., Wang, G.N., Wang, T., Xia, D.D., & Ma, B.H. (2016). Transcriptome analysis for identification of genes related to gonad differentiation, growth, immune response and marker discovery in the turbot (*Scophthalmus maximus*). *PLoS One, 11*(2), e0149414. https://doi.org/10.1371/journal.pone.0149414

Malachowicz, M., Wenne, R., & Burzynski, A. (2017). *De novo* assembly of the sea trout (*Salmo trutta* m. *trutta*) skin transcriptome to identify putative genes involved in the immune response and epidermal mucus secretion. *PLoS One, 12*(2), e0172282. https://doi.org/10.1371/journal.pone.0172282

Metzgar, D., Bytof, J., & Wills, C. (2000). Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome Research*, *10*, 72-80. https://doi.org/10.1101/gr.10.1.72

Munday, P.L., Buston, P.M., & Warner, R.R. (2006). Diversity and flexibility of sex-change strategies in animals. *Trends in Ecology and Evolution, 21*(2), 89-95. https://doi.org/10.1016/j.tree.2005.10.020

Ohta, K., Mine, T., Yamaguchi, A., & Matsuyama, M. (2008). Sexually dimorphic expression of pituitary glycoprotein hormones in a sex-changing fish (*Pseudolabrus sieboldi*). *Journal of Experimental Zoology Part A Ecological and Integrative Physiology, 309*(9), 534-541. https://doi.org/10.1002/jez.485

Pereiro, P., Balseiro, P., Romero, A., Dios, S., Forn-Cuni, G., Fuste, B., ... Figueras, A. (2012). High-throughput sequence analysis of turbot (*Scophthalmus maximus*) transcriptome using 454-pyrosequencing for the discovery of antiviral immune genes. *PLoS One, 7*(5), e35369. https://doi.org/10.1371/journal.pone.0035369

Planas, J.V., Swanson, P., Rand-Weaver, M., & Dickhoff, W.W. (1992). Somatolactin stimulates in vitro gonadal steroidogenesis in coho salmon, *Oncorhynchus kisutch*. *General and Comparative Endocrinology, 87*(1), 1-5. https://doi.org/10.1016/0016-6480(92)90142-7

Sadovy de Mitcheson, Y., & Liu, M. (2008). Functional hermaphroditism in teleosts. *Fish and Fisheries, 9*(1), 1-43. https://doi.org/10.1111/j.1467-2979.2007.00266.x

Shi, Q. (2005). Melatonin is involved in sex change of the ricefield eel, *Monopterus albus* Zuiew. *Reviews in Fish Biology and Fisheries, 15*(1-2), 23-36. https://doi.org/10.1007/s11160-005-7848-2

Shi, C.Y., Yang, H., Wei, C.L., Yu, O., Zhang, Z.Z., Jiang, C.J., ... Wan, X.C. (2011). Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. *BMC Genomics, 12*, 131. https://doi.org/10.1186/1471-2164-12-131

Thiel, T., Michalek, W., Varshney, R., & Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics, 106*(3), 411-422. https://doi.org/10.1007/s00122-002-1031-0

Wang, S., Hou, R., Bao, Z.M., Du, H.X., He, Y., Su, H.L., ... Hu, X.L. (2013). Transcriptome sequencing of zhikong scallop (*Chlamys farreri*) and comparative transcriptomic analysis with yesso scallop (*Patinopecten yessoensis*). *PLoS One, 8*(5), e63927. https://doi.org/10.1371/journal.pone.0063927

Wu, G.C., Tomy, S., & Chang, C.F. (2008). The expression of *nr0b1* and *nr5a4* during gonad development and sex change in protandrous black porgy fish, *Acanthopagrus schlegeli*. *Biology of Reproduction, 78*(2), 200-210. https://doi.org/10.1095/biolreprod.107.062612

Wu, G.C., Tomy, S., Lee, M.F., Lee, Y.H., Yueh, W.S., Lin, C.J., ... Chang, C.F. (2010). Sex differentiation and sex change in the protandrous black porgy, *Acanthopagrus schlegeli*. *General and Comparative Endocrinology, 167*(3), 417-421. https://doi.org/10.1016/j.ygcen.2009.11.003

Xie, C., Mao, X.Z., Huang, J.J., Ding, Y., Wu, J.M., Dong, S., ... Wei, L.P. (2011). KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Research, 39*(suppl 2), W316-W322. https://doi.org/10.1093/nar/gkr483

Xie, Z.Z., Xiao, L., Wang, D.D., Fang, C., Liu, Q.Y., Li, Z.H., ... Lin, H.R. (2014). Transcriptome analysis of the *Trachinotus ovatus*: identification of reproduction, growth and immune-related genes and microsatellite markers. *PLoS One, 9*(10), e109419. https://doi.org/10.1371/journal.pone.0109419

Yang, W., Chen, H.P., Cui, X.F., Zhang, K.W., Jiang, D.N., Deng, S.P., ... Li, G.L. (2017). Sequencing, *de novo* assembly and characterization of the spotted scat *Scatophagus argus* (Linnaeus 1766) transcriptome for discovery of reproduction related genes and SSRs. *Chinese Journal of Oceanology and Limnology*, published online on September 2017. https://doi.org/10.1007/s00343-018-7090-0

Yasir, I., & Qin, G.J. (2007). Embryology and early ontogeny of an anemonefish *Amphiprion ocellaris*. *Journal of the Marine Biological Association of the United Kingdom, 87*(4), 1025-1033. https://doi.org/10.1017/S0025315407054227

Ye, J., Fang, L., Zheng, H.K., Zhang, Y., Chen, J., Zhang, Z.J., ... Wang, J. (2006). WEGO: a web tool for plotting GO annotations. *Nucleic Acids Research, 34*(suppl 2), W293-W297. https://doi.org/10.1093/nar/gkl031

Zhang, J., Ma, W., Song, X., Lin, Q., Gui, J.F., & Mei, J. (2014). Characterization and development of EST-SSR markers derived from transcriptome of yellow catfish. *Molecules, 19*(10), 16402-16415. https://doi.org/10.3390/molecules191016402

Zhu, J.Y., Li, Y.H., Yang, S., & Li, Q.W. (2013). *De novo* assembly and characterization of the global transcriptome for *Rhyacionia leptotubula* using Illumina paired-end sequencing. *PLoS One, 8*(11), e81096. https://doi.org/10.1371/journal.pone.0081096

Zhu, Y.F., Wang, C.L., Chen, X.W., & Guan, G.J. (2016). Identification of gonadal soma-derived factor involvement in Monopterus albus (protogynous rice field eel) sex change. Molecular Biology Reports, 43(7), 629-637. https://doi.org/10.1007/s11033-016-3997-8