Comparative transcriptome analysis of female and male finepatterned puffer (Takifugu poecilonotus): identification of candidate genes and pathways associated with growth and gonad differentiation

Binbin Shan
South China Sea Fisheries Research Institute, Chinese Academy of Fisheries Sciences
Yan Liu
South China Sea Fisheries Research Institute, Chinese Academy of Fisheries Sciences
Changping Yang
South China Sea Fisheries Research Institute, Chinese Academy of Fisheries Sciences
Liangming Wang
South China Sea Fisheries Research Institute, Chinese Academy of Fisheries Sciences
Gongjun Zhang
South China Sea Fisheries Research Institute, Chinese Academy of Fisheries Sciences
Dianrong Sun (sundianrong@yeah.net)
South China Sea Fisheries Research Institute, Chinese Academy of Fisheries Sciences

Research Article

Keywords: De novo assembly, Takifugu poecilonotus, transcriptome, growth-related genes, sex-related genes, differentially expressed genes

Posted Date: April 13th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-396640/v1

Abstract

Background: As a common phenomenon in metazoans, sexual growth dimorphism is evident in a wide range of teleost fish. The growth rate is found to be different between the sexes; however, the mechanism of this complex phenomenon remains unclear. The finepatterned puffer (*Takifugu poecilonotus*; Temminck & Schlegel, 1850) is a typical female heterogamete species that exhibits female-biased sexual size dimorphism. As a commercial fish, *T. poecilonotus* has high exploitability and aquaculture prospects; however, little genomic information on the species is available.

Results: In the present study, the transcriptomes of female and male *T. poecilonotus* were sequenced with RNA-seq technology. A total of 285.95 million raw read pairs were generated from sequence libraries. After identification and assembly, a total of 149,814 nonredundant unigenes were obtained with an N50 length of 3,538 bp. Of these candidates, 122,719 unigenes (81.91% of the total) were successfully annotated with multiple public databases. A comparison of the unigenes of different sexes of *T. poecilonotus* revealed that 10,385 unigenes (2,034 in females and 8,351 in males) were differentially expressed between females and males. Then, we identified many candidate growth- and sex-related genes, including *Dmrt1*, *Sox3*, *Spatas Prl/ Prlr*, *fabps*, Ghr, and *Igf1r*. In addition to these well-known genes, *Fabp*4 was identified for the first time in fish. Furthermore, the expression levels of ten unigenes were confirmed by qRT-PCR to validate the transcriptome data. Moreover, 68,281 simple sequence repeats (SSRs) were identified in SSR-containing sequences.

Conclusions: This informative transcriptome analysis provides valuable data to increase the genomic resources of *T. poecilonotus*. The results will be useful for clarifying the molecular mechanism of sexual growth dimorphism and for future functional analyses of growth- and sex-associated genes.

1. Introduction

Sexual growth dimorphism, growth rate differing between the genders, is common among animals, e.g., in mammals, birds, insects and fish [1–3]. In teleost fish, sexual growth dimorphism is widely encountered, and many species display different growth rates between the sexes. For some species, males grow faster than females to an adult size, such as many tilapia species [4]; silurids, such as the American and European catfishes *Ictalurus punctatus* [5] and *Silurus glanis* [6]; and salmonids, such as *Oncorhynchus nerka* [7] and *Salmo trutta* [8]. However, in most other fish species, females outgrow males. For instance, flatfish [9], sea bass [10], eel [11], cyprinids [12] and other fish [13, 14]. The mechanism of sexual growth dimorphism is complex, and numerous factors are involved in gender growth differences [15].

As a complex polygenic trait, fish growth is regulated by numerous factors, including the environment, energy metabolism, reproductive activity and nutrients [16]. Fortunately, sexually dimorphic growth can be used to explore candidate networks and genes for enhancing body size or growth speed, which may offer

rapid and significant economic gains [13]. An increasing number of growth-related genes has been identified based on the analysis of differentially expressed genes between females and males [16–18].

Pufferfishes of the genus *Takifugu* are mainly distributed in the northwestern Pacific Ocean; they are extremely similar in external morphology, and natural hybrids within the genus have frequently been found in the natural environment [19, 20]. Due to their excellent taste, abundant nutrients, and ability to serve as senior food ingredients in some cultures, these fish are important aquaculture species in East Asia [21, 22]. Although the finepatterned puffer (*Takifugu poecilonotus* Temminck & Schlegel, 1850) has not been developed as an aquaculture species, this species is commercially viable and highly traded [23]. Finepatterned puffer exhibits sexual growth dimorphism; similar to other pufferfishes; the male fish tend to be shorter and lighter than the female fish [24–26]. These biological features make *T. poecilonotus* an ideal non-model marine fish species for growth-related and sex-related gene studies. In addition, considering the fast growth rate, taste and sexual differences in growth, *T. poecilonotus* has tremendous future potential for future aquaculture and the gene breeding industry. However, genomic resources of *T. poecilonotus* are insufficient. Only 42 sequences in the National Center for Biotechnology Information (NCBI) database can be used. Therefore, identifying growth-related and sex-related genes and their functions may not only enable the characterization of candidate genes and pathways associated with the growth and sex determination of *T. poecilonotus* but may also lead to an expanded genome database for *T. poecilonotus* and support future molecular biology research.

With the rapid development of next-generation sequencing (NGS) technologies, high-throughput RNA sequencing (RNA-seq) technology is becoming readily available for non-model organisms [17, 27]. Comparative transcriptome analysis is useful for studying gene expression in many samples simultaneously [28]. This methodology is ideal for investigating the molecular mechanisms of gene or pathway responses to an experimental hypothesis [29]. Furthermore, for species whose genome is not yet available, comparing complex genomes is an attractive option to sequence the transcriptome for detecting genes. In addition, transcriptomics allows simultaneous analyses of multiple processes (metabolism, protein homeostasis and other bioprocesses) [30, 31]. In this study, muscle, liver, gonads and heart of females and males were used for transcriptome analysis. Then, the genes expression profiles of different *T. poecilonotus* individuals were generated, differentially expressed genes (DEGs) between *T. poecilonotus* males and females were detected. Subsequently, candidate important genes and pathways involved in growth-related and sex-related genes were identified. Moreover, abundant SSRs markers were detected in deep coverage sequence region reads. The present study's objective was to detect the growth-related and sex-related genes and provide a foundation for discovering sexual growth dimorphism mechanisms of teleost fishes. Furthermore, the transcriptomic resource in this study would facilitate research on the molecular regulatory mechanisms of teleost fishes.

2. Materials And Methods

2.1 *Ethics statement*

Before the experiment, we have read the policies relating to animal experiments and confirmed the present study complied. All applicable international, national, and/or institutional guidelines for the care and use of animals were followed. In accordance to the directive of Animal Research: Reporting of In Vivo Experiments (ARRIVE) guidelines 2.0, we have designed this experiment. The methods involving animals in this study were conducted in accordance with the ARRIVE guidelines 2.0. Meanwhile, our study did not require specific authorization because our observations did not cause any pain, suffering, distress, or lasting harm. In addition, all the experimental procedures in the present study were approved by the ethics committee of Laboratory Animal Welfare and Ethics of South China Sea Fisheries Research Institute (project identification code: nhdf 2021-05, date of approval: 15th Jan. 2021).

2.2 *Sample collection*

Three males and three females *T. poecilonotus* were sampled from the Beibu Gulf of China. Then, the six specimens were temporarily acclimated in an aerated seawater for three days. After eliminating confounding effects, all samples were anaesthetized and killed by severing the spinal cord. Then, heart, liver, gonad, and muscle from each *T. poecilonotus* specimen were collected and frozen in liquid nitrogen for total RNA extraction.

2.3 *RNA extraction, cDNA library building, and sequencing*

Total RNA was extracted from each tissue of the six samples using TRIzol reagent. RNA integrity was assessed by using an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA), and a sample with an RNA integrity number (RIN) ≥ 7 was subjected to subsequent analysis [32]. The RNA of tissues from every individual was pooled in equal amounts. Six cDNA libraries were constructed using 3 μg of RNA from each sample via a conventional protocol. Then, the prepared cDNA libraries were sequenced on a BGISEQ-500 platform (BGI, Shenzhen, China).

2.4 *De novo assembly and functional annotation*

Raw read quality control was performed using SOAPnuke v 0.36. Trimmomatic v0.35 [33] was used for raw read trimming. Clean data were obtained after removing adaptors and low-quality reads. Subsequently, the Trinity software package (version: 2.0.6) [34] was applied to assemble the high-quality clean reads *de novo* (-min_kmer_cov: 3). During this period, the k-mer value was set to 25. TGICL clustering software (version 2.1) [35] was then used to remove the redundancy and acquire nonredundant unigenes of maximum length with a minimum of 70% similarity. To assess the completeness of the assembly and annotation, BUSCO (Benchmarking Universal Single-Copy Orthologs) was then used to evaluate transcriptome completeness. BUSCO detects the presence of single-copy orthologues universal to all *Actinopterygii* (ray-finned fishes) as an overall measure of transcriptome quality. The orthologue dataset was provided by BUSCO, and BUSCO v3, Augustus v3.2.3, HMMER v3.1b2 and BLAST + v2.6.0 were also used [36].

All unigenes were then annotated by comparison using different public databases, including the NCBI nonredundant protein (NR) database, NCBI nucleotide database (NT), Swiss-Prot protein database, Gene

Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) database and the eukaryotic orthologous group (KOG) database (BLASTx 2.5.0 with E-value threshold of 1×10^{-5}).

2.5 *Analysis of potential growth-related genes*

To characterize gene expression variation between female samples and male samples, the *de novo* assembled transcriptome served as a reference for read mapping. Clean reads were aligned to the references with Bowtie2 (with end-to-end alignment, 22 base seed length, two reseed attempts with no mismatches allowed in the seed, a seed interval of 17 bases, and all other settings at default values) [37]. To determine fragments per kilobase of exon model per million reads mapped (FPKM), RSEM was applied to normalize the expression levels of all the unigenes [38]. To identify differentially expressed genes (DEGs), the DEG-seq package was used [39]. The thresholds of differentially expressed genes were defined as $|\log_2$ old change$| \geq 2$ and a Q-value ≤ 0.01. Among the DEGs, unigenes specifically expressed in female or male crimson seabream were regarded as specifically expressed genes (SEGs). Furthermore, enrichment analyses (KEGG and GO) of the DEGs were performed based on the hypergeometric distribution test. A significance test was applied, and a P value ≤ 0.05 indicated significantly enriched terms.

2.6 *Potential simple sequence repeat (SSR) marker detection*

The MIcroSAtellite identification tool (MISA, version 1.0) was applied to identify the simple sequence repeats (SSRs) in the assembled transcript reference. The threshold of the minimum repeat number for various unit types was twelve for mononucleotide microsatellites, six for dinucleotides, five for trinucleotide and tetranucleotide microsatellites, and four for pentanucleotide and hexanucleotide microsatellites. Then, SSR-containing unigenes were used as candidates to design primers on flanking sequences of the repeats with Primer 3 software [40].

2.7 *Validation of the transcriptome data by quantitative real-time PCR*

To validate the confidence of the high-throughput transcriptome sequencing, 10 differentially expressed genes were selected and analysed via quantitative real-time PCR (qPCR). All the selected genes exhibited significantly different expression between the male and female specimens, including five female-biased genes and five male-biased genes. Primer design was carried out using Primer Premier 6.0 software (PREMIER Biosoft International, Palo Alto, Calif). In addition, 18S and β-actin were chosen as reference genes for internal standardization, and the PCR procedure and amplification reaction system were the same as those described by Lou [17]. ABI 7300 SDS software (Applied Biosystems) was used to analyse the data after the PCR programme was complete. We calculated the relative expression levels of ten target unigenes based on the $2^{-\Delta\Delta Ct}$ method ($\Delta CT = CT_{\text{target unigene}} - CT_{\text{reference gene}}$, $-\Delta\Delta CT = \Delta CT_{\text{female}} - \Delta CT_{\text{male}}$), and the $\log_2$ fold change was then used for comparison with the $\log_2$ fold change of RNA-seq.

3. Results

3.1 *Transcriptome sequencing and de novo assembly*

In the present study, six cDNA libraries were constructed and subsequently sequenced on the BGISEQ-500 platform. In total, 285.95 million raw read pairs were generated, and after preprocessing the raw data, 264.02 million clean paired-end sequence reads with a Q30 percentage of 86.46% were obtained. The statistical data of six *T. poecilonotus* cDNA libraries are shown in Table 1. Transcriptome assembly was completed after clustering by removing redundancy with Trinity and TGICL. A total of 149,814 unigenes were identified with an N50 length of 3,538 bp and a mean length of 2,066 bp (Table 2). The length distribution of the unigenes revealed that 23,132 (15.44%) were between 200 and 300 bp, and 38,528 (25.72%) were 3,000 bp and longer (Figure S1). The assembly assay indicated that the length distribution pattern and mean length of the unigenes were similar to those in previous fish transcriptome studies [16, 18].

Table 1
Summary statistics of clean transcriptome sequencing data from each sample.

Sample	Raw Reads (M)	Clean Reads (M)	Clean Bases (Gb)	Clean Reads Q30 (%)	Clean Reads Ratio (%)
Female-1	52.59	48.33	7.25	87.49	91.90
Female-2	44.04	40.55	6.08	86.17	92.06
Female-3	47.33	44.00	6.60	86.04	92.98
Male-1	47.33	43.93	6.59	86.37	92.81
Male-2	47.33	43.63	6.54	86.48	92.19
Male-3	47.33	43.58	6.54	86.18	92.09

Table 2
Statistics for the assembled unigenes.

Sample	Total Number	Total Length	Mean Length	N50	GC (%)
Female-1	81,249	145,145,039	1,786	3,128	48.91
Female-2	72,625	118,111,439	1,626	2,823	49.00
Female-3	72,194	120,065,634	1,663	2,836	49.06
Male-1	93,484	144,668,988	1,547	2,725	48.62
Male-2	95,350	142,802,517	1,497	2,627	48.55
Male-3	84,292	131,779,534	1,563	2,735	48.81
All-Unigene	149,814	309,566,984	2,066	3,538	48.64

Since a reference genome is unavailable for *T. poecilonotus*, to assist in transcriptome reconstruction, an examination of the completeness of the reference assembly was performed by comparing it to the Benchmarking Universal Single-Copy Orthologues gene database. The results showed that 99.67% of the genome (30.36% as single genes and 69.31% as duplicated genes) was complete, and 0.03% of the genome was missing, suggesting that the *de novo* assembly was relatively successful (Figure S2).

All raw reads in the present study are archived in the National Center for Biotechnology Information (NCBI) Short Read Archive (SRA) databases under BioProject PRJNA683736, with accession numbers SRR13236436 - SRR13236441. This Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GIXS00000000. The version described in this paper is the first version, GIXS01000000.

3.2 *Functional annotation of the T. poecilonotus transcriptome*

All the unigenes were functionally characterized based on the description of their similar sequences as obtained from databases. The results showed that 122,719 unigenes matched to various databases (Table 3). From the annotated information of six databases, 63,250 unigenes had significant hits in the NR database, and most unigenes were matched with genes from *Takifugu rubripes* (74.38%) (Figure S3). In summary, 110,737 unigenes (73.92%) were annotated with at least one of the five protein databases used in this study, while 52,989 unigenes (35.37%) were annotated with all five protein databases that were searched in this study, namely, the KEGG, KOG, Swiss-Prot, NR and GO databases (Fig. 1).

Table 3
Summary of unigene annotations.

Database	Number of unigenes	Percentage (%)
Total	122,719	81.91%
NT	70,243	46.89%
NR	63,250	42.22%
SwissProt	67,197	44.85%
KEGG	100,035	66.77%
KOG	90,961	60.72%
GO	63,518	42.40%

A total of 63,518 unigenes were annotated to GO categories: biological process, cellular component and molecular function. The terms ‘binding’ (31,409 unigenes), ‘catalytic activity’ (24,082 unigenes) and ‘cellular process’ (14,736 unigenes) were dominant in these GO categories, respectively (Figure S4). In addition, 90,961 unigenes were annotated with the KOG database and classified into 25 subcategories, and ‘general function prediction only’ (15,183 unigenes) and ‘signal transduction mechanisms’ (14,975 unigenes) were the top two subcategories (Fig. 2). In total, 100,035 unigenes were assigned to pathways based on the KEGG pathway analysis (Fig. 3); the top three pathways were ‘human diseases’ (51,859 unigenes), ‘organism system’ (40,564 unigenes) and ‘metabolism’ (29,797 unigenes). Furthermore, the predominant pathway subcategories were ‘signal transduction’ (16,907 unigenes), ‘global and overview maps’ (11,183 unigenes), ‘immune system’ (10,598 unigenes) and ‘cancers: overview’ (10,177 unigenes). These annotations offer a valuable resource for investigating specific functions, pathways and processes in teleost research.

3.3 *Potential growth-related genes and functional enrichment analysis results*

To analyse the potential growth-related genes, unigenes meeting the differentially expression criteria, $|\log_2 \text{fold change}| \geq 2$ and Q-values < 0.01, were determined to be DEGs between female *T. poecilonotus* and male *T. poecilonotus*. After filtering, 10,385 DEGs were obtained between the two sex samples; these genes might be associated with sex determination or be growth-related genes (Fig. 4). Among these genes, 8,351 unigenes were male-biased DEGs, and 2,034 unigenes were female-biased DEGs. In addition, sex-specific genes were detected, including 4,416 male SEGs and 599 female SEGs.

All the DEGs between the two sexes were used as input to perform KEGG and GO enrichment analyses. Consequently, 3,942 unigenes were assigned to 2,146 GO terms. The top 20 statistically significant KEGG classifications of the DEGs are shown in Fig. 5. The results suggested that the three most-enriched KEGG classifications were ‘neuroactive ligand-receptor interaction’ (ko04080), ‘PI3K-Akt signalling pathway’

(ko04151) and ‘cellular senescence’ (ko04218). The results in GO terms showed that sex-biased genes were predominantly associated with ‘nucleus’ (GO:0005634), ‘ATP binding’ (GO:0005524), ‘integral component of membrane’ (GO:0016021) and other terms (Table S2).

The results of the sequence annotation showed that numerous differentially expressed genes were well-known as sex control and gonadal development genes, including the zona pellucida sperm-binding protein gene (*Zps*), sperm acrosome membrane-associated protein gene (*Spacas)*, cytochrome P450 aromatase (*Cyp19a*), androgen-induced gene 1 protein, StAR-related lipid transfer protein gene (*Start*), double-sex and mab-3-related transcription gene (*Dmrt*) and other potential candidate protein-encoding genes (Table 4). Growth is a complex trait that is controlled by multiple genes. After annotation, a variety of DEGs involved in regulating growth were identified from the sequence database, including genes in somatotropic axis growth, controlling growth at the muscle tissue level and other genes related to growth (Table 5).

Table 4
Genes related to the sex determination of *Takifugu poecilonotus*.

Gene	Unigene	Annotation	log_2(⍰/⍰)	Q-value
Dmtr1	Unigene1322	Doublesex and mab-3-related transcription factor 1	6.66	2.34E-07
Sox3	Unigene1791	SRY-box containing transcription factor 3	5.72	8.44E-04
Dnali1	Unigene12399	Axonemal dynein light intermediate polypeptide 1	7.63	3.40E-07
Ropn1l	CL4645.Contig3	ropporin-1-like protein	9.37	1.32E-18
Cyp19a	CL5324.Contig8	Cytochrome P450 aromatase	-7.54	2.15E-07
Zp1	CL6076.Contig3	zona pellucida sperm-binding protein 1	-12.66	1.43E-110
Zp2	CL8003.Contig1	zona pellucida sperm-binding protein 2	-7.70	1.84E-20
Zp3	CL7875.Contig2	zona pellucida sperm-binding protein 3	-7.93	4.04E-16
Start-7	CL3580.Contig10	stAR-related lipid transfer protein 7	7.57	2.66E-06
Gdf3	CL9083.Contig1	growth differentiation factor 3	-8.90	3.99E-14
Gdf9	CL6347.Contig1	growth differentiation factor 9	-9.23	1.61E-12
Gtf3a	Unigene10317	Transcription factor IIIA-like	-2.42	2.11E-08
42sp43	Unigene8014	P43 5S RNA-binding protein-like	-7.38	6.57E-24
Mnd1	Unigene817	meiotic nuclear division protein 1 homolog	3.25	8.39E-03
Star	Unigene31813	steroidogenic acute regulatory protein	-2.49	3.42E-04
Wee1	CL12054.Contig1	wee1-like protein kinase	2.05	6.42E-03
Spaca6	CL4045.Contig1	sperm acrosome membrane-associated protein 6	5.97	4.05E-04
Spaca4	CL12530.Contig2	sperm acrosome membrane-associated protein 4	3.53	2.18E-15
Spata5	CL8799.Contig2	spermatogenesis-associated protein5	6.47	7.85E-05
Spata7	CL4586.Contig1	spermatogenesis-associated protein 7	2.93	1.40E-03
Spata17	CL4320.Contig15	spermatogenesis-associated protein 17	8.17	2.09E-08
Spata32	Unigene13081	spermatogenesis-associated protein 32	4.78	1.72E-06

Table 5
Growth related genes in *Takifugu poecilonotus*.

Gene	Unigene	Annotation	log_2(□/□)	Q-value
Igfbp1	CL6482.Contig4	insulin-like growth factor binding protein 1	-8.52	2.51E-09
Igfbp3	Unigene7287	Insulin-like growth factor binding protein 3	-3.36	9.25E-06
Igf1r	Unigene24558	Insulin-like growth factor 1 receptor	-1.06	1.75E-02
Ghr	CL9297.Contig1	Growth hormone receptor	-4.79	3.10E-03
Sstr2	Unigene6248	Somatostatin receptor 2	7.45	5.93E-07
Fabp1	CL7463.Contig1	fatty acid-binding protein 1	3.85	8.11E-12
Fabp4	CL10654.Contig1	fatty acid-binding protein 4	-8.28	6.17E-56
Fabp6	CL6561.Contig1	fatty acid-binding protein 6	6.70	2.41E-04
Fabp7	Unigene40265	fatty acid-binding protein 7	5.45	2.86E-05
Grb1	CL5906.Contig1	Growth factor receptor-bound proteins 1	2.16	6.38E-03
Grb7	Unigene6338	Growth factor receptor-bound proteins 7	-2.98	8.15E-03
Grb10	CL11116.Contig1	Growth factor receptor-bound proteins 10	3.07	1.43E-04
Htr1	Unigene31616	5-hydroxytryptamine receptor 1D-like	4.59	1.31E-04
Htr2	Unigene28993	5-hydroxytryptamine receptor 2C-like	5.27	3.86E-03
Htr3	CL1712.Contig2	5-hydroxytryptamine receptor 3	5.61	1.44E-03
Htr6	Unigene18275_	5-hydroxytryptamine receptor 6	4.33	7.54E-03
Prlr	CL1256.Contig1	Prolactin receptor	-8.19	3.29E-11
Prl	CL12326.Contig4	Prolactin	-5.65	4.17E-06
Myhc	Unigene12897	myosin heavy chain	6.29	5.15E-05

3.4 *Discovery of molecular markers*

In the present study, we detected a large number of transcripts within which 68,281 potential SSRs were discovered. Dinucleotide SSRs represented the largest fraction (30,255 SSRs, 44.31%), followed by mononucleotide (20,211 SSRs, 29.60%) and trinucleotide (13,570 SSRs, 19.87%), as shown in Fig. 6. Only a small fraction of quad-nucleotide (2,382), penta-nucleotide (924) and hexa-nucleotide (936) SSRs were identified. Furthermore, SSRs with six tandem repeats were the most common, followed by SSRs with five, twelve, seven, thirteen, fourteen and eight tandem repeats, whereas the remaining tandem repeats each accounted for fewer than 5% of the SSRs.

3.5 *Transcriptome data validation*

We validated the expression patterns of 10 DEGs using qRT-PCR. Of these genes, 5 genes were predicted to be overexpressed in female specimens, and 5 genes were predicted to be overexpressed in male specimens. The specific primers for each gene are shown in Table S1. Similar up- and downregulation patterns of these genes were observed in the qRT-PCR and RNA-seq results (Fig. 7). Thus, the high correlation between qPCR and RNA-seq data indicated the reliability and accuracy of the transcriptome expression analysis.

4. Discussion

As a valuable commercial fish, *Takifugu poecilonotus* has high exploitability and aquaculture prospects. However, there were few studies regarding growth-related and sex-related genes of *T. poecilonotus.* In order to explore the mechanisms of sexual growth dimorphism and sex determination, we identified sex-related and growth-related genes and their biological pathways. It is first time that the transcriptome data of finepatterned puffer (*Takifugu poecilonotus*) has been reported. The result of annotation revealed that numerous sequences of finepatterned puffer could be annotated to well-known genes and biological pathways regarding growth and sex. In the present study, 149,814 unigenes were sequenced and assembled. In total, 110,737 (73.92%) unigenes were significantly matched to protein databases. In the present study, 10,385 DEGs were identified between different sexes of *T. poecilonotus*. Some of these DEGs play important roles in sex growth and determination. Our results may provide fundamental resources for further research on the molecular mechanisms of biological processes of *T. poecilonotus*.

In the present study, numerous growth- and sex-related genes were detected from the transcriptome data of *T. poecilonotus*. Obviously, it is widely accepted that genes encoding components of the somatotropic axis play critical roles in regulating the formation of skeletal muscles in finfish, including the growth hormone gene (*Gh*), insulin-like growth factors (*Igf*), somatostatin and their carrier proteins and receptors [41]. Jia et al. [42,43] showed that specific growth rate, body weight and plasma growth hormone of female tiger puffers (*Takifugu rubripes*) reared in an offshore sea cage aquaculture system (OSCS) were significantly higher than those of fish reared in other aquaculture systems. The mRNA expression results indicated that female individuals reared in the OSCS showed higher somatic growth axis-related gene (*Igf1, Igf2, Igf1r, Igf2r, Ghr1,* and *Ghr2*) expression levels. In the present study, compared to male *T. poecilonotus*, female *T. poecilonotus* showed significantly higher *Igf1r, Igfbp1, Igfbp3* and *Ghr* expression levels.

Fatty acid-binding proteins (FABPs) belong to the protein superfamily of lipid-binding proteins [44]. They are 126-137 amino acids in length depending on species, with an average molecular mass of 14-16 kDa [45-47]. Fabps have been well studied for decades [44,48]. The first teleost FABPs were identified in the hearts of sea ravens (*Hemitripterus americanus*) and ocean pouts (*Macrozoarces americanus*) [49]. Subsequently, teleost FABPs/*Fabps* were identified in different tissues of many teleosts, including striped bass (*Morone saxatilis*) [50], rainbow trout (*Oncorhynchus mykiss*) [51], Atlantic salmon (*Salmo salar* L.)

[52]. Fatty acid-binding protein genes are involved in many processes of cell physiology, including development, growth, and cell differentiation [53]. In many respects, *Fabps* participates in the binding, sequestration and metabolism of long-chain fatty acids, eicosanoids, bile salts, and other hydrophobic ligands [54-56]. *Fabps* may also be possible carriers of certain hydrophobic reactants in their passage from the cytosol to chromatin, and thus, they may have a direct or indirect effect on cell growth [57]. Previous studies identified and characterized the *Fabps* genes of pufferfish (*Tetraodon nigroviridis*), including *Fabp*1-3, *Fabp*6, *Fabp*7, *Fabp*10, and *Fabp*11 [58-59]. In the present study, compared to female *T. poecilonotus*, male *T. poecilonotus* exhibited higher *Fabp*1, *Fabp*6 and *Fabp*7 expression levels, showing that these sex-specific genes may play critical roles in reproduction and growth.

More importantly, we also found *Fabp*4 in the transcriptome of *T. poecilonotus*. This is the first time that the expression of *Fabp*4 has been found in fish. By binding with long-chain fatty acids, *Fabp*4 can affect the uptake, transportation, esterification, and b-oxidation of fatty acids and regulate the energy balance and lipid signalling within cells [60,61]. Furthermore, *fabp4* has also been found to be associated with growth, fat deposition and carcass traits in mammals. It was reported that *Fabp4* is a potential candidate gene for obesity, as it is located within a quantitative trait locus (QTL) region for serum leptin levels in mice [62]. In addition, FABP4 protein content may be a marker of intramuscular fat accretion in the longissimus thoracis muscle in pigs [63,64]. Obviously, *Fabp4* has become a strong candidate gene for fat metabolism. However, the biological function of the *Fabp4* gene in fish is unknown, and further work is required to fully characterize the *Fabp4* gene in fish.

Prolactin (PRL) plays critical role in multiple biological function by binding to its receptor (PRLR) in fish [65]. In adult fish, the major action of PRL is freshwater osmoregulation. In addition, it has also been reported that PRL associated with reproduction, behaviour, growth, and immunoregulation [65-67]. Prolactin and PRLR are also present in embryos and exhibit widespread tissue distribution in fish larvae. Yang et al. [68] identified transcripts of PRL and PRLR in the early embryo of rainbow trout (*O. mykiss*) and indicated PRL and PRLR was associated with the post-hatching development of larvae. In early development and metamorphosis of amphibians and mammals, the role of PRL has been well understood. However, the role of PRL in fish is not clearly established. The potential roles of PRL in fish embryos and larvae are considered in relation to their physiological status, and the spectrum of activities differs by species (Power, 2005). Relatively few reports support somatotropic action for PRL in fish. Some researchers indicated that PRL influences the growth of Mozambique tilapia by stimulating liver IGF-I production [69]; however, the way in which PRL stimulates IGF-I production is unclear. Hence, further work is required to fully characterize the biological function of PRL and PRLR in fish.

Sex determination is the process of establishing individual gender and regulating the differentiation of sex characteristics [70,71]. In contrast to other vertebrates (birds or humans), the sex determination of teleost fishes is diverse [72-74]. According to the annotations, many sex-related genes have been identified in the *T. poecilonotus* transcriptome. As a sex-determining gene, doublesex and mab-3-related transcription factor 1 (*Dmrt1*) is highly conserved and has been identified in many teleost fish [72,75,76]. The *dmrt1* gene plays an important role in the testis differentiation and maintenance of male-specified

germ cells by encoding a transcription factor [77]. In the present study, *Dmrt1* showed higher expression levels in male *T. poecilonotus* than in female *T. poecilonotus*, indicating that *Dmrt1* may be crucial for the development and maturation of gonads in male *T. poecilonotus*.

As extracellular matrix surrounding oocytes, *zona pellucida* plays a protective role in fish oocytes, and it is important in sperm binding [78]. *Zona pellucida* comprises four kinds of glycoproteins (ZP1-4), they are incorporated into long filaments [79,80]. The zona pellucida proteins were originally detected in the egg envelope of mammalian, and have also been found in the inner layer of the fish chorion [81]. Previous study has shown that *Zp2* plays an important role in the early formation of oocyte envelope, and *Zp3* could be treated as a major class of female-specific reproductive molecules [82]. In the present study, *Zp1*, *Zp2*, and *Zp3* showed higher expression levels in *T. poecilonotus* females than that in males, suggesting that these genes may also play critical roles in folliculogenesis and reproduction.

Declarations

Ethics approval

All the experimental procedures were approved by the ethics committee of Laboratory Animal Welfare and Ethics of South China Sea Fisheries Research Institute. The procedures involving animals in this study were conducted in accordance with the Laboratory Animal Management Principles of China. In the present, all samples were anaesthetized and killed by severing the spinal cord.

Consent for publication

Not applicable.

Availability of data and materials

All raw reads in the present study are archived in the National Center for Biotechnology Information (NCBI) Short Read Archive (SRA) databases under BioProject PRJNA683736, with accession numbers SRR13236436 - SRR13236441. This Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GIXS00000000. The version described in this paper is the first version, GIXS01000000.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded by the Science and Technology Project of Guangdong Province (2019B121201001); the Central Public-Interest Scientific Institution Basal Research Fund, South China Sea Fisheries Research Institute, CAFS (No.2020TD01) and the China-ASEAN Maritime Cooperation Fund (China-Vietnam Fishery Stock Enhancement and Conservation in Beibu Gulf).

Authors' contributions

BS: Conceptualization, Data curation, Methodology, Investigation, Formal analysis, Funding acquisition, Writing-Original draft preparation. YL: Investigation, Methodology, Data curation. CY: Writing-review & editing, Resources. LW: Methodology, Visualization. GZ: Investigation. DS: Supervision, Resources, Funding acquisition, Writing- review & editing.

Acknowledgments

The authors would like to thank Qijian Xie and Teng Li for sample collection.

References

1. L. Isaac, Potential causes and life-history consequences of sexual size dimorphism in mammals, Mammal Rev. 35 (1) (2005) 101-115.
2. J. Loonstra, M.A. Verhoeven, T. Piersma, Sex-specific growth in chicks of the sexually dimorphic Black-tailed Godwit, Ibis. 160 (1) (2018) 89-100.
3. Pauly, Female fish grow bigger-let's deal with it, Trends Ecol. Evol. 34 (3) (2019) 181-182.
4. Toguyeni, B. Fauconneau, T. Boujard, Fostier, E.R. Kuhn, K.A. Mol, J. Baroiller, Feeding behaviour and food utilisation in tilapia, *Oreochromis niloticus*: effect of sex ratio and relationship with the endocrine status, Physiol. Behav. 62 (2) (1997) 273-279.
5. A. Simco, C.A. Goudie, G.T. Klar, N.C. Parker, K.B. Davies, Influence of sex on growth of channel catfish, Trans. Am. Fish. Soc. 118 (1989) 427-434.
6. Haffray, C. Vauchez, M. Vandeputte, O. Linhart, Different growth and processing traits in males and females of European catfish, *Silurus glanis*, Aquat. Living Resour. 11 (5) (1998) 341-345.
7. P. Quinn, C.J. Foote, The effects of body size and sexual dimorphism on the reproductive behavior of sockeye salmon, *Oncorhynchus nerka*, Anim. Behav. 48 (4) (1994) 751-761.
8. Bonnet, P. Haffray, J.M. Blanc, F. Vallée, C. Vauchez, A. Fauré, B. Fauconneau, Genetic variation in growth parameters until commercial size in diploid and triploid freshwater rainbow trout (*Oncorhynchus mykiss*) and seawater brown trout (*Salmo trutta*), Aquaculture 173 (1-4) (1999) 359-375.
9. Imsland, A. Folkvord, G.L. Grung, S.O. Stefansson, G.L. Taranger, Sexual dimorphism in growth and maturation of turbot, *Scophthalmus maximus* (Rafinesque, 1810), Aquac. Res. 28 (2) (1997) 101-114.
10. Saillant, A. Fostier, B. Menu, P. Haffray, B. Chatain, Sexual growth dimorphism in sea bass *Dicentrarchus labrax*, Aquaculture 202 (3-4) (2001) 371-387.
11. Roncarati, P. Melotti, O. Mordenti, L. Gennari, Influence of stocking density of European eel (*Anguilla anguilla*, L) elvers on sex differentiation and zootechnical performances, J. Appl. Ichtyol. 13 (3) (1997) 131-136.

12. M. Tyus, Ecology and management of Colorado squawfish, in: W.L. Minckley (Eds.), Battle Against Extinction: Native Fish Management in the American West, Tucson, 1993, pp. 379-402.
13. Dutney, A. Elizur, P. Lee, Analysis of sexually dimorphic growth in captive reared cobia (*Rachycentron canadum*) and the occurrence of intersex individuals, Aquaculture 468 (2017) 348-355.
14. Pongthana, D.J. Penman, P. Baoprasertkul, M.G. Hussain, M.S. Islam, S.F. Powell, B.J. McAndrew, Monosex female production in the silver barb (*Puntius gonionotus* Bleeker), Aquaculture 173 (1999) 247-256.
15. D. Rennie, C.F. Purchase, N. Lester, N.C. Collins, B.J. Shuter, P.A. Abrams, Lazy males? Bioenergetic differences in energy acquisition and metabolism help to explain sexual size dimorphism in percids, J. Anim. Ecol. 77 (5) (2008) 916-926.
16. Wang, R. Wang, R. Wang, S. Chen, Transcriptomics analysis revealing candidate networks and genes for the body size sexual dimorphism of Chinese tongue sole (*Cynoglossus semilaevis*), Funct. Integr. Genomics 18 (3) (2018) 327-339.
17. Lou, T. Yang, Z. Han, T. Gao, Transcriptome analysis for identification of candidate genes related to sex determination and growth in *Charybdis japonica*, Gene 677 (2018) 10-16.
18. Ma, A. Ma, Z. Huang, G. Wang, T. Wang, D. Xia, B. Ma, 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) (2016) e0149414.
19. Masuda, N. Shinohara, Y. Takahashi, O. Tabeta, K. Matsuura, Occurrenc of natural hybrid between pufferfishes, *Takifugu xanthopterus* and *T. vermicularis*, in Ariake Bay, Kyushu, Japan, Nippon Suisan Gakkaishi 57 (7) (1991) 1247-1255.
20. Miyaki, O. Tabeta, H. Kayano, Karyotypes in six species of pufferfishes genus *Takifugu* (Tetraodontidae, Tetraodontiformes), Fish. Sci. 61 (4) (1995) 594-598.
21. Kikuchi, N. Iwata, T. Furuta, T. Kawabata, T. Yanagawa, Growth of tiger puffer *Takifugu rubripes* in closed recirculating culture system, Fish. Sci. 72 (5) (2006) 1042-1047.
22. L. Wang, H.T. Zhang, Y.Q. Ren, Q. Zhou, Comparison of growth parameters of tiger puffer *Takifugu rubripes* from two culture systems in China, Aquaculture 453 (2016) 49-53.
23. Masuda, K. Amaoka, C.K. Muzik, T.T. Uyeno, T. Yoshimo, The fishes of the Japanese Archipelago, Tokai University Press, 1984, pp. 437.
24. Yang, Y. Chen, Length-weight relationship of obscure puffer (*Takifugu obscurus*) during spawning migration in the Yangtze River, China, J. Freshwater Ecol. 18 (3) (2003) 349-352.
25. Ueda, J. Sano, H. Uchida, C. Amano, Y. Matsumura, T. Katayama, Growth and age-length key of the tiger puffer *Takifugu rubripes* in the East China Sea, Sea of Japan, and Seto Inland Sea, Japan, Nippon Suisan Gakkaishi 76 (5) (2010) 803-811.
26. Zhou, Z. Zhuang, R. Zhang, Temperature-control-induced masculinization in tiger puffer *Takifugu rubripes*, J. Oceanol. Limnol. 37 (3) (2019) 1125-1135.

27. Chatchaiphan, P. Srisapoome, J.H. Kim, R.H. Devlin, U. Na-Nakorn, *De novo* transcriptome characterization and growth-related gene expression profiling of diploid and triploid bighead catfish (*Clarias macrocephalus* Günther 1864), Mar. Biotechnol. 19 (1) (2017) 36-48.
28. Morozova, M. Hirst, M.A. Marra, Applications of new sequencing technologies for transcriptome analysis, Annu. Rev. Genom. Hum. G. 10 (2009) 135-151.
29. B. Wolf, Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial, Mol. Ecol. Resour. 13 (4) (2013) 559-572.
30. Casas, F. Saborido-Rey, T. Ryu, C. Michell, T. Ravasi, X. Irigoien, Sex change in clownfish: molecular insights from transcriptome analysis, Sci Rep. 6 (2016) 35461.
31. Yang, M. Ikhwanuddin, X. Li, F. Lin, Q. Wu, Y. Zhang, C. You, W. Liu, Y. Cheng, X. Shi, S. Wang, H. Ma, Comparative transcriptome analysis provides insights into differentially expressed genes and long non-coding RNAs between ovary and testis of the mud crab (*Scylla paramamosain*), Mar. Biotechnol. 20 (2018) 20-34.
32. Schroeder, O. Mueller, S. Stocker, R. Salowsky, M. Leiber, M. Gassmann, S. Lightfoot, W. Menzel, M. Granzow, T. Ragg, The RIN: an RNA integrity number for assigning integrity values to RNA measurements, BMC Mol. Biol. 7 (1) (2006) 1-14.
33. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (15) (2014) 2114-2120.
34. G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data, Nat. Biotechnol. 29 (7) (2011) 644-652.
35. Pertea, X. Huang, F. Liang, V. Antonescu, R. Sultana, S. Karamycheva, Y. Lee, J. White, F. Cheung, B. Parvizi, J. Tsai, J. Quackenbush, TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets, Bioinformatics 19 (5) (2003) 651-652.
36. A. Simão, R.M. Waterhouse, P. Ioannidis, E.V. Kriventseva, E.M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, Bioinformatics 31 (2015) 3210-3212.
37. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods 9 (4) (2012) 357.
38. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, BMC Bioinformatics 12 (1) (2011) 323.
39. Wang, Z. Feng, X. Wang, X. Wang, X. Zhang, DEGseq: an R package for identifying differentially expressed genes from RNA-seq data, Bioinformatics 26 (1) (2010) 136-138.
40. Untergasser, I. Cutcutache, T. Koressaar, J. Ye, B.C. Faircloth, M. Remm, S.G. Rozen, Primer3-new capabilities and interfaces, Nucleic Acids Res. 40 (15) (2012) e115-e115.
41. De-Santis, D.R. Jerry, Candidate growth genes in finfish-Where should we be looking? Aquaculture 272 (2007) 22-38.

42. Jia, Q. Jing, Z. Xing, X. Gao, J. Zhai, C. Guan, B. Huang, Effects of two different culture systems on the growth performance and physiological metabolism of tiger pufferfish (*Takifugu rubripes*), Aquaculture 495 (2018) 267-272.
43. Jia, Q. Jing, J. Zhai, C. Guan, B. Huang, Alternations in oxidative stress, apoptosis, and innate-immune gene expression at mRNA levels in subadult tiger puffer (*Takifugu rubripes*) under two different rearing systems, Fish Shellfish Immunol. 92 (2019) 756-764.
44. Hanhoff, C. Lucke, F. Spener, Insights into binding of fatty acids by fatty-acid binding proteins, Mol. Cell. Biochem. 239 (2002) 45-54.
45. Kanda, S. Iseki, M. Hitomi, H. Kimura, S. Odani, H. Kondo, Y. Matsubara, T. Muto, T. Ono, Purification and characterization of a fatty-acid-binding protein from the gastric mucosa of rats: Possible identity with heart fatty-acid-binding protein and its parietal cell localization, Eur. J. Biochem. 185 (1) (1989) 27-33.
46. Börchers, F. Spener, Fatty acid binding proteins, Curr. Top. Membr. 40 (1994) 261-294.
47. M. Pelsers, W.T. Hermens, J.F. Glatz, Fatty acid-binding proteins as plasma markers of tissue injury, Clin. Chim. Acta 352 (2005) 15-35.
48. K. Ockner, J.A. Manning, R.B. Poppenhausen, W.K. Ho, A binding protein for fatty acids in cytosol of intestinal mucosa, liver, myocardium, and other tissues, Science 177 (1972) 56-58.
49. M. Stewart, W.R. Driedzic, Fatty acid binding proteins in teleost fish, Can. J. Zool. 66 (12) (1988) 2671-2675.
50. L. Londraville, B.L. Sidell, Cold acclimation increases fatty acidbinding protein concentration in aerobic muscle of striped bass, *Morone saxatilis*, J. Exp. Zool. 275 (1) (1996) 36-44.
51. Ando, X.H. Xue, G.F. Tibbits, N.H. Haunerland, Cloning and sequencing of complementary DNA for fatty acid binding protein from rainbow trout heart, Comp. Biochem. Physiol. B 119 (1) (1998) 213-217.
52. E.O. Jordal, I. Hordvik, M. Pelsers, D.A. Bernlohr, B.E. Torstensen, FABP3 and FABP10 in Atlantic salmon (*Salmo salar* L.)-General effects of dietary fatty acid composition and life cycle variations, Comp. Biochem. Physiol. B 145 (2) (2006) 147-158.
53. Lücke, L.H. Gutiérrez-González, J.A. Hamilton, Intracellular lipid binding proteins: evolution, structure, and ligand binding, in: A.S. Duttaroy (Eds.), Cellular proteins and their fatty acids in health and disease, 2003, pp. 95-118.
54. M. Bass, The cellular fatty acid binding proteins: Aspects of structure, regulation, and function, Int. Rev. Cytol. 3 (1988) 143-184.
55. Binas, H. Danneberg, J. McWhir, L. Mullins, A.J. Clark, Requirement for the heart-type fatty acid binding protein in cardiac fatty acid utilization, FASEB J. 13 (8) (1999) 805-812.
56. J. Leaver, E. Boukouvala, E. Antonopoulou, A. Diez, L. Favre-Krey, M.T. Ezaz, J.M. Bautista, D.R. Tocher, G. Krey, Three peroxisome proliferators activated receptor isotypes from each of two species of marine fish, Endocrinology 146 (7) (2005) 3150-3162.

57. Wolfrum, T. Borchers, J.C. Sacchettini, F. Spener, Binding of fatty acids and peroxisome proliferators to orthologous fatty acid binding proteins from human, murine, and bovine liver, Biochemistry 39 (6) (2000) 1469-1474.
58. B. Parmar, J.M. Wright, Comparative genomic organisation and tissue-specific transcription of the duplicated fabp7 and fabp10 genes in teleost fishes, Genome 56 (11) (2013) 691-701.
59. Thirumaran, J.M. Wright, Fatty acid-binding protein (fabp) genes of spotted green pufferfish (*Tetraodon nigroviridis*): comparative genomics and spatial transcriptional regulation, Genome 57 (5) (2014) 289-301.
60. Furuhashi, G.S. Hotamisligil, Fatty acid-binding proteins: role in metabolic diseases and potential as drug targets, Nat. Rev. Drug Discov. 7 (6) (2008) 489-503.
61. S. Hotamisligil, D.A. Bernlohr, Metabolic functions of FABPs-mechanisms and therapeutic implications, Nat. Rev. Endocrinol. 11 (10) (2015) 592.
62. Ogino, D.H. Moralejo, H. Kose, T. Yamada, K. Matsumoto, Serum leptin concentration is linked to chromosomes 2 and 6 in the OLETF rat, an animal model of type 2 diabetes with mild obesity, Mamm. Genome 14 (12) (2003) 239-244.
63. Gerbens, F.J. Verburg, H.T.B. Van Moerkerk, B. Engel, W. Buist, J.H. Veerkamp, M.F.W Te Pas, Associations of heart and adipocyte fatty acid-binding protein gene expression with intramuscular fat content in pigs, J. Anim. Sci. 79 (2) (2001) 347-354.
64. Mercadé, M. Pérez-Enciso, L. Varona, E. Alves, J.L. Noguera, A. Sánchez, J.M. Folch, Adipocyte fatty-acid binding protein is closely associated to the porcine FAT_1 locus on chromosome 4, J. Anim. Sci. 84 (11) (2006) 2907-2913.
65. M. Power, Developmental ontogeny of prolactin and its receptor in fish, Gen. Comp. Endocrinol. 142 (1-2) (2005) 25-33.
66. A. Bern, Functional evolution of prolactin and growth hormone in lower vertebrates, Am. Zool. 23 (3) (1983) 663-671.
67. C. Clarke, H.A. Bern, Comparative endocrinology of prolactin, in: C.H. Li (Eds.), Hormonal Proteins and Peptides, New York, 1980, pp. 105-197.
68. Y. Yang, M. Greene, T.T. Chen, Early embryonic expression of the growth hormone family protein genes in the developing rainbow trout, *Oncorhynchus mykiss*, Mol. Reprod. Dev. 53 (2) (1999) 127-134.
69. S. Shepherd, T. Sakamoto, R.S. Nishioka, N.H. Richman III, I. Mori, S.S. Madsen, T.T. Chen, T. Hirano, H.A. Bern, E.G. Grau, Somatotropic actions of the homologous growth hormone and prolactins in the euryhaline teleost, the tilapia, *Oreochromis mossambicus*, Proc. Natl. Acad. Sci. USA 94 (5) (1997) 2068-2072.
70. Yang, Y. Wang, Z. Zhang, S. He, Comprehensive transcriptome analysis reveals accelerated genic evolution in a Tibet Fish *Gymnodiptychus pachycheilus*, Genome Biol. Evol. 7 (1) (2015) 251-261.

71. Chen, G. Xiao, X. Chai, X. Lin, J. Fang, S. Teng, Transcriptome analysis of sex-related genes in the blood clam *Tegillarca granosa*, PLoS One 12 (9) (2017) e0184584.
72. Ribas, D. Robledo, A. Gómez-Tato, A. Viñas, P. Martínez, F. Piferrer, Comprehensive transcriptomic analysis of the process of gonadal sex differentiation in the turbot (*Scophthalmus maximus*), Mol. Cell Endocrinol. 422 (2016) 132-149.
73. Kobayashi, Y. Nagahama, M. Nakamura, Diversity and plasticity of sex determination and differentiation in fishes, Sex. Dev. 7 (2013) 115-125.
74. Casas, F. Saborido-Rey, T. Ryu, C. Michell, T. Ravasi, X. Irigoien, Sex change in clownfish: molecular insights from transcriptome analysis, Sci. Rep. 6 (2016) 35461.
75. Du, B. Wang, X. Liu, X. Liu, Y. He, Q. Zhang, X. Wang, Comparative transcriptome analysis of ovary and testis reveals potential sex-related genes and pathways in spotted knifejaw *Oplegnathus punctatus*, Gene 637 (2017) 203-210.
76. Cui, Y. Liu, W. Wang, Q. Wang, N. Zhang, F. Lin, N. Wang, C. Shao, Z. Dong, Y. Li, Y. Yang, M. Hu, H. Li, F. Gao, Z. Wei, L. Meng, Y. Liu, M. Wei, Y. Zhu, H. Guo, C.H.K. Cheng, M. Schartl, S. Chen, Genome editing reveals dmrt1 as an essential male sex-determining gene in Chinese tongue sole (*Cynoglossus semilaevis*), Sci. Rep. 7 (2017) 42213.
77. A. Webster, U. Schach, A. Ordaz, J.S. Steinfeld, B.W. Draper, K.R. Siegfried, Dmrt1 is necessary for male sexual development in zebrafish, Dev. Biol. 422 (1) (2017) 33-46.
78. N. Dumont, A.R. Brummett, Egg envelopes in vertebrates, Dev. Biol. 1 (1985) 235-288.
79. M. Wassarman, Zona pellucida glycoproteins, Annu. Rev. Biochem. 57 (1988) 415-442.
80. S. Litscher, P. M. Wassarman, The fish Egg's zona Pellucida, Curr. Top. Dev. Biol. 130 (2018) 275-305.
81. S. Litscher, P.M. Wassarman, Egg extracellular coat proteins: from fish to mammals, Histol. Histopathol. 22 (2007) 337-347.
82. Liu, H. Wang, Z. Gong, Tandem-repeated zebrafish *zp3* genes possess oocyte-specific promoters and are insensitive to estrogen induction1, Biol. Reprod. 74 (2006) 1016-1025.

Tables

Table 1. Summary statistics of clean transcriptome sequencing data from each sample.

Sample	Raw Reads (M)	Clean Reads (M)	Clean Bases (Gb)	Clean Reads Q30 (%)	Clean Reads Ratio (%)
Female-1	52.59	48.33	7.25	87.49	91.90
Female-2	44.04	40.55	6.08	86.17	92.06
Female-3	47.33	44.00	6.60	86.04	92.98
Male-1	47.33	43.93	6.59	86.37	92.81
Male-2	47.33	43.63	6.54	86.48	92.19
Male-3	47.33	43.58	6.54	86.18	92.09

Table 2. Statistics for the assembled unigenes.

Sample	Total Number	Total Length	Mean Length	N50	GC (%)
Female-1	81,249	145,145,039	1,786	3,128	48.91
Female-2	72,625	118,111,439	1,626	2,823	49.00
Female-3	72,194	120,065,634	1,663	2,836	49.06
Male-1	93,484	144,668,988	1,547	2,725	48.62
Male-2	95,350	142,802,517	1,497	2,627	48.55
Male-3	84,292	131,779,534	1,563	2,735	48.81
All-Unigene	149,814	309,566,984	2,066	3,538	48.64

Table 3. Summary of unigene annotations.

Database	Number of unigenes	Percentage (%)
Total	122,719	81.91%
NT	70,243	46.89%
NR	63,250	42.22%
SwissProt	67,197	44.85%
KEGG	100,035	66.77%
KOG	90,961	60.72%
GO	63,518	42.40%

Table 4. Genes related to the sex determination of *Takifugu poecilonotus*.

Gene	Unigene	Annotation	$\log_2$(□/□)	Q-value
		Doublesex and mab-3-related transcription factor		2.34E-07
Dmtr1	Unigene1322	1	6.66	
Sox3	Unigene1791	SRY-box containing transcription factor 3	5.72	8.44E-04
Dnali1	Unigene12399	Axonemal dynein light intermediate polypeptide 1	7.63	3.40E-07
Ropn1l	CL4645.Contig3	ropporin-1-like protein	9.37	1.32E-18
Cyp19a	CL5324.Contig8	Cytochrome P450 aromatase	-7.54	2.15E-07
Zp1	CL6076.Contig3	zona pellucida sperm-binding protein 1	-12.66	1.43E-110
Zp2	CL8003.Contig1	zona pellucida sperm-binding protein 2	-7.70	1.84E-20
Zp3	CL7875.Contig2	zona pellucida sperm-binding protein 3	-7.93	4.04E-16
Start-7	CL3580.Contig10	stAR-related lipid transfer protein 7	7.57	2.66E-06
Gdf3	CL9083.Contig1	growth differentiation factor 3	-8.90	3.99E-14
Gdf9	CL6347.Contig1	growth differentiation factor 9	-9.23	1.61E-12
Gtf3a	Unigene10317	Transcription factor IIIA-like	-2.42	2.11E-08
42sp43	Unigene8014	P43 5S RNA-binding protein-like	-7.38	6.57E-24
Mnd1	Unigene817	meiotic nuclear division protein 1 homolog	3.25	8.39E-03
Star	Unigene31813	steroidogenic acute regulatory protein	-2.49	3.42E-04
Wee1	CL12054.Contig1	wee1-like protein kinase	2.05	6.42E-03
Spaca6	CL4045.Contig1	sperm acrosome membrane-associated protein 6	5.97	4.05E-04
Spaca4	CL12530.Contig2	sperm acrosome membrane-associated protein 4	3.53	2.18E-15
Spata5	CL8799.Contig2	spermatogenesis-associated protein5	6.47	7.85E-05
Spata7	CL4586.Contig1	spermatogenesis-associated protein 7	2.93	1.40E-03
Spata17	CL4320.Contig15	spermatogenesis-associated protein 17	8.17	2.09E-08
Spata32	Unigene13081	spermatogenesis-associated protein 32	4.78	1.72E-06

Table 5. Growth related genes in *Takifugu poecilonotus*.

Gene	Unigene	Annotation	$\log_2$(□/□)	Q-value
Igfbp1	CL6482.Contig4	insulin-like growth factor binding protein 1	-8.52	2.51E-09
Igfbp3	Unigene7287	Insulin-like growth factor binding protein 3	-3.36	9.25E-06
Igf1r	Unigene24558	Insulin-like growth factor 1 receptor	-1.06	1.75E-02
Ghr	CL9297.Contig1	Growth hormone receptor	-4.79	3.10E-03
Sstr2	Unigene6248	Somatostatin receptor 2	7.45	5.93E-07
Fabp1	CL7463.Contig1	fatty acid-binding protein 1	3.85	8.11E-12
Fabp4	CL10654.Contig1	fatty acid-binding protein 4	-8.28	6.17E-56
Fabp6	CL6561.Contig1	fatty acid-binding protein 6	6.70	2.41E-04
Fabp7	Unigene40265	fatty acid-binding protein 7	5.45	2.86E-05
Grb1	CL5906.Contig1	Growth factor receptor-bound proteins 1	2.16	6.38E-03
Grb7	Unigene6338	Growth factor receptor-bound proteins 7	-2.98	8.15E-03
Grb10	CL11116.Contig1	Growth factor receptor-bound proteins 10	3.07	1.43E-04
Htr1	Unigene31616	5-hydroxytryptamine receptor 1D-like	4.59	1.31E-04
Htr2	Unigene28993	5-hydroxytryptamine receptor 2C-like	5.27	3.86E-03
Htr3	CL1712.Contig2	5-hydroxytryptamine receptor 3	5.61	1.44E-03
Htr6	Unigene18275_	5-hydroxytryptamine receptor 6	4.33	7.54E-03
Prlr	CL1256.Contig1	Prolactin receptor	-8.19	3.29E-11
Prl	CL12326.Contig4	Prolactin	-5.65	4.17E-06
Myhc	Unigene12897	myosin heavy chain	6.29	5.15E-05

Figures

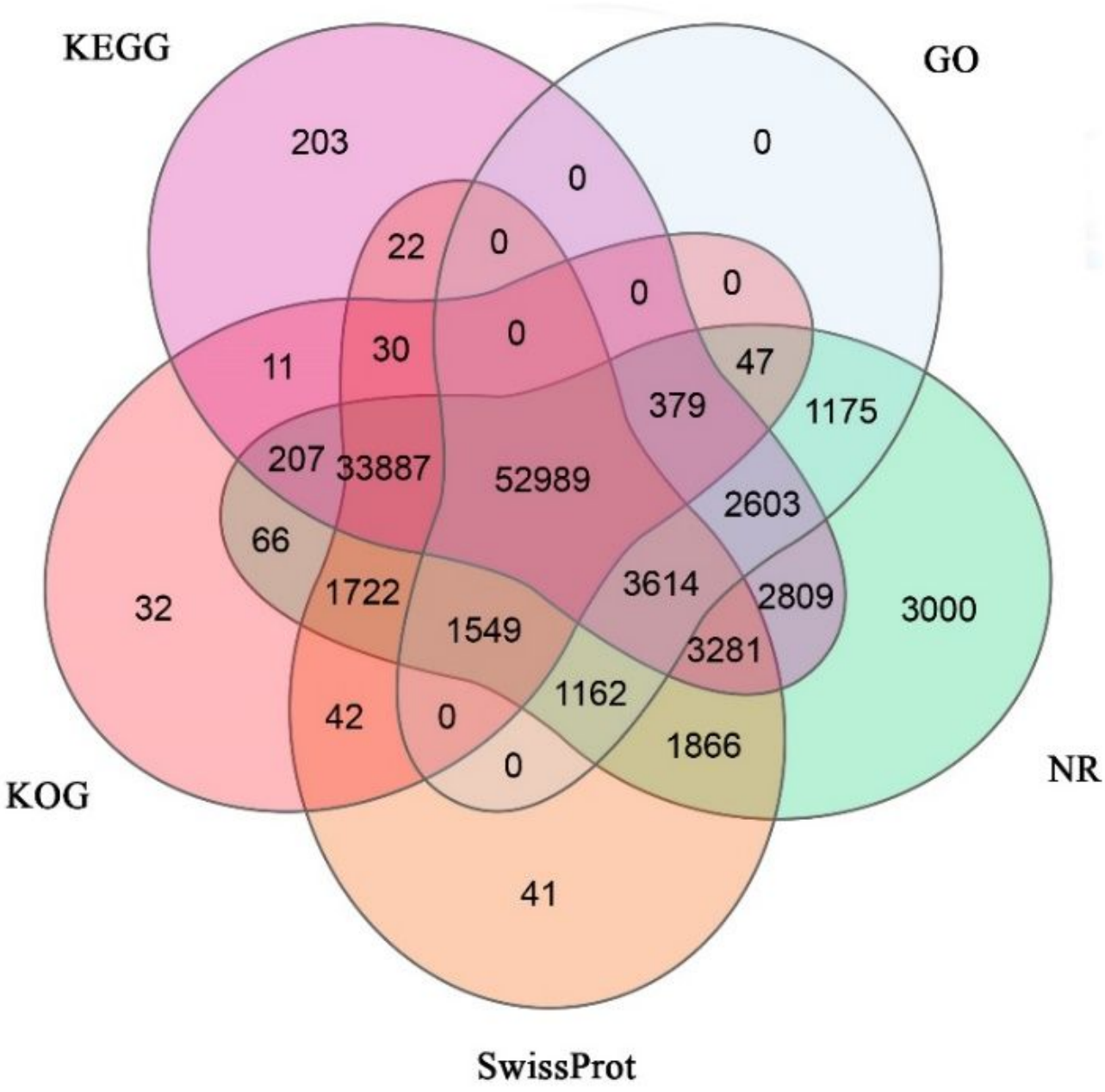

Figure 1

Venn diagram of the functional annotations.

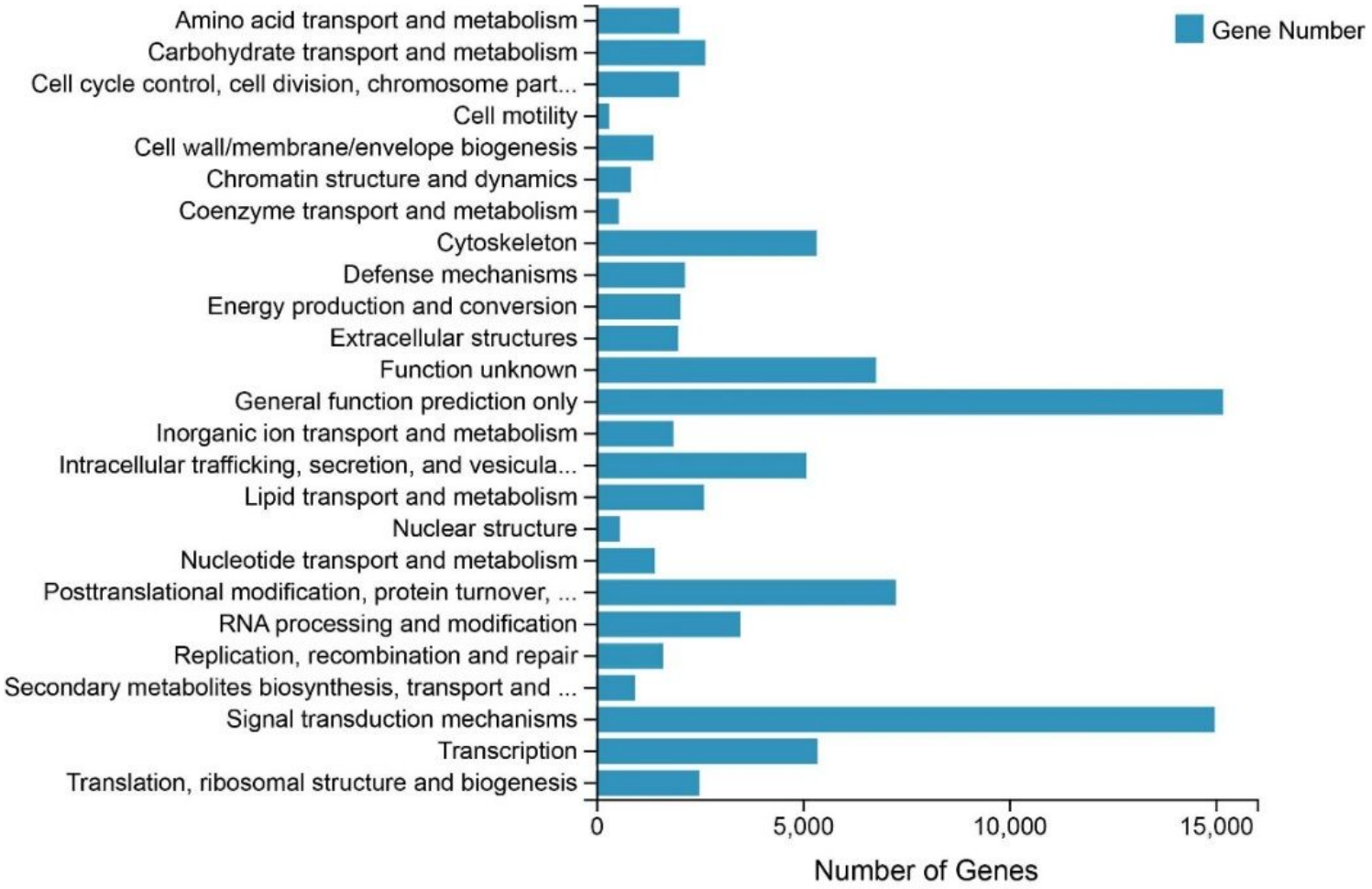

Figure 2

Clusters of KOG functional classifications.

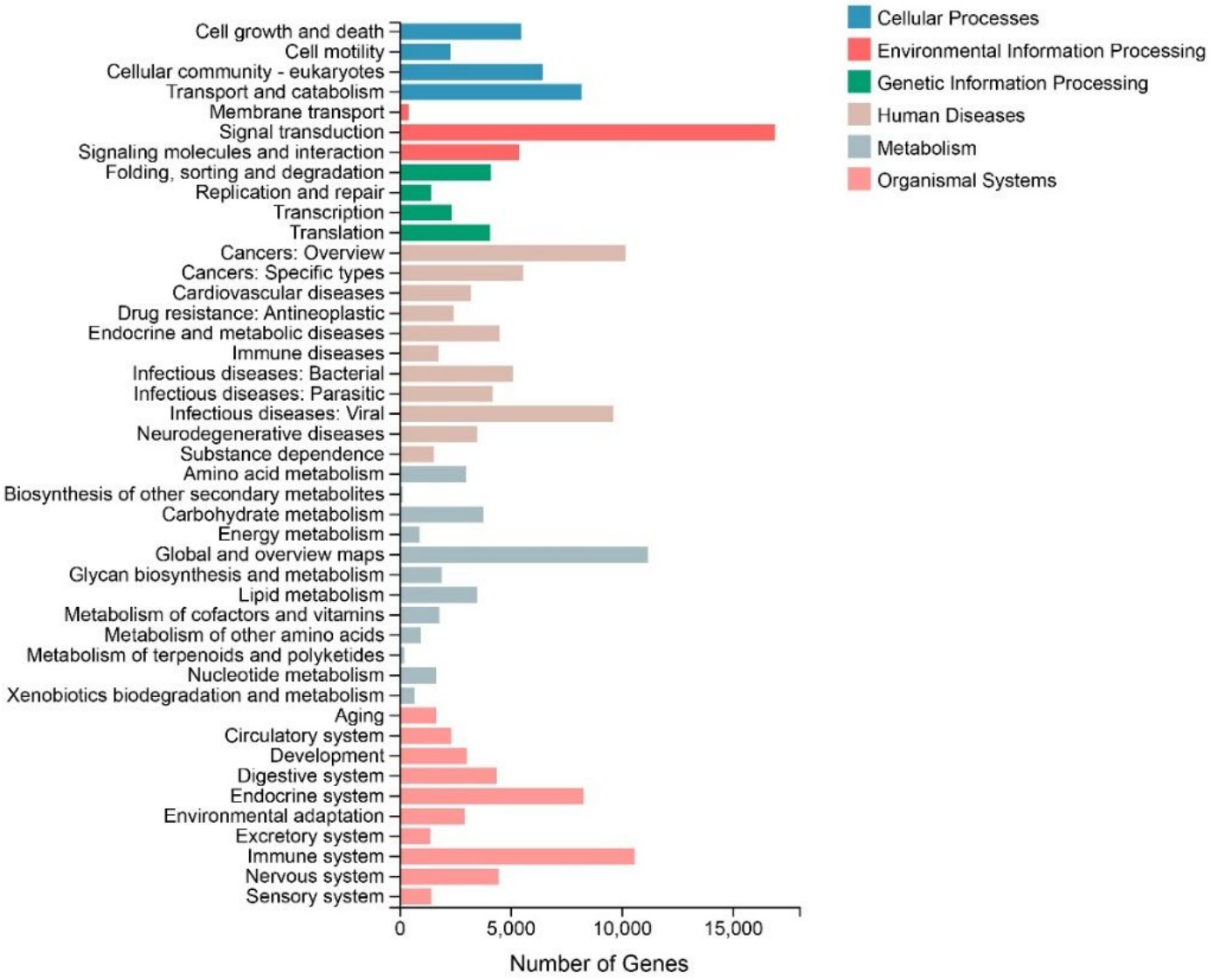

Figure 3

KEGG pathway classifications of the unigenes.

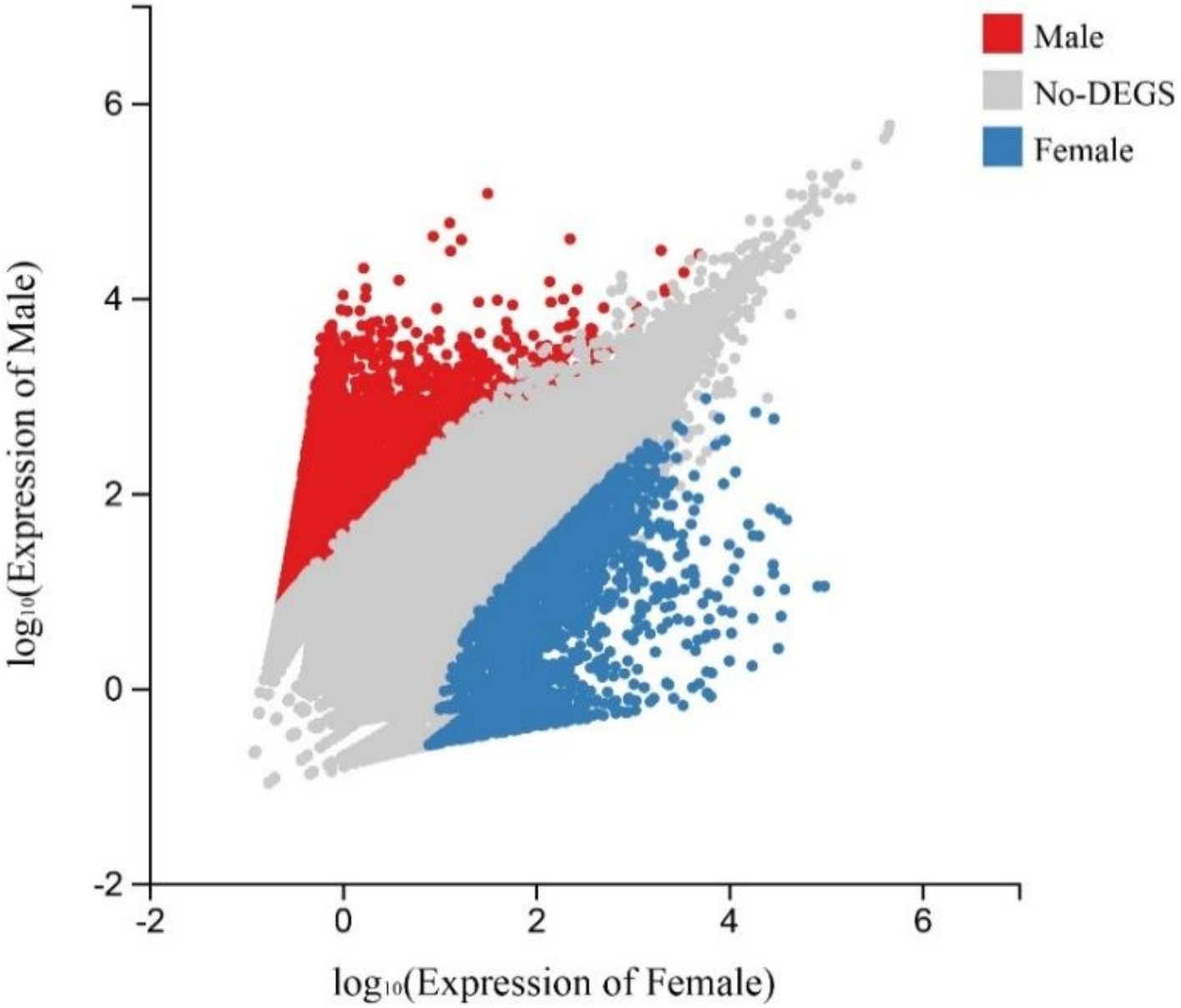

Figure 4

Scatter plots showing gene expression profiles in female T. poecilonotus and male T. poecilonotus. Thresholds for inclusion were defined by Q-values < 0.01 and |log2 fold change| ≥ 2.

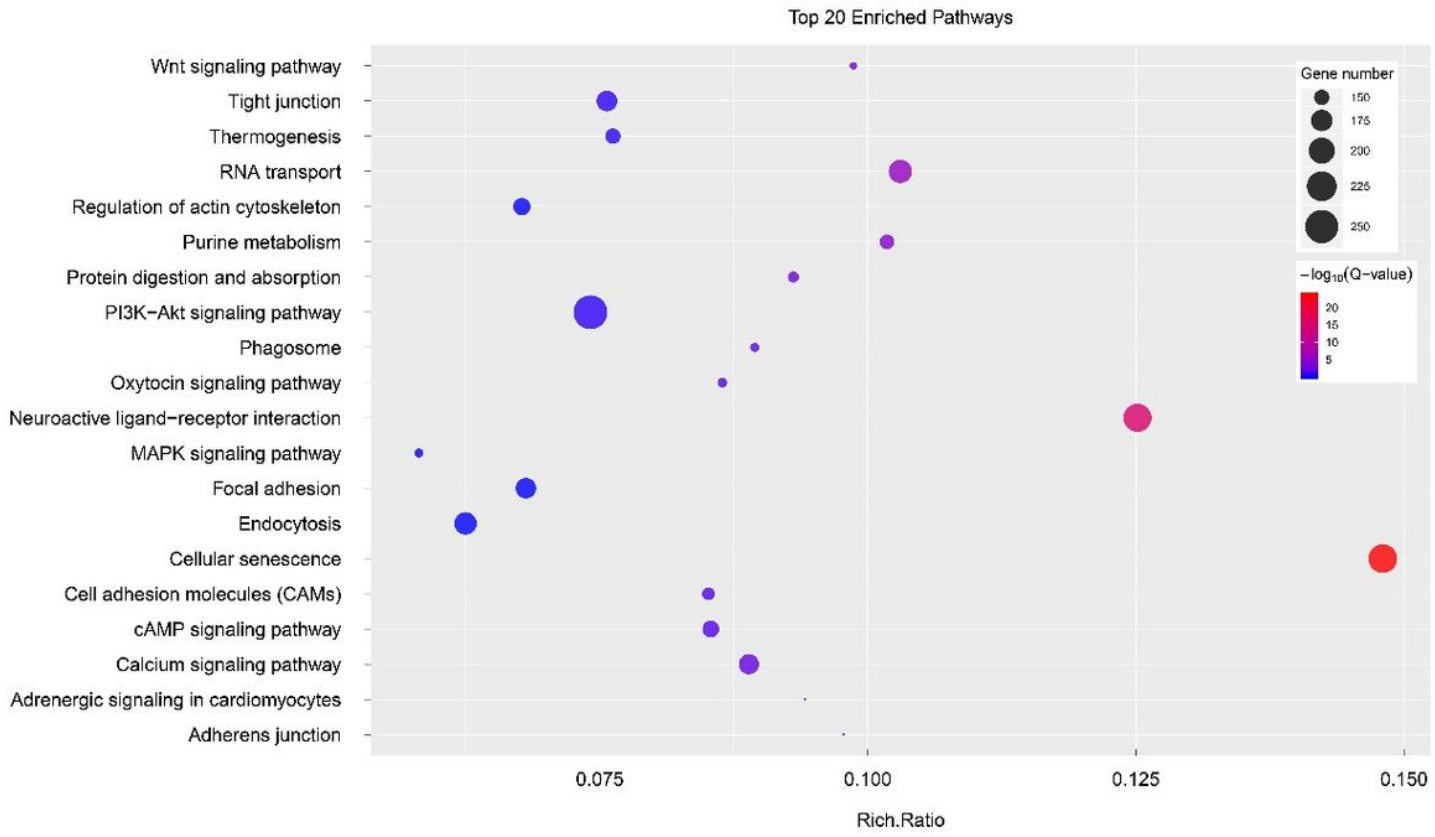

Figure 5

KEGG pathway enrichment analyses of the DEGs

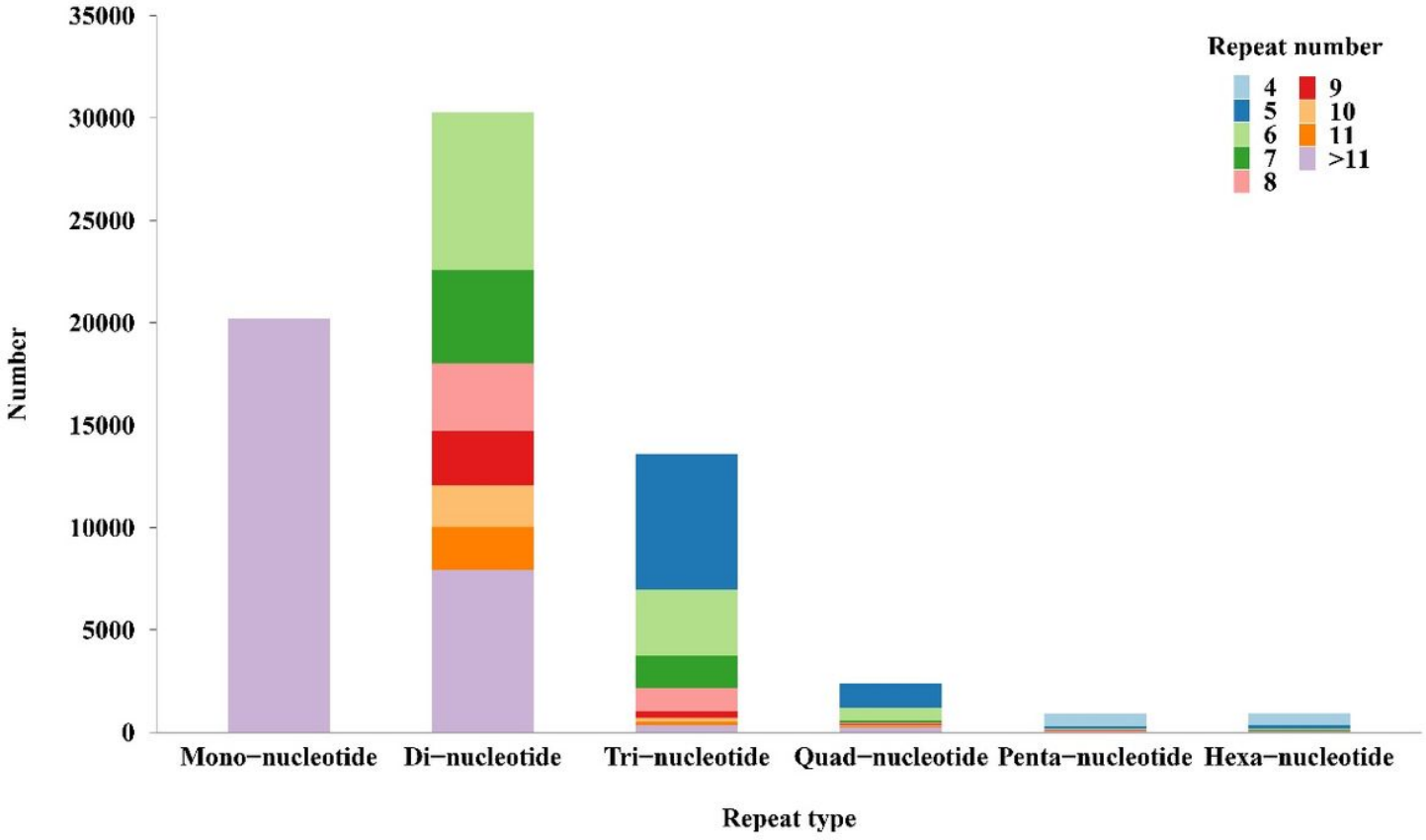

Figure 6

The distribution of different SSRs.

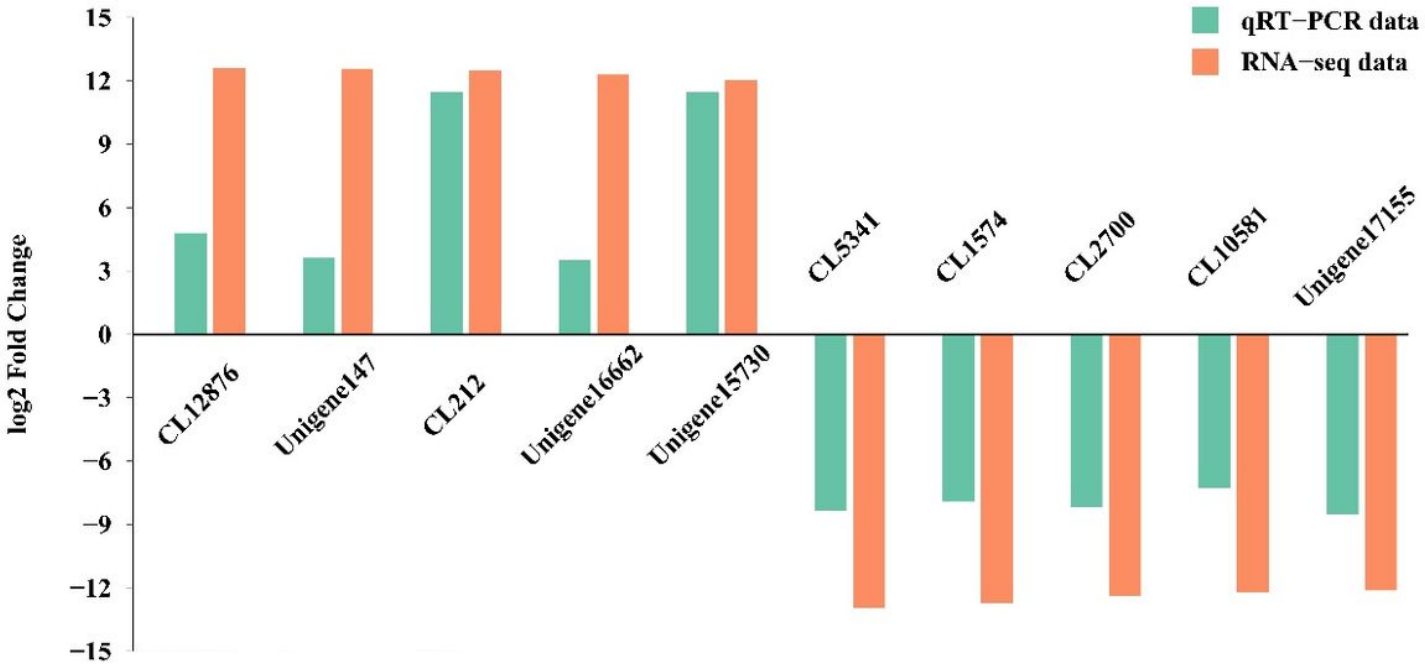

Figure 7

The expression levels of 10 selected unigenes by RNA-seq and qRT-PCR.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementfileTakifugupoecilonotus.docx
- SupplementfileTakifugupoecilonotus.docx