Whole-genome sequencing of leopard coral grouper (*Plectropomus leopardus*) and exploration of regulation mechanism of skin color and adaptive evolution

Yang Yang[1], Li-Na Wu[1], Jing-Fang Chen[2], Xi Wu[1], Jun-Hong Xia[1,3], Zi-Ning Meng[1,3,*], Xiao-Chun Liu[1,3,*], Hao-Ran Lin[1,3]

[1] *State Key Laboratory of Biocontrol, Institute of Aquatic Economic Animals and Guangdong Provincial Key Laboratory for Aquatic Economic Animals, Life Sciences School, Sun Yat-Sen University, Guangzhou, Guangdong* 510275, *China*

[2] *State Key Laboratory of Biocontrol and Guangdong Provincial Key Laboratory of Plant Resources, School of Life Sciences, Sun Yat-Sen University, Guangzhou, Guangdong* 510275, *China*

[3] *Southern Laboratory of Ocean Science and Engineering, Zhuhai, Guangdong* 519000, *China*

ABSTRACT

Leopard coral groupers belong to the *Plectropomus* genus of the Epinephelidae family and are important fish for coral reef ecosystems and the marine aquaculture industry. To promote future research of this species, a high-quality chromosome-level genome was assembled using PacBio sequencing and Hi-C technology. A 787.06 Mb genome was assembled, with 99.7% (784.57 Mb) of bases anchored to 24 chromosomes. The leopard coral grouper genome size was smaller than that of other groupers, which may be related to its ancient status among grouper species. A total of 22 317 protein-coding genes were predicted. This high-quality genome of the leopard coral grouper is the first genomic resource for *Plectropomus* and should provide a pivotal genetic foundation for further research. Phylogenetic analysis of the leopard coral grouper and 12 other fish species showed that this fish is closely related to the brown-marbled grouper. Expanded genes in the leopard coral grouper genome were mainly associated with immune response and movement ability, which may be related to the adaptive evolution of this species to its habitat. In addition, we also identified differentially expressed genes (DEGs) associated with carotenoid metabolism between red and brown-colored leopard coral groupers. These genes may play roles in skin color decision by regulating carotenoid content in these groupers.

Keywords: Leopard coral grouper; Genome; Evolution; Skin color; Immune

INTRODUCTION

Groupers (Epinephelidae, Perciformes) are predators found in coral reef ecosystems, and include more than 160 species in 16 genera (Zhuang et al., 2013). They are commercially important in aquaculture, with approximately 47 grouper species cultured in East and Southeast Asia (Rimmer &

Glamuzina, 2019) and market demand increasing sharply worldwide. In China, the annual grouper output reached 160 000 tons in 2018 (Yang et al., 2020). Groupers are protogynous hermaphrodites, whereby they firstly differentiate as females, with some later changing to males (Adams, 2003). They have a complicated social system and the occurrence of sex reversal can be triggered by a change in social status (Chen et al., 2019). Thus, groupers are a good model for studies on sex inversion, speciation, and social systems.

The leopard coral grouper (*Plectropomus leopardus*), also called common coral trout, is a valuable marine fish belonging to the *Plectropomus* genus of the Epinephelidae family (Wang et al., 2015a). This fish is mainly distributed in the western Pacific regions of Western Australia, eastward to the Caroline Islands and Fiji and from southern Japan to Queensland, Australia (https://www.iucnredlist.org/species/44684/ 100462709). Wild populations have decreased due to the destruction of spawning aggregations and overfishing, and international protection measures (e.g., IUCN) for this fish are being considered. In recent years, artificial breeding technology has been successfully established for the leopard coral grouper. Moreover, these fish are considered an important resource for intensive industrial farming in recirculating aquaculture systems due to their high nutritional value, tender flesh, beautiful skin color, and high breeding density.

The substantial commercial and research value of leopard coral groupers has attracted the attention of aquatic scientists, including exploration on breeding (Burgess et al., 2020; Chen et al., 2016; Melianawati et al., 2013), feed additives (Yu et al., 2018), reproductive biology (Adams, 2003; Bunt & Kingsford, 2014; Khasanah et al., 2019; Metcalfe et al., 2018), phylogenetic analysis (Chen et al., 2018; Craig & Hastings, 2007; Ma et al., 2016; Xie et al., 2016), and skin color (Maoka et al., 2017; Wang et al., 2015a; Xie et al., 2016). However, the lack of genomic resources severely hinders deeper investigation of leopard coral groupers. So far, only two transcriptome analyses have been carried out with relatively limited gene sequences and expression information (Mekuchi et al., 2017; Wang et al., 2015a). Furthermore, only two grouper species genomes, i.e., red-spotted grouper *Epinephelus akaara* (Ge et al., 2019) and giant grouper *Epinephelus lanceolatus* (Zhou et al., 2019a), have been published. However, *Epinephelus* and *Plectropomus* species diverged from their most recent common ancestor (MRCA) a long time ago. At present, no valid genome has been reported for *Plectropomus* species. The assembly of a leopard coral grouper genome is essential for further research on evolutionarily status, molecular-assisted selection, sex reversal, and quantitative trait locus (QTL) mapping of traits of interest in *Plectropomus* species.

Body color is an important economic trait in aquatic animals, especially for the leopard coral grouper. Its integument is mainly divided into black, brown, or red color (Maoka et al., 2017), which is an important factor in determining fish quality as Chinese markets prefer those fish with a bright red color,

which sell for twice the price of black fish. Recent research has shown that different proportions of carotenoids, such as tunaxanthin, astaxanthin, and tunaxanthin, determine skin color in this species (Maoka et al., 2017). Several studies related to skin color in this fish have also been carried out in our laboratory. For example, 38 candidate genes underlying the mechanism of color and pigmentation were detected based on comparative transcriptome analysis (Wang et al., 2015a); a total of 74 single nucleotide polymorphisms and one Indel were identified in the complete mitochondrial genome of red and gray fish (Xie et al., 2016); and clear differences between light-red and gray fish in *COI*, *ND2*, and D-loop sequences were observed via mitochondrial DNA analysis (Cai et al., 2013; Van Herwerden et al., 2009). Thus, to improve leopard coral grouper traits, it is essential that the molecular mechanism regulating skin color is explored.

An elaborate reference genome could provide a strong foundation for further research on the leopard coral grouper. Third-generation sequencing technologies have improved the extension of sequencing reads and provide superior platforms to produce complete and exact genomes (Quail et al., 2012). Here, we used PacBio long-read sequencing (Eid et al., 2009) and high-throughput chromosome conformation capture (Hi-C) technology (Lieberman-Aiden et al., 2009) to assemble a high-quality chromosome-level genome of the leopard coral grouper. This should provide important genomic resources for subsequent investigations on sex reversal, genetic evolution, physiological regulation, and molecular breeding. Based on the high-quality genome, comparative genomes and transcriptomes were used to analyze the adaptive evolution and regulation of skin color in leopard coral groupers.

MATERIALS AND METHODS

Ethics statement
All experiments in the present study were approved by the Animal Care and Use Committee of the School of Life Sciences, Sun Yat-Sen University, Guangdong, China.

Animal preparation for genome sequencing
One leopard coral grouper (Figure 1A) sampled from the Tropical Marine Aquatic Breeding Center, Wenchang, Hainan Province, China, was used for genome sequencing and assembly. The fish was immediately dissected after anesthesia with MS-222. White muscle tissue was sampled for DNA extraction, which was then used for genomic DNA sequencing and Hi-C library construction. To assist annotation of genes, 10 tissues, including skin, muscle, liver, kidney, brain, intestine, fat, spleen, heart, and gill, were collected, and then stored in RNAlater for RNA isolation.

Library construction and genome assembly
Total DNA was extracted from white muscle tissue with a TIANamp Marine Animals DNA Kit (Tiangen Biotech Co., Ltd., China). Quality and quantity of total DNA were determined by NanoDrop 2000 (Thermo Fisher Scientific Inc., USA). A paired-end sequencing library was constructed using a
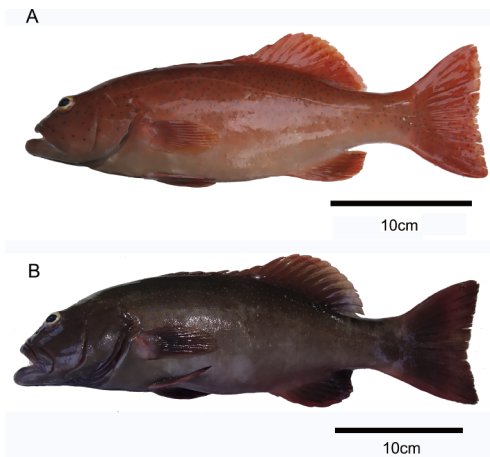
Figure 1 Leopard coral grouper (*Plectropomus leopardus*)
A: Bright red fish (used for genome assembly). B: Brown fish. Scale bars: 10 cm.

TruSeq Nano DNA LT Library Preparation Kit (Illumina, USA). The obtained library was then sequenced using the Illumina HiSeq X Ten platform. We estimated the main genome characteristics of the leopard coral grouper through K-mer methods (Liu et al., 2013). A 20 kb SMRTbell library was constructed for sequencing using the PacBio Bioscience Sequel platform (Pacific Biosciences, USA).

Total RNA of the 10 tissues (approximately 80 mg each) was extracted using RNAiso reagents (Takara, China) following the manufacturer's instructions. The quantity and quality of RNA samples were determined using a microplate spectrophotometer (BioTek Company, USA) and electrophoresis was conducted using 1% agarose gel. For predicting protein-coding genes, a mRNA library was constructed. Total RNA of the 10 tissues was mixed at equal amounts to generate a mixed RNA pool. An RNA-seq library was prepared based on (Yang et al., 2019b). Library quality and quantity were measured using an Agilent 2100 Bioanalyzer (Agilent Technologies, USA). Finally, the library was sequenced using the Illumina Hiseq 2000 platform with the PE150 approach. The RNA and DNA libraries were sent to the Berry Genomics Company, China.

K-mer frequency distribution analysis was used to estimate genome characteristics (Li et al., 2010). The raw data were filtered to obtain high-quality clean reads according to three stringent filtering standards: i.e., removing (a) reads linked with the barcode adapter; (b) reads with ≥10% unidentified nucleotides (N); and (c) low-quality reads with >50% bases harboring Phred quality scores of ≤20. The remaining clean data were used to estimate genome size and heterozygosity using 19-mer. For the PacBio sequencing data, after removing short polymerase reads, low-quality polymerase reads, and polymerase reads linked with the barcode adapter, clean polymerase reads (subreads) were used for assembly with Canu v1.8 (Koren et al., 2017). Simply, the longest subreads were set as seed reads, other PacBio data were aligned to the seed reads, and low-quality reads were trimmed; the revised

seed reads were assembled into contigs based on overlap of seed reads. The contigs were corrected by alignment of PacBio data using pbmm2 v1.1.0 (https://github.com/PacificBiosciences/pbmm2) and GenomicConsensus: Arrow (https://github.com/PacificBiosciences/GenomicConsensus). Consensus correction was performed based on Illumina data using BWA v0.7.17 (Li & Durbin, 2010) and Pilon v1.16 (Walker et al., 2014). Lastly, redundant contigs were detected and selectively removed from *de novo* assembly using Redundans (Pryszcz & Gabaldón, 2016).

For estimating genome completeness, Illumina data were first mapped back to the leopard coral grouper genome to calculate the mapping rate using BWA v0.7.17 (Li & Durbin, 2010). The orthologs of the leopard coral grouper were then aligned to a reference gene set from the *Actinopterygii* database (Actinopterygii_odb9), which was constructed from 20 fish species consisting of 4 584 orthologs using Benchmarking Universal Single-Copy Orthologs (BUSCO).

Pseudochromosome construction

Hi-C was performed to assist in the construction of the chromosome-level leopard coral grouper genome. Muscle samples were fixed with fresh paraformaldehyde and then DNA-protein bonds were created. The Mbo I restriction enzyme was used to digest the DNA and the overhanging 5' ends of the DNA fragments were repaired with a biotinylated residue. The fragments close to each other in the nucleus during fixation were ligated and the denatured proteins were removed. The Hi-C fragments were further sheared by sonication into 350 bp fragments, which were then pulled-down with streptavidin beads. The library was sequenced on the Illumina HiSeq X Ten platform with PE150. To anchor scaffolds into pseudochromosomes, JUICER v1.6.2 (https://github.com/aidenlab/juicer) was first used to map and process the reads. The misjoins in the scaffolds were eliminated using 3D *de novo* assembly (3D-DNA) software (Dudchenko et al., 2017).

Annotation of repeats

Tandem repeats were predicted using TRF v4.07b (Benson, 1999) and transposable elements (TEs) were identified using LTR Finder (http://tlife.fudan.edu.cn/tlife/ltr_finder/) (Xu et al., 2010), RepeatScout v1.05 (http://www.repeatmasker.org), and PILER v1.0 (Edgar & Myers, 2005). Coding sequences were removed from the predicted repeat sequences through alignment to the SwissProt database using blastx with e-value ≤1e–4, identity ≥30, coverage ≥30%, and length ≥90 bp. In addition, non-coding RNA (ncRNA) was removed from the predicted repeat sequences via alignment in Rfam 11.0 using blastn with e-value ≤1e–10, identity ≥80, and coverage ≥ 50%. The predicted TEs were aligned to the RepBase and TE protein database using WU-BLAST to remove simple repeats, satellites, and ncRNAs. The TEs were classified using RepeatClassifier. For removing redundant repeat sequences, the predicted repeat sequences were aligned with each other using blastn with e-value ≤1e–10, identity ≥80, coverage ≥ 80%, and length ≥80bp. The repeat regions were searched

against RepBase using RepeatMasker v3.3.0 and against the TE protein database using RepeatProteinMask v3.3.0. Finally, the repeat sequences in the scaffold were masked using RepeatMasker v3.3.0 and RepeatProteinMask v3.3.0.

ncRNA and mRNA annotation

Non-coding RNA includes ribosomal RNA (rRNA), transfer RNA (tRNA), long non-coding RNA (lncRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), micro RNA (miRNA), and other unknown RNA. Here, tRNA was predicted using tRNAscan-SE v1.3.1 (Lowe & Eddy, 1996) and rRNA, snRNA, snoRNA, and microRNA were predicted following alignment to Rfam (ftp://ftp.ebi.ac.uk/pub/databases/Rfam).

Protein-coding genes were predicted via alignment of homologous sequences and RNA-seq assisted methods. For predicting homologous sequences, protein sequences of *Oreochromis niloticus*, *Cyprinus carpio*, *Salmo salar*, *Larimichthys crocea*, and *Oryzias latipes* were downloaded from the NCBI database and aligned to the repeat masked genome using GeMoMa v1.4.2 (Keilwagen et al., 2016). Moreover, RNA-seq data from the 10 tissues were aligned to the genome using HISAT2 v2.0.4 (Kim et al., 2015) and assembled using Cufflinks v2.2.1 (Trapnell et al., 2012). Open reading frames (ORFs) were predicted using PASA v2.0.1. The gene structure was predicted based on assembled RNA-seq data using Augustus v3.2.2 (Stanke et al., 2006), SNPAP (Leskovec & Sosič, 2016), GlimmerHMM v3.01 (Majoros et al., 2004), and GeneMark-ET v4.21 (Lomsadze et al., 2014). Gene predictions based on the above results were merged using EVM (http://evidencemodeler.sourceforge.net/) (Haas et al., 2008).

Functional annotations of the predicted genes in the leopard coral grouper genome were performed based on several public databases, including NCBI-NR, KEGG, eggNOG, and SwissProt using blast+ (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/) with e-value ≤1e–05. We performed gene ontology (GO) annotation using Blast2GO (Conesa et al., 2005). We also used InterProScan v4.7 (Jones et al., 2014) to obtain protein domain annotations in the InterPro database, which contains multiple protein databases including CATH-Gene3D, CDD, HAMAP, PANTHER, Pfam, PIRSF, PRINTS, ProDom, PROSITE, SMAT, SUPERFAMILY, and TIGRFAMs. Lastly, functional annotations of the best alignments in each database were used as the final consensus gene annotation results.

Genome evolution analysis

For identification of gene families, we downloaded 12 fish species genomes, including zebra fish (*Danio rerio*, GCF_000002035.6), large yellow croaker (*Larimichthys crocea*, GCF_000972845.2), Atlantic salmon (*Salmo salar*, GCF_000233375.1), Asian swamp eel (*Monopterus albus*, GCF_001952655.1), Atlantic cod (*Gadus morhua*, GCF_902167405.1), rainbow trout (*Oncorhynchus mykiss*, GCF_002163495.1), barramundi (*Lates calcarifer*, GCF_001640805.1), tilapia (*Oreochromis niloticus*, GCF_001858045.2), coelacanth (*Latimeria chalumnae*,

GCF_000225785.1), medaka (*Oryzias latipes* GCF_002234675.1), lancelet (*Branchiostoma belcheri*, GCF_001625305.1), and brown-marbled grouper (*E. fuscoguttatus*, not published) for comparison using all to all BLAST with an e-value of 1e–5. The Markov Cluster Algorithm (MCL) was then used to cluster the alignments into gene families using OrthoMCL with inflation 1.5.

A phylogenetic tree was constructed using shared single-copy genes of the leopard coral grouper and other fish species mentioned above. Protein sequences of these single-copy genes were aligned using MUSCLE v3.7 (Edgar, 2004). The phylogenetic tree was constructed using PhyML software v3.0 (Guindon et al., 2010) with the maximum-likelihood (ML) algorithm. Collinearity analysis of the leopard coral grouper and brown-marbled grouper genomes was carried out using MUMmer v4 (Kurtz et al., 2004). The divergence times in each tree node were predicted using MCMCtree in the PAML package (Yang, 1997). Three calibration times were obtained from the TimeTree database (Kumar et al., 2017).

The expansion and contraction of orthologous gene families between the ancestor and other species were predicted using CAFÉ v1.6 (De Bie et al., 2006). A random death model was used to study changes in gene families along each lineage of the phylogenetic tree. A probabilistic graphical model (PGM) was introduced to the probability of transitions in gene family size from parent to child node phylogeny. Functional prediction of expanded and contracted genes was carried out based on the GO and KEGG pathway databases.

Detection of differentially expressed genes (DEGs) among leopard coral groupers with different body color using RNA-seq

The high-quality chromosome-level leopard coral grouper genome provides an important reference for transcriptome analysis. To determine transcriptome differences between bright red (Figure 1A, 517.7±82.47 g) and brown fish (Figure 1B, 504.3±95.6 g), the skins of three bright red and three brown leopard coral groupers were sampled from the Tropical Marine Aquatic Breeding Center, Wenchang, Hainan Province, China. Isolation of total RNA and library construction were performed using the same methods as described above. Trimmomatic v0.39 (Bolger et al., 2014) was used to filter raw data. The clean data were mapped to the leopard coral grouper genome using Tophat2 v2.1.1 (Trapnell et al., 2009). Read count statistics were determined using HTSeq v0.11.1 (Anders et al., 2015). R package, Edger v3.1.0 (Robinson et al., 2010), and Limma v3.42.2 (Ritchie et al., 2015) were applied to identify DEGs between the two groups.

RESULTS

Genome assembly and assessment
A total of 94.77 Gb of Illumina raw reads were obtained. After removal of low-quality reads, adapter sequences, and polymerase chain reaction (PCR)-duplicates, the size of the leopard coral grouper genome using the K-mer approach was
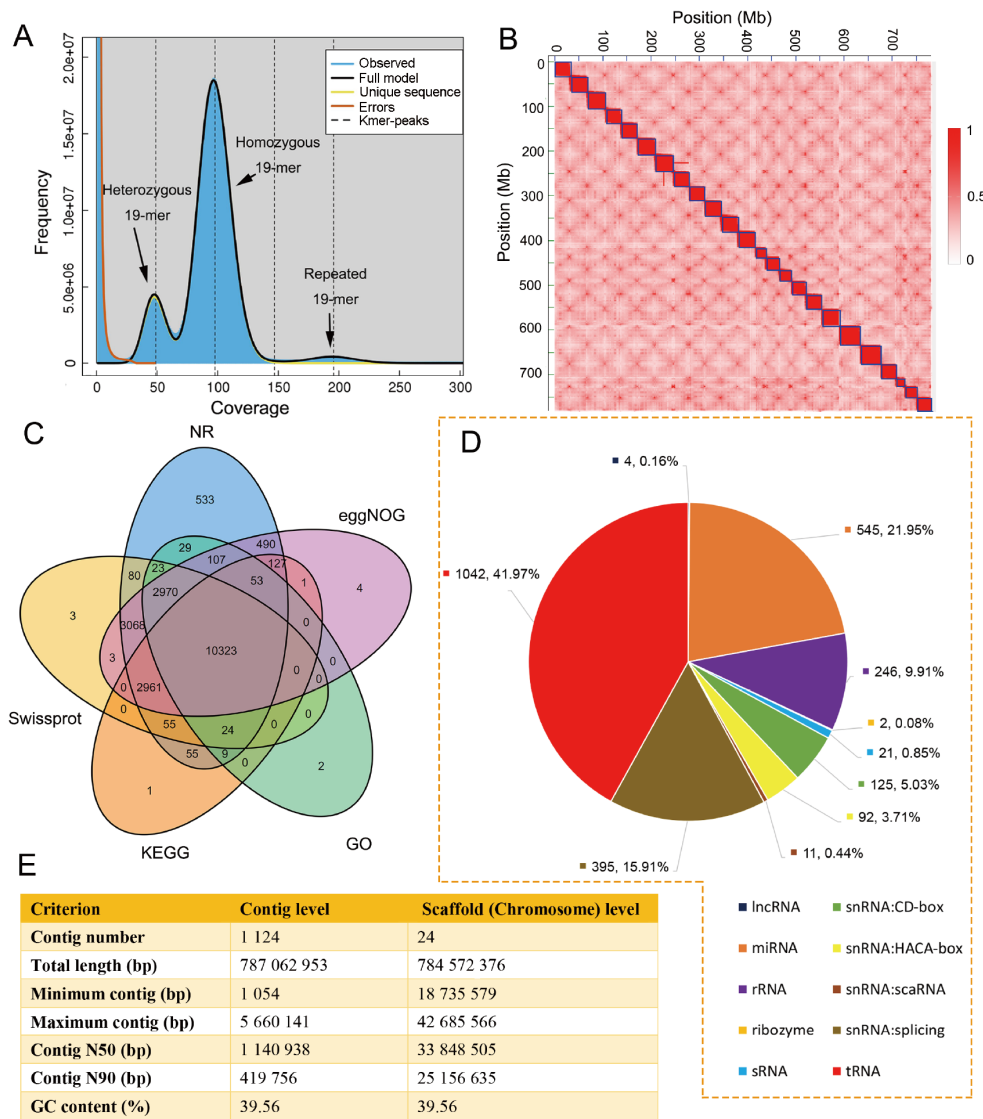
estimated to be 858 Mb, with 39.54% GC content (Figure 2A). The heterozygosity and repeats were approximately 0.42% and 23%, respectively.

A total of 106.10 Gb (×145) of PacBio sequencing data were used for *de novo* assembly of the genome (Supplementary Table S1 and Figure S1). The mean length and N50 were 12 893 bp and 18 054 bp, respectively. The reads and base number of fragments longer than 5 Kb were 6 831 058 and 101.47 Gb, respectively. The genome assembly spanned 787.06 Mb with a contig N50 of 1.14 Mb and GC content of 39.56% (Figure 2E), and was deposited in the DDBJ database. BUSCO was used to verify the integrality of the

assembled genome. BUSCO alignment showed that our final assembly contained 4 149 complete BUSCOs (90.5%), including 4 034 single-copy and 115 duplicated BUSCOs (Supplementary Table S2).

Pseudochromosome construction

Hi-C was applied for genome assembly. A total of 99.5 Gb of clean bases were obtained with Q20 of 98.26% and Q30 of 94.77%. The mapped ratio of Hi-C reads in the genome was 78.96%. Based on the Hi-C results, a total of 1 091 contigs were anchored into 24 pseudochromosomes (Chr) of the leopard coral grouper, and a total length of 784.57 Mb was assembled, which comprised 99.62% of assembled genome

Criterion	Contig level	Scaffold (Chromosome) level
Contig number	1 124	24
Total length (bp)	787 062 953	784 572 376
Minimum contig (bp)	1 054	18 735 579
Maximum contig (bp)	5 660 141	42 685 566
Contig N50 (bp)	1 140 938	33 848 505
Contig N90 (bp)	419 756	25 156 635
GC content (%)	39.56	39.56

Figure 2 Statistics on genome assembly and annotation of leopard coral grouper (*Plectropomus leopardus*)

A: Estimation of genome size, repeat content, and heterozygosity by survey using 19-mers. B: Genome contig contact matrix using Hi-C data, color bar indicates contact density from red (high) to white (low). C: Functional annotated protein-coding genes. D: Summary statistics of non-coding RNA. E: Statistics of leopard coral grouper genome assembly.

sequences (Supplementary Table S3 and Figure 2B). The longest and shortest pseudochromosome lengths were 42 685 566 bp and 18 735 579 bp, respectively.

Repetitive element annotation

We used *de novo* and homologue prediction to detect repeat sequences (Supplementary Table S4). In the leopard coral grouper, a total of 1 705 874 repeat sequences with a length of 272 473 600 bp were detected, accounting for 34.596% of the whole genome. Among them, we found 808 677 DNA repeat elements (DNA, 111 693 098 bp, 14.181%); 135 729 long interspersed nuclear elements (LINE, 26 212 322 bp, 3.329%); 57 382 long terminal repeat elements (LTR, 9 810 107 bp, 1.245%); and 16 333 short interspersed nuclear elements (SINE, 2 350 545 bp, 0.298%).

ncRNA and mRNA annotation

A total of 2 483 ncRNAs were detected, including 1 042 tRNAs, 623 snRNAs, 545 miRNAs, and 246 rRNAs (Figure 2D and Supplementary Table S5).

Based on the repeat-masked genome of the leopard coral grouper, a total of 22 317 genes were detected using *de novo* prediction and homology searching. Based on the homologous protein sequences of *Oreochromis niloticus*, *Cyprinus carpio*, *Salmo salar*, *Larimichthys crocea*, and *Oryzias latipes*, a total of 10 600–21 861 proteins were obtained. Based on RNA-seq data, a total of 28 329 genes were detected. Based on *de novo* methods, a total of 28 713–51 359 genes were detected (Supplementary Table S6). After removing redundant genes and merging, a total of 22 317 genes were found with a total length of 463 272 851 bp and mean length of 20 758 bp; a total of 25 558 transcripts were obtained with a total length of 79 212 336 bp and mean length of 3 099 bp. There were 22 000 complete ORFs (98.58%), 25 172 complete transcripts (98.49%), and 286 860 exons and 261 302 introns detected in the leopard coral grouper genome. The mean lengths of exons and introns were 276 bp and 1 890 bp, respectively. The average exon number for each gene was 11.2 and the average length of coding sequence (CDS) was 1 836 bp (Supplementary Table S6).

To understand gene function, the predicted genes were mapped against several public databases. A total of 20 921 predicted genes were mapped to these databases, including 20 107, 20 907, 13 540, 13 609, and 19 510 genes mapped to the eggNOG, NR, GO, KEGG, and SwissProt databases, respectively. A total of 10 323 genes were mapped to all these databases (Figure 2C). For KEGG, a total of 41 terms were enriched, with the top two terms being signal transduction and infectious diseases (Supplementary Figure S2). For GO, a total of 60 terms were enriched, with the top two terms being integral component of membrane and ATP binding (Supplementary Figure S3). For eggNOG, a total of 24 terms were enriched, with the top two terms being cell wall/membrane/envelope biogenesis and intracellular trafficking, secretion, and vesicular transport (Supplementary Figure S4).

Genome evolution analysis

Orthologs of the leopard coral grouper and 12 other fish species (mentioned above) were predicted. A total of 16 974 gene families were obtained (Figure 3A), including 1 211 (9.7%) common single-copy gene families. In the leopard coral grouper, the number of single-copy, multi-copy, and unique genes were 4 200, 4 040, and 384, respectively (Figure 3B). The common single-copy genes were used to construct the phylogenetic tree. Brown-marbled groper was most closely related to the leopard coral grouper, followed by the large yellow croaker. Two other Perciformes species, tilapia and *L. calcarifer*, were relatively distant from the leopard coral grouper. Tilapia and medaka were clustered into one clade, whereas *L. calcarifer* and Asian swamp eel were clustered into another clade (Figure 3D). The genomes of brown-marbled groper and leopard coral grouper were compared (Figure 3C) and showed relatively high genomic synteny.

Divergence times were predicted using single-copy orthologous genes. The divergence time of leopard coral grouper and brown-marbled groper was about 49.3 (32.5–65.9) million years ago. The ancestor of the grouper separated from the large yellow croaker about 75.7 (62.9–96.9) million years ago (Figure 4A). Changes in gene families were also predicted. A total of 39 expanded and 86 contracted gene families were observed in the leopard coral grouper (Figure 4A). The expanded gene families were mainly enriched in acetylcholine biosynthetic process, chitin catabolic process, and immune response-related terms based on the GO database (Figure 4B and Supplementary Table S7). The gene families enriched in acetylcholine biosynthetic process were high affinity choline transporters. The genes families associated with chitin catabolic process were acidic mammalian chitinases (AMCase). The gene families enriched in immune response were immunoglobulin heavy chain (IGH) and beta-2-microglobulin (B2M), PYPAF1, NACHT, LRR, and PYD domain-containing protein (NLRP) (Figure 4C).

Detection of DEGs associated with carotenoid between fish with different body colors using RNA-seq

To analyze the differences in transcriptome profiles among leopard coral groupers with different body colors, we obtained the skin transcriptome profiles of three bright red and three brown leopard coral groupers using Illumina sequencing. A total of 47.42 Gb of raw data were obtained with an average Q20 of 96.7% and Q30 of 96.2% (Table S8). Based on the edgeR and Limma methods, a total of 323 DEGs were discovered, including 31 and 302 up-regulated DEGs in brown and bright red fish, respectively (Supplementary Figure S5). Potential genes associated with skin color and carotenoid metabolism were collected (Table 1). In the up-regulated DEGs of brown fish, an important gene associated with oxidoreduction of carotenoid, beta-carotene 9',10'-oxygenase (BCO2), was discovered. In red fish, several genes associated with carotenoid and liquid transportation were upregulated, such as low-density lipoprotein receptor-related protein 11 (LRP11) and angiopoietin-related proteins (ANGPTLs).
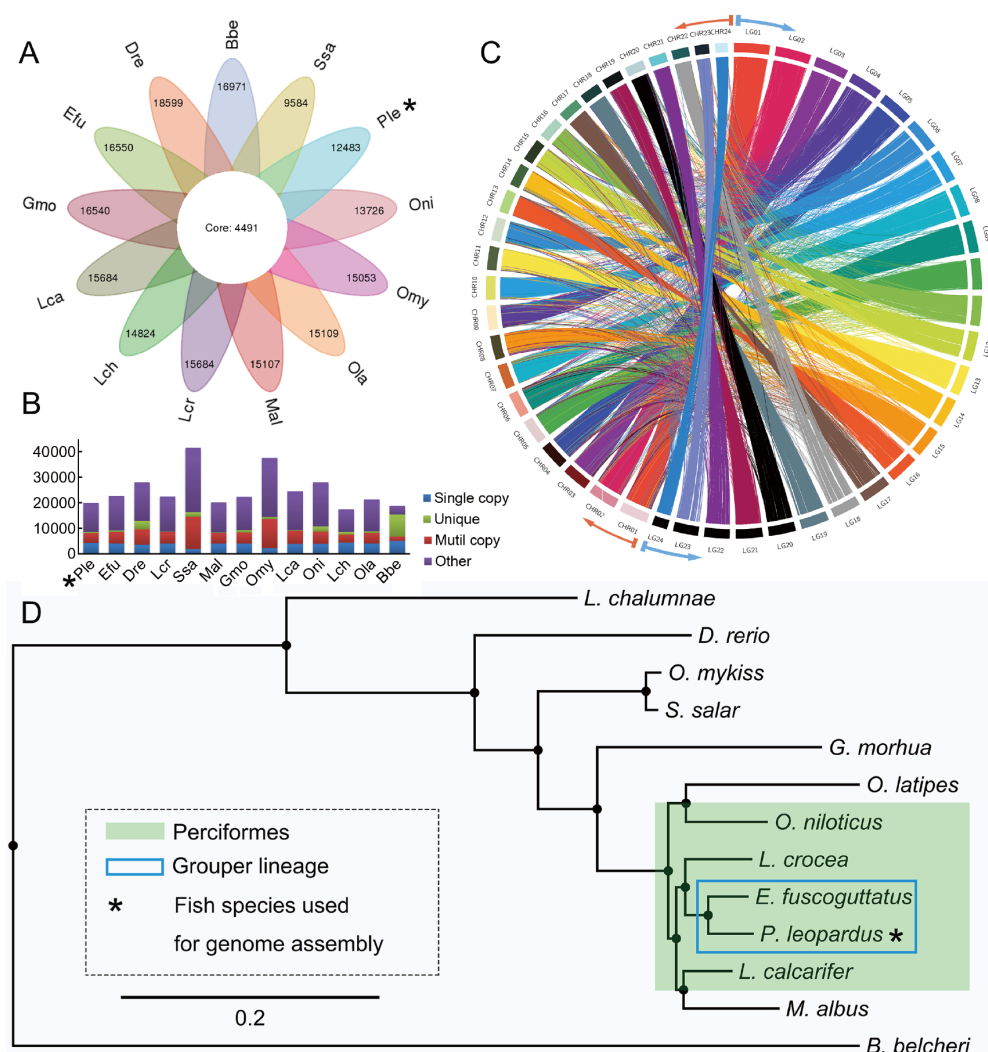
Figure 3 Prediction of gene families and analysis of genetic relationship

A: Flower chart of gene family numbers in 13 fish species. Ple: *P. leopardus*, Efu: *E. fuscoguttatus*, Dre: *D. rerio*, Lcr: *L. crocea*, Ssa: *S. Salar*, Mal: *M. albus*, Gmo: *G. morhua*, Omy: *O. mykiss*, Lca: *L. calcarifer* Oni: *O. niloticus*, Lch: *L. chalumnae*, Ola: *O. latipes*, Bbe: *B. belcheri*. B: Gene family clustering for leopard coral grouper and other 12 fish species. C: Collinearity analysis of leopard coral grouper and brown-marbled grouper, orange arrows indicate genome of leopard coral grouper and blue arrows indicate genome of brown-marbled grouper. D: Phylogenetic tree of leopard coral grouper and other 12 fish species, with *Branchiostoma belcheri* set as outgroup.

DISCUSSION

Leopard coral groupers are an important marine fish species in coral reef ecosystems and have huge commercial value. This species is an excellent research model due to its special biological characteristics and evolutionary status. In this study, we first assembled a high-quality chromosome-level leopard coral grouper genome, thus providing pivotal genomic resources for further studies on the evolution, molecular-assisted selection, and genetic mechanisms of the special biological characteristics of groupers.

Genome assembly and assessment

Estimation of basic genome characteristics is important for

adoption of appropriate assembly strategies. The K-mer method is an efficient tool for estimating genome properties (Liu et al., 2013). The estimated genome size of the leopard coral grouper was 858 Mb, higher than the final genome size (784.57 Mb) based on PacBio sequencing data. This difference may be due to collapse of repetitive elements and statistical differences in the methods. Similar results have also been reported in previous research, e.g., scallop (*Patinopecten yessoensis*) (Wang et al., 2017), Murray cod (Austin et al., 2017), and pike-perch (Nguinkal et al., 2019). In addition, the leopard coral grouper genome was smaller than that of other grouper genomes, including 1.135 Gb for red-spotted grouper (Ge et al., 2019), 1.086 Gb for giant grouper
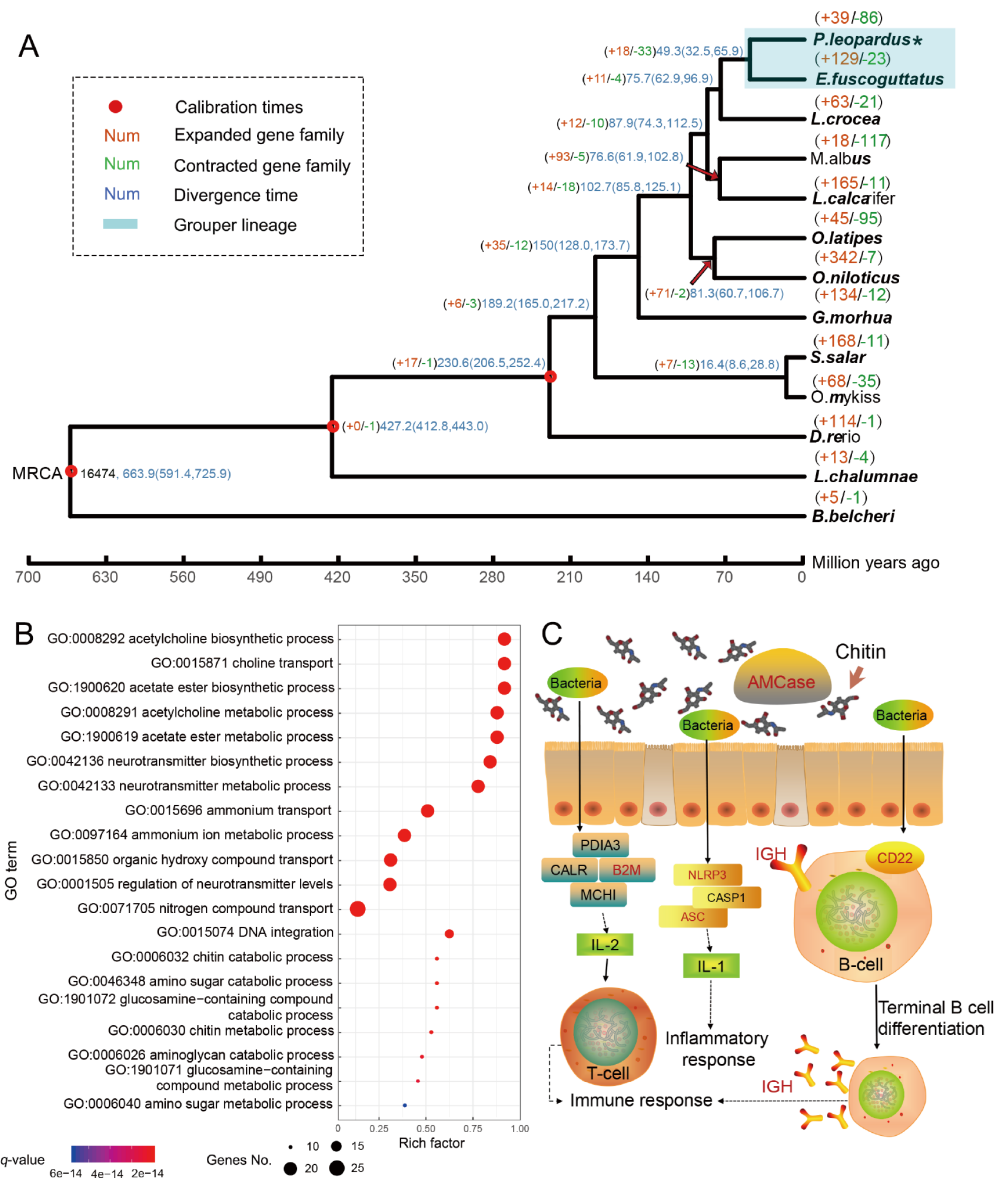
Figure 4 Analysis of divergence time and gene family of leopard coral grouper with other fish species

A: Phylogenetic tree with dynamic evolution of gene families and divergence time. MRCA: Most recent common ancestor. B: GO enrichment of expanded gene families in leopard coral grouper. C: Sketch map of expanded gene families related to immune response. Red font indicates expanded gene families.

(Zhou et al., 2019a), 1.047 Gb for brown-marbled grouper (not published), and 1.02 Gb for orange-spotted grouper (*Epinephelus coioides*, unpublished data). The genome sizes of most Perciformes species were smaller than that of the leopard coral grouper (Supplementary Table S9), including two closely related species, i.e., large yellow croaker and three-spine stickleback (*Gasterosteus aculeatus*). Therefore, our study suggests that the leopard coral grouper is in a basal evolutionary status in the grouper lineage, which agrees with previous studies onf molecular and morphological phylogeny (Craig & Hastings, 2007). Moreover, we also observed lower

repeat content in the leopard coral grouper (34.6%) than in the red-spotted grouper (43.02%) and giant grouper (41.01%) (Ge et al., 2019, Zhou et al., 2019a). Strong correlations between repeat content and the genome have been observed in Perciformes (Nguinkal et al., 2019). Heterozygosity was approximately 0.42% in the leopard coral grouper, higher than that in the red-spotted grouper (*Epinephelus akaara*, 0.375%) (Ge et al., 2019), Murray cod (*Maccullochella peelii*, 0.113%) (Austin et al., 2017), pike-perch (*Sander lucioperca*, 0.14%) (Nguinkal et al., 2019), *Triplophysa tibetana* (0.1%) (Yang et al., 2019a), and male (*Ctenopharyngodon idellus*, 0.25%) and

Table 1 Differentially expressed genes (DEGs) between brown and red leopard coral grouper transcriptomes

ID	log2 (Foldchange)	*q*-value	Description	Gene symbol
Up-regulated candidate DEGs associated with skin color in brown fish				
Chr04.g04131.m1	−2.64	2.12E−04	Beta-carotene 9',10'-oxygenase	BCO2
Chr12.g11897.m1	−2.41	1.14E−03	Eosinophil peroxidase	EPX
Up-regulated candidate DEGs associated with skin color in red fish				
Chr16.g15722.m1	2.60	2.08E−04	Low-density lipoprotein receptor-related protein 11	LRP11
Chr05.g04888.m1	2.13	7.79E−04	Angiopoietin-related protein 4	ANGPTL4
Chr15.g14859.m1	3.80	8.48E−05	High affinity cationic amino acid transporter 1	SLC7A1
Chr10.g09989.m1	2.68	7.40E−04	Protein Wnt-2b	WNT2B

female grass carp (0.09%) (Wang et al., 2015b). The relatively high heterozygosity implies high population genetic diversity for the leopard coral grouper and a lack of selective breeding in aquaculture.

Though the relatively high heterozygosity increased the difficulty of genome sequencing and assembly in the leopard coral grouper, we obtained the high-quality chromosome-level genome based on PacBio sequencing and Hi-C technology. The mean length and N50 of PacBio reads were 12 893 bp and 18 054 bp, respectively. The contig N50 was 1.14 Mb, higher than that of most published aquaculture species, e.g., arapaima (*Arapaima gigas*, female: 325 Kb and male: 285 Kb) (Du et al., 2019); *Glyptosternon maculatum* (993 Kb) (Shao et al., 2018), sea bass (*Lateolabrax maculatus*, 31K) (Shao et al., 2018), northern snakehead (*Channa argus,* 81.4K) (Liu et al., 2017), and blunt snout bream (*Megalobrama amblycephala*, 49.4K) (Liu et al., 2017). Lastly, a total of 1 091 contigs were anchored to 24 pseudochromosomes of the leopard coral grouper, consistent with the number of chromosomes in other groupers (Ge et al., 2019; Zhou et al., 2019a).

Genome annotation

Through *de novo* prediction and homologue annotation, a total of 22 317 protein-coding genes were detected in the leopard coral grouper genome. The number of predicted genes was slightly less than that in other groupers, such as red-spotted grouper (23 808) (Ge et al., 2019), giant grouper (24 718) (Zhou et al., 2019b), and orange grouper (24 593). These differences may result from the different statistical methods or may be induced by gene family expansion in *Epinephelus* species after dividing from a common ancestor. The number of protein-coding genes was similar to other Perciformes fish such as pike-perch (21 249) (Nguinkal et al., 2019), Chinese sillago (*Sillago sinica*, 22 122) (Zhou et al., 2018), spotted sea bass (*Lateolabrax maculatus*, 22 015), northern snakehead (19 877) (Liu et al., 2017), and Eurasian perch (*Perca fluviatilis*, 23 397) (Ozerov et al., 2018). The comprehensive annotated genes provide essential information for multi-omics analysis of the leopard coral grouper.

Genome evolution analysis

Based on the phylogenetic tree, the leopard coral grouper was found to be most closely related to the brown-marbled grouper, followed by the large yellow croaker. There was significant genomic synteny between the leopard coral grouper and brown-marbled grouper. Several genes in the brown-marbled grouper did not match the leopard coral grouper genome, which may be due to gene family expansion in the brown-marbled grouper. Compared to the leopard coral grouper genome, we observed 129 gene families were expanded after divergence of the two fish species, and just 23 gene families were contracted in the brown-marbled grouper. In the leopard coral grouper, we identified 39 and 86 expanded and contracted gene families, respectively.

The expanded gene families were mainly associated with immune response. Chitin is an important structural component of parasites, bacteria, and fungi, and can fix a multitude of antigenic glycoproteins, and thus its degradation plays a crucial role in determining subsequent immune responses (Van Dyken & Locksley, 2018). Structural chitin presented in microorganisms can be degraded using acidic mammalian chitinases (Elieh Ali Komi et al., 2018). NLRP3s are important regulatory factors in inflammatory responses, which increase cleavage of caspase-1 and levels of ASC and mature IL-1 (Ising et al., 2019). Beta-2-microglobulin is an important component of the class I major histocompatibility complex (MHC) and is involved in the presentation of peptide antigens to the immune system (Sreejit et al., 2014). Moreover, CD22 is an inhibitory receptor in B cell-specific transmembrane and a negative regulator of B cell antigen receptor signaling (Kelm et al., 2002). Immunoglobulins, also called antibodies, are membrane-bound or secreted glycoproteins produced by B lymphocytes with important functions in humoral immunity. Immunoglobulin heavy chains participate in antigen recognition and presentation (McHeyzer-Williams et al., 2012). These expanded gene families suggest that evolutionary changes in immune response and movement ability may have contributed to the adaptation of leopard coral groupers to their habitat. In addition, high-affinity choline transporters were also expended, which are important proteins in regulation of movement (Banerjee et al., 2019). The mutant of this protein can induce severe dyskinesia in humans (Banerjee et al., 2019, Pardal-Fernández et al., 2018). A greater number of these genes were found in the leopard coral grouper than in the brown-marbled grouper, suggesting that the brown-marbled grouper may be more adaptable to benthic life in coral reef ecosystems than the leopard coral grouper.

Detection of DEGs between leopard coral groupers with different body colors using RNA-seq

Skin color in leopard coral groupers is an important commercial trait, which can substantially affect market price. Carotenoids, including astaxanthin, tunaxanthin, α-cryptoxanthin, and adonirubin, are the direct molecules for regulation of body color in leopard coral groupers (Maoka et al., 2017). Individuals with more astaxanthin tend to possess red skin color.

To explore the molecular mechanism of body color differences in leopard coral groupers, three red and three brown fish were used for comparative transcriptome analysis. Among the highly expressed DEGs in brown fish, BCO2 is an important gene in the regulation of carotenoid metabolism, which catalyzes the degradation of carotenoid to induce differences in skin color (Strychalski et al., 2015). This gene may influence the proportion of carotenoid, resulting in the diversity of skin color observed in leopard coral groupers. Several studies have demonstrated that BCO2 is associated with body color. For example, mutation in BCO2 can change skin color in chickens (Eriksson et al., 2008); and nonsense mutation in BCO2 can increase the concentration of carotenoids and change white fat to yellow in sheep (Våge & Boman, 2010). Here, in the red fish, several genes related to carotenoid transport, including LRP11, ANGPTL4, and ANGPTL2, were up-regulated. *De novo* synthesis of carotenoid is difficult in fish (Ho et al., 2013), and thus uptake efficiency can affect accumulation of carotenoids. Carotenoid transport in organisms is important for the digestion of carotenoid. Genes associated with lipid transport are major regulators of skin color and can affect the deposition of pigment (Chintala et al., 2005). Low-density lipoprotein (LDL) and chylomicron are important carriers of carotenoid in plasma (Harrison, 2019). LDL receptor-related protein (LDP) can bind with LDL and chylomicron (Lillis et al., 2015); in addition, ANGPTLs are associated with lipid utilization and storage via regulation of hydrolysis of lipids from chylomicrons and very low density lipoproteins (VLDL) (Lei et al., 2011; Lutz et al., 2001). Therefore, changes in these genes may influence the accumulation of carotenoids and differences of skin color in leopard coral groupers.

CONCLUSIONS

In the present study, we assembled a high-quality chromosome-level genome of the leopard coral grouper using PacBio sequencing and Hi-C technology. A 787.06 Mb genome with contig N50 of 1.14 Mb and GC content of 39.56% was assembled, with 784.57 Mb (99.7%) of the genome anchored to 24 chromosomes with N50 length of 33.8 Mb. A total of 22 317 protein-coding genes were predicted, among which 22 238 (99.65%) were anchored to chromosomes and 20 921 (93.74%) were functionally annotated. This high-quality leopard coral grouper genome provides important genomic resources for further research on this species. Based on the high-quality genome, we also performed genome evolution analyses, which should help to improve our understanding of the adaptive evolution of this species. The comparative transcriptome showed that carotenoid metabolism may play an important role in the regulation of skin color in leopard coral groupers.

DATA AVAILABILITY

The leopard coral grouper genome was deposited in the DNA Data Bank of Japan (DDBJ) database. The accession No. is PRJDB9154. The accession No. of chromosomes are AP022700-AP022723; accession Nos. of contigs unanchored into chromosomes are AP022724-AP022809. Accession No. of transcriptome data for annotation is PRJDB9167. The PacBio sequencing data used for assembly of leopard coral grouper genome was deposited in Genome Sequence Archive (GSA) and accession No. is PRJCA002371. The raw data of RNA-seq for skin analysis of leopard coral grouper was deposited in GSA and accession No. is PRJCA002372.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

Y.Y., Z.N.M., and X.C.L. designed the study. Y.Y. and L.N.W. collected the samples. Y.Y., X.W., and Z.N.M. performed the laboratory work. Y.Y. and J.F.C. performed analyses. Y.Y., Z.N.M., X.C.L., J.H.X., and H.R.L. drafted and revised the paper. All authors read and approved the final version of the manuscript.

REFERENCES

Adams S. 2003. Morphological ontogeny of the gonad of three plectropomid species through sex differentiation and transition. *Journal of Fish Biology*, **63**(1): 22−36.

Anders S, Pyl PT, Huber W. 2015. HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**(2): 166−169.

Austin CM, Tan MH, Harrisson KA, Lee YP, Croft LJ, Sunnucks P, Pavlova A, Gan HM. 2017. De novo genome assembly and annotation of Australia's largest freshwater fish, the Murray cod (*Maccullochella peelii*), from Illumina and Nanopore sequencing read. *GigaScience*, **6**(8): gix063.

Banerjee M, Arutyunov D, Brandwein D, Janetzki-Flatt C, Kolski H, Hume S, Leonard NJ, Watt J, Lacson A, Baradi M, Leslie EM, Cordat E, Caluseriu O. 2019. The novel p.Ser263Phe mutation in the human high-affinity choline transporter 1 (CHT1/SLC5A7) causes a lethal form of fetal akinesia syndrome. *Human Mutation*, **40**(10): 1676−1683.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, **27**(2): 573−580.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**(15): 2114−2120.

Bunt CM, Kingsford MJ. 2014. Movement, habitat utilization and behaviour

of coral trout *Plectropomus leopardus* during and after the reproductive period on the southern Great Barrier Reef. *Marine Ecology Progress Series*, **496**: 33−45.

Burgess AI, Callan CK, Touse R, Delos Santos M. 2020. Increasing survival and growth in larval leopard coral grouper (*Plectropomus leopardus*) using intensively cultured *Parvocalanus crassirostris* nauplii. *Journal of the World Aquaculture Society*, **51**(1): 171−182.

Cai XP, Qu M, Ding SX, Wang HJ, Wang HJ, Hu LY, Su YQ. 2013. Differentiation of coral trout (*Plectropomus leopardus*) based on an analysis of morphology and complete mitochondrial DNA: are cryptic species present?. *Acta Oceanologica Sinica*, **32**(6): 40−46.

Chen JX, Xiao L, Peng C, Ye ZF, Wang DD, Yang YQ, Zhang HF, Zhao M, Li SS, Lin HR, Zhang Y. 2019. Socially controlled male-to-female sex reversal in the protogynous orange-spotted grouper, *Epinephelus coioides*. *Journal of Fish Biology*, **94**(3): 414−421.

Chen SY, Peng XL, Lin SQ, Xu QB. 2018. Identification of *Plectropomus leopardus*, *Promicrops lanceolatus* and *Cromileptes altivelis* using species-specific TaqMan real-time PCR. *Food Control*, **84**: 408−412.

Chen WQ, Liu ZM, Wu L, Wang YH, Ma JZ, Wu HX. 2016. The influences of light color on growth, haematological and biochemical indices of juvenile *Plectropomus leopardus*. *Chinese Journal of Ecology*, **35**(7): 1889−1895. (in Chinese)

Chintala S, Li W, Lamoreux ML, Ito S, Wakamatsu K, Sviderskaya EV, Bennett DC, Park YM, Gahl WA, Huizing M, Spritz RA, Ben ST, Novak EK, Tan J, Swank RT. 2005. *Slc7a11* gene controls production of pheomelanin pigment and proliferation of cultured cells. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(31): 10964−10969.

Conesa A, Götz S, García-Gómez JM, 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.

Craig MT, Hastings PA. 2007. A molecular phylogeny of the groupers of the subfamily Epinephelinae (Serranidae) with a revised classification of the Epinephelini. *Ichthyological Research*, **54**(1): 1−17.

De Bie T, Cristianini N, Demuth JP, Hahn MW. 2006. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics*, **22**(10): 1269−1271.

Du K, Wuertz S, Adolfi M, Kneitz S, Stöck M, Oliveira M, Nóbrega R, Ormanns J, Kloas W, Feron R, Klopp C, Parrinello H, Journot L, He SP, Postlethwait J, Meyer A, Guiguen Y, Schartl M. 2019. The genome of the arapaima (*Arapaima gigas*) provides insights into gigantism, fast growth and chromosomal sex determination system. *Scientific Reports*, **9**(1): 5293.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, Aiden EL. 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*, **356**(6333): 92−95.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**(5): 1792−1797.

Edgar RC, Myers EW. 2005. PILER: identification and classification of genomic repeats. *Bioinformatics*, **21**(S1): i152−i158.

Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, deWinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong XX, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma CC, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen GE, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S. 2009. Real-time DNA sequencing from single polymerase molecules. *Science*, **323**(5910): 133−138.

Elieh Ali Komi D, Sharma L, Dela Cruz CS. 2018. Chitin and its effects on inflammatory and immune responses. *Clinical Reviews in Allergy & Immunology*, **54**(2): 213−223.

Eriksson J, Larson G, Gunnarsson U, Bed'hom B, Tixier-Boichard M, Strömstedt L, Wright D, Jungerius A, Vereijken A, Randi E, Jensen P, Andersson L. 2008. Identification of the *Yellow skin* gene reveals a hybrid origin of the domestic chicken. *PLoS Genetics*, **4**(2): e1000010.

Ge H, Lin KB, Shen M, Wu SQ, Wang YL, Zhang ZP, Wang ZY, Zhang Y, Huang Z, Zhou C, Lin Q, Wu JS, Liu L, Huang ZC, Zheng LY. 2019. De novo assembly of a chromosome-level reference genome of red-spotted grouper (*Epinephelus akaara*) using nanopore sequencing and Hi-C. *Molecular Ecology Resources*, **19**(6): 1461−1469.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**(3): 307−321.

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology*, **9**(1): R7.

Harrison EH. 2019. Mechanisms of transport and delivery of vitamin a and carotenoids to the retinal pigment epithelium. *Molecular Nutrition and Food Research*, **63**(15): 1801046.

Ho ALFC, O'shea SK, Pomeroy HF. 2013. Dietary esterified astaxanthin effects on color, carotenoid concentrations, and compositions of clown anemonefish, *Amphiprion ocellaris*, skin. *Aquaculture International*, **21**(2): 361−374.

Ising C, Venegas C, Zhang SS, Scheiblich H, Schmidt SV, Vieira-Saecker A, Schwartz S, Albasset S, McManus RM, Tejera D, Griep A, Santarelli F, Brosseron F, Opitz S, Stunden J, Merten M, Kayed R, Golenbock DT, Blum D, Latz E, Buée L, Heneka MT. 2019. NLRP3 inflammasome activation drives tau pathology. *Nature*, **575**(7784): 669−673.

Jones P, Binns D, Chang HY, Fraser M, Li WZ, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics*, **30**(9): 1236−1240.

Keilwagen J, Wenk M, Erickson JL, Schattat MH, Grau J, Hartung F. 2016. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Research*, **44**(9): e89.

Kelm S, Gerlach J, Brossmer R, Danzer CP, Nitschke L. 2002. The ligand-binding domain of CD22 is needed for inhibition of the B cell receptor signal, as demonstrated by a novel human CD22-specific inhibitor compound. *Journal of Experimental Medicine*, **195**(9): 1207−1213.

Khasanah M, Kadir NN, Jompa J. 2019. Reproductive biology of three important threatened/near-threatened groupers (*Plectropomus leopardus*, *Epinephelus polyphekadion* and *Plectropomus areolatus*) in eastern Indonesia and implications for management. *Animals*, **9**(9): 643.

Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, **12**(4): 357−360.

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *κ*-mer weighting and repeat separation. *Genome Research*, **27**(5): 722−736.

Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: a resource for timelines, timetrees, and divergence times. *Molecular Biology and Evolution*, **34**(7): 1812−1819.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. *Genome Biology*, **5**(2): R12.

Lei X, Shi FJ, Basu D, Huq A, Routhier S, Day R, Jin WJ. 2011. Proteolytic processing of angiopoietin-like protein 4 by proprotein convertases modulates its inhibitory effects on lipoprotein lipase activity. *Journal of Biological Chemistry*, **286**(18): 15747−15756.

Leskovec J, Sosič R. 2016. SNAP: A general purpose network analysis and graph mining library. *ACM Transactions on Intelligent Systems and Technology*, **8**(1): 1.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, **26**(5): 589−595.

Li RQ, Fan W, Tian G, Zhu HM, He L, Cai J, Huang QF, Cai QL, Li B, Bai YQ, Zhang ZH, Zhang YP, Wang W, Li J, Wei FW, Li H, Jian M, Li JW, Zhang ZL, Nielsen R, Li DW, Gu WJ, Yang ZT, Xuan ZL, Ryder OA, Leung FCC, Zhou Y, Cao JJ, Sun X, Fu YG, Fang XD, Guo XS, Wang B, Hou R, Shen FJ, Mu B, Ni PX, Lin RM, Qian WB, Wang GD, Yu C, Nie WH, Wang JH, Wu ZG, Liang HQ, Min JM, Wu Q, Cheng SF, Ruan J, Wang MW, Shi ZB, Wen M, Liu BH, Ren XL, Zheng HS, Dong D, Cook K, Shan G, Zhang H, Kosiol C, Xie XY, Lu ZH, Zheng HC, Li YR, Steiner CC, Lam TTY, Lin SY, Zhang QH, Li GQ, Tian J, Gong TM, Liu HD, Zhang DJ, Fang L, Ye C, Zhang JB, Hu WB, Xu AL, Ren YY, Zhang GJ, Bruford MW, Li QB, Ma LJ, Guo YR, An N, Hu YJ, Zheng Y, Shi YY, Li ZQ, Liu Q, Chen YL, Zhao J, Qu N, Zhao SC, Tian F, Wang XL, Wang HY, Xu LZ, Liu X, Vinar T, Wang YJ, Lam TW, Yiu SM, Liu SP, Zhang HM, Li DS, Huang Y, Wang X, Yang GH, Jiang Z, Wang JY, Qin N, Li L, Li JX, Bolund L, Kristiansen K, Wong GKS, Olson M, Zhang XQ, Li SG, Yang HM, Wang J, Wang J. 2010. The sequence and *de novo* assembly of the giant panda genome. *Nature*, **463**(7279): 311−317.

Lieberman-Aiden E, Van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**(5950): 289−293.

Lillis AP, Muratoglu SC, Au DT, Migliorini M, Lee MJ, Fried SK, Mikhailenko I, Strickland DK. 2015. LDL receptor-related protein-1 (LRP1) regulates cholesterol accumulation in macrophages. *PLoS One*, **10**(6): e0128903.

Liu BH, Shi YJ, Yuan JY, Hu XS, Zhang H, Li N, Li ZY, Chen YX, Mu DS, Fan W. 2013. Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. arXiv: 1308.2012.

Liu H, Chen CH, Gao ZX, Min JM, Gu YM, Jian JB, Jiang XW, Cai HM, Ebersberger I, Xu M, Zhang XH, Chen JW, Luo W, Chen BX, Chen JH, Liu H, Li J, Lai RF, Bai MZ, Wei J, Yi SK, Wang HL, Cao XJ, Zhou XY, Zhao YH, Wei KJ, Yang RB, Liu BN, Zhao SC, Fang XD, Schartl M, Qian XQ, Wang WM. 2017. The draft genome of blunt snout bream (*Megalobrama amblycephala*) reveals the development of intermuscular bone and adaptation to herbivorous diet. *GigaScience*, **6**(7): gix039.

Lomsadze A, Burns PD, Borodovsky M. 2014. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Research*, **42**(15): e119.

Lowe TM, Eddy SR. 1996. TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, **25**(5): 955−964.

Lutz EP, Merkel M, Kako Y, Melford K, Radner H, Breslow JL, Bensadoun A, Goldberg IJ. 2001. Heparin-binding defective lipoprotein lipase is unstable and causes abnormalities in lipid delivery to tissues. *Journal of Clinical Investigation*, **107**(9): 1183−1192.

Ma KY, Craig MT, Choat JH, Van Herwerden L. 2016. The historical biogeography of groupers: clade diversification patterns and processes. *Molecular Phylogenetics and Evolution*, **100**: 21−30.

Majoros WH, Pertea M, Salzberg SL. 2004. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics*, **20**(16): 2878−2879.

Maoka T, Sato W, Nagai H, Takahashi T. 2017. Carotenoids of red, brown, and black specimens of *plectropomus leopardus*, the coral trout (Suziara in Japanese). *Journal of Oleo Science*, **66**(6): 579−584.

McHeyzer-Williams M, Okitsu S, Wang N, McHeyzer-Williams L. 2012. Molecular programming of B cell memory. *Nature Reviews Immunology*, **12**(1): 24−34.

Mekuchi M, Sakata K, Yamaguchi T, Koiso M, Kikuchi J. 2017. Trans-omics approaches used to characterise fish nutritional biorhythms in leopard coral grouper (*Plectropomus leopardus*). *Scientific Reports*, **7**(1): 9372.

Melianawati R, Astuti NWW, Suwirya K. 2013. The use of copepods to improve juveniles production of coral trout *Plectropomus leopardus* (Lacepède, 1802). *Middle East Journal of Scientific Research*, **16**(2): 237−244.

Metcalfe SS, Kroon FJ, Beale DJ, Miller G. 2018. Development of a validation protocol of enzyme immunoassay kits used for the analysis of steroid hormones in fish plasma. *Journal of Experimental Marine Biology and Ecology*, **499**: 26−34.

Nguinkal JA, Brunner RM, Verleih M, Rebl A, Ríos-Pérez L, Schäfer N, Hadlich F, Stüeken M, Wittenburg D, Goldammer T. 2019. The first highly contiguous genome assembly of pikeperch (*Sander lucioperca*), an emerging aquaculture species in Europe. *Genes*, **10**(9): 708.

Ozerov MY, Ahmad F, Gross R, Pukk L, Kahar S, Kisand V, Vasemägi A. 2018. Highly continuous genome assembly of eurasian perch (*Perca fluviatilis*) using linked-read sequencing. *G3: Genes, Genomes, Genetics*, **8**(12): 3737−3743.

Pardal-Fernández JM, Carrascosa-Romero MC, Álvarez S, Medina-Monzón MC, Caamaño MB, De Cabo C. 2018. A new severe mutation in the SLC5A7 gene related to congenital myasthenic syndrome type 20. *Neuromuscular Disorders*, **28**(10): 881−884.

Pryszcz LP, Gabaldón T. 2016. Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Research*, **44**(12): e113.

Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, **13**(1): 341.

Rimmer MA, Glamuzina B. 2019. A review of grouper (Family Serranidae:

Subfamily Epinephelinae) aquaculture from a sustainability science perspective. *Reviews in Aquaculture*, **11**(1): 58−87.

Ritchie ME, Phipson B, Wu D, Hu YF, Law CW, Shi W, Smyth GK. 2015. *Limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, **43**(7): e47.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**(1): 139−140.

Shao CW, Li C, Wang N, Qin YT, Xu WT, Liu Q, Zhou Q, Zhao Y, Li XH, Liu SS, Chen XW, Mahboob S, Liu X, Chen SL. 2018. Chromosome-level genome assembly of the spotted sea bass, *Lateolabrax maculatus*. *GigaScience*, **7**(11): giy114.

Sreejit G, Ahmed A, Parveen N, Jha V, Valluri VL, Ghosh S, Mukhopadhyay S. 2014. The ESAT-6 protein of *Mycobacterium tuberculosis* interacts with Beta-2-Microglobulin (β2M) affecting antigen presentation function of macrophage. *PLoS Pathogens*, **10**(10): e1004446.

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. Augustus: *ab* initio prediction of alternative transcripts. *Nucleic Acids Research*, **34**(S2): W435−W439.

Strychalski J, Brym P, Czarnik U, Gugołek A. 2015. A novel AAT-deletion mutation in the coding sequence of the BCO2 gene in yellow-fat rabbits. *Journal of Applied Genetics*, **56**(4): 535−537.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, **25**(9): 1105−1111.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, **7**(3): 562−578.

Våge DI, Boman IA. 2010. A nonsense mutation in the *beta-carotene oxygenase 2* (*BCO2*) gene is tightly associated with accumulation of carotenoids in adipose tissue in sheep (*Ovis aries*). *BMC Genetics*, **11**(1): 10.

Van Dyken SJ, Locksley RM. 2018. Chitins and chitinase activity in airway diseases. *Journal of Allergy and Clinical Immunology*, **142**(2): 364−369.

Van Herwerden L, Howard Choat J, Newman SJ, Leray M, Hillersøy G. 2009. Complex patterns of population structure and recruitment of *Plectropomus leopardus* (Pisces: Epinephelidae) in the Indo-West Pacific: implications for fisheries management. *Marine Biology*, **156**(8): 1595−1607.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng QD, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, **9**(11): e112963.

Wang L, Yu CP, Guo L, Lin HR, Meng ZN. 2015a. In silico comparative transcriptome analysis of two color morphs of the common coral trout (*Plectropomus leopardus*). *PLoS One*, **10**(12): e0145868.

Wang S, Zhang JB, Jiao WQ, Li J, Xun XG, Sun Y, Guo XM, Huan P, Dong B, Zhang LL, Hu XL, Sun XQ, Wang J, Zhao CT, Wang YF, Wang DW, Huang XT, Wang RJ, Lv J, Li YL, Zhang ZF, Liu BZ, Lu W, Hui YY, Liang J, Zhou ZC, Hou R, Li X, Liu YC, Li HD, Ning XH, Lin Y, Zhao L, Xing Q, Dou JZ, Li YP, Mao JX, Guo HB, Dou HQ, Li TQ, Mu C, Jiang WK, Fu Q, Fu XT, Miao Y, Liu J, Yu Q, Li RJ, Liao H, Li X, Kong YF, Jiang Z, Chourrout D, Li RQ, Bao ZM. 2017. Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nature Ecology & Evolution*, **1**(5):

0120.

Wang YP, Lu Y, Zhang Y, Ning ZM, Li Y, Zhao Q, Lu HY, Huang R, Xia XQ, Feng Q, Liang XF, Liu KY, Zhang L, Lu TT, Huang T, Fan DL, Weng QJ, Zhu CR, Lu YQ, Li WJ, Wen ZR, Zhou CC, Tian QL, Kang XJ, Shi MJ, Zhang WT, Jang SH, Du FK, He S, Liao LJ, Li YM, Gui B, He HH, Ning Z, Yang C, He L, Luo LF, Yang R, Luo Q, Liu XC, Li SS, Huang W, Xiao L, Lin HR, Han B, Zhu ZY. 2015b. The draft genome of the grass carp (*Ctenopharyngodon idellus*) provides insights into its evolution and vegetarian adaptation. *Nature Genetics*, **47**(6): 625−631.

Xie ZZ, Yu CP, Guo L, Li MM, Yong Z, Liu XC, Meng ZN, Lin HR. 2016. Ion Torrent next-generation sequencing reveals the complete mitochondrial genome of black and reddish morphs of the Coral Trout *Plectropomus leopardus*. *Mitochondrial DNA*, **27**(1): 609−612.

Xu L, Zhang Y, Su Y, Liu L, Yang J, Zhu YY, Li CY. 2010. Structure and evolution of full-length LTR retrotransposons in rice genome. *Plant Systematics and Evolution*, **287**(1): 19−28.

Yang XF, Liu HP, Ma ZH, Zou Y, Zou M, Mao YZ, Li XM, Wang H, Chen TS, Wang WM, Yang RB. 2019a. Chromosome-level genome assembly of *Triplophysa tibetana*, a fish adapted to the harsh high-altitude environment of the Tibetan Plateau. *Molecular Ecology Resources*, **19**(4): 1027−1036.

Yang Y, Wu LN, Wu X, Li BJ, Huang WH, Weng ZY, Lin ZX, Song LL, Guo Y, Meng ZN, Liu XC, Xia JH. 2020. Identification of candidate growth-related SNPs and genes using GWAS in brown-marbled grouper (*Epinephelus fuscoguttatus*). *Marine Biotechnology*, **22**(2): 153−166.

Yang Y, Zhou HQ, Hou LP, Xing K, Shu H. 2019b. Transcriptional profiling of skeletal muscle reveals starvation response and compensatory growth in *Spinibarbus hollandi*. *BMC Genomics*, **20**(1): 938.

Yang ZH. 1997. Paml: a program package for phylogenetic analysis by maximum likelihood. *Bioinformatics*, **13**(5): 555−556.

Yu W, Wen GL, Lin HZ, Yang YK, Huang XL, Zhou CP, Zhang ZW, Duan YF, Huang Z, Li T. 2018. Effects of dietary *Spirulina platensis* on growth performance, hematological and serum biochemical parameters, hepatic antioxidant status, immune responses and disease resistance of Coral trout *Plectropomus leopardus* (Lacepede, 1802). *Fish & Shellfish Immunology*, **74**: 649−655.

Zhou Q, Gao HY, Zhang Y, Fan GY, Xu H, Zhai JM, Xu WT, Chen ZF, Zhang H, Liu SS, Niu YP, Li WS, Li WM, Lin HR, Chen SL. 2019a. A chromosome-level genome assembly of the giant grouper (*Epinephelus lanceolatus*) provides insights into its innate immunity and rapid growth. *Molecular Ecology Resources*, **19**(5): 1322−1332.

Zhou S, Song DL, Zhou XF, Mao XL, Zhou XF, Wang SL, Wei JG, Huang YH, Wang WX, Xiao SM, Qin QW. 2019b. Characterization of *Bacillus subtilis* from gastrointestinal tract of hybrid Hulong grouper (*Epinephelus fuscoguttatus* × *E. lanceolatus*) and its effects as probiotic additives. *Fish & Shellfish Immunology*, **84**: 1115−1124.

Zhou ZX, Chen L, Dong CJ, Peng WZ, Kong SN, Sun JS, Pu F, Chen BH, Feng JX, Xu P. 2018. Genome-scale association study of abnormal scale pattern in Yellow River carp identified previously known causative gene in European Mirror Carp. *Marine Biotechnology*, **20**(5): 573−583.

Zhuang X, Qu M, Zhang X, Ding SX. 2013. A comprehensive description and evolutionary analysis of 22 grouper (Perciformes, *Epinephelidae*) mitochondrial genomes with emphasis on two novel genome organizations. *PLoS One*, **8**(8): e73561.