



Chromosome Genome Assembly of the Leopard Coral Grouper (*Plectropomus leopardus*) With Nanopore and Hi-C Sequencing Data

Yongbo Wang^{1†}, Xin Wen^{2†}, Xinhui Zhang^{3†}, Shuyuan Fu^{1†}, Jinye Liu¹, Wei Tan¹, Ming Luo¹, Longlong Liu¹, Hai Huang¹, Xinxin You³, Jian Luo^{2*} and Fuxiao Chen^{1*}

¹ Key Laboratory of Utilization and Conservation for Tropical Marine Bioresources of Education of Ministry, Hainan Academy of Ocean and Fisheries Sciences, Hainan Tropical Ocean University, Haikou, China, ² State Key Laboratory of Marine Resource Utilization in South China Sea, Hainan Aquaculture Breeding Engineering Research Center, Hainan University, Haikou, China, ³ Shenzhen Key Lab of Marine Genomics, Guangdong Provincial Key Lab of Molecular Breeding in Marine Economic Animals, BGI Academy of Marine Sciences, BGI Marine, Shenzhen, China

OPEN ACCESS

Edited by:

Peter Dovc,
University of Ljubljana, Slovenia

Reviewed by:

Xiaozhu Wang,
Auburn University, United States
Yulin Jin,
Emory University, United States

*Correspondence:

Jian Luo
luojian@hainanu.edu.cn
Fuxiao Chen
cfx69@163.com

[†]These authors have contributed equally to this work

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 06 April 2020

Accepted: 17 July 2020

Published: 02 September 2020

Citation:

Wang Y, Wen X, Zhang X, Fu S, Liu J, Tan W, Luo M, Liu L, Huang H, You X, Luo J and Chen F (2020) Chromosome Genome Assembly of the Leopard Coral Grouper (*Plectropomus leopardus*) With Nanopore and Hi-C Sequencing Data. *Front. Genet.* 11:876. doi: 10.3389/fgene.2020.00876

Keywords: chromosome-level genome, *Plectropomus leopardus*, leopard coral grouper, Nanopore sequencing, Hi-C technologies

INTRODUCTION

As the storehouse of life information, the genome of an organism harbours all of its biological aspects and evolutionary history. Research conducted at the genomic level has become more common, providing important breakthroughs for the comprehensive interpretation of species. In this respect, numerous, large populations of fish species live in the diverse habitats worldwide, and their genome information presents a valuable genetic resource for fisheries. Exploring the massive genetic information contained in the genomes of fishes can not only reveal the adaptive mechanisms of these organisms to various aquatic habitats, but also help to clarify the gene regulatory networks and mechanisms of the economically relevant traits and important life history phenomena.

Over the past decade, researchers have revealed much fish genome information and associated characteristics (You et al., 2020). For example, whole genome sequencing of Atlantic cod (*Gadus morhua*) revealed its special immune mechanism (Star et al., 2011), and likewise demonstrated the doubling mechanism of the Atlantic salmon (*Salmo salar*) genome (Lien et al., 2016). Analysis of the *Paralichthys olivaceus* genome shows that retinoic acid plays an important role in its eye movement and metamorphic development, achieved via the double antagonistic regulation of thyroxine and retinoic acid (Shao et al., 2017). In constructing the whole genome fine map of channel catfish (*Ictalurus punctatus*), Liu et al. (2016) uncovered the mechanism of its scale formation, and a study of the whole genome of *Leuciscus waleckii* elucidated its alkaline environment adaptation mechanism (Xu et al., 2017), to name a few impressive cases. Besides providing insight to molecular mechanisms underpinning biological characteristics, decoding genome information of fish could be used to lay a sound theoretical foundation for distinguishing the genomic location of key economic traits. For example, based on genome-wide association analysis, disease resistance characters of channel catfish were mapped (Geng et al., 2015), and the SNP (single nucleotide polymorphism) loci related to fat content characters of carp were found by GWAS (Genome-Wide Association Studies) analysis (Zheng et al., 2016). In terms of their breeding, genomic selection and breeding technologies on growth and disease resistance, respectively, have been carried out for economic fish species such as Atlantic salmon (*Salmo salar*) (Ødegård et al., 2014), rainbow trout (*Oncorhynchus mykiss*) (Vallejo et al., 2016) and European sea

bass (*Dicen trarachus labrax*) (Palaiokostas et al., 2018). In sum, harnessing genomic information can provide an efficient platform for the in-depth study of the biological and economic characteristics of fish.

The leopard coral grouper, *Plectropomus leopardus*, belongs to the Serranidae family of Perciformes (Morris et al., 2000). It is an important commercial marine fish, being both delicious as sea food and colourful as an aquarium fish (Greenfiel, 2002). This species, due to the high economic price it commands, has been overfished and is now considered under threat by the International Union for Conservation of Nature (IUCN) (Morris et al., 2000). Currently, the genetic resources of the fish are still scarcely known, which greatly hinders both the study and conservation of this species (Wang et al., 2015). Like other coral reef fishes, the leopard coral grouper is capable of displaying a variety of body colours (Wu et al., 2016), which can change rapidly in response to light, food,

disease and other stresses (Kingsford, 1992; Wang et al., 2015). In our view, it is a perfect representative model for studying the genetic mechanism of body colouring in coral reef fishes. Moreover, the leopard coral grouper can be used as a material for better understanding the mechanism of melanoma (Lerebours et al., 2016), and for gauging the impact of global warming on coral reef ecosystems (Messmer et al., 2017). The decoding of *P. leopardus*'s genome information could yield insight into its ecological significance and accelerate its genetic breeding applications.

In this study, we provide the chromosome-level genome assembly of leopard coral grouper by using Nanopore sequencing and high-throughput chromosome conformation capture (Hi-C) technologies. Our intent is to illustrate and decipher the genome information of a leopard coral grouper and lay a theoretical foundation for the analysis of its body-colour mechanism. This genome resource will be useful for the future conservation,

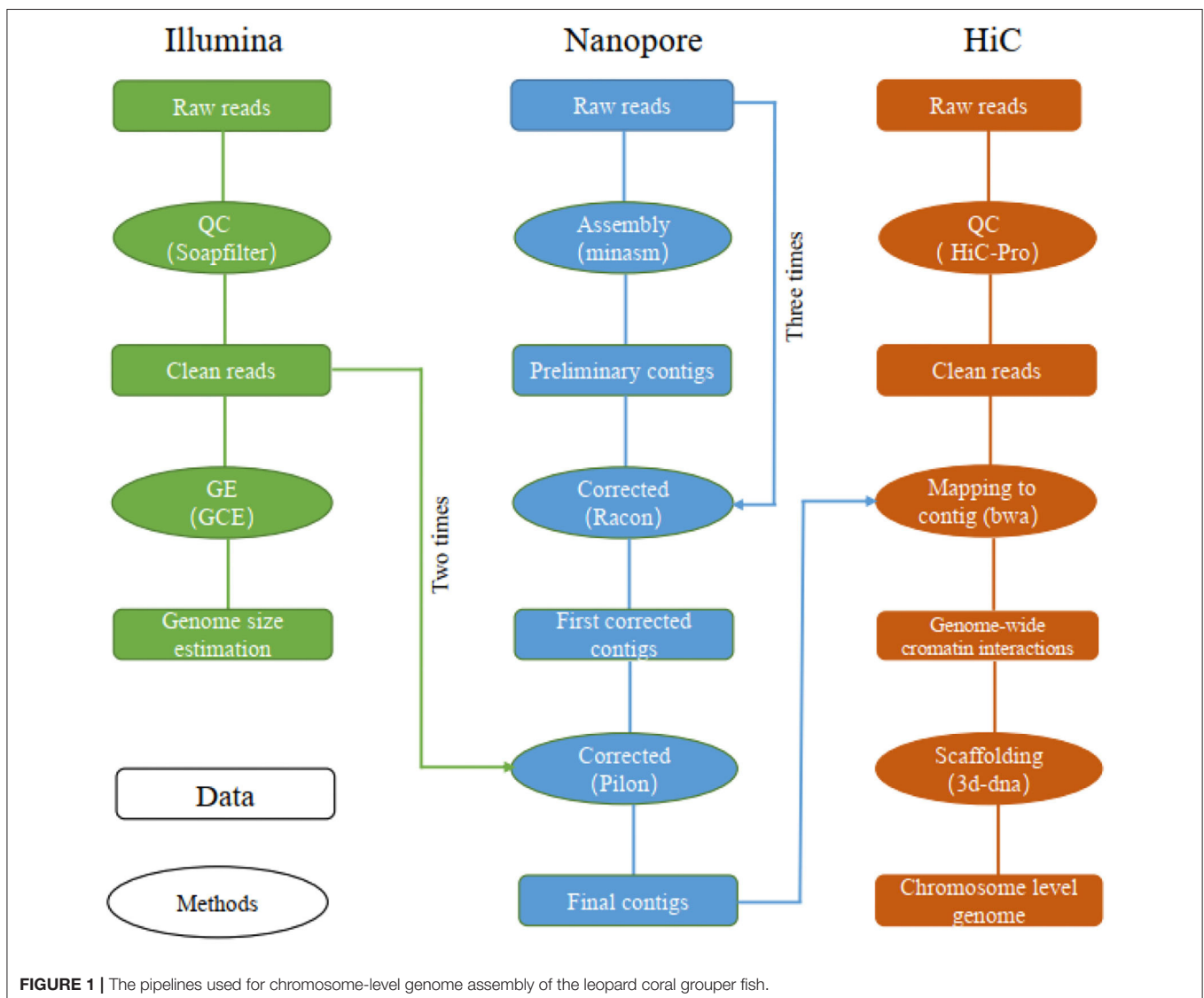


FIGURE 1 | The pipelines used for chromosome-level genome assembly of the leopard coral grouper fish.

molecular breeding, and population genetics of the leopard coral grouper.

MATERIALS AND METHODS

Sample Collection, Library Construction, and Sequencing

We collected a female leopard coral grouper from the Qionghai Breeding Base of the Hainan Academy of Ocean and Fisheries Sciences, in Qionghai, China.

To extract DNA from its muscle tissue and blood, a DNA Extraction Kit was used following the manufacturer's protocols. Both the quantity and quality of DNA were determined using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA).

Two paired-end libraries (insert sizes of 500 and 800 bp) were constructed according to standard Illumina procedures. These libraries were sequenced using the HiSeq 2,500 platform (Illumina, San Diego, CA, USA) with the PE 150 bp model. The raw data had any adapters and low-quality reads removed by SOAPfilter (Luo et al., 2012). All the ensuing clean reads were then applied to estimate the genome size of the leopard coral grouper through a k-mer analysis, done in the Genome Characteristics Estimation (GCE) software (Liu et al., 2013).

For each Nanopore library, the gDNA was size-selected (10–50 kb) with a Blue Pippin system (Sage Science, USA) and processed using the Ligation Sequencing 1D kit (SQKLSK109, Oxford Nanopore Technologies, UK) according to the manufacturer's instructions. Library construction and sequencing were done by the GridION X5/PromethION sequencer (Oxford Nanopore Technologies, UK) at the Genome Center of Nextomics (Wuhan, China). Base calling was performed on fast 5 files by using the ONT Albacore software (v1.2.6) (Sutton et al., 2019), and only those “passed filter” reads representing data of generally higher quality were used for further analyses.

The Hi-C technique has been used to construct a chromosome-level scaffold (Dudchenko et al., 2017). Our Hi-C library was constructed according to previously reported procedures (Rao et al., 2014). First, we used formaldehyde to fix the conformation of the HMW gDNA. Then, the fixed DNA was sheared with the MboI restriction enzyme; the 5' overhangs induced in that shearing step were then repaired using biotinylated residues. Following the ligation of blunt-end fragments *in situ*, the isolated DNA was reverse-crosslinked, purified and filtered to remove biotin-containing fragments. Next, DNA fragment end repair, adaptor ligation and polymerase chain reaction (PCR) were performed successively. Finally, the Hi-C raw data were sequenced on the Illumina HiSeq X platform in its 150 bp PE mode.

For the gene annotation of leopard coral grouper genome, transcriptome sequencing was carried out with the muscle tissue of *P. leopardus*. The total RNA was extracted using a Trizol reagent (Invitrogen, Carlsbad, CA, USA) and purified using an RNeasy Animal Mini Kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. Agilent 2,100 (Agilent Technologies,

Palo Alto, CA) was applied to determine the RNA concentration and the RNA integrity number (RIN). The cDNA library was constructed following the manufacturer's instructions (Illumina, San Diego, CA). Finally, the library was sequenced on a HiSeq 2,500 platform (Illumina, San Diego, CA) using paired-end 150 bp reads. The clean data were obtained by removing reads containing adapters and low-quality reads (e.g., N more than 5% and the quality value <20) from the raw data.

Genome Assembly and Chromosome Anchoring

Reads obtained from the Illumina reads, Nanopore sequencing data, and Hi-C reads of libraries were used separately for different assembly stages (Figure 1). Specifically, the Illumina reads, Nanopore sequencing data and Hi-C reads were obtained for genome size estimation, *de novo* contig assembly, primary scaffolding, genome survey and sequence error-correction, contig assembly and chromosome anchoring, respectively.

The long reads were assembled using “minimap2” (v2.14) and “miniasm” (v0.3) under their default parameters (Li, 2016, 2018). The assembled contigs were corrected by “racon” v1.3.1 by using long reads (repeated three times) (Vaser et al., 2017), followed by two rounds of polishing by “pilon” v1.22 using the Illumina reads (Walker et al., 2014; Michael et al., 2018).

To obtain the chromosome-level genome, we constructed an interaction matrix with the cleaned reads from the Hi-C library by using HiC-Pro (v2.8.0, default parameters and LIGATION_SITE=GATC) (Servant et al., 2015); this was mapped to the *de novo* assembled contigs to construct contacts among the contigs in “bwa” (v0.7.15) with its default parameters (Li and Durbin, 2009). The bam files containing Hi-C linking messages were processed by another round of filtering, in which any reads were removed if they did not map to the assembled

TABLE 1 | Sequencing data used for the leopard coral grouper's genome assembly.

Sequencing strategy	Insert size	Total data (bp)	Reads length (bp)	Sequence coverage (X)
Illumina	500 bp	19,478,316,000	150	20.5
	800 bp	21,927,815,500	125	23.1
Nanopore	–	82,601,731,312	30,352	86.9
Hi-C	400 bp	56,639,825,400	150	59.9
Total	–	180,647,688,212	–	190.5

The sequence coverage values were calculated based on the genome size estimated by the k-mer-based method.

TABLE 2 | Genome assembly statistics for the leopard coral grouper.

Type	Scaffold	Contig
Total length (bp)	913,382,752	912,657,252
Contig N50 length (bp)	40,038,452	1,412,334
Contig N90 length (bp)	28,399,218	217,765
Maximum length (bp)	43,747,248	7,817,432
GC content	39.4%	39.4%

genome within 500 bp from the nearest restriction enzyme site (“juicer” v1.7) (Durand et al., 2016). To assemble the chromosome-level genome based on genomic proximity signals in the Hi-C data, the 3d-dna (v170123) pipeline was used with parameters set to ‘-m haploid -s 0 -c 24’ (Sutton et al., 2019).

Genomic Quality Assessment

To evaluate the quality of the assembled genome, its completeness and accuracy were assessed by

TABLE 3 | Summary of the assembled chromosomes of the leopard coral grouper.

Chromosomes	Length (bp)	Number of contigs	Gene number
Chr1	43,747,248	50	1,182
Chr2	43,338,506	56	987
Chr3	43,327,923	107	1,193
Chr4	41,481,500	58	1,182
Chr5	41,352,528	76	1,163
Chr6	41,116,667	69	1,043
Chr7	41,088,328	76	1,073
Chr8	41,017,068	80	1,033
Chr9	40,373,582	49	1,123
Chr10	40,147,833	80	1,187
Chr11	40,038,452	89	1,191
Chr12	38,795,244	67	1,148
Chr13	35,959,459	57	979
Chr14	35,875,708	36	902
Chr15	35,801,114	55	998
Chr16	34,844,746	42	877
Chr17	34,135,762	39	969
Chr18	33,903,067	68	980
Chr19	32,808,368	38	1,017
Chr20	29,357,699	40	782
Chr21	28,567,991	47	690
Chr22	28,399,218	40	644
Chr23	24,312,335	35	657
Chr24	18,601,276	30	453
Linked total	868,391,622	1,346	23,453
Unlinked total	44,991,130	180	1,246
Linked percent	95.2%	88.2%	94.9%

TABLE 4 | The genome assembly statistics of several Perciformes fish.

Species	Assembled genome size	Scaffold N50	Contig N50	Gene number	Mapping rate of chromosome	References
<i>Dicentrarchus labrax</i>	675 Mb	5.1 Mb	53 kb	26,719	86%	Tine et al., 2014
<i>Epinephelus akaara</i>	1.135 Gb	46.03 Mb*	5.25 Mb	23,808	95.55%	Zhou et al., 2019
<i>Epinephelus lanceolatus</i>	1.086 Gb	46.2 Mb*	119.9 kb	24,718	98.4%	Ge et al., 2019
<i>Trachinotus ovatus</i>	647.5 Mb	5.05 Mb	1.80 Mb	21,915	99.4%	Zhang et al., 2019
<i>Larimichthys crocea</i>	669.78 Mb	6.55 Mb	282.69 kb	26,100	90%	Mu et al., 2018
<i>Lates calcarifer</i>	668.5 Mb	25.8 Mb*	1.06 Mb	22,184	87.8%	Vij et al., 2016
<i>Plectropomus leopardus</i>	913 Mb	40.04 Mb*	1.42 Mb	24,700	95.2%	This study

The asterisk represents chromosome-level scaffold N50.

using short-read mapping and BUSCO (v3.1) (Simão et al., 2015). We aligned Illumina short reads to the genome by using “bwa” (v0.7.15) (Li and Durbin, 2009).

De novo Repeat Sequences and Gene Annotation

The repeat sequences in the leopard coral grouper genome were identified using a combination of homology-based and *de novo* approaches. First, the homology-based approach was detected repeat sequences, after which the Tandem Repeats Finder (version 4.07) was applied to search for tandem repeats (Benson, 1999). Then, RepeatMasker (v4.0.6) and RepeatProteinMask (v4.0.6), with updated software from the RepeatMasker package), were used jointly to detect known transposable elements (TEs) based on the Repbase TE library (v 21.01) (Tarailo-Graovac and Chen, 2009; Bao et al., 2015). Next, RepeatModeler (v1.0.8) and LTR_FINDER (v1.0.6) set to their default parameters were used to generate the *de novo* repeat library (Xu and Wang, 2007), after which RepeatMasker (Tarailo-Graovac and Chen, 2009) was relied upon again to search for repeat regions against the built repeat library.

Gene Structure Prediction

For gene structure annotations, we utilized three different approaches to annotate the structures of predicted genes in our assembly genome, including *de novo* prediction, homology-based prediction, and transcriptome-based prediction. For *de novo* predictions, both AUGUSTUS (v3.2.1) (Burge and Karlin, 1997; Stanke et al., 2006) and SNAP (v1.0) (Korf, 2004) software packages were used to identify pro-coding genes within the leopard coral grouper genome. For homology-based predictions, we aligned the homologous proteins of eight fish species—*Danio rerio*, *Gasterosteus aculeatus*, *Oreochromis niloticus*, *Oryzias latipes*, *Takifugu rubripes*, *Lepisosteus oculatus*, *Epinephelus tauvina* and *Lates calcarifer* (from the “ensembl 97” release) (Hubbard et al., 2005)—to the repeat-masked genome by using the “tblastn” tool (Blastall v2.2.26) (Mount, 2007) at threshold cut-off E-value $\leq 1e-5$. Next, the Solar (v0.9.6) (Li et al., 2010) and

GeneWise (version 2.4.1) (Birney et al., 2004) programs were executed to distinguish and delineate the potential gene structures for all alignments made. Additionally, the RNA-Seq data from muscle tissues were aligned to the assembled genome by using “tophat” (v2.0.13) (Trapnell et al., 2009), and their corresponding gene structures were predicted by “cufflinks” (v2.1.1) (Trapnell et al., 2012). The above three datasets were combined to generate a

consistent and comprehensive gene set in “maker” (v1.0) (Cantarel et al., 2008; Thrasher et al., 2014).

Comparison of Genome

To compare the assembled leopard coral grouper genome to other already known Serranidae fish genomes, we used Lastz (v1.02) (Harris, 2007). These results were then plotted in the “circus” (v0.69) software (Krzywinski et al., 2009).

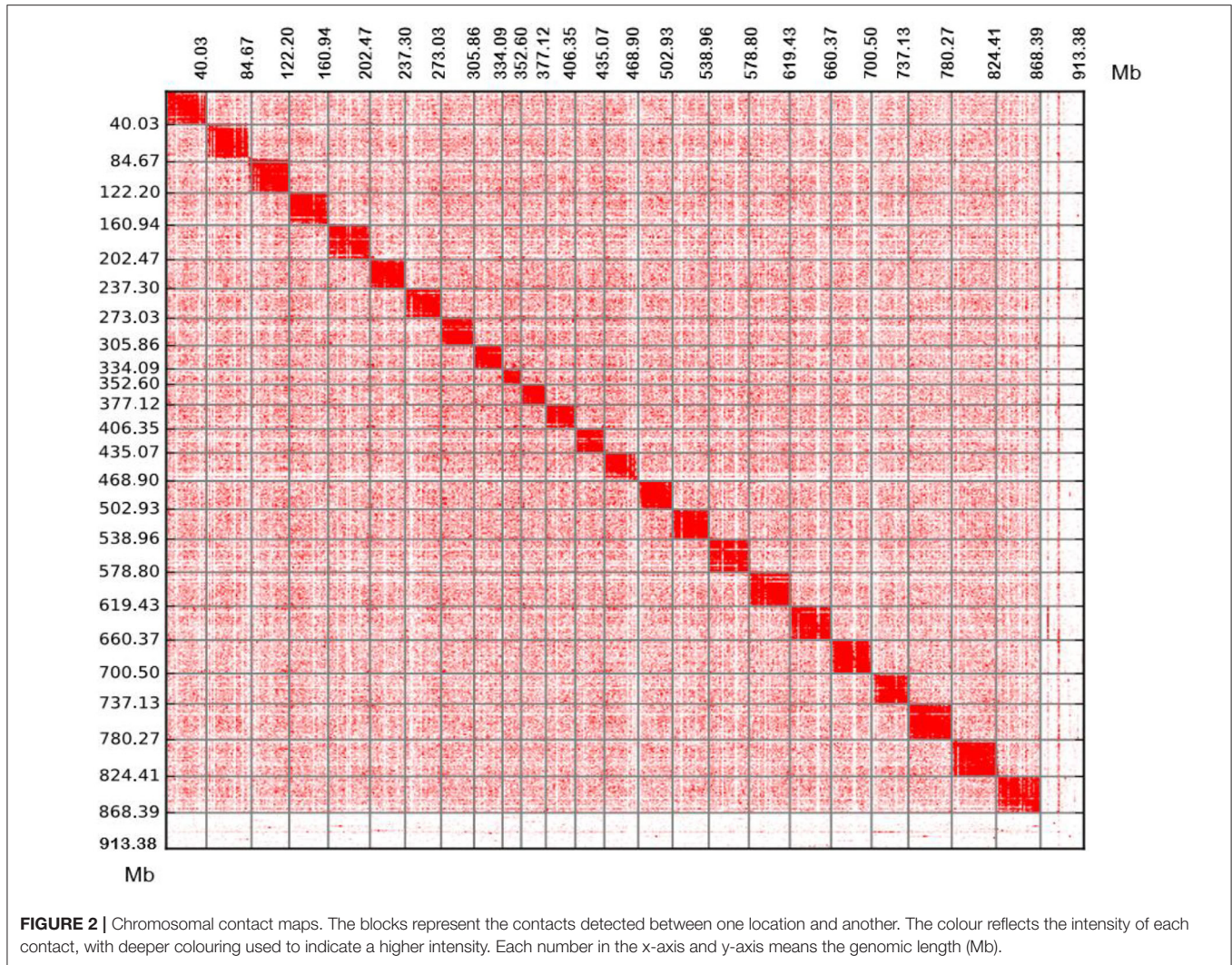


FIGURE 2 | Chromosomal contact maps. The blocks represent the contacts detected between one location and another. The colour reflects the intensity of each contact, with deeper colouring used to indicate a higher intensity. Each number in the x-axis and y-axis means the genomic length (Mb).

TABLE 5 | Summary statistics for the annotated repeat sequences.

Type	Repbse TEs		TE proteins		De novo		Combined TEs	
	Length (bp)	% in genome	Length (bp)	% in genome	Length (bp)	% in genome	Length (bp)	% in genome
DNA	34,078,348	3.73	1,715,926	0.19	147,891,061	16.20	160,562,959	17.59
LINE	17,746,417	1.91	13,612,807	1.49	60,770,280	6.65	82,561,848	9.04
SINE	2,149,356	0.23	0	0.00	1,640,938	0.18	2,841,026	0.31
LTR	10,757,322	1.17	4,273,499	0.47	61,117,147	6.69	71,106,208	7.79
Other	27,455	0.003	72	0.00	0	0.00	27,527	0.003
Unknown	0	0.00	0	0.00	63,867,205	6.99	63,867,025	6.99
Total	58,320,329	6.39	19,596,947	2.14	285,544,317	31.28	315,750,864	34.59

Usage Notes

All contig sequences were assembled into chromosomes by using interaction information from the Hi-C sequencing data. Hence, we used 500 bp to represent the unknown gap sizes among contigs in the obtained chromosome sequences.

Code Availability

The execution of this work involved using many advanced software tools. The settings and parameters for these are provided below.

Genome assembly: (1) *minimap2+miniasm*: all parameters were set to their defaults; (2) *racon*: all parameters were set to their defaults; (3) *pilon*: all parameters were set to their defaults; (4) *3d-dna*: -m haploid -s 4 -c 24 -j 10.

Genome annotation: (1) *ProteinMask*: -engine ncbi -noLowSimple -pvalue 0.0001; (2) *RepeatMasker*: -nolow -no_is -norma -engine ncbi -parallel 1; (3) *LTR_FINDER*: -w 2; (4) *RepeatModeler*: -database genome -engine ncbi -pa 9; (5) *TRF*: matching weight = 2, mismatching penalty = 7, INDEL penalty = 7, match probability = 80, INDEL probability = 10, minimum alignment score to report = 50, maximum period size to report = 2,000, -d -h; (6) *Augustus*: -uniqueGeneId = true -noInFrameStop = true -gff3 = on -strand = both -species = zebrafish; (7) *SNAP*: all parameters were set to the defaults; (8) *BLAST*: -p tblastn -e 1e-05 -F T -m 8 -d; (9) *tophat*: -max-intron-length 20000 -m 1 -solexa-quals -r 20 -no-coverage-search -mate-std-dev 20 -microexon-search -p 8; (10) *cufflinks*: -I 20000 -p 4; (11) *maker*: all parameters were set to the defaults.

Genome alignment: (1) *Lastz*: T = 2, C = 2, H = 2,000, Y = 3,400, L = 6,000, K = 2,200, -format = axt; (2) *Mcscan*: -a -e 1e-5 -u 1 -s 5.

RESULTS AND DISCUSSION

Library Construction and Sequencing

After removing any redundant and low-quality reads, a total of 38.56 Gb (43.6X) clean reads were left, which included

18.14 and 20.42 Gb of reads from the 500- and 800-bp reads length via Illumina sequencing, respectively. After the k-mer analysis, all the clean reads were estimated to be 945 Mbp using the Genome Characteristics Estimation (GCE) software. A nanopore library was constructed and sequenced using the GridION X5/PromethION sequencer, which yielded 76.93 Gb of final contigs. The high-throughput chromosome conformation capture (Hi-C) library was sequenced by the Illumina HiSeq X10 platform (with 150 bp PE model). This Hi-C sequencing was done for chromosome-level scaffold constructions, yielding a total of 52.75 Gb of paired-end Hi-C reads generated, whose average sequencing coverage was 59.9X (Table 1).

Genome Assembly and Chromosome Anchoring

We obtained an assembled genome of leopard coral grouper containing 1,526 contigs, whose total length was 912.66 Mb. The assembly covered 96.5% of the estimated genome regions. The contig N50 length was 1.42 Mb (Table 2). Through Hi-C data, 1,346 contigs were found anchored and orientated on 24 chromosomes, including 95.2% of genomic sequences; the results were consistent with previous karyotype analyses of the leopard coral grouper (Gao et al., 2015). The respective lengths of the 24 chromosomes ranged from 18.6 to 43.74 Mb (Tables 2, 3;

TABLE 7 | Statistics for the functional annotation of protein-coding genes.

Database	Number	Percent (%)
NR	23,986	97.10
SwissProt	19,994	80.95
Interpro	21,518	87.11
KEGG	21,277	86.14
At least one database	24,014	97.22
Total	24,700	100

Note that "at least one database" refers to those genes with at least one hit among the multiple databases searched.

TABLE 6 | Summary statistics of predicted protein-coding genes.

	Gene set	Number	Average transcript length (bp)	Average CDS length (bp)	Average exons per gene	Average exon length (bp)	Average intron length (bp)
De novo	Augustus	27,740	15,893	1,417	8.3	170	1,980
	SNAP	72,047	19,338	967	6.85	141	3,137
Homolog	<i>Takifugu rubripes</i>	19,388	16,833	1,638	10	163	1,682
	<i>Lepisosteus oculatus</i>	15,494	21,220	1,714	10.8	157	1,973
	<i>Gasterosteus aculeatus</i>	28,411	14,597	1,475	9.3	157	1,572
	<i>Oreochromis niloticus</i>	26,405	19,590	1,786	10.7	165	1,817
	<i>Oryzias latipes</i>	20,987	15,042	1,474	9.1	160	1,656
	<i>Danio rerio</i>	17,176	22,791	1,576	9.6	163	2,448
	<i>Epinephelus tauvina</i>	23,026	17,077	1,636	9.7	168	1,766
	<i>Lates calcarifer</i>	24,253	16,883	1,776	10	182	1,732
RNA-seq	Cufflinks	19,890	6,176	1,398	5.1	274	1,166
	Maker	24,700	16,883	1,777	9.7	182	1,732

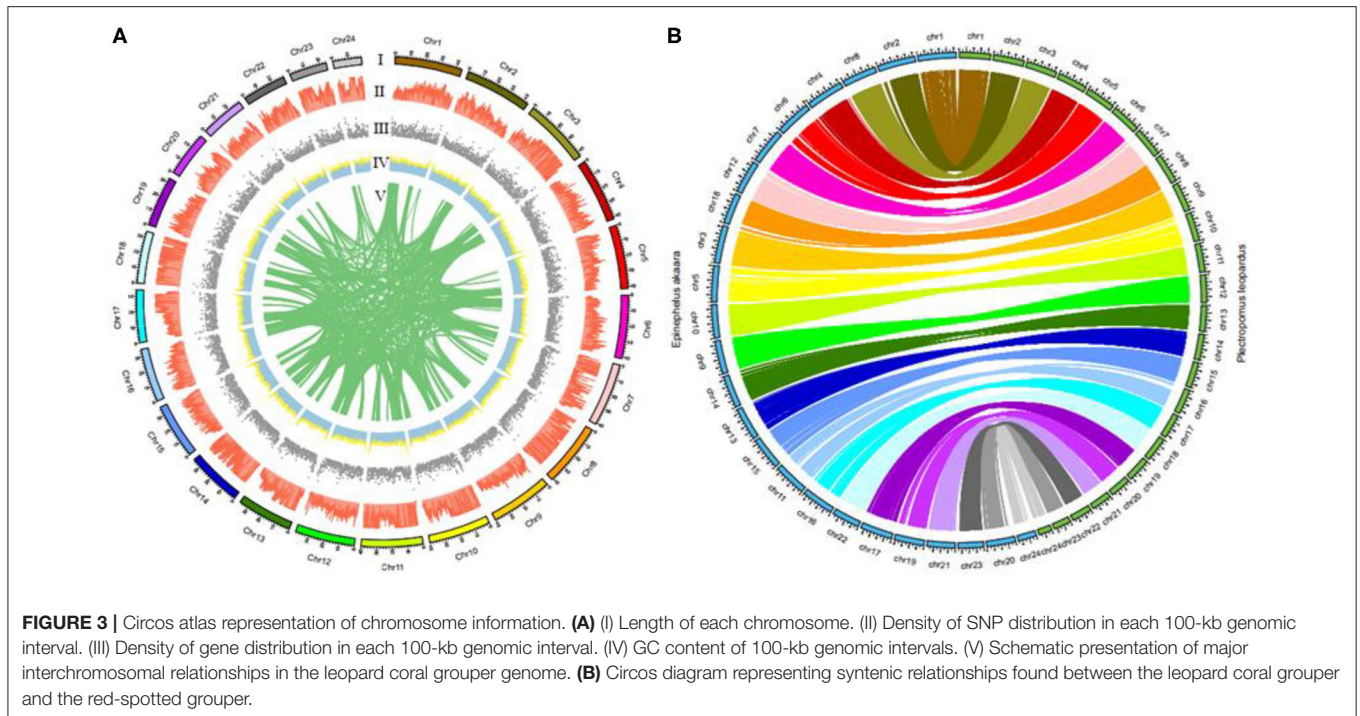


Figure 2). Compared with other Perciformes fish, the genome assembly of *P. leopardus* shows a higher level (**Table 4**).

Using the vertebrata_odb9 database, we found that 92.5% BUSCO genes were completely within the leopard coral grouper genome. We then aligned Illumina short reads to the genome using “bwa” (v0.7.15), finding that more than 94.25% of the reads were aligned to the reference genome, which demonstrated a high mapping ratio for the short-read sequencing data.

Repeat Sequences and Gene Annotation

A total of 315.75 Mb (34.59% of the assembled genome) repeat sequences were thus identified. Among these repeat elements, DNA transposons were more abundant than any other types, accounting for 17.59% (160.56 Mb) (**Table 5**).

A total of 24,700 protein-coding genes were predicted. The average number of exons per gene and average gene length were 9.7 and 1,777 bp, respectively (**Table 6**). In all, we were able to annotate 24,014 genes in at least one of the databases; hence, in this way, 97.22% of leopard coral grouper genes were functionally annotated (**Table 7**).

Comparison With Other Serranidae Fish Genomes

We recently used Lastz (v1.02) to successfully compare the leopard coral grouper genome to the red-spotted grouper (*Epinephelus akaara*) genome (Ge et al., 2019). **Figure 3A** summarizes the distribution of SNPs, genes, GC content on 100-kb genomic intervals, as well as the interchromosomal relationships of our assembled leopard coral grouper chromosomes. The genomic sequences of the red-spotted grouper showed evidence of synteny to the leopard coral grouper’s genome. We found that the 24 chromosomes of the

red-spotted grouper had a clear one-to-one relationship to the leopard coral grouper’s chromosomes (**Figure 3B**). According to these results, we therefore anticipate that the leopard coral grouper genome will contribute to the study of genome evolution in the Serranidae family members.

DATA AVAILABILITY STATEMENT

The genome assembly sequences and predicted gene were deposited in at CNGB under the accession CNA0007316. The Illumina genomic sequencing reads, Nanopore long reads, Hi-C data, and RNA-seq reads were deposited in CNGB under the accession CNP0000859.

ETHICS STATEMENT

The animal study was reviewed and approved by Institutional Review Board on Bioethics and Biosafety of BGI (No. FT 18134).

AUTHOR CONTRIBUTIONS

FC, JLu, and XY contributed to the study design. YW, SF, JLi, WT, ML, LL, and HH contributed to the fish culture and sample preparation. XZ, YW, and XW performed the bioinformatics analysis. JLu, XY, XZ, and XW wrote the paper. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Key R&D Program China (2019YFD0900900), the Major Scientific Research

Project of Hainan (ZDKJ2019011), the Provincial Budget project of Hainan (2018-2020), the Key R&D Project in Hainan (ZDYF2018066, ZDYF2019074), the National Natural

Science Foundation of China (31872572) and the Natural Science Foundation for Fundamental Research of Shenzhen (JCYJ20190812105801661).

REFERENCES

- Bao, W., Kojima, K. K., and Kohany, O. (2015). Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* 6:11. doi: 10.1186/s13100-015-0041-9
- Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580. doi: 10.1093/nar/27.2.573
- Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and genomewise. *Genome Res* 14, 988–995. doi: 10.1101/gr.1865504
- Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* 268, 78–94. doi: 10.1006/jmbi.1997.0951
- Cantarel, B. L., Korf, I., Robb, S. M., Parra, G., Ross, E., Moore, B., et al. (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* 18, 188–196. doi: 10.1101/gr.6743907
- Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., et al. (2017). *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* 356, 92–95. doi: 10.1126/science.aal3327
- Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S., Huntley, M. H., Lander, E. S., et al. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Systems* 3, 95–98. doi: 10.1016/j.cels.2016.07.002
- Gao, N., Cai, Y., Zhou, Y. C., Gen-Lang, L. I., Yuan, W., Liu, H. T., et al. (2015). Karyotype and chromosome localization of nucleolar organizer region in wild leopard coraltrout *Plectropomus leopardus* from Hainan. *J. Dalian Ocean University* 30, 257–260. doi: 10.16535/j.cnki.dlhyxb.2015.03.004
- Ge, H., Lin, K., Shen, M., Wu, S., Wang, Y., Zhang, Z., et al. (2019). *De novo* assembly of a chromosome-level reference genome of red-spotted grouper (*Epinephelus akaara*) using nanopore sequencing and Hi-C. *Mol. Ecol. Resour.* 19, 1461–1469. doi: 10.1111/1755-0998.13064
- Geng, X., Sha, J., Liu, S., Bao, L., Zhang, J., Wang, R., et al. (2015). A genome-wide association study in catfish reveals the presence of functional hubs of related genes within QTLs for columnaris disease resistance. *BMC Genomics* 16:196. doi: 10.1186/s12864-015-1409-4
- Greenfield, D. W. (2002). Coral Reef Fishes: Dynamics and Diversity in a Complex Ecosystem. *Academic Press*. The American Society of Ichthyologists and Herpetologists.
- Harris, R. S. (2007). *Improved Pairwise Alignment of Genomic DNA*, College of Engineering. University Park, TX: Pennsylvania State University.
- Hubbard, T., Andrews, D., Caccamo, M., Cameron, G., Chen, Y., Clamp, M., et al. (2005). Ensembl 2005. *Nucleic Acids Res.* 33, D447–D453. doi: 10.1093/nar/gki138
- Kingsford, M. (1992). Spatial and temporal variation in predation on reef fishes by coral trout (*Plectropomus leopardus*, Serranidae). *Coral Reefs* 11, 193–198. doi: 10.1007/BF00301993
- Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics* 5:59. doi: 10.1186/1471-2105-5-59
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109
- Lerebours, A., Chapman, E. C., Sweet, M. J., Heupel, M. R., and Rotchell, J. M. (2016). Molecular changes in skin pigmented lesions of the coral trout *Plectropomus leopardus*. *Mar. Environ. Res.* 120, 130–135. doi: 10.1016/j.marenvres.2016.07.009
- Li, H. (2016). Minimap and miniasm: fast mapping and *de novo* assembly for noisy long sequences. *Bioinformatics* 32, 2103–2110. doi: 10.1093/bioinformatics/btw152
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, R., Fan, W., Tian, G., Zhu, H., He, L., Cai, J., et al. (2010). The sequence and *de novo* assembly of the giant panda genome. *Nature* 463, 311–317. doi: 10.1038/nature08696
- Lien, S., Koop, B. F., Sandve, S. R., Miller, J. R., Kent, M. P., Nome, T., et al. (2016). The Atlantic salmon genome provides insights into rediploidization. *Nature* 533, 200–205. doi: 10.1038/nature17164
- Liu, B., Shi, Y., Yuan, J., Hu, X., Zhang, H., Li, N., et al. (2013). *Estimation of Genomic Characteristics by Analyzing k-mer Frequency in de novo Genome Projects*, arXiv, Ithaca, NY: Cornell University Press.
- Liu, Z., Liu, S., Yao, J., Bao, L., Zhang, J., Li, Y., et al. (2016). The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts. *Nat. Commun.* 7, 1–13. doi: 10.1038/ncomms11757
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *Gigascience* 1, 18–23. doi: 10.1186/2047-217X-1-18
- Messmer, V., Pratchett, M. S., Hoey, A. S., Tobin, A. J., Coker, D. J., Cooke, S. J., et al. (2017). Global warming may disproportionately affect larger adults in a predatory coral reef fish. *Glob. Chang. Biol.* 23, 2230–2240. doi: 10.1111/gcb.13552
- Michael, T. P., Jupe, F., Bemm, F., Motley, S. T., Sandoval, J. P., Lanz, C., et al. (2018). High contiguity *Arabidopsis thaliana* genome assembly with a single nanopore flow cell. *Nat. Commun.* 9, 1–8. doi: 10.1038/s41467-018-03016-2
- Morris, A. V., Roberts, C. M., and Hawkins, J. P. (2000). The threatened status of groupers (Epinephelinae). *Biodivers. Conserv.* 9, 919–942. doi: 10.1023/A:1008996002822
- Mount, D. W. (2007). *Using the Basic Local Alignment Search Tool (BLAST)*. New York, NY: Cold Spring Harbor Protocols. doi: 10.1101/pdb.top17
- Mu, Y., Huo, J., Guan, Y., Fan, D., Xiao, X., Wei, J., et al. (2018). An improved genome assembly for *Larimichthys crocea* reveals hepcidin gene expansion with diversified regulation and function. *Commun. Biol.* 1:195. doi: 10.1038/s42003-018-0207-3
- Ødegård, J., Moen, T., Santi, N., Korsvoll, S. A., and Kjøglum, S., Meuwissen, T. H. (2014). Genomic prediction in an admixed population of Atlantic salmon (*Salmo salar*). *Front. Genet.* 5:402. doi: 10.3389/fgene.2014.00402
- Palaiokostas, C., Cariou, S., Bestin, A., Bruant, J.-S., Haffray, P., Morin, T., et al. (2018). Genome-wide association and genomic prediction of resistance to viral nervous necrosis in European sea bass (*Dicentrarchus labrax*) using RAD sequencing. *Genet. Sel. Evol.* 50:30. doi: 10.1186/s12711-018-0401-2
- Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680. doi: 10.1016/j.cell.2014.11.021
- Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C.-J., Vert, J.-P., et al. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* 16:259. doi: 10.1186/s13059-015-0831-x
- Shao, C., Bao, B., Xie, Z., Chen, X., Li, B., Jia, X., et al. (2017). The genome and transcriptome of Japanese flounder provide insights into flatfish asymmetry. *Nat. Genet.* 49, 119–124. doi: 10.1038/ng.3732
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351
- Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* 34, W435–W439. doi: 10.1093/nar/gkl200
- Star, B., Nederbragt, A. J., Jentoft, S., Grimholt, U., Malmstrøm, M., Gregers, T. F., et al. (2011). The genome sequence of Atlantic cod reveals a unique immune system. *Nature* 477, 207–210. doi: 10.1038/nature10342

- Sutton, M. A., Burton, A. S., Zaikova, E., Sutton, R. E., Brinckerhoff, W. B., Bevilacqua, J. G., et al. (2019). Radiation tolerance of nanopore sequencing technology for life detection on mars and europa. *Sci. Rep.* 9, 1–10. doi: 10.1038/s41598-019-41488-4
- Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics*. 25, 1–14. doi: 10.1002/0471250953.bi0410s25
- Thrasher, A., Musgrave, Z., Kachmarck, B., Thain, D., and Emrich, S. (2014). Scaling up genome annotation using MAKER and work queue. *Int. J. Bioinformatics Res. Applications*. 10, 447–460. doi: 10.1504/IJBRA.2014.062994
- Tine, M., Kuhl, H., Gagnaire, P. A., Louro, B., Desmarais, E., Martins, R. S., et al. (2014). European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nat. Commun.* 5:5770. doi: 10.1038/ncomms6770
- Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111. doi: 10.1093/bioinformatics/btp120
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–578. doi: 10.1038/nprot.2012.016
- Vallejo, R. L., Leeds, T. D., Fragomeni, B. O., Gao, G., Hernandez, A. G., Misztal, I., et al. (2016). Evaluation of genome-enabled selection for bacterial cold water disease resistance using progeny performance data in rainbow trout: insights on genotyping methods and genomic prediction models. *Front. Genet.* 7:96. doi: 10.3389/fgene.2016.00096
- Vaser, R., Sović, I., and Nagarajan, N., Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 27, 737–746. doi: 10.1101/gr.214270.116
- Vij, S., Kuhl, H., Kuznetsova, I. S., Komissarov, A., Yurchenko, A. A., Van Heusden, P., et al. (2016). Chromosomal-level assembly of the asian seabass genome using long sequence reads and multi-layered scaffolding. *PLoS Genet.* 12:e1005954. doi: 10.1371/journal.pgen.1005954
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE*. 9:e112963. doi: 10.1371/journal.pone.0112963
- Wang, L., Yu, C., Guo, L., Lin, H., and Meng, Z. (2015). *In silico* comparative transcriptome analysis of two color morphs of the common coral trout (*Plectropomus leopardus*). *PLoS ONE*. 10:e0145868. doi: 10.1371/journal.pone.0145868
- Wu, L., Wu, H., Chen, W., Liu, Z., and Guan, X. (2016). Effects of different types of light on habitat of *Plectropomus leopardus* juveniles. *Fisheries Sci.* 35, 14–20. doi: 10.16378/j.cnki.1003-1111.2016.01.003
- Xu, J., Li, J.-T., Jiang, Y., Peng, W., Yao, Z., Chen, B., et al. (2017). Genomic basis of adaptive evolution: the survival of amur ide (*Leuciscus waleckii*) in an extremely alkaline environment. *Mol. Biol. Evol.* 34, 145–159. doi: 10.1093/molbev/msw230
- Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* 35, W265–W268. doi: 10.1093/nar/gkm286
- You, X., Shan, X., and Shi, Q. (2020). Research advances in the genomics and applications for molecular breeding of aquaculture animals. *Aquaculture* 526:735357. doi: 10.1016/j.aquaculture.2020.735357
- Zhang, D. C., Guo, L., Guo, H. Y., Zhu, K. C., Li, S. Q., Zhang, Y., et al. (2019). Chromosome-level genome assembly of golden pompano (*Trachinotus ovatus*) in the family Carangidae. *Sci Data* 6:216. doi: 10.1038/s41597-019-0238-8
- Zheng, X., Kuang, Y., Lv, W., Cao, D., Sun, Z., and Sun, X. (2016). Genome-wide association study for muscle fat content and abdominal fat traits in common carp (*Cyprinus carpio*). *PLoS ONE*. 11:e0169127. doi: 10.1371/journal.pone.0169127
- Zhou, Q., Gao, H., Zhang, Y., Fan, G., Xu, H., Zhai, J., et al. (2019). A chromosome-level genome assembly of the giant grouper (*Epinephelus lanceolatus*) provides insights into its innate immunity and rapid growth. *Mol. Ecol. Resour.* 19, 1322–1332. doi: 10.1111/1755-0998.13048

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Wang, Wen, Zhang, Fu, Liu, Tan, Luo, Liu, Huang, You, Luo and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.