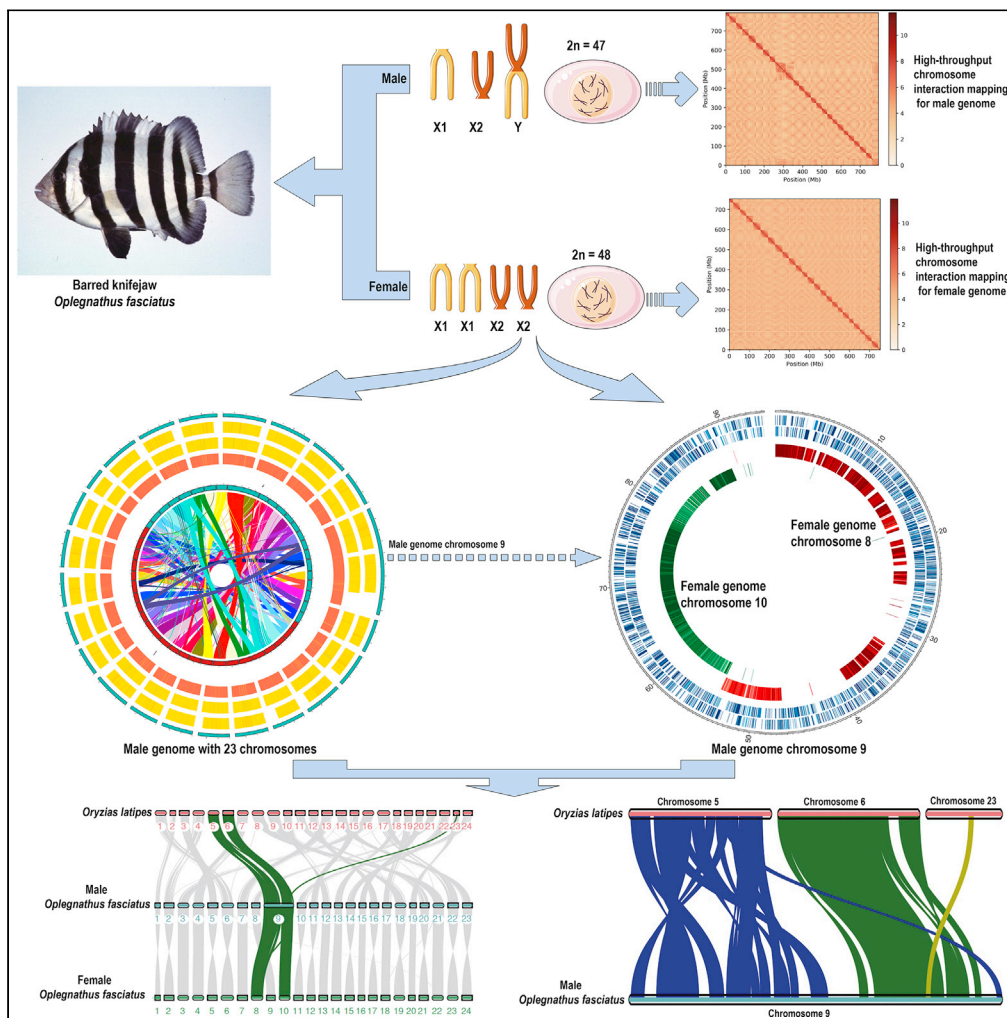


Article

# Chromosome-Level Genome Reveals the Origin of Neo-Y Chromosome in the Male Barred Knifejaw *Oplegnathus fasciatus*



Yongshuang Xiao,  
Zhizhong Xiao,  
Daoyuan Ma, ...,  
Jing Liu, Jun Li,  
Angel Herrera-  
Ulloa

shijun\_xiao@163.com (S.X.)  
jliu@qdio.ac.cn (J.L.)  
junli@qdio.ac.cn (J.L.)

**HIGHLIGHTS**

Construction of a chromosome-level reference genome for the male *O. fasciatus*

Identification of the origin of neo-Y chromosome to the  $X_1X_2Y$  system

Accurate comparisons of sequences and genes between female  $X_1X_1X_2X_2$  and male  $X_1X_2Y$

Xiao et al., iScience 23, 101039  
April 24, 2020 © 2020 The Author(s).  
<https://doi.org/10.1016/j.isci.2020.101039>

## Article

Chromosome-Level Genome Reveals the Origin of Neo-Y Chromosome in the Male Barred Knifejaw *Oplegnathus fasciatus*

Yongshuang Xiao,<sup>1,2,8</sup> Zhizhong Xiao,<sup>1,2,8</sup> Daoyuan Ma,<sup>1,2</sup> Chenxi Zhao,<sup>3</sup> Lin Liu,<sup>4</sup> Hao Wu,<sup>4</sup> Wenchao Nie,<sup>4</sup> Shijun Xiao,<sup>5,6,\*</sup> Jing Liu,<sup>1,2,\*</sup> Jun Li,<sup>1,2,9,\*</sup> and Angel Herrera-Ulloa<sup>7</sup>

## SUMMARY

The barred knifejaw, *Oplegnathus fasciatus*, is characterized by an  $X_1X_2Y$  system with a neo-Y chromosome for males. Here, a chromosome-level genome was assembled to investigate the origin of neo-Y chromosome to the male *O. fasciatus*. Twenty-three chromosomes corresponding to the male karyotypes were scaffolded to 762-Mb genome with a contig N50 length of 2.18 Mb. A large neo-Y chromosome (Ch9) in the male *O. fasciatus* genome was also assembled and exhibited high identity to those of the female chromosomes Ch8 and Ch10. Chromosome rearrangements events were detected in the neo-chromosome Ch9. Our results suggested that a centric fusion of acrocentric chromosomes Ch8 and Ch10 should be responsible for the formation of the  $X_1X_2Y$  system. The high-quality genome will not only provide a solid foundation for further sex-determining mechanism research in the  $X_1X_2Y$  system but also facilitate the artificial breeding aiming to improve the yield and disease resistance for *Oplegnathus*.

## INTRODUCTION

The barred knifejaw (*Oplegnathus fasciatus*) (FishBase ID: 1709; NCBI Taxonomy ID: 163134) (Temminck and Schlegel, 1844), a member of the Oplegnathidae family of the Centrarchiformes, is a commercially important rocky reef fish native to East Asia. *O. fasciatus* has become an important fishery resource in offshore cage aquaculture and fish stocking for marine ranching in China, Japan, and Korea (Schembri et al., 2010; Xiao et al., 2016, 2019). This fish is also a valuable species for sashimi and recreational fishing, and its ex-factory price has reached up to \$30 per kilogram in China (Xiao, 2015; Park et al., 2018). It has been reported that the male of *O. fasciatus* has  $2n = 47$  chromosomes ( $1m + 2m/sm + 44t$ ), whereas the females possess  $2n = 48$  chromosomes ( $2m/sm + 46t$ ) (Xu et al., 2012, 2019). Similar chromosome karyotypes have also been reported in male and female individuals of *Oplegnathus punctatus* (Xue et al., 2016; Xu et al., 2019). A large metacentric Y chromosome was found in male individuals of *O. fasciatus* and *O. punctatus* based on karyotypes and microsatellite DNA motif analyses, and it was suggested that the sex-determining types of *O. fasciatus* and *O. punctatus* should belong to the multiple  $X_1X_1X_2X_2/X_1X_2Y$  sexual system (Xu et al., 2012, 2019; Xue et al., 2016). Sexual dimorphism in growth has been detected in *O. fasciatus*, with male fish exhibiting faster growth than females, possibly due to the sex chromosome system in *Oplegnathus* (Xiao, 2015). *O. fasciatus* is vulnerable to viruses (e.g., iridovirus) due to inbreeding in aquaculture industry (Li et al., 2011; Zhang et al., 2014). Its high aquaculture value, multiple  $X_1X_1X_2X_2/X_1X_2Y$  sex chromosome system, and susceptibility to widespread biotic diseases have led to increasing research interests in *O. fasciatus*. Although the previous reports provided a preliminary description of the multiple sex chromosome system, the exact origin and molecular composition of the large metacentric Y chromosome of the  $X_1X_2Y$  system at the genomic level remain unclear.

Approximately 37 cases of multiple sex chromosomes with  $X_1X_1X_2X_2/X_1X_2Y$  system have been reported across the teleost phylogeny (Sember et al., 2015; Bitencourt et al., 2016; Zhang et al., 2018; Krysanov and Demidova, 2018; Cai et al., 2019; Xu et al., 2019). A preliminary description of those multiple sex chromosome systems, including karyotypes, C-bands, rDNA locations, karyotype diversification, and identification of sex-specific regions at the cytogenetic level, has been carried out based on conventional cytogenetic (Giemsa-staining and C-banding) and molecular cytogenetic protocols (repetitive DNA markers, comparative genomic hybridization, and whole-chromosome painting) (Parise-Maltempi et al., 2007; Cioffi

<sup>1</sup>Center for Ocean Mega-Science, The Key Laboratory of Experimental Marine Biology, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China

<sup>2</sup>Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China

<sup>3</sup>BGI Education Center, University of Chinese Academy of Sciences, Shenzhen, China

<sup>4</sup>Wuhan Frasergen Bioinformatics Co., Ltd. East Lake High-Tech Zone, Wuhan, China

<sup>5</sup>College of Plant Protection, Jilin Agriculture University, Changchun, Jilin, China

<sup>6</sup>School of Computer Science and Technology, Wuhan University of Technology, Wuhan, China

<sup>7</sup>Escuela de Ciencias Biológicas, Universidad Nacional, San José, Costa Rica

<sup>8</sup>These authors contributed equally

<sup>9</sup>Lead Contact

\*Correspondence: shijun\_xiao@163.com (S.X.), jliu@qdio.ac.cn (J.L.), junli@qdio.ac.cn (J.L.)  
<https://doi.org/10.1016/j.isci.2020.101039>



and Bertollo, 2012; Blanco et al., 2013; Sember et al., 2015; Ferreira et al., 2016; Bitencourt et al., 2016). However, adequate genome resources to support more comprehensive descriptions of the multiple sex chromosome system and the origin of the large metacentric Y chromosome of male *O. fasciatus* have been lacking. The recent release of the chromosome-level reference genome of female *O. fasciatus* has provided valuable resource for sex-determination studies; however, a female genome is still needed to investigate the origin of the unique  $X_1X_2Y$  system for male *O. fasciatus* (Xiao et al., 2019). Using PacBio sequencing and high-throughput chromosome interaction mapping (Hi-C), Xiao et al. (2019) obtained a chromosome-level reference genome of the female *O. fasciatus* with a final size of 768.8 Mb and a contig and scaffold N50 length of 2.1 Mb and 33.5 Mb, respectively (Xiao et al., 2019). Twenty-four chromosomes corresponding to the female karyotype ( $2n = 48$ ) were assembled at the genome level. Although the high-quality genome of female *O. fasciatus* provides a valuable genomic resource for further study of breeding systems, it could not be used to identify the origin of the large metacentric Y chromosome of male *O. fasciatus* without a male genome.

Here, we report the chromosome-level genome assembly of male *O. fasciatus* based on PacBio long-read sequencing and Hi-C. Genomic comparisons between male and female *O. fasciatus* were carried out to provide insights into the origin of the  $X_1X_2Y$  system of male *O. fasciatus* based on the chromosome-level genome, which has excellent continuity at the contig and scaffold levels. The genome of male *O. fasciatus* can lay a solid foundation for further study of sex-determining mechanisms of the  $X_1X_2Y$  system and will provide valuable genomic data for conservation genetics and resistance breeding of *Oplegnathus*.

## RESULTS

### PacBio Sequencing and Genome Assembly

Two 20-kb PacBio long-read DNA libraries were constructed using the standard protocol provided by the PacBio Sequel platform. A total of ~39.79 Gb of subreads were obtained using SMRT LINK 5.0 to remove the adaptor sequences from the raw data derived from the zero-mode waveguide (Tables 1 and S1). Approximately 4.71 million sequences with an average length of 8.45 kb were obtained for the draft genome assembly of male *O. fasciatus* (Table 1).

To increase the continuity and completeness of the genome assembly, four processes were carried out for the contig assembly. First, the Canu v1.4 software was used to assemble an initial genome of male *O. fasciatus* (Koren et al., 2017). As a result, a total length of 866.9 Mb, including 4,453 contigs with an N50 length of 1.73 Mb, was obtained (Table S2). Second, Redundans v0.13c software was employed to remove sequence redundancies in the initial assembled genome to obtain a 794.8-Mb genome with a contig N50 length of 2.13 Mb (Table S2). Third, Arrow tool implemented in SMRT Link 5.0 software and Pilon v.122 was applied to perform error correction using long read data and Illumina NGS data mentioned in the genome survey analysis (Table S2) (Walker et al., 2014; Xiao et al., 2019). The final contig assembly of 795.1 Mb with a contig N50 length of 2.13 Mb was obtained. The genome contained 2,295 contigs with a longest contig of 9.8 Mb (Tables S2 and S3). A total of 881 contigs were longer than 100 kb, representing 92.6% of the total 794.8 Mb for the male *O. fasciatus* genome (Table 1). The GC content of the contig assembly genome was 40.87% (Figures S1 and S2).

To obtain the chromosome-level genome of male *O. fasciatus*, the Illumina HiSeq X Ten platform was used to generate ~95.9 Gb clean data from the Hi-C library (Tables S1 and S4). According to the abovementioned mapping strategy, more than 95.5% of total reads mapped to the assembled genome in pairs and ~32.5% of read pairs mapped to different contigs. Lachesis software was used to cluster, order, and orientate contigs along chromosomes based on their interaction frequencies. As a result, 1,355 contigs were successfully anchored and oriented into 23 chromosomes, which was consistent with the previous karyotype analyses of male *O. fasciatus* ( $X_1X_2Y$  system) (Table S4 and Figure 1) (Xu et al., 2012). The total length of anchored contigs was ~762.2 Mb, representing 95.9% of all assembled contigs. Finally, we obtained the chromosome assembly with a contig N50 length of 2.18 Mb and a scaffold N50 length of 32.4 Mb (Table 1). Obviously, a large neo-chromosome (*Ch9*) showed strong interaction signals from two genomic blocks, corresponding to the *Ch8* and *Ch10* in female *O. fasciatus* ( $X_1X_1X_2X_2$  system) (Figure 1) (Xu et al., 2012). Therefore, *Ch9* was likely to be the large metacentric Y chromosome of male *O. fasciatus*. This chromosome (*Ch9*) was scaffolded from 444 contigs and was 94.2 Mb, more than three times larger than any other chromosome (Figures 1, 2, and 3). More than 99.7% of contigs longer than 100kb were anchored on chromosomes, exhibiting the excellent anchoring rate for male *O. fasciatus* chromosome assembly (Table S4).

|  | Draft Scaffolds for Male <i>O. fasciatus</i> | Chromosome-Length Scaffolds Based on Hi-C for Male <i>O. fasciatus</i> | <sup>a</sup> Chromosome-Length Scaffolds Based on Hi-C for Female <i>O. fasciatus</i> |
|--|--|--|---|
| Genome assembly                        |  |  |   |
| Length of genome (bp)                  | 795,074,755                                  | 762,267,613  | 768,808,243   |
| Number of contigs                      | 2,295  | 1,355  | 1,372   |
| Contigs N50 (bp)                       | 2,127,178                                    | 2,183,645  | 2,130,780   |
| Number of scaffold                     | /  | 23   | 24  |
| Scaffold N50 (bp)                      | /  | 32,431,321   | 33,548,962  |
| Genome coverage (X)                    | 251.1  |  | 314.6   |
| Number of contigs (≥ 100 kb)           | 881  | 891  | 708   |
| Total length of contigs (≥ 100 kb)     | 736,155,642                                  | 733,715,954  | 732,827,446   |
| Mapping rate of contigs (≥ 100 kb) (%) | /  | 99.67  | 99.67   |
| Genome annotation                      |  |  |   |
| Protein-coding gene number             | 24,835                                       |  | 24,003  |
| Mean transcript length (kb)            | 15.8   |  | 16.1  |
| Mean exons per gene                    | 10.0   |  | 10.1  |
| Mean exon length (bp)                  | 220.1  |  | 217.7   |
| Mean intron length (bp)                | 1,511.7                                      |  | 1,527.4   |

**Table 1. Summary of Male *O. fasciatus* Genome Assembly and Annotation**

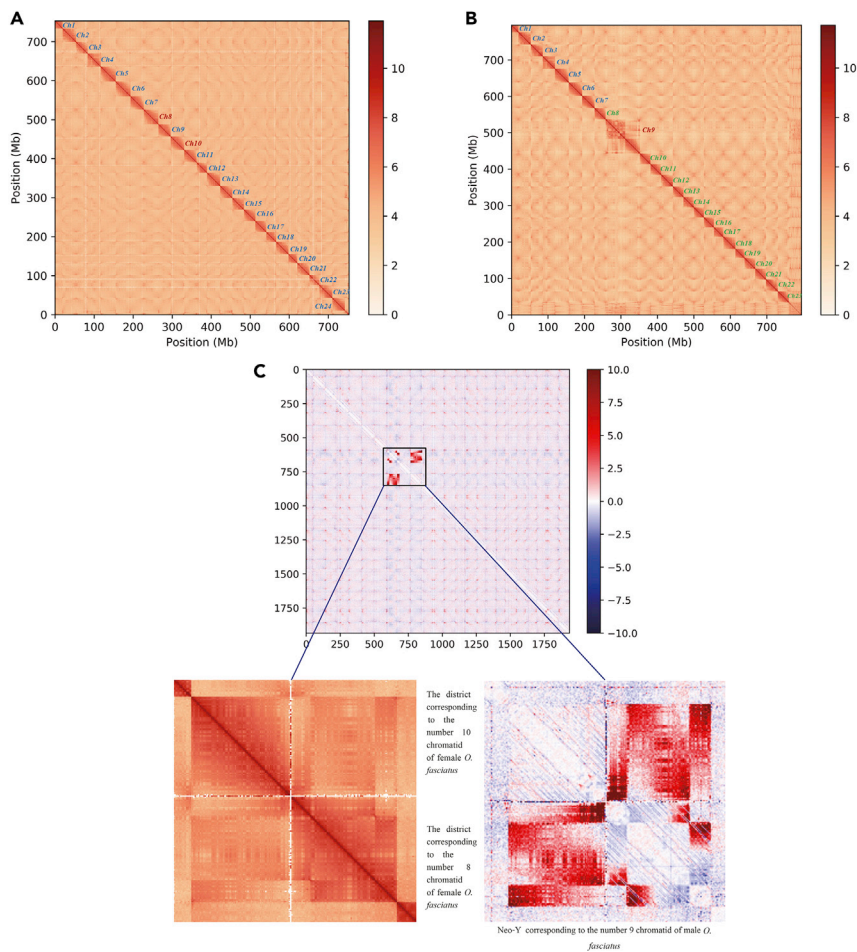
<sup>a</sup>The data were cited from the reference (Yongshuang Xiao, Zhizhong Xiao, Daoyuan Ma, Jing Liu, Jun Li. Genome sequence of the barred knifejaw *Oplegnathus fasciatus* (Temminck and Schlegel, 1844): the first chromosome-level draft genome in the family Oplegnathidae, GigaScience, Volume 8, Issue 3, March 2019, giz013, doi.org/10.1093/gigascience/giz013).

### Genome Quality Evaluation

The Minimap2 software was employed to evaluate the completeness and homogeneity of the assembled genome of male *O. fasciatus* by using the CLR subreads (Table S5). The mapping rate and the coverage of the assembled genome reached 87.6% and 99.9%, respectively (Table S5). These results showed the high completeness and homogeneity of the genome assembly of male *O. fasciatus*. BUSCO v3.0 software with the actinopterygii\_odb9 database was employed to further evaluate the completeness of the assembled genome (Simão et al., 2015). The result showed that 97.2% and 1.0% of the 4,456 conserved single-copy orthologous genes were identified as complete BUSCO and fragmented BUSCO profiles in the genome assembly, respectively (Table 2). Among the 4,456 conserved single-copy orthologous genes, 4,210 (91.8%) and 246 (5.4%) genes were identified as single-copy and duplicated BUSCOs, respectively (Table 2). Approximately 80 single-copy orthologous genes were not detected in the actinopterygii\_odb9 database. Then, SNP calling data were used to evaluate the accuracy of the male *O. fasciatus* genome assembly, which was generated from the alignment of next-generation sequencing-based short reads to the assembled genome by using BWA and GATK software. Approximately 1.87 million SNP loci were identified, including 1.86 and 0.01 million heterozygous homologous SNPs, respectively (Table S6). The heterozygous SNPs accounted for 0.23% of the male *O. fasciatus* genome, which was comparable with our previous study of the heterozygosity for the female *O. fasciatus* genome (Table S6) (Xiao et al., 2019).

### Repetitive Element Identification and Protein Gene Annotation

Approximately 33.5% of the assembled genome was identified as repetitive elements, including repetitive sequences accounting for 23.16% of the male *O. fasciatus* genome based on the *de novo* repeat library



**Figure 1. Genome Assembly of Female and Male *O. fasciatus* Based on the Hi-C Interaction Analyses**

(A) The heatmap of interactions among genomic bins of 500 kb along 24 chromosomes for female *O. fasciatus* (The data were cited from Xiao et al., 2019).

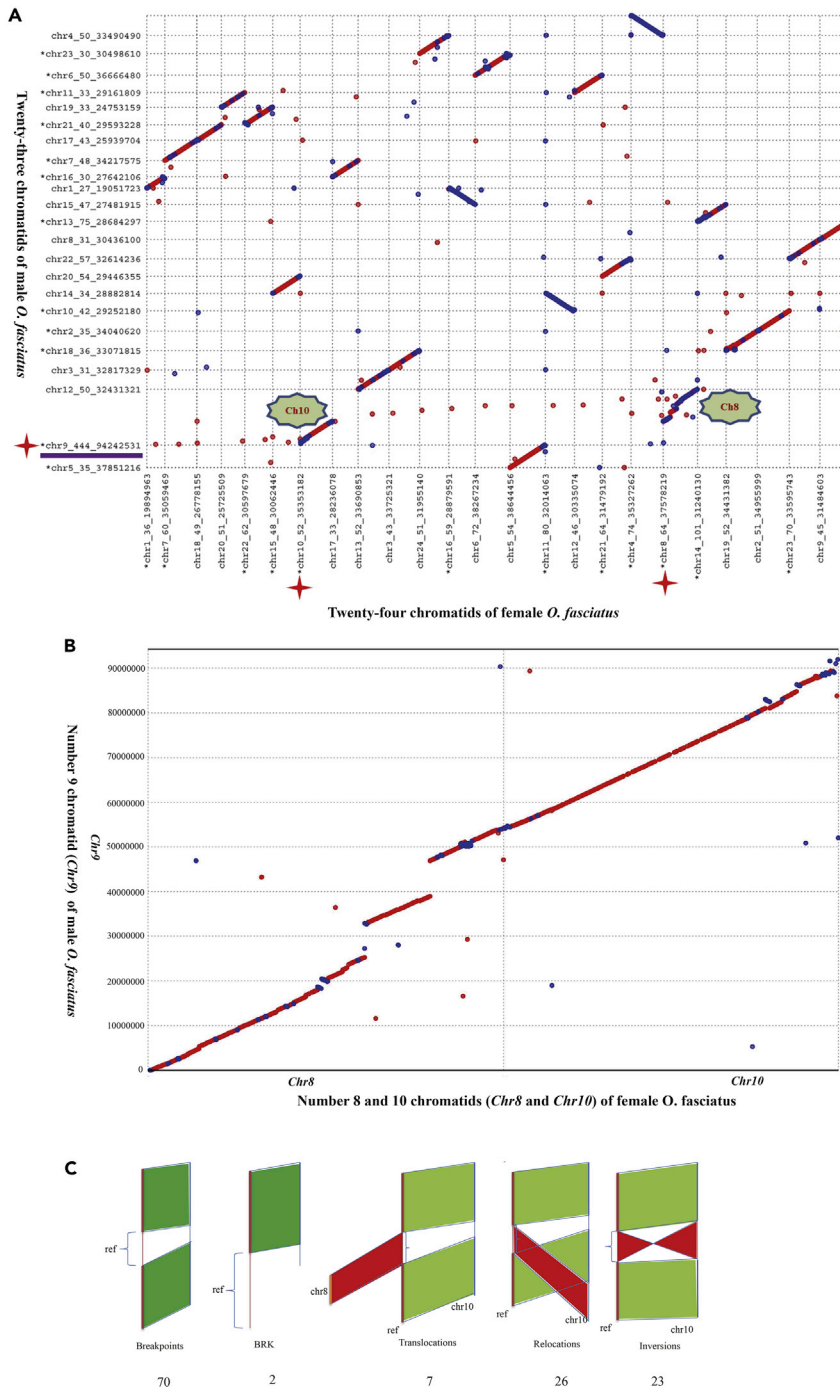
(B) The heatmap of interactions among genomic bins of 500 kb along 23 chromosomes for male *O. fasciatus*.

(C) The cumulative distribution of subtraction Hi-C Z-scores for interactions between 400- and 40-kb bins from the whole-genome and chromosome levels. Blocks represent the interactions among genomic bins, and the interaction strength was represented by the color scheme from deep (strong interactions) to light (weak interactions). A large neo-chromosome (*Ch9*) was assembled in the male *O. fasciatus* reference genome.

(Table 3). The estimation of repetitive element content for the male *O. fasciatus* genome was comparable to the result of the *k*-mer analysis (38.4%) (Table 3) (Xiao et al., 2019). Interspersed repetitive elements accounting for 22.0% of the male *O. fasciatus* genome were identified, including DNA transposons (10.55%), long interspersed nuclear elements (LINEs, 7.08%), and long terminal repeats (LTRs, 4.11%) (Tables 3 and Table S7 and Figure S3). The repetitive contents of the male *Ch9* and the female *Ch8/Ch10* were also identified as 23.79%, 26.07%, and 22.70%, respectively (Table S8). Although the frequency of DNA transposons, LINEs, and LTRs was higher than that in *Larimichthys crocea*, *Gasterosteus aculeatus*, *Oryzias latipes*, and *Dicentrarchus labrax*, the top three categories of TEs were significantly less frequent than in *Epinephelus lanceolatus* and *Triplophysa tibetana* (Table S7).

Homology-based, *de novo*, and transcriptome sequencing-based approaches were integrated to predict protein-coding genes. As a result, 24,835 genes were annotated with an average of 10.0 exons per gene in the male *O. fasciatus* genome (Tables S9 and S10). The distribution statistics of average gene length, average coding sequence (CDS) length, average exons per gene, average exon length, and average intron length of protein-coding genes were also compared with those of six related species (*Lates calcarifer*, *L. crocea*, *G. aculeatus*, *Gadus morhua*, *Panaeolus olivaceus* and *Cynoglossus semilaevis*) and showed a



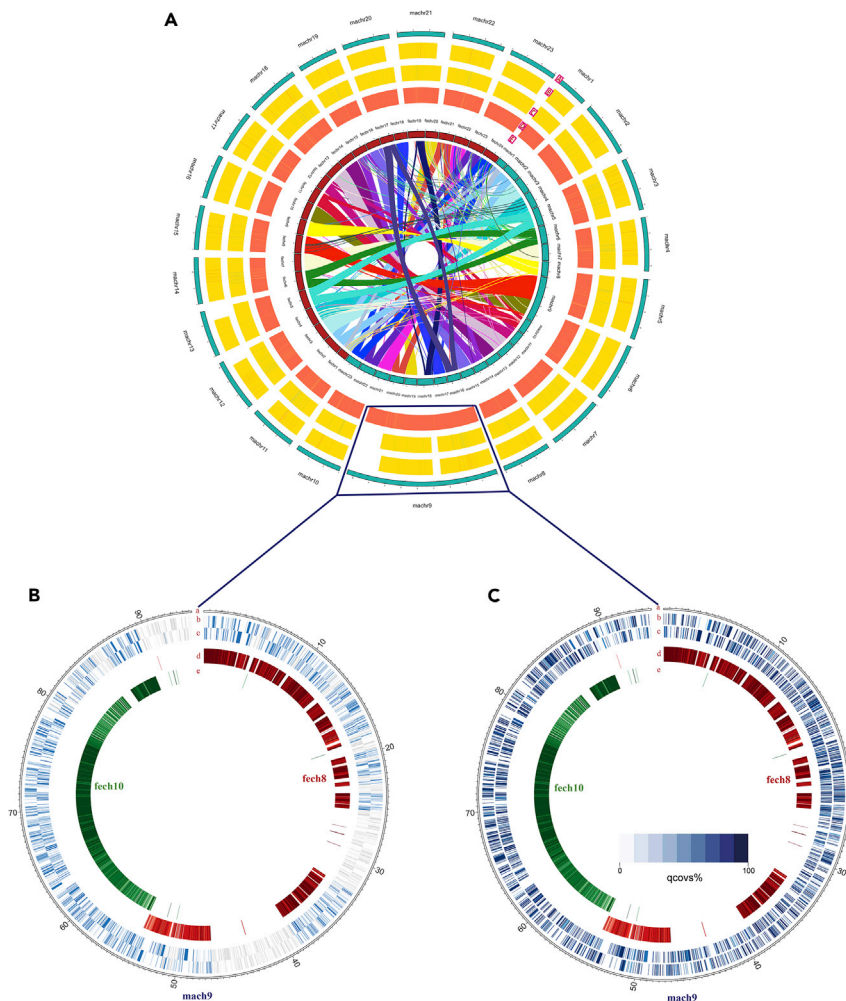


**Figure 2. Genomic Comparisons between Female and Male *O. fasciatus***

(A) Genomic comparisons of the whole genome by direct sequence alignment. The majority of female and male *O. fasciatus* chromosomes exhibited 1:1 correspondence except for the large neo-chromosome (Ch9).

(B) Detailed genomic comparisons between Ch9 and Ch8/Ch10 from male and female genomes. The large neo-chromosome (Ch9) of the male *O. fasciatus* genome showed largely synteny with the Ch8 and Ch10 of the female *O. fasciatus* genome.

(C) The statistics of structure variants with length more than 10 kb between Ch9 and Ch8/Ch10 from male and female genomes.



**Figure 3. Genome Comparisons Among Male/Female *O. fasciatus* and *L. crocea***

(A) From outer to inner circles: A, 23 chromosomes of male *O. fasciatus*; B, 24 chromosomes of *L. crocea*. The yellow color represents the whole chromosomes and the red lines in the yellow areas represent the common chromosomal region with the male *O. fasciatus*. C, 24 chromosomes of female *O. fasciatus*. The yellow color represents the whole chromosome, and the blue lines in the yellow areas represent the common chromosomal region with the male *O. fasciatus*. D, 23 chromosomes of male *O. fasciatus*. The red color represents the whole chromosome, and the yellow lines in the red areas represent the common chromosomal region with the female *O. fasciatus* and *L. crocea*. E, the red color represents the chromosomes of female *O. fasciatus*, and the green color represents the male *O. fasciatus*. Each line precisely joined pairs of genes originated from the male and female *O. fasciatus* chromosomes.

(B) and (C) From outer to inner circles: a, the ninth chromosome (Ch9) of male *O. fasciatus* with coordinate; b, the distribution of forward protein-coding genes in male Ch9; c, the distribution of reverse protein-coding genes in male Ch9; d, the chromosomal region of male Ch9 aligned to Ch8 of female *O. fasciatus* with red color; e, the chromosomal region of male Ch9 aligned to Ch10 of female *O. fasciatus* with green color. The gray color represents the distribution of protein-coding genes in unique genomic regions of male Ch9 (b, c tracks of figure (B)). The color gradient corresponds to the degree of similarities for male Ch9 genes with the female genes in tracks b and c of figure (C).

similar distribution with those of other teleosts (Figure S4 and Table S10). The average gene length and CDS reached 15,819.4 bp and 1,707.0 bp, respectively (Table S10). Functional annotation of predicted genes in the male *O. fasciatus* genome was further performed using the InterPro, Swiss-Prot, TrEMBL, NR, GO, and KEGG databases (Table 4). Approximately 23,364 of the 24,835 genes (97.34%) in the male *O. fasciatus* genome could be functionally annotated by at least one of the abovementioned databases (Table 4). We used BUSCO v3.0 software to further evaluate the completeness of the annotated genome against actinopterygii\_odb9 in the OrthoDB database (Simão et al., 2015). Approximately 96.8% complete BUSCO genes were successfully identified (Table 3). We also used tRNAscan-SE software to annotate the

| Type                            | Assembly |                | Annotation |                |
|---------------------------------|----------|----------------|------------|----------------|
|                                 | Proteins | Percentage (%) | Proteins   | Percentage (%) |
| Complete BUSCOs                 | 4,456    | 97.2           | 4,435      | 96.8           |
| Complete and single-copy BUSCOs | 4,210    | 91.8           | 4,143      | 90.4           |
| Complete and duplicated BUSCOs  | 246      | 5.4            | 292        | 6.4            |
| Fragmented BUSCOs               | 48       | 1.0            | 66         | 1.4            |
| Missing BUSCOs                  | 80       | 1.8            | 83         | 1.8            |
| Total BUSCOs groups searched    | 4,584    | 100.0          | 4,584      | 100.0          |

**Table 2. Genome Quality of *O. fasciatus* Based on the BUSCO Assessment**

non-coding RNAs against the Rfam database, and four types of non-coding RNAs (miRNAs [0.006%], tRNAs [0.009%], rRNAs [0.007%], and small nucleolar RNAs [0.015%]), including 2,666 copies with a total length of 291,392 bp (0.037% of the whole genome) were identified (Table S11).

### Chromosome Comparison of Female/Male *O. fasciatus*

According to the synteny-based chromosome comparison between the male and female *O. fasciatus* genomes using the program MUMmer, we found excellent consistency of genome sequences in corresponding chromosomes (Figures 2 and 3). The genome sequences from male *O. fasciatus* had high identity (~99.0%) to those from female *O. fasciatus*, as follows: male Ch1/female Ch1 (99.0%), male Ch2/female Ch2 (99.1%), male Ch3/female Ch3 (99.2%), male Ch4/female Ch4 (99.0%), male Ch5/female Ch5 (99.2%), male Ch6/female Ch6 (99.2%), male Ch7/female Ch7 (99.2%), male Ch8/female Ch9 (99.2%), male Ch10/female Ch11 (99.1%), male Ch11/female Ch12 (99.1%), male Ch12/female Ch13 (99.1%), male Ch13/female Ch14 (99.1%), male Ch14/female Ch15 (99.1%), male Ch15/female Ch16 (99.3%), male Ch16/female Ch17 (99.1%), male Ch17/female Ch18 (99.2%), male Ch18/female Ch19 (99.2%), male Ch19/female Ch20 (99.1%), male Ch20/female Ch21 (99.1%), male Ch21/female Ch22 (99.2%), male Ch22/female Ch23 (99.2%), and male Ch23/female Ch24 (99.2%). The comparisons of chromosomal sequences of female Ch8/Ch10 and the male Ch9 were further performed (Figure 3B). A total of ~31.3 Mb homology sequences for female Ch8 were aligned to male Ch9 with a high identity (~99.0%), representing 83.4% of the whole Ch8 length (37.5 Mb) (Table S12). Similarly, more than 90.1% of Ch10 sequences exhibited a high identity with male Ch9 (Table S12). Meanwhile, approximately 67.0% sequences (63.1 Mb) of the male Ch9 (94.2 Mb) showed high identity (~99.0%) with those from the female Ch8 and Ch10 using nucmer with minimum match length of 1,000 bp (Table S12). After reducing the parameter of minimum match length to 100 bp, we observed that more than 89.5% of the male Ch9 could align to the female Ch8 and Ch10, suggesting that the male Ch9 might undergo massive rearrangements during the neo-Y chromosome formation (Table S13 and Figure S5). Indeed, structure variations were identified in sequences for the male Ch9 lacking homolog to the female Ch8 and Ch10, including 72 breakpoints, 7 translocations, 26 relocations, and 23 inversions (Figure 2C).

According to homology searching of genes in the male genome to the female genome, we identified 172 male-specific genes in the male Ch9. Several genes were involved in the chromosome organization and nucleosome assembly processes for fish, such as chromosome transmission fidelity protein 8 (*ctf8*), centromere protein P (*cenpp*), synaptonemal complex protein 1 (*symp1*), and caveolin 3 (*cav3*) (Table S14). The *ctf8* could regulate sister chromatid cohesion and fidelity of chromosome transmission (Bermudez et al., 2003). The *cenpp* is involved in assembly of kinetochore proteins, mitotic progression, and chromosome segregation (Okada et al., 2006). The *symp1* is a major component of the transverse filaments of synaptonemal complexes and formed between homologous chromosomes during meiotic prophase (Bisig et al., 2012). The functions of *cav3* could serve as a component of the caveolae plasma membranes in most cell types (Shang et al., 2019).

The conservation synteny analysis for male-female *O. fasciatus* comparison and *O. fasciatus* (the X<sub>1</sub>X<sub>2</sub>Y system)-*O. latipes* (the normal XY system) comparison using homolog gene pairs between two species were also performed. As a result, twenty-two female *O. fasciatus* chromosomes harbored an excellent



| Type          | Rebase TEs  |             | TE Proteins |             | De Novo     |             | Combined TEs |             |
|---------------|-------------|-------------|-------------|-------------|-------------|-------------|--------------|-------------|
|               | Length (bp) | % in Genome | Length (bp) | % in Genome | Length (bp) | % in Genome | Length (bp)  | % in Genome |
| DNA           | 39,085,328  | 4.92        | 5,858,619   | 0.74        | 83,843,085  | 10.55       | 115,535,672  | 14.53       |
| LINE          | 24,759,524  | 3.11        | 17,460,721  | 2.20        | 56,286,293  | 7.08        | 85,210,163   | 10.72       |
| SINE          | 889,332     | 0.11        | 0           | 0.00        | 1,986,947   | 0.25        | 2,817,685    | 0.35        |
| LTR           | 10,536,213  | 1.33        | 6,615,817   | 0.83        | 32,670,415  | 4.11        | 44,682,943   | 5.62        |
| Satellite     | 1,910,832   | 0.24        | 0           | 0.00        | 733,763     | 0.09        | 2,480,580    | 0.31        |
| Simple_repeat | 1,304,732   | 0.16        | 0           | 0.00        | 7,478,505   | 0.94        | 8,578,237    | 1.08        |
| Other         | 6,957       | 0.00        | 171         | 0.00        | 0           | 0.00        | 7,128        | 0.00        |
| Unknown       | 338,847     | 0.04        | 0           | 0.00        | 30,384,129  | 3.82        | 30,719,924   | 3.86        |
| Total         | 74,440,379  | 9.36        | 29,922,429  | 3.76        | 184,141,930 | 23.16       | 252,879,666  | 31.81       |

**Table 3. The Detailed Classification of Repeat Sequences for Male *O. fasciatus***

one-to-one correspondence to those of the male *O. fasciatus* genome except for female *Ch8* and *Ch10*. The male *Ch9* showed strong conserved synteny with female *Ch8* and *Ch10*, consistent with the abovementioned results that *Ch9* might be the neo-Y chromosome (Table S15 and Figures 3, 4, and S6). Furthermore, we found that the synteny of chromosomes for *O. fasciatus* and *O. latipes* was largely conserved. Fourteen chromosomes of male *O. fasciatus* genome were unambiguously aligned to single chromosomes of *O. latipes* genome (Table S15 and Figure 4). Other nine chromosomes of the male *O. fasciatus* genome exhibited several small inter-chromosome conservation synteny to *O. latipes* chromosomes, suggesting that massive inter-chromosome rearrangements occurred after divergence of two species (Table S15 and Figure 4). We found that *Ch5* and *Ch6* in the *O. latipes* genome exhibited excellent synteny with female *Ch8* and *Ch10*, as well as with *Ch9* of male *O. fasciatus* (Table S15 and Figures 4 and S6).

The syntenic blocks of the chromosomes were also evaluated among the male/female *O. fasciatus* and *L. crocea* genomes using the program MUMmer. The consistency of chromosomes, with 24 blocks between female *O. fasciatus* and *L. crocea* and 23 blocks among male *O. fasciatus*, *L. crocea*, and female *O. fasciatus*, was detected (Figure 3). Precise pairings of protein-coding genes originating from the male and female *O. fasciatus* chromosomes were established using BLASTP software with identity  $\geq 95\%$  (coverage  $\geq 90\%$ ) and e-value  $\leq 1 \times 10^{-5}$ . The results showed that 10,919 protein-coding gene pairs were identified, 1,459 of which were located on the large neo-chromosome (*Ch9*) in the male genome, corresponding to 809 genes of *Ch8* and 628 genes of *Ch10* in the female genome (Figure 3).

### Gene Family Identification and Phylogenetic Tree Construction

According to the homolog searching of protein-coding genes for male *O. fasciatus* and other species, including *Salmo salar*, *L. crocea*, *G. morhua*, *P. olivaceus*, *C. semilaevis*, *Nolothenia coriiceps*, *Boleophthalmus pectinirostris*, *Branchiostoma floridae*, *G. aculeatus*, *Callorhynchus milii*, *Danio rerio* and *O. latipes*, approximately 23,302 gene families were identified based on their H-scores (Figure S7). The specific and common gene families of male *O. fasciatus* and other teleosts (*L. crocea*, *G. morhua*, and *S. salar*) were further analyzed, which yielded 551 specific gene families in the male genome and 11,484 common gene families among the four teleosts (Figure S8). Using the MCL program implemented in the OrthoMCL pipeline with a coefficient setting of 1.5 to cluster the abovementioned gene families, we obtained 810 single-copy genes, which were employed to reconstruct the phylogenetic relationships among male *O. fasciatus* and the other species. Based on a length filter that retained protein sequences  $\geq 100$  amino acids, 759 single-copy orthogroups were obtained using ClustalW software to extract and align single-copy genes from the 810 single-copy orthogroups (Figure S7). The multiple sequence alignment for the filtered single-copy genes was performed using MUSCLE software, and a super-alignment dataset for each species was obtained and used to construct a phylogenetic tree of the male *O. fasciatus* and the other species based on the maximum-likelihood method implemented in the RAxML package (Figure 5). The results of the phylogenetic tree showed that *O. fasciatus* from the Oplegnathidae family of the Centrarchiformes (Eupercaria) was close to *L. crocea* in the order Perciformes (Eupercaria), consistent with the new

| Type        |           | Number | Percent (%) |
|-------------|-----------|--------|-------------|
| Total       |           | 24,835 |             |
| Annotated   | InterPro  | 21,696 | 87.36       |
|             | GO        | 16,494 | 66.41       |
|             | KEGG_ALL  | 23,916 | 96.30       |
|             | KEGG_KO   | 15,260 | 61.45       |
|             | Swissprot | 22,380 | 90.11       |
|             | TrEMBL    | 23,953 | 96.45       |
|             | NR        | 24,072 | 96.93       |
| Annotated   |           | 24,105 | 97.06       |
| Unannotated |           | 730    | 2.94        |

**Table 4. Functional Annotation of the Protein-Coding Genes in Male *O. fasciatus* Genome**

phylogenetic classification of bony fishes (Figure 5) (Betancur-R et al., 2017). The divergence times among clades were evaluated using the MCMCTree program with calibration times based on the TimeTree database, showing that *O. fasciatus* diverged from its common ancestor with *L. crocea* approximately 62.8–73.4 million years ago (Figure 5).

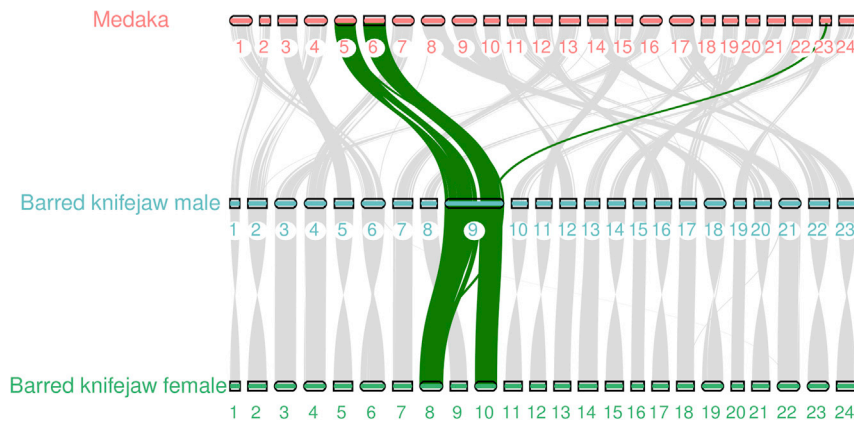
## DISCUSSION

*O. fasciatus* is an important fishery species in offshore cage aquaculture and fish stocking for marine ranching in East Asia (Schembri et al., 2010; Xiao et al., 2016, 2019). The male *O. fasciatus* genome was characterized by an  $X_1X_2Y$  system with a neo-Y chromosome based on male karyotype analyses. The species could be used as an excellent model to address the sex determination, origin, and evolution of the  $X_1X_2Y$  system. The chromosome-level genome of male *O. fasciatus* assembled in the present study, combined with the released reference genome of female *O. fasciatus*, will provide valuable genomic resources to gain insights into the origin of the  $X_1X_2Y$  system (Xiao et al., 2019).

To assess the quality of the assembly, the continuity and completeness of the genome was evaluated. The final contig assembly was 795.1 Mb with a contig N50 length of 2.13 Mb for male *O. fasciatus*, which was comparable to those of the female (Xiao et al., 2019). The contig N50 values of the male/female *O. fasciatus* genomes were also larger than those of many reported teleost genomes, which indicated that high genome continuity existed in *O. fasciatus* genomes (Table S3); 1,355 ordered contigs were scaffolded into 23 chromosomes, yielding a final chromosome-level genome of approximately 762 Mb with a scaffold N50 length of 32.43 Mb (Table 1). The completeness of the assembled genome was evaluated using BUSCO. The high continuity and completeness of the male *O. fasciatus* genome will lay a solid foundation for further studies of population genetics, evolutionary of genome comparisons, neo-chromosome structure, and sex-determining mechanisms (Sun et al., 2017, 2019; Yang et al., 2018).

So far, 37 species have been reported to possess multiple sex chromosomes with  $X_1X_1X_2X_2/X_1X_2Y$  system among teleosts. Although techniques, such as Giemsa staining, C-banding, repetitive DNA markers, CGH and WCP, have been used for chromosomal studies, the origination of the neo-Y chromosome of *O. fasciatus* in previous studies was still largely hindered by the lack of reference genome resources, especially for the neo-Y chromosome (Sember et al., 2015; Bitencourt et al., 2016; Zhang et al., 2018; Krysanov and Demidova, 2018; Cai et al., 2019; Xu et al., 2019). In this work, a large neo-chromosome (Ch9) was assembled into 94.2 Mb, corresponding to the large metacentric Y chromosome of male *O. fasciatus*. The neo-chromosome could be responsible for the genome size discrepancy between male and female *O. fasciatus* (Table S4 and Figure 1).

Three proposed mechanisms for the origin of an  $X_1X_2Y$  multiple sex chromosome system have been postulated, which included fusion between the Y chromosome and an autosome, fission of the X chromosome, and reciprocal translocation between the X chromosome and an autosome (White, 1983; Kitano and Peichel, 2012). All those



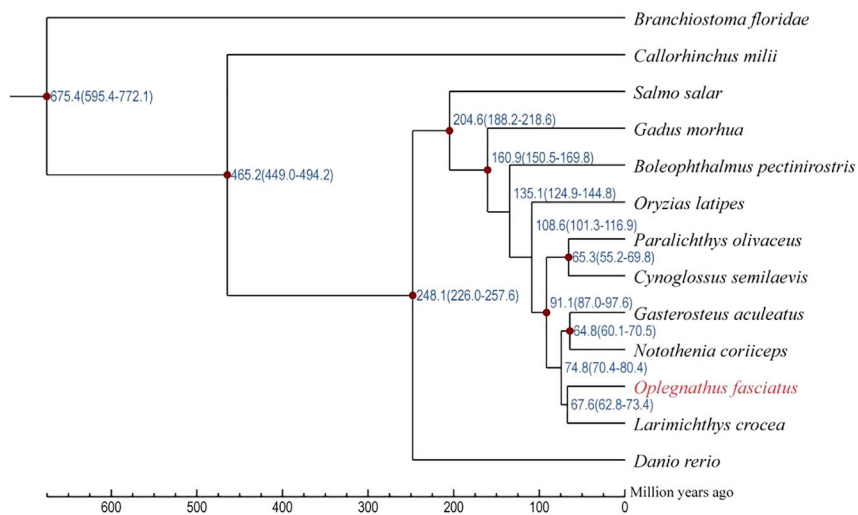
**Figure 4. Chromosome Conserved Synteny between *Oryzias latipes* Genome (the Normal XY system) and *O. fasciatus* Genome (the  $X_1X_2Y$  System)**

Ribbons between two genomes represented chromosomal conservation synteny blocks.

mechanisms (fusion and fission) could induce remarkable genome size changes for sex chromosomes (Y, X), leading to a large neo-Y chromosome or a small neo-X chromosome (Kitano and Peichel, 2012). Previous studies have shown that the  $X_1X_2Y$  systems of teleosts mainly originate from chromosomal fusions, leading to large metacentric chromosomes (neo-Y chromosomes) through a Robertsonian fusion of two acrocentric chromosomes (the Y chromosome and an autosome) (Uyeno and Miller, 1971; Bertollo et al., 2000, 2004; Ueno and Takai, 2008; Kitano and Peichel, 2012). Although the formation of the  $X_1X_2Y$  multiple sex chromosome system could be achieved through fission of the X chromosome, this process would lead to an increase in the diploid number of chromosomes (e.g., female  $2n = 50$ , male  $2n = 49$ ) compared with the ancestral karyotype of marine teleost ( $2n = 48$ ) (White, 1983; Kitano and Peichel, 2012). Our genome assembly for males and females of *O. fasciatus* led to 23 and 24 chromosomes, directly corresponding to the male and female karyotypes ( $2n = 47/48$ ), respectively (Figures 1 and 2) (Xu et al., 2012; Xiao et al., 2019). The comparative analysis showed excellent chromosomal synteny between the male and female *O. fasciatus* genomes (Figures 1, 2, and 3). No small neo-X chromosome was observed at the genome level; however, a large neo-Y chromosome (Ch9) (63.1 Mb/94.2 Mb) in the male *O. fasciatus* genome exhibited high identity (~99.0%) to those of the female chromosomes Ch8 and Ch10. From the chromosomal comparison, *O. fasciatus* female Ch8 and Ch10 exhibited excellent synteny with those of Ch5 and Ch6 in *O. latipes* genome, indicating Ch8 and Ch10 in *O. fasciatus* likely separated in their common ancestor (Table S12). According to the sequence and synteny comparison and previous karyotypes results, we suggested that a centric fusion of acrocentric chromosomes Ch8 and Ch10 should be responsible for the formation of the  $X_1X_2Y$  system of male *O. fasciatus*.

Neo-sex chromosome systems are always derived from rearrangements between original sex chromosomes and autosomes through chromosomal fissions, fragmentations, and fusions (Uyeno and Miller, 1971; Bertollo et al., 2000, 2004; Ueno and Takai, 2008; Kitano and Peichel, 2012). Indeed, although the male large neo-chromosome Ch9 showed general excellent synteny with the female Ch8/Ch10, several obvious rearrangements were also observed at the middle of the chromosomes between Ch9 and Ch8/Ch10, especially for flanking regions of breakpoints around 18 Mb–20 Mb, 28 Mb–32 Mb, and 39 Mb–47 Mb in male Ch9 (Figure 2). Cross-chromosome synteny was also identified between the male Ch9 of *O. fasciatus* and the Ch5/Ch6 of *O. latipes* genome, which exhibited excellent synteny with female Ch8 and Ch10 of *O. fasciatus* (Figure 4). These results showed that chromosome rearrangements events have occurred in the neo-chromosome Ch9 of male *O. fasciatus*.

The neo-chromosomes of the  $X_1X_2Y$  system could also participate in sex regulation and determination (Ueno and Takai, 2008). Previous studies showed that neo-chromosomes might harbor important genes or regulatory elements responsible for the behaviors, phenotype, and speciation (Kitano et al., 2009). Some male-specific genes involved in chromosome and nucleosome assembly (*ctf8*, *cenpp*, *sycp1*) and steroid hormone synthesis (*nr4a1*) were identified in the present study, which might be responsible for the fidelity of homologous chromosome pairing between Ch9 and Ch8/Ch10 during meiotic prophase and male sex determination (Table S14). This high-quality chromosome-level genome will enable us to explore the fusion mechanism and biological functions of neo-chromosomes by analyses of the genetic composition



**Figure 5. Phylogenetic Analysis of Male *O. fasciatus* and Other Related 12 Species**

A total of 21,528 gene families were identified by clustering the homologous gene sequences, and 810 single-copy orthogroups were obtained; 719 filtered single-copy orthogroups were used to construct the phylogenetic relationship between *O. fasciatus* and other species (*S. salar*, *L. crocea*, *G. morhua*, *P. olivaceus*, *C. semilaevis*, *N. coriiceps*, *B. pectinirostris*, *B. floridae*, *G. aculeatus*, *C. milii*, *D. rerio*, and *O. latipes*). Divergence times among the species (red dots) from TimeTree database were used as the calibration divergence times. Blue values on branches indicated the estimated divergence time in millions of years ago (Mya), and numbers in parentheses showed the interval of 95% confidence.

and chromosome conformation studies based on the Hi-C. The multiple sex chromosome system with sexual dimorphism could also lead to growth differences. Sexual dimorphism in growth has been detected in *O. fasciatus*—male fish grow faster than females (Chen et al., 2014; Xiao, 2015). A total of 24,105 protein-coding genes were functionally annotated for the chromosome-level genome of male *O. fasciatus*, and these genes will serve as a framework combined with quantitative trait locus (QTL) and bulked segregant analysis (BSA) techniques for studies of growth regulation and breeding.

In summary, we have successfully completed a chromosome-level genome assembly for male *O. fasciatus* and first assembled the large neo-chromosome corresponding to the karyotype of male *O. fasciatus* with high continuity and completeness. This study demonstrated for the first time that the  $X_1X_2Y$  system of male *O. fasciatus* originated from the fusions of the non-homologous chromosomes *Ch8* and *Ch10* via significant homology and chromosomal interactions at the genome level. This high-quality genome assembly will not only provide a solid foundation for further sex-determining mechanism research in the  $X_1X_2Y$  system but also facilitate the artificial breeding aiming to improve the yield and disease resistance for *Oplegnathus*.

### Limitations of the Study

Chromosome fusion was suggested to be responsible for the formation of the  $X_1X_2Y$  system only from the male *O. fasciatus*; the extremely limited genome information of the fishes with multiple sex chromosomes led to difficulties in accurately determining the dynamics and mechanism of chromosome fusions.

### METHODS

All methods can be found in the accompanying [Transparent Methods supplemental file](#).

### DATA AND CODE AVAILABILITY

The RNA sequencing data of the *Oplegnathus fasciatus* have been deposited in the SRA under Bioproject number PRJNA486572. The whole-genome sequencing data are available in the NCB with the accession number SRP220007.

### SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.isci.2020.101039>.

## ACKNOWLEDGMENTS

This study was supported by a grant from the National Natural Science Foundation of China (No. 41506170, No. 31672672, and No. 31872195), National Key Research and Development Program (2018YFD0901204), STS project (KFZD-SW-106, ZSSD-019, 2017T3017, KFJ-STQ-QYZX-020), Qingdao National Laboratory for Marine Science and Technology (2018SDKJ0502-2, 2015ASKJ02), China Agricultural Research System (CARS-47), Shandong Province Key Research and Invention Program (2017GHY15102, 2017GHY15106), Qingdao Source Innovation Program (17-1-1-57-jch), Marine Fishery Institute of Zhejiang Province, and Key Laboratory of Mariculture and Enhancement of Zhejiang Province (2016KF002). The project of Guangdong laboratory of southern ocean science and engineering (GML2019ZD0402).

## AUTHOR CONTRIBUTIONS

Y.X. conceived the project. Z.X., D.M., Jun Li collected the samples and extracted the genomic DNA. Y.X., Jing Liu, C.Z., H.W., L.L., W.N., S.X., and Jun Li performed the genome assembly and data analysis. Y.X., Z.X., S.X., Jing Liu, D.M., and Jun Li wrote the paper. A.H.-U. polished the paper.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 11, 2019

Revised: March 5, 2020

Accepted: April 1, 2020

Published: April 24, 2020

## REFERENCES

- Bertollo, L.A.C., Born, G.G., Dergam, J.A., Fenocchio, A.S., and Moreira-Filho, O. (2000). A biodiversity approach in the neotropical Erythrinidae fish, *Hoplias malabaricus*. Karyotypic survey, geographic distribution of cytotypes and cytotoxicological considerations. *Chromosome Res.* 8, 603–613.
- Bertollo, L.A.C., Oliveira, C., Molina, W.F., Margarido, V.P., Fontes, M.S., Pastori, M.C., and Fenocchio, A.S. (2004). Chromosome evolution in the erythrinid fish, *Erythrinus erythrinus* (Teleostei: Characiformes). *Heredity* 93, 228–233.
- Betancur-R, R., Wiley, E.O., Arratia, G., Acero, A., Bailly, N., Miya, M., Leconte, G., and Ortí, G. (2017). Phylogenetic classification of bony fishes. *BMC Evol. Biol.* 17, 162.
- Bitencourt, J.A., Sampaio, I., Ramos, R.T.C., Vicari, M.R., and Affonso, P.R.A.M. (2016). First report of sex chromosomes in Achiridae (Teleostei: Pleuronectiformes) with inferences about the origin of the multiple X<sub>1</sub>X<sub>2</sub>X<sub>2</sub>/X<sub>1</sub>X<sub>2</sub>Y system and dispersal of ribosomal genes in *Achirus achirus*. *Zebrafish* 14, 90–95.
- Blanco, D.R., Vicari, M.R., Lui, R.L., Bertollo, L.A.C., Traldi, J.B., and Moreira-Filho, O. (2013). The role of the Robertsonian rearrangements in the origin of the XX<sub>1</sub>XY<sub>1</sub>Y<sub>2</sub> sex chromosome system and in the chromosomal differentiation in *Harttia* species (Siluriformes, Loricariidae). *Rev. Fish Biol. Fisher.* 23, 127–134.
- Bermudez, V.P., Maniwa, Y., Tappin, I., Ozato, K., Yokomori, K., and Hurwitz, J. (2003). The alternative Ctf18-Dcc1-Ctf8-replication factor C complex required for sister chromatid cohesion loads proliferating cell nuclear antigen onto DNA. *Proc. Natl. Acad. Sci. U S A* 100, 10237–10242.
- Bisig, C.G., Guiraldelli, M.F., Kouznetsova, A., Scherthan, H., Höög, C., Dawson, D.S., and Pezza, R.J. (2012). Synaptonemal complex components persist at centromeres and are required for homologous centromere pairing in mouse spermatocytes. *PLoS Genet.* 8, e1002701.
- Cai, M.Y., Xiao, S.J., Li, W.B., Han, Z.F., Han, F., Xiao, J.Z., Liu, F.L., and Wang, Z.Y. (2019). Chromosome assembly of *Collichthys lucidus*, a fish of Sciaenidae with a multiple sex chromosome system. *Sci. Data* 6, 132.
- Chen, S.L., Zhang, G.J., Shao, C.W., Huang, Q.F., Liu, G., Zhang, P., Song, W.T., An, N., Chalopin, D., Volff, J.N., et al. (2014). Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nature* 46, 253–260.
- Cioffi, M.B., and Bertollo, L.A.C. (2012). Chromosomal distribution and evolution of repetitive DNAs in fish. *Repetit. DNA* 7, 197–221.
- Ferreira, M., Garcia, C., Matoso, D.A., de Jesus, I.S., and Feldberg, E. (2016). A new multiple sex chromosome system X<sub>1</sub>X<sub>2</sub>X<sub>2</sub>/X<sub>1</sub>Y<sub>1</sub>X<sub>2</sub>Y<sub>2</sub> in Siluriformes: cytogenetic characterization of *Bunocephalus coracoides* (Aspredinidae). *Genetica* 144, 591–599.
- Kitano, J., and Peichel, C. (2012). Turnover of sex chromosomes and speciation in fishes. *Environ. Biol. Fish.* 94, 549–558.
- Kitano, J., Ross, J.A., Mori, S., Kume, M., Jones, F.C., Chan, Y.F., Absher, D.M., Grimwood, J., Schmutz, J., Myers, R.M., et al. (2009). A role for a neo-sex chromosome in stickleback speciation. *Nature* 461, 1079–1083.
- Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27, 722–736.
- Krysanov, E., and Demidova, T. (2018). Extensive karyotype variability of African fish genus *Nothobranchius* (Cyprinodontiformes). *Comp. Cytogenet.* 12, 387–402.
- Li, H., Sun, Z.P., Li, Q., and Jiang, Y.L. (2011). Characterization of an iridovirus detected in rock bream (*Oplegnathus fasciatus*; Temminck and Schlegel). *Chin. J. Virol.* 27, 158–164.
- Okada, M., Cheeseman, I.M., Hori, T., Okawa, K., McLeod, I.X., Yates, J.R., Desai, A., and Fukagawa, T. (2006). The CENP-H-I complex is required for the efficient incorporation of newly synthesized CENP-A into centromeres. *Nat. Cell Biol.* 8, 446–457.
- Parise-Maltempi, P.P., Martins, C., Oliveira, C., and Foresti, F. (2007). Identification of a new repetitive element in the sex chromosomes of *Leporinus elongatus* (Teleostei: Characiformes: Anostomidae): new insights into the sex chromosomes of *Leporinus*. *Cytogenet. Genome Res.* 116, 218–223.
- Park, H.S., Kim, C.G., Kim, S., Park, Y.J., Choi, H.J., Xiao, Z.Z., Li, J., Xiao, Y.S., and Lee, Y.H. (2018). Population genetic structure of rock bream (*Oplegnathus fasciatus* Temminck & Schlegel, 1844) revealed by mtDNA COI sequence in Korea and China. *Ocean. Sci. J.* 53, 261–274.
- Schembri, P.J., Bodilis, P., Evans, J., and Francour, P. (2010). Occurrence of barred knifejaw, *Oplegnathus fasciatus* (Actinopterygii: Perciformes: Oplegnathidae), in Malta (Central Mediterranean) with a discussion on possible modes of entry. *Acta Ichthyol. Piscat.* 40, 101–104.



- Sember, A., Bohlen, J., Šlechtová, V., Altmanová, M., Symonová, R., and Ráb, P. (2015). Karyotype differentiation in 19 species of river loach fishes (Nemacheilidae, Teleostei): Extensive variability associated with rDNA and heterochromatin distribution and its phylogenetic and ecological interpretation. *BMC Evol. Biol.* 15, 251–272.
- 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.
- Shang, L., Chen, T., Xian, J., Deng, Y., Huang, Y., Zhao, Q., Liang, G., Liang, Z., Lian, F., Wei, H., and Huang, Q. (2019). The caveolin-3 P104L mutation in LGMD-1C patients inhibits non-insulin-stimulated glucose metabolism and growth but promotes myocyte proliferation. *Cell Biol. Int.* 43, 669–677.
- Sun, J., Zhang, Y., Xu, T., Zhang, Y., Mu, H., Zhang, Y., Lan, Y., Fields, C.J., Hui, J.H.L., Zhang, W., et al. (2017). Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. *Nat. Ecol. Evol.* 1, 121.
- Sun, J., Mu, H.W., CH Ip, J., Li, R.S., Xu, T., Accorsi, A., Sánchez Alvarado, A., Ross, E., Lan, Y., Sun, Y.N., et al. (2019). Signatures of divergence, invasiveness, and terrestrialization revealed by four apple snail genomes. *Mol. Biol. Evol.* 36, 1507–1520.
- Temminck, C.J., and Schlegel, H. (1844). In: P.F. von Siebold. *Fauna Japonica*. P, 62.
- Ueno, K., and Takai, A. (2008). Multiple sex chromosome system of  $X_1X_1X_2X_2/X_1X_2Y$  type in lutjanid fish, *Lutjanus quinquelineatus* (Perciformes). *Genetica* 132, 35–41.
- Uyeno, T., and Miller, R.R. (1971). Multiple sex chromosomes in a Mexican Cyprinodontid fish. *Nature* 231, 452–453.
- Walker, B., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q.D., Wortman, J., Yong, S.K., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS one* 9, e112963.
- White, M.J.D. (1983). *Animal Cytology and Evolution* (Cambridge University Press).
- Xiao, Y.S., Li, J., Ren, G.J., Ma, G.Y., Wang, Y.F., Xiao, Z.Z., and Xu, S.H. (2016). Pronounced population genetic differentiation in the rock bream *Oplegnathus fasciatus* inferred from mitochondrial DNA sequences. *Mitochondrial DNA A DNA Mapp. Seq. Anal.* 27, 2045–2052.
- Xiao, Y.S., Xiao, Z.Z., Ma, D.Y., Liu, J., and Li, J. (2019). Genome sequence of the barred knifejaw *Oplegnathus fasciatus* (Temminck & Schlegel, 1884): the first chromosome-level draft genome in the family Oplegnathidae. *GigaScience* 8, giz013.
- Xiao, Z.Z. (2015). Study on Population Genetics and Culture Biology of *Oplegnathus fasciatus* Doctor thesis (China National Knowledge Internet), pp. 162–176.
- Xu, D.D., Sember, A., Zhu, Q.H., de Oliveira, E.A., Liehr, T., Al-Rikabi, A.B.H., Xiao, Z.Z., Song, H.B., and de Bello Cioffi, M. (2019). Deciphering the origin and evolution of the  $X_1X_2Y$  system in two closely-related *Oplegnathus* species (Oplegnathidae and Centrarchiformes). *Int. J. Mol. Sci.* 20, 3571.
- Xu, D.D., You, F., Lou, B., Geng, Z., Li, J., and Xiao, Z.Z. (2012). Comparative analysis of karyotype and C-banding in male and female *Oplegnathus fasciatus*. *Acta Hydrobiol. Sin.* 36, 552–557.
- Xue, R., An, H., Liu, Q.H., Xiao, Z.Z., Wang, Y.F., and Li, J. (2016). Karyotype and Ag-NORs in male and female of *Oplegnathus punctatus*. *Oceanol. Limnol.* Sin. 47, 626–632.
- Yang, X.F., Liu, H.P., Ma, Z.H., Zou, Y., Zou, M., Mao, Y.Z., Li, X.M., Wang, H., Chen, T.S., and Wang, W.M. (2018). Chromosome-level genome assembly of *Triplophysa tibetana*, a fish adapted to the harsh high-altitude environment of the Tibetan Plateau. *Mol. Ecol. Resour.* 19, 1027–1036.
- Zhang, B.C., Zhang, J., Xiao, Z.Z., and Sun, L. (2014). Rock bream (*Oplegnathus fasciatus*) viperin is a virus-responsive protein that modulates innate immunity and promotes resistance against megalocytivirus infection. *Dev. Comp. Immunol.* 45, 35–42.
- Zhang, S., Zheng, J., Zhang, J., Wang, Z.Y., Wang, Y., and Cai, M. (2018). Cytogenetic characterization and description of an  $X_1X_1X_2X_2/X_1X_2Y$  sex chromosome system in *Collichthys lucidus* (Richardson, 1844). *Acta Oceanol. Sin.* 37, 34–39.

iScience, Volume 23

## Supplemental Information

### **Chromosome-Level Genome Reveals the Origin of Neo-Y Chromosome in the Male Barred Knifejaw *Oplegnathus fasciatus***

**Yongshuang Xiao, Zhizhong Xiao, Daoyuan Ma, Chenxi Zhao, Lin Liu, Hao Wu, Wenchao Nie, Shijun Xiao, Jing Liu, Jun Li, and Angel Herrera-Ulloa**

## Supplemental Figures

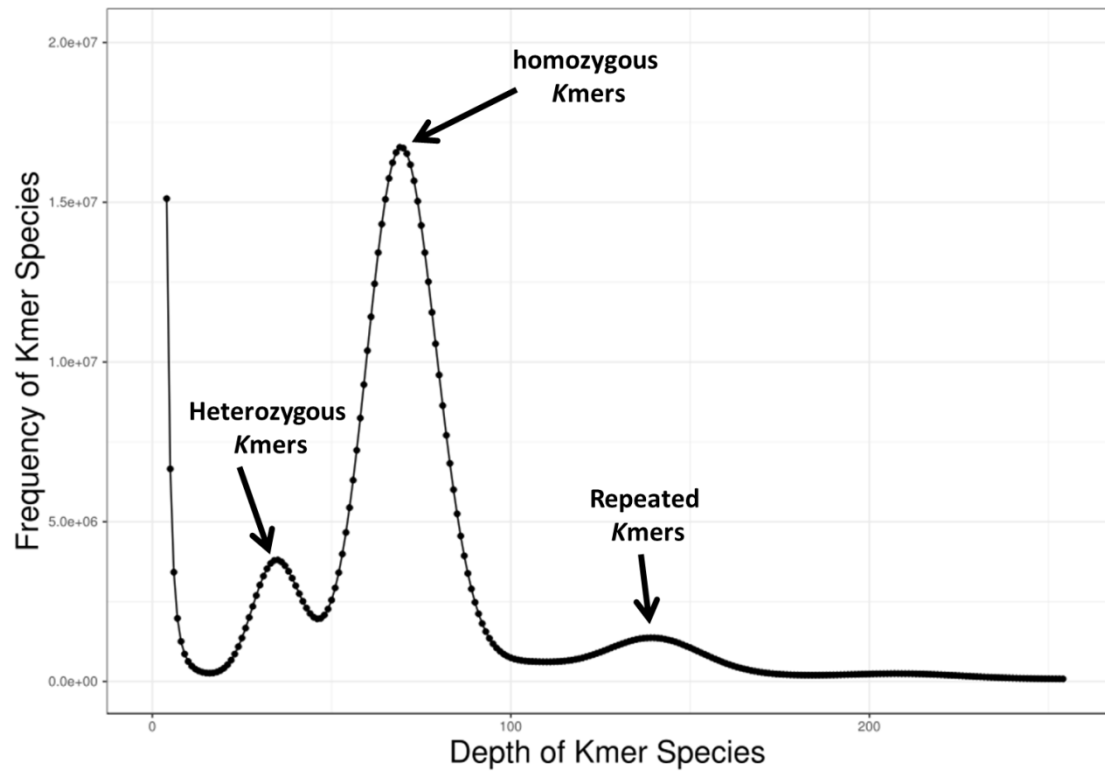


Figure S1. *K*-mer distribution of male *O. fasciatus* genome, Related to Table 1.

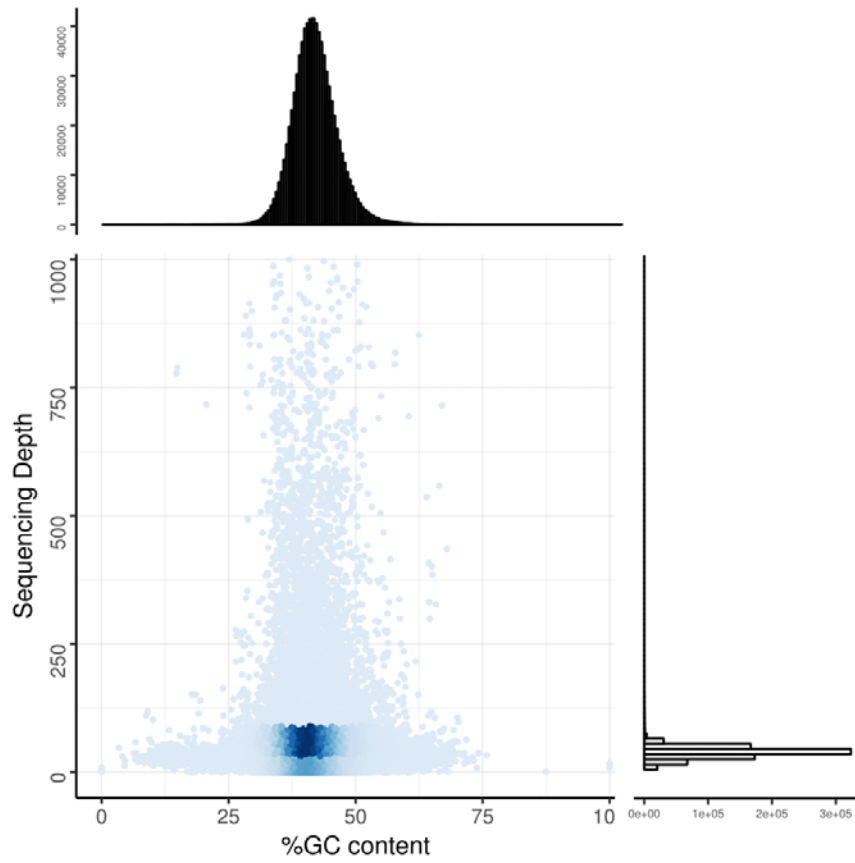


Figure S2. The GC content of male *O. fasciatus* base on the PacBio platform for genome assembly GC statistics based on the CLR (continuous long reads) mapping information using a sliding window approach (10kb windows), Related to Table 1.

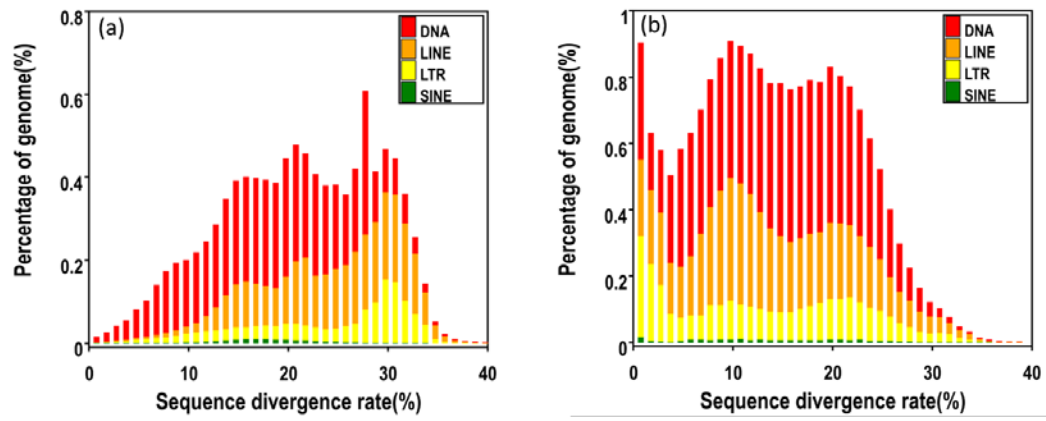


Figure S3. Interspersed repeats statistics of genome assembly for male *O. fasciatus*. (a) Rebase library, (b) *de novo* library, Related to Table 2 and Table 3.



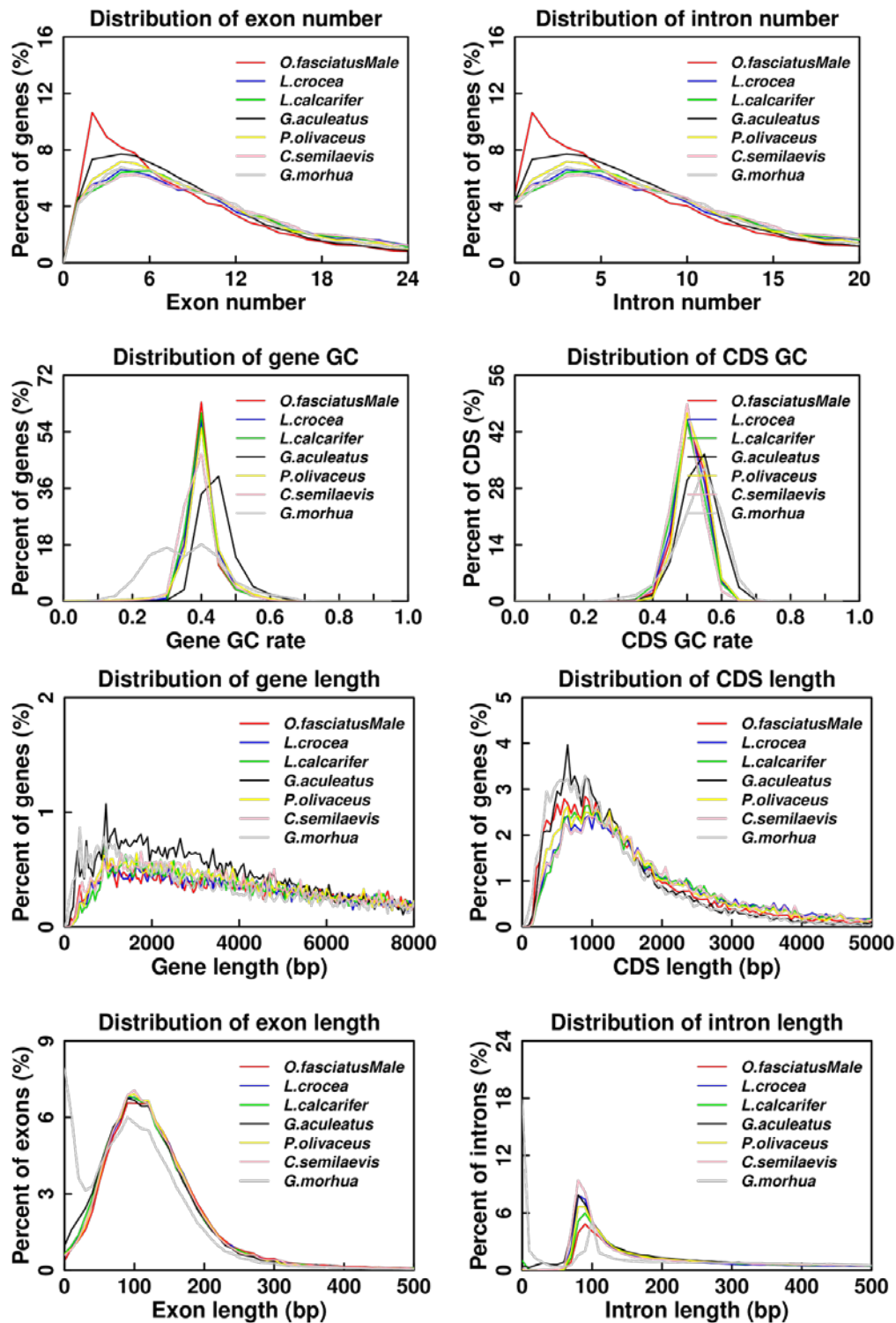


Figure S4. The gene number, GC content, gene length distribution, CDS length distribution, exon length distribution and intron length distribution for male *O. fasciatus* were all comparable with those in other teleost fish species (*Larimichthys crocea*, *Lates calcarifer*, *Gasterosteus aculeatus*, *Paralichthys olivaceus*, *Cynoglossus semilaevis* and *Gadus morhua*), Related to Table 1 and Table 2.

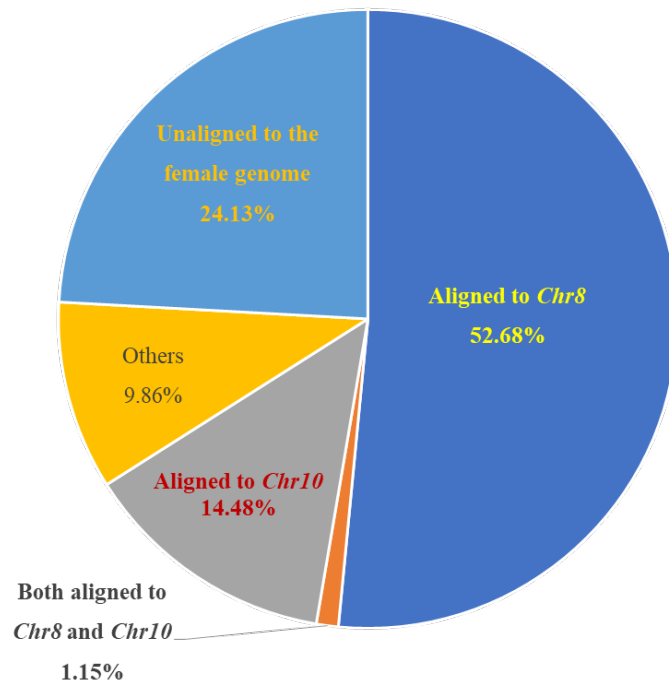


Figure S5. Distribution of the rest chromosomal sequences of the male *Ch9* compared to the whole female genome of *O. fasciatus*, Related to Figure 3 and Table S13.

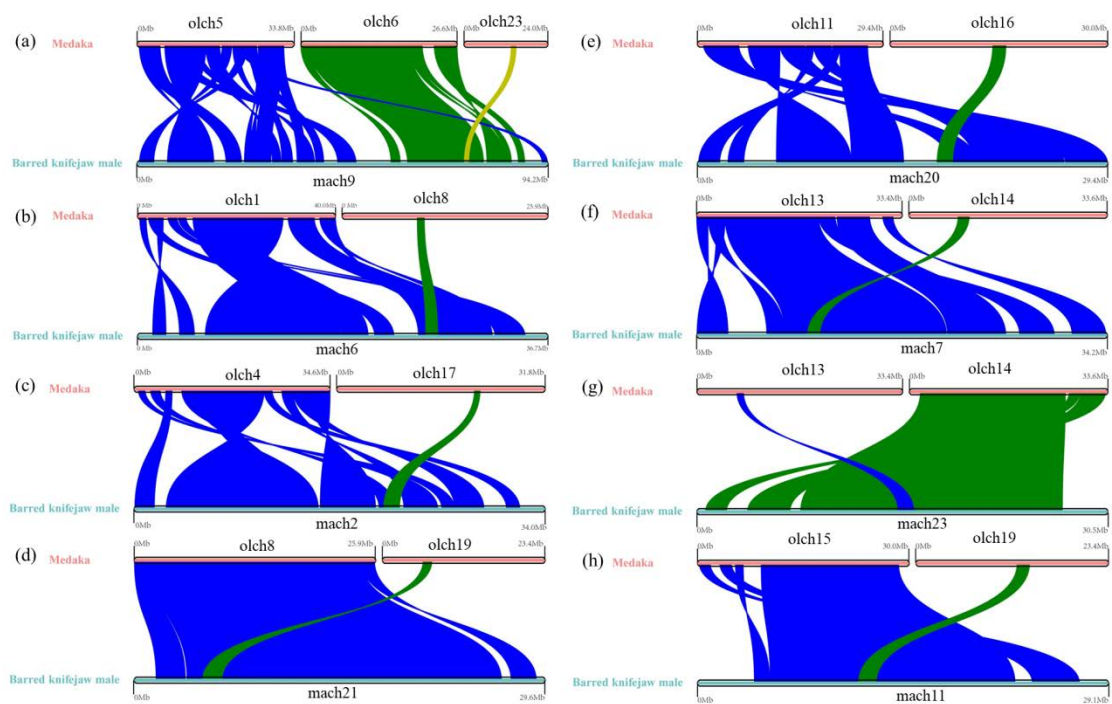


Figure S6. (a)-(h) Chromosome conserved synteny with one-to-two or one-to-three collinear alignment relationships between *Oryzias latipes* (the normal XY system) and *O. fasciatus* (the  $X_1X_2Y$  system), Related to Figure 4 and Table S15..

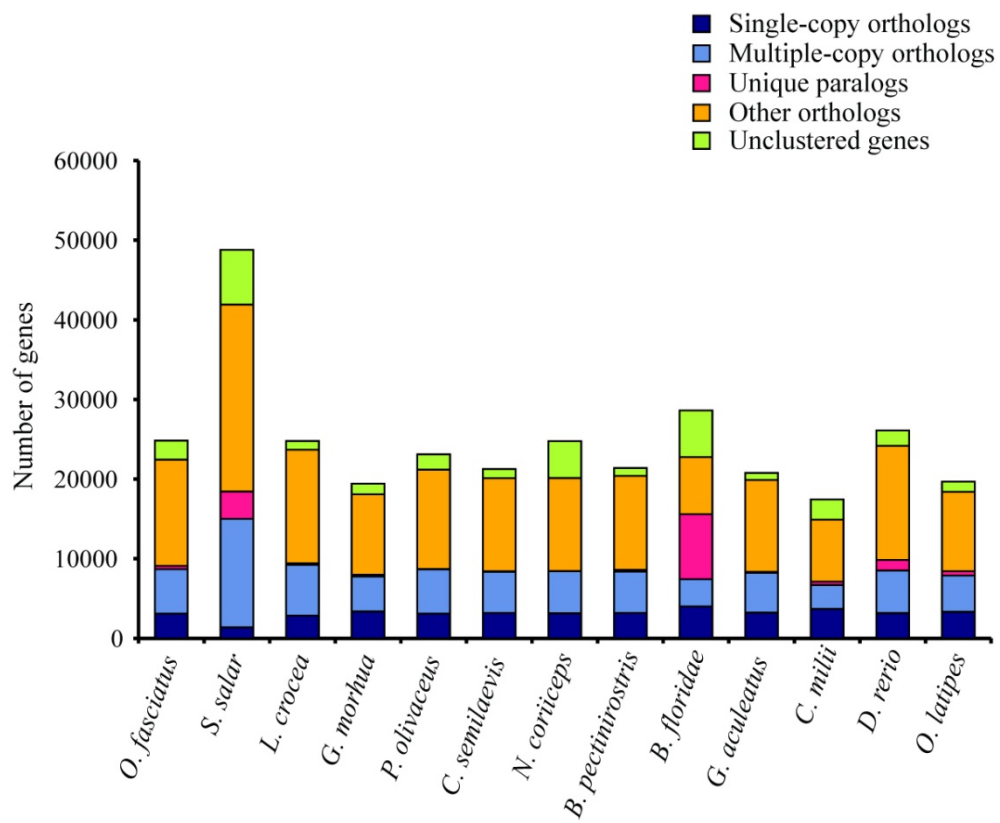


Figure S7. Comparing genome assemblies between male *O. fasciatus* and other 12 species, Related to Figure 5.

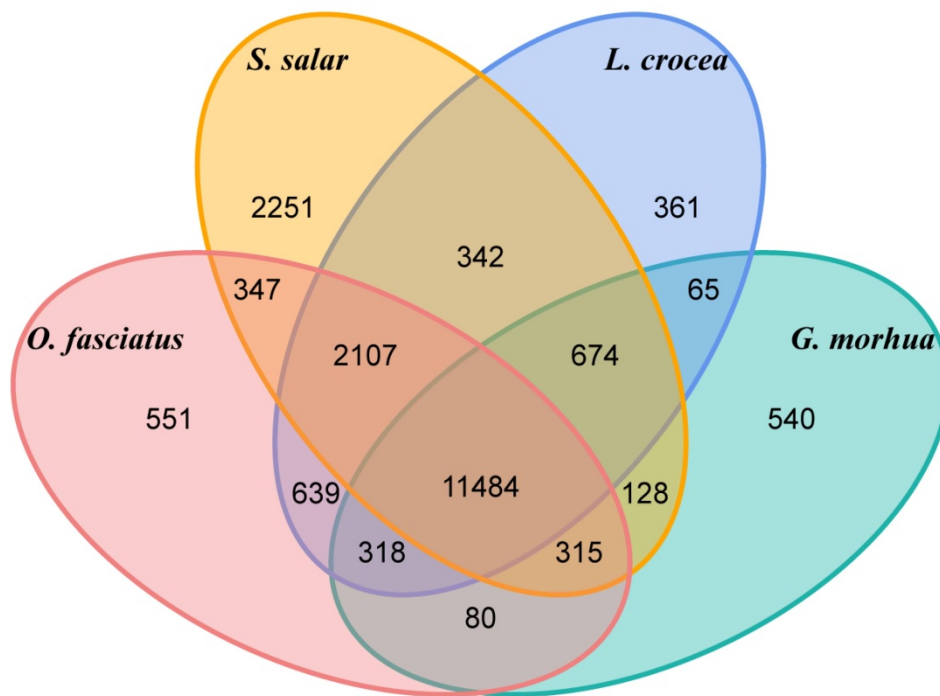


Figure S8. The Venn Diagram of orthologous gene families between male *Oplegnathus fasciatus* and the other four fish genomes (*Larimichthys crocea*, *Gadus morhua* and *Salmo salar*), Related to Figure 5 and Table 4.



## Supplemental Tables

Table S1. Summary of sequence data from male *O. fasciatus*, Related to Figure 1 and Table 1.

| Type       | Method        | Library size (bp) | Data size (Gb) | Read N50 (bp) |
|------------|---------------|-------------------|----------------|---------------|
| DNA        | HiSeq 2000    | 300~350           | 63.9           | 150           |
| DNA        | PacBio Sequel | 22,000            | 39.8           | 18,850        |
| DNA (Hi-C) | HiSeq 2000    | 300~350           | 95.9           | 150           |
| RNA        | HiSeq 2000    | 150               | 42.2           | 150           |

Table S2. Genome assembly statistics for male *O. fasciatus*, Related to Figure 1, Figure 2, Figure 3 and Table 1.

| Method    | Type   | Genome size (Mb) | Longest sequence (Mb) | Sequences number | Sequence N50 (Mb) |
|-----------|--------|------------------|-----------------------|------------------|-------------------|
| Canu      | Contig | 866.9            | 9.820                 | 4,453            | 1.7331            |
| Redundans | Contig | 794.8            | 9.820                 | 2,295            | 2.1271            |
| Smrtlink  | Contig | 794.8            | 9.820                 | 2,295            | 2.1271            |
| Pilon     | Contig | 795.1            | 9.825                 | 2,295            | 2.1271            |
| Final     | Contig | 795.1            | 9.825                 | 2,295            | 2.1271            |

Table S3. Comparing genome assemblies between *O. fasciatus* and other fish species  
Some data were cited from the reference, Related to Figure 5 and Table 1.

| Species                                | Estimated<br>Size (Mb) | Genome<br>Contig N50 (Mb) | Scaffold N50 (Mb) |
|--|------------------------|---------------------------|-------------------|
| <i>Oplegnathus fasciatus</i><br>male   | 795                    | 2.13                      | 32.43             |
| <i>Oplegnathus fasciatus</i><br>female | 768                    | 2.13                      | 33.55             |
| <i>Larimichthys crocea</i> *           | 679                    | 0.06                      | 1.03              |
| <i>Ictalurus punctatus</i> *           | 845                    | 0.05                      | 7.20              |
| <i>Paralichthys olivaceus</i> *        | 546                    | 0.03                      | 3.90              |
| <i>Hippocampus comes</i> *             | 501                    | 0.04                      | 1.80              |
| <i>Gadus morhua</i> *                  | 643                    | 0.12                      | 1.15              |
| <i>Channa argus</i> *                  | 615                    | 0.08                      | 4.50              |
| <i>Anguilla rostrata</i> *             | 1,410                  | 0.01                      | 0.086             |
| <i>Anguilla anguilla</i> *             | 860                    | 0.001                     | 0.077             |
| <i>Lates calcarifer</i> *              | 700                    | 1.06                      | 25.80             |
| <i>Clupea harengus</i> *               | 850                    | 0.02                      | 1.84              |
| <i>Scophthalmus<br/>maximus</i> *      | 568                    | 0.03                      | 4.30              |
| <i>Müichthys miuiy</i> *               | 636                    | 0.07                      | 1.15              |
| <i>Lepisosteus oculatus</i> *          | 945                    | 0.07                      | 6.90              |
| <i>Sillago sinica</i>                  | 534                    | 2.60                      | -                 |
| <i>Lates calcarifer</i> *              | 586                    | 1.07                      | 25.85             |
| <i>Oreochromis niloticus</i> *         | 868                    | 3.30                      | 37.00             |

\* The data was cited from the reference Gong et al., 2018.

Table S4. Alignment of clean reads for Hi-C data and Hi-C libraries for chromosome-scale assembly, Related to Figure 1, Figure 2, Figure 3 and Table 1.

| Clean data  | 650,652,366          |                   |                      |
|---|----------------------|-------------------|----------------------|
| Paired  | 621,961,380 (95.59%) |                   |                      |
| With mate mapped to a different contig (or scaffold)                  | 235,931,390 (37.93%) |                   |                      |
| With mate mapped to a different contig (or scaffold) (map Q $\geq$ 5) | 202,370,680 (32.54%) |                   |                      |
| Chromosome  | Number of contigs    | Length of contigs | Length of chromosome |
| Chr1  | 27                   | 19,038,723        | 19,051,723           |
| Chr2  | 35                   | 34,023,620        | 34,040,620           |
| Chr3  | 31                   | 32,802,329        | 32,817,329           |
| Chr4  | 50                   | 33,465,990        | 33,490,490           |
| Chr5  | 35                   | 37,834,216        | 37,851,216           |
| Chr6  | 50                   | 36,641,980        | 36,666,480           |
| Chr7  | 48                   | 34,194,075        | 34,217,575           |
| Chr8  | 31                   | 30,421,100        | 30,436,100           |
| Chr9  | 444                  | 94,021,031        | 94,242,531           |
| Chr10   | 42                   | 29,231,680        | 29,252,180           |

|       |       |             |             |
|-------|-------|-------------|-------------|
| Chr11 | 33    | 29,145,809  | 29,161,809  |
| Chr12 | 50    | 32,406,821  | 32,431,321  |
| Chr13 | 75    | 28,647,297  | 28,684,297  |
| Chr14 | 34    | 28,866,314  | 28,882,814  |
| Chr15 | 47    | 27,458,915  | 27,481,915  |
| Chr16 | 30    | 27,627,606  | 27,642,106  |
| Chr17 | 43    | 25,918,704  | 25,939,704  |
| Chr18 | 36    | 33,054,315  | 33,071,815  |
| Chr19 | 33    | 24,737,159  | 24,753,159  |
| Chr20 | 54    | 29,419,855  | 29,446,355  |
| Chr21 | 40    | 29,573,728  | 29,593,228  |
| Chr22 | 57    | 32,586,236  | 32,614,236  |
| Chr23 | 30    | 30,484,110  | 30,498,610  |
| Total | 1,355 | 761,601,613 | 762,267,613 |

Table S5. The completeness and coverage statistics of genome assembly for male *O. fasciatus* using CLR mapping information, Related to Table 2.

| Mapping rate (%) | Average sequencing depth | Coverage (%) | Coverage ( $\geq 5X$ , %) | Coverage ( $\geq 10X$ , %) | Coverage( $\geq 20X$ , %) |
|------------------|--------------------------|--------------|---------------------------|----------------------------|---------------------------|
| <b>87.63</b>     | 50.05                    | 99.89        | 99.55                     | 98.94                      | 94.24                     |

Table S6. The accuracy estimation of genome assembly for male *O. fasciatus* based on SNP calling, Related to Table 2.

| SNP                        | Number    | Percentage of SNP (%) | Percentage of genome (%) |
|----------------------------|-----------|-----------------------|--------------------------|
| <b>All SNP</b>             | 1,873,725 | 100.00                | 0.24                     |
| <b>Heterozygosis SNP</b>   | 1,863,418 | 99.45                 | 0.23                     |
| <b>Homology SNP</b>        | 10,307    | 0.55                  | 0.00                     |
| <b>All indel</b>           | 608,220   | 100.00                | 0.08                     |
| <b>Heterozygosis indel</b> | 573,975   | 94.37                 | 0.07                     |
| <b>Homology indel</b>      | 34,245    | 5.63                  | 0.00                     |

Table S7. Repeat content comparison of male *O. fasciatus* with several other teleost, Related to Table 3 and Figure S3.

|                           | <i>Gasterosteus</i><br><i>aculeatus</i> | <i>Larimichthys</i><br><i>crocea</i> | <i>Oryzias</i><br><i>latipes</i> | <i>Dicentrarchus</i><br><i>labrax</i> | <i>Triplophysa</i><br><i>tibetana</i> | <i>Epinephelus</i><br><i>lanceolatus</i> | <i>Oplegnathus</i><br><i>fasciatus</i> male | <i>Oplegnathus</i><br><i>fasciatus</i> female |
|---------------------------|---|--------------------------------------|----------------------------------|---------------------------------------|---------------------------------------|--|---|---|
| <b>SINE (%)</b>           | 0.42                                    | 0.63                                 | 0.89                             | 0.53                                  | 1.01                                  | 1.33                                     | 0.25  | 0.12  |
| <b>LINE (%)</b>           | 2.76                                    | 2.76                                 | 5.1                              | 2.72                                  | 12.39                                 | 11.31                                    | 7.08  | 7.34  |
| <b>LTR (%)</b>            | 2.85                                    | 1.79                                 | 1.41                             | 0.9                                   | 10.99                                 | 5.20                                     | 4.11  | 4.00  |
| <b>DNA (%)</b>            | 4.74                                    | 5.24                                 | 10.06                            | 7.25                                  | 16.63                                 | 20.06                                    | 10.55                                       | 11.95   |
| <b>Tandem repeats (%)</b> | 2.03                                    | 2.7                                  | 0.92                             | 2.81                                  | 1.06                                  | /  | 1.03  | 1.01  |
| <b>Total (%)</b>          | 16.12                                   | 18.63                                | 28.59                            | 21.88                                 | 39.75                                 | 38.59                                    | 23.16                                       | 23.62   |

Table S8. Repeat content in female *Ch8*, *Ch10* and male *Ch9* of *O. fasciatus*, Related to Figure S5 and Table 3.

| <b>Repeat types</b>       | <b>Female <i>Ch8</i></b> | <b>Female <i>Ch10</i></b> | <b>Male <i>Ch9</i></b> |
|---------------------------|--------------------------|---------------------------|------------------------|
| <b>SINE (%)</b>           | 0.06                     | 0.1                       | 0.17                   |
| <b>LINE (%)</b>           | 7.24                     | 5.98                      | 7.10                   |
| <b>LTR (%)</b>            | 3.52                     | 3.75                      | 5.02                   |
| <b>DNA (%)</b>            | 13.31                    | 11.06                     | 10.23                  |
| <b>Tandem repeats (%)</b> | 0.81                     | 0.91                      | 0.87                   |
| <b>Total (%)</b>          | 26.07                    | 22.70                     | 23.79                  |



Table S9. Transcriptome data from RNA-seq for male *O. fasciatus*, Related to Table 4.

| Sample | Before filter           |         |         | After filter           |         |         | Gene No. | Read N50 | Max length | Average length |
|--------|-------------------------|---------|---------|------------------------|---------|---------|----------|----------|------------|----------------|
|        | Before filter data (bp) | Q20 (%) | Q30 (%) | After filter data (bp) | Q20 (%) | Q30 (%) |          |          |            |                |
|        |                         |         |         |                        |         |         | 62,467   | 2,450 bp | 59,651 bp  | 1,214 bp       |
| CK-1   | 4 193 837 760           | 96.13   | 91.49   | 3,946,007,277          | 98.13   | 94.56   |          |          |            |                |
| CK-2   | 4 542 887 548           | 96.23   | 91.66   | 4,287,840,962          | 98.16   | 91.61   |          |          |            |                |
| CK-3   | 3 707 145 734           | 97.34   | 93.32   | 3,557,038,640          | 98.45   | 95.12   |          |          |            |                |
| VC-1   | 3 855 448 572           | 97.42   | 93.49   | 3,703,645,388          | 98.49   | 95.23   |          |          |            |                |
| VC-2   | 4 565 074 582           | 97.40   | 93.44   | 4,386,344,462          | 98.47   | 95.19   |          |          |            |                |
| VC-3   | 4 180 130 282           | 97.34   | 93.30   | 4,012,959,725          | 98.45   | 95.10   |          |          |            |                |
| TP-1   | 3 379 514 994           | 97.35   | 93.32   | 3,242,985,938          | 98.45   | 95.13   |          |          |            |                |
| TP-2   | 3 281 341 740           | 97.43   | 93.47   | 3,157,433,918          | 98.47   | 95.16   |          |          |            |                |
| TP-3   | 3 647 614 890           | 97.45   | 93.55   | 3,509,897,398          | 98.50   | 95.25   |          |          |            |                |
| NC-1   | 3 398 536 162           | 97.44   | 93.52   | 3,268,777,185          | 98.49   | 95.22   |          |          |            |                |
| NC-2   | 3 404 172 992           | 97.44   | 93.53   | 3,276,779,782          | 98.48   | 95.24   |          |          |            |                |
| NC-3   | 3 488 720 308           | 97.51   | 93.66   | 3,359,820,359          | 98.52   | 95.31   |          |          |            |                |

\*The data was cited from the reference Xiao et al., 2019.

Table S10. Gene annotation of male *O. fasciatus* genome, Related to Figure 1, Figure 3 and Table 4.

| Gene set       |                               | Number | Average gene length (bp) | Average CDS length (bp) | Average exon per gene | Average exon length (bp) | Average intron length (bp) |
|----------------|-------------------------------|--------|--------------------------|-------------------------|-----------------------|--------------------------|----------------------------|
| <i>De novo</i> | AUGUSTUS                      | 29,784 | 13,418.44                | 1,453.16                | 8.34                  | 174.34                   | 1,631.18                   |
|                | Genscan                       | 35,112 | 16,257.97                | 1,524.37                | 8.61                  | 177.15                   | 1,937.33                   |
| Homolog        | <i>Larimichthys crocea</i>    | 46,228 | 9,281.13                 | 1,115.50                | 5.93                  | 188.27                   | 1,657.97                   |
|                | <i>Lates calcarifer</i>       | 47,999 | 9,284.06                 | 1,128.68                | 5.94                  | 190                      | 1,650.76                   |
|                | <i>Gasterosteus aculeatus</i> | 38,627 | 8,821.88                 | 1,063.55                | 6.14                  | 173.25                   | 1,509.74                   |
|                | <i>Paralichthys olivaceus</i> | 40,557 | 8,970.29                 | 1,114.68                | 6.08                  | 183.25                   | 1,545.55                   |
|                | <i>Cynoglossus semilaevis</i> | 40,155 | 9,646.94                 | 1,132.68                | 6.16                  | 183.83                   | 1,649.56                   |
|                | <i>Gadus morhua</i>           | 12,282 | 8,739.86                 | 966.93                  | 6.03                  | 160.24                   | 1,543.95                   |
| RNAseq         |                               | 20,297 | 15,195.19                | 1,242.78                | 8.74                  | 340.38                   | 1,578.61                   |
| BUSCO          |                               | 4,781  | 13,202.83                | 1,956.20                | 12.3                  | 159.09                   | 995.60                     |
| MAKER          |                               | 24,357 | 17,192.56                | 1,740.06                | 10.31                 | 251.18                   | 1,569.42                   |
| HICESAP        |                               | 24,835 | 15,819.40                | 1,707.02                | 10.01                 | 220.05                   | 1,511.67                   |

Table S11. The annotation of non-coding RNAs of male *O. fasciatus* genome, Related to Table 2.

| Type  |          | Copy | Average length (bp) | Total length (bp) | % of genome |
|-------|----------|------|---------------------|-------------------|-------------|
| miRNA |          | 593  | 85.66               | 50,797            | 0.0064      |
| tRNA  |          | 921  | 75.28               | 69,329            | 0.0087      |
| rRNA  | rRNA     | 134  | 193.52              | 25,932            | 0.0033      |
|       | 18S      | 7    | 1,668.86            | 11,682            | 0.0015      |
|       | 28S      | 0    | 0.00                | 0                 | 0.0000      |
|       | 5.8S     | 5    | 156.00              | 780               | 0.0001      |
|       | 5S       | 122  | 110.41              | 13,470            | 0.0017      |
|       | 8S       | 0    | 0.00                | 0                 | 0.0000      |
| snRNA | snRNA    | 442  | 135.07              | 59,701            | 0.0075      |
|       | CD-box   | 127  | 109.98              | 13,967            | 0.0018      |
|       | HACA-box | 75   | 144.15              | 10,811            | 0.0014      |
|       | Splicing | 231  | 143.28              | 33,098            | 0.0042      |
|       | scaRNA   | 9    | 202.78              | 1,825             | 0.0002      |

Table S12. Homology comparisons of chromosomal sequences and genes between female  $X_1X_1X_2X_2$  and male  $X_1X_2Y$  corresponding to the female *Ch8/Ch10* and the male *Ch9* of *O. fasciatus*, Related to Figure 3 and Table S4.

|   | Female <i>Ch8</i> vs male <i>Ch9</i> | Female <i>Ch10</i> vs male <i>Ch9</i> | Male <i>Ch9</i> vs female <i>Ch8/Ch10</i> |
|---|--------------------------------------|---------------------------------------|---|
| *Total length of homology sequences                             | 31,327,076 bp                        | 31,843,995 bp                         | 63,130,076 bp                             |
| **Percentages of whole chromosome length for homology sequences | 83.36%                               | 90.14%                                | 66.99%                                    |
| Total numbers of genes  | 1194                                 | 1139                                  | 3064                                      |
| Numbers of homologous genes                                     | 1041                                 | 968                                   | 2809                                      |
| Coverage ratios of homologous genes                             | 87.19%                               | 84.99%                                | 91.68%                                    |

\*Aligned by using MUMmer with `nucmer -num -l 100 -c 1000 -D5; delta-filter -i 95 -o 95 -l`.

\*\*Aligned by `ncbi-blast-2.2.29+` with parameter `1e-5` using the gene CDS sequences.

Table S13. Statistics of the rest chromosomal sequences (33.0%) of the male *Ch9* compared to the whole female genome of *O. fasciatus* by using MUMmer with parameter setting of `mum -l 100 -c 100 -D 5`, Related to Figure 3 and Figure S5.

|              | The rest chromosomal sequences of the male <i>Ch9</i> with length more than 10 kb | Alignment to the whole female genome | Unaligned to the whole female genome | Only aligned to the female <i>Ch8</i> | Only aligned to the female <i>Ch10</i> | Both aligned to the female <i>Ch8</i> and <i>Ch10</i> | Others       |
|--------------|---|--------------------------------------|--------------------------------------|---------------------------------------|--|---|--------------|
| Total length | 29,925,193 bp   | 22,703,944 bp                        | 7,221,249 bp                         | 15,763,128 bp                         | 4,333,338 bp                           | 343,922 bp  | 2,951,400 bp |
| Ratios       |   | 75.87%                               | 24.13%                               | 52.68%                                | 14.48%                                 | 1.15%   | 9.86%        |

Table S14. Representatives of 172 male-specific genes with location and annotation details of *O. fasciatus*, Related to Figure 3.

| Male_Chr9 | Gene_start | Gene_end | Gene_ID   | NR_def   | GO_ID   | GO_term  | KO_ID  | KO_gene | KEGG Pathway               |
|-----------|------------|----------|-----------|--|---|--|--------|---------|----------------------------|
| Chr9      | 392143     | 395375   | Ofa000011 | Uncharacterized protein<br>LOC109083330<br>[Cyprinus carpio]   | GO:0005654;<br>GO:0007010;<br>GO:0000981;<br>GO:0022604;<br>GO:0007275;<br>GO:0008270;<br>GO:0005737;<br>GO:0003677 | Nucleoplasm; cytoskeleton organization; RNA polymerase II transcription factor activity, sequence-specific DNA binding; regulation of cell morphogenesis; multicellular organism development; zinc ion binding; cytoplasm; DNA binding | -      | -       | -                          |
| Chr9      | 5305798    | 5328719  | Ofa000102 | [Larimichthys crocea]  | GO:0031110  | Regulation of microtubule polymerization or depolymerization   | -      | -       | -                          |
| Chr9      | 10475126   | 10483190 | Ofa000279 | Synaptonemal complex protein 1 [Oreochromis niloticus] &gt;<br>XP_013132081.1<br>PREDICTED:<br>synaptonemal complex protein 1 [Oreochromis niloticus];<br>XP_005458454.1<br>PREDICTED:<br>synaptonemal complex protein 1 [Oreochromis niloticus] &gt;<br>XP_005458453.1<br>PREDICTED:<br>synaptonemal complex protein 1 [Oreochromis niloticus] &gt; | -   | -  | K19533 | SYCP1   | -                          |
| Chr9      | 16925807   | 16926699 | Ofa000496 | Caveolin-3-like<br>[Scleropages formosus]  | GO:0016020;<br>GO:0000139;<br>GO:0016021;<br>GO:0032947;<br>GO:0005901;   | Membrane; Golgi membrane; integral component of membrane; protein complex scaffold activity; caveola; regulation of signaling; muscle cell cellular homeostasis;   | K12959 | CAV3    | Focal adhesion;Endocytosis |

|             |                 |                 |                  |  |  |  |        |            |   |
|-------------|-----------------|-----------------|------------------|--|--|--|--------|------------|---|
|             |                 |                 |                  |  | GO:0023051; caveola assembly; plasma membrane; ion channel binding; Golgi apparatus  |  |        |            |   |
|             |                 |                 |                  |  | GO:0046716;  |  |        |            |   |
|             |                 |                 |                  |  | GO:0070836;  |  |        |            |   |
|             |                 |                 |                  |  | GO:0005886;  |  |        |            |   |
|             |                 |                 |                  |  | GO:0044325;  |  |        |            |   |
|             |                 |                 |                  |  | GO:0005794   |  |        |            |   |
| <b>Chr9</b> | <b>19585888</b> | <b>19589330</b> | <b>Ofa000610</b> | Microfibrillar-associated protein 2 isoform X1 [Larimichthys crocea] > XP_019128312.1 PREDICTED: microfibrillar-associated protein 2-like isoform X1 [Larimichthys crocea] |  |  |        |            |   |
| <b>Chr9</b> | 20468886        | 20471223        | Ofa000667        | Splicing factor 3 [Larimichthys crocea]; serine/arginine-rich splicing factor 3 isoform X1 [Larimichthys crocea]   | GO:0000166; GO:0003676   | Nucleotide binding; nucleic acid binding   | K12892 | SFRS3      | Spliceosome   |
| <b>Chr9</b> | <b>46210055</b> | <b>46211646</b> | <b>Ofa001651</b> | Protein P [Larimichthys crocea]  | GO:0034080; GO:0000775   | CENP-A containing nucleosome assembly; chromosome, centromeric region  | K11508 | CENPP      | -   |
| <b>Chr9</b> | <b>23563653</b> | <b>23568436</b> | <b>Ofa000846</b> | Nuclear receptor subfamily 4 group A member 1-like, partial [Larimichthys crocea]  | GO:0043565; GO:0003707; GO:0046872; GO:0008270; GO:0003677; GO:0005634; GO:0006355; GO:0043401; GO:0004879; GO:0003700; GO:0006351 | Sequence-specific DNA binding; steroid hormone receptor activity; metal ion binding; zinc ion binding; DNA binding; nucleus; regulation of transcription, DNA-templated; steroid hormone mediated signaling pathway; RNA polymerase II transcription factor activity, ligand-activated sequence-specific DNA binding; transcription factor activity, sequence-specific DNA binding; transcription, DNA-templated | K04465 | NR4A1,H MR | MAPK signaling pathway; PI3K-Akt signaling pathway; Aldosterone synthesis and secretion |
| <b>Chr9</b> | 67479363        | 67485472        | Ofa002284        | Protein regulator of cytokinesis 1-like isoform X1 [Lates calcarifer]  | GO:0008017; GO:0001578; GO:0000910; GO:0005856;  | Microtubule binding; microtubule bundle formation; cytokinesis; cytoskeleton; microtubule cytoskeleton organization  | K16732 | PRC1       | -   |

---

|             |                 |                 |                  |  |   |  |        |      |   |
|-------------|-----------------|-----------------|------------------|--|---|--|--------|------|---|
| <b>Chr9</b> | <b>85600993</b> | <b>85602200</b> | <b>Ofa002902</b> | Transmission fidelity<br>protein 8 [Anoplopoma<br>fimbria] | GO:0000226<br>GO:0031390;<br>GO:0007064 | Ctf18 RFC-like complex; mitotic<br>sister chromatid cohesion | K11270 | CTF8 | - |
|-------------|-----------------|-----------------|------------------|--|---|--|--------|------|---|

---

Note: The details of 172 male-specific genes with location and annotation of *O. fasciatus* in the suppinformation file.

Table S15. Coverage ratios of homologous genes from the syntenic blocks in chromosomes of Barred knifejaw and Medaka (female Barred knifejaw vs male Barred knifejaw and Medaka vs male Barred knifejaw), Related to Figure 4, Figure S6.

| <b>Chromosome ID of female knifejaw</b> | <b>Chromosome ID of Barred male knifejaw</b> | <b>Chromosome ID of Barred Medaka</b> | <b>Coverage ratio of female Barred knifejaw (female <i>O. fasciatus</i> vs male <i>O. fasciatus</i>)</b> | <b>Coverage ratio of male Barred knifejaw ( male <i>O. fasciatus</i> vs <i>O. latipes</i> )</b> | <b>Coverage ratio of Medaka (<i>O. latipes</i> vs male <i>O. fasciatus</i> )</b> |
|---|--|---------------------------------------|--|---|--|
| fechr_1                                 | machr_1                                      | olch2                                 | 81.02%   | 44.07%  | 43.30%   |
| fechr_2                                 | machr_2                                      | olch4/olch17                          | 86.97%   | 56.57%/0.70%  | 50.99%/0.59%   |
| fechr_3                                 | machr_3                                      | olch16                                | 87.52%   | 58.19%  | 61.14%   |
| fechr_4                                 | machr_4                                      | olch7                                 | 86.43%   | 58.52%  | 63.60%   |
| fechr_5                                 | machr_5                                      | olch3                                 | 86.16%   | 53.82%  | 57.96%   |
| fechr_6                                 | machr_6                                      | olch1/olch8                           | 79.35%   | 46.44%/0.90%  | 52.25%/0.97%   |
| fechr_7                                 | machr_7                                      | olch13/olch14                         | 81.97%   | 52.34%/0.78%  | 53.07%/0.82%   |
| fechr_9                                 | machr_8                                      | olch21                                | 84.13%   | 48.32%  | 47.38%   |
| fechr_8                                 | machr_9                                      | olch5/olch6/olch23                    | 82.63%/82.18%  | 23.79%/19.06%/0.16%   | 51.67%/55.77%/0.89%  |
| fechr_10                                |  |                                       |  |   |  |
| fechr_11                                | machr_10                                     | olch24                                | 76.53%   | 45.31%  | 44.87%   |
| fechr_12                                | machr_11                                     | olch15/olch19                         | 85.27%   | 51.39%/2.41%  | 50.77%/2.62%   |
| fechr_13                                | machr_12                                     | olch10                                | 84.79%   | 55.27%  | 61.34%   |
| fechr_14                                | machr_13                                     | olch18                                | 63.49%   | 46.92%  | 54.35%   |
| fechr_15                                | machr_14                                     | olch22                                | 85.82%   | 50.87%  | 49.47%   |
| fechr_16                                | machr_15                                     | olch20                                | 83.86%   | 58.69%  | 61.00%   |
| fechr_17                                | machr_16                                     | olch12                                | 87.07%   | 54.93%  | 54.50%   |
| fechr_18                                | machr_17                                     | olch19/olch15                         | 84.49%   | 53.84%/1.72%  | 61.47%/1.79%   |
| fechr_19                                | machr_18                                     | olch9                                 | 84.86%   | 60.54%  | 57.34%   |
| fechr_20                                | machr_19                                     | olch23                                | 82.26%   | 55.26%  | 62.97%   |



|          |          |               |        |              |              |
|----------|----------|---------------|--------|--------------|--------------|
| fechr_21 | machr_20 | olch11/olch16 | 83.44% | 50.40%/1.36% | 48.63%/1.12% |
| fechr_22 | machr_21 | olch8/olch19  | 84.89% | 62.40%/0.62% | 62.49%/0.92% |
| fechr_23 | machr_22 | olch17        | 84.42% | 63.70%       | 58.26%       |
| fechr_24 | machr_23 | olch14/olch13 | 78.13% | 53.08%/1.11% | 53.64%/1.09% |

---

Note: The details of homology locations between Barred knifejaw and Medaka in the suppinformation file.

## Transparent Methods

### METHOD DETAILS

#### Sample collection and sequencing

A male *O. fasciatus* captured from the location (35.49N, 119.69E) of Qingdao city (Yellow Sea), Shandong province was used for genome sequencing and assembly. High-quality genomic DNA was extracted from the muscle and blood samples of the same male *O. fasciatus* individual using the PacBio sequel platform (Pacific Biosciences of California, Menlo Park, CA, USA) and the Illumina HiSeq X Ten platform (Illumina Inc., San Diego, CA, USA), respectively. Two genomic DNA libraries (20 kb) were constructed using the manufacturer's protocol provided by the Pacific Biosciences. Sequencing was conducted using Sequel Binding Kit 2.0, Sequel Sequencing Kit 2.1 and Sequel SMRT Cell 1M v2 on the PacBio sequel platform. Five SMRT Cells were used for the male genome sequencing and SMRT LINK 5.0 was used to filter the raw data from the zero-mode waveguide.

To obtain a chromosome-level genome assembly for male *O. fasciatus*, a Hi-C library was constructed for sequencing. The whole blood from the same male *O. fasciatus* individual was extracted for Hi-C library construction. The blood cells were cross-linked using formaldehyde and then lysed. A restriction enzyme (MboI) was used to digest the cross-linked DNA, and then, the sticky ends were filled and marked with biotin. Then, the biotin-labelled chromosomes in proximity were ligated to create chimeric junctions representing their physical interactions in the chromosome system. Finally, the cross-linked DNA was unlinked using a protease, and the chimeric junctions of the genome were sheared to a size of 300-500 bp (Belaghzal et al., 2017). A PE library with an approximately 300 bp insert size was constructed according to the Illumina library preparation protocol (Illumina Inc., San Diego, CA, USA). The sequencing of the Hi-C library was conducted using the Illumina HiSeq X Ten platform (Illumina) with a read length of 150 bp for each end.

High-quality RNA was also extracted from the eggs of *O. fasciatus* for genome annotation, and sequencing was performed using the Illumina HiSeq X Ten in 150PE

mode (Xiao et al. 2019).

### **Genome assembly of the male *O. fasciatus* genome**

The genome size of female *O. fasciatus* was estimated using a 17-mer frequency distribution analysis (Marcais & Kingsford 2011; Xiao et al. 2019). The genome size of female *O. fasciatus* was calculated to be 777.5 Mb with a heterozygous ratio of 0.29% and a repeat percentage of 38.46% (Xiao et al. 2019). According to the genome survey of female *O. fasciatus* combined with the subreads obtained using the PacBio Sequel platform, the Canu v1.4 with the Corrected-Error-Rate parameter set to 0.040 was first used to complete the *de novo* assembly of the male *O. fasciatus* genome (Koren et al. 2017). And then, the Redundans v0.13c was used to remove sequence redundancy with the minCoverage parameter of 15 and the Arrow tool in SMRT Link 5.0 software was used to polish genome sequence using long-read sequencing data (Pryszcz & Gabaldón, 2016; Chin et al., 2013). Finally, the Pilon software was applied to polish the assembly genome using the Illumina NGS short reads mentioned in the genome survey of female *O. fasciatus* (Walker et al., 2014).

The high-quality Hi-C data was further used to assist the genome assembly of the male *O. fasciatus* to a chromosome level. The reads from the Hi-C library sequencing were first mapped to the polished genome using BWA software with default parameters (Li & Durbin, 2009). To avoid the influence of long reads, reads apart from each restriction sites more than 500bp were filtered out using BWA and Lachesis software (Burton et al., 2013). And then the Lachesis software with tuned parameters was used to cluster, order and orientate the contigs using their contact frequencies. Finally, the Juicer software with JucieBox was combined to construct a physical map of chromosome interactions and complete the visual correction with the interaction map (Durand et al., 2016). LOWESS method (alpha = 0.5%, ignore zeros, IQR filter) was explored to model the overall chromatin interaction decay with distance (Sanyal et al., 2012). Cumulative distribution of subtraction Hi-C using Z-scores values for interactions between 400 kb and 40 kb bins were carried out from the whole genome and chromosome levels.

### **Genome quality evaluation for the male *O. fasciatus***

First, the Minimap2 software using the CLR (continuous long reads) data was employed to evaluate the completeness and homogeneity of the assembled genome with default settings (Li, 2018). Then, the completeness of the assembled genome was evaluated using BUSCO v3 software with the actinopterygii\_odb9 database (Simão et al., 2015). Finally, the BWA software (BWA, RRID: SCR\_010910) with the default settings was applied to align the NGS-based short reads from the whole-genome sequencing data to the reference genome and evaluated the accuracy of the assembled genome based on the called SNP loci using GATK (Li & Durbin, 2009; Mckenna et al., 2010).

### **Repetitive element annotation and protein-coding gene prediction**

We used The Tandem Repeat Finder was used to detect repetitive elements in the male *O. fasciatus* genome (Benson, 1999). A *de novo* repeat library of the male *O. fasciatus* genome was constructed using Repeat Modeler (version 1.04) and LTR\_FINDER with default parameters. The known and novel transposable elements (TEs) were identified using Repeat Masker (version 3.2.9) from the Repbase TE database (version 14.04) and the *de novo* repeat library (Tarailo-Graovac & Chen, 2009). The TE-relevant proteins in the assembled genome were also identified using Repeat Protein Mask software (version 3.2.2) (Tarailo-Graovac & Chen, 2009).

The prediction of protein-coding genes in the assembled genome of male *O. fasciatus* was conducted using three approaches: homology, *de novo* and transcriptome sequencing. Protein sequences of *Lates calcarifer*, *Larimichthys crocea*, *Gasterosteus aculeatus*, *Gadus morhua*, *Paralichthys olivaceus* and *Cynoglossus semilaevis* were obtained from Ensembl and used for the homology annotation (Flicek et al., 2014). Then, the TBLASTN software with an e-value threshold of 1E-5 was used to align protein sequences to the assembled genome of male *O. fasciatus*. Finally, the Genewise2.2.0 software was employed to predict potential gene structures based on all alignments. Augustus (version 2.5.5) and Genscan (version 1.0) software were used for *de novo* prediction of protein-coding genes (Burge & Karlin, 1997; Stanke et al., 2004). Augustus was trained with the gene set from the homology annotation. Simultaneously, the TopHat 1.2 software was used to map the transcriptome

sequencing reads to the assembled genome (Trapnell et al., 2009). The Cufflinks software was then applied to predicate the potential gene structures based on the sorted and integrated alignments using TopHat software. Finally, the MAKER software was employed to obtain a comprehensive and non-redundant gene set by merging the annotation results of the abovementioned annotation pipelines (Campbell et al., 2014).

### **Functional annotation of predicted protein-coding genes**

To complete the functional annotation of predicted genes in the male *O. fasciatus* genome, all protein sequences were aligned against the Swiss-Prot, the non-redundant nucleotide (NT) and non-redundant protein (NR) database using local BLASTX and BLASTN programs with an e-value threshold of 1E-5 (Lobo, 2008; Altschul et al., 2012). The Blast2GO software was then used to perform the annotation of the Gene Ontology (GO) annotation (Harris et al., 2004; Conesa et al., 2005). The best hit for each protein genes coupled with a biological pathway was also identified using the Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway database (Ogata et al., 2000). In addition, tRNAscan-SE and the Rfam database were also used to annotate the non-coding RNAs (microRNAs, transfer RNAs, ribosomal RNAs, and small nuclear RNAs) with default settings (Lowe & Eddy, 1997; Griffiths-jones et al., 2003).

### **Chromosome comparison with synteny analysis**

The synteny relationships between the male and female *O. fasciatus* genomes were determined using the program MUMmer 3.23 (<http://mummer.sourceforge.net/>) with the default settings (Kurtz et al., 2004). The alignment results were filtered for more than 85% identity and visualized using the mummer plot (Delcher et al., 2003). Synteny analyses were also performed among the male/female *O. fasciatus* genomes and *L. crocea* genomes to detect syntenic blocks. According to the annotated protein-coding genes of male and female *O. fasciatus*, precise gene pairs from the male and female *O. fasciatus* chromosomes was identified based on their synteny and homological searching with BLASTP software with identity  $\geq 95\%$  and e-value  $\leq 1E-5$ . The homology comparison of chromosomal sequences between the female

$X_1X_1X_2X_2$  and the male  $X_1X_2Y$  were also analyzed using nucmer software with parameter settings (mum -l 100 -c 1000/100 -D 5). The BLASTN software with an e-value threshold of  $1E-5$  was further used to compare genes between male *Ch9* and female genome based on the CDS. The mapping results mentioned above were merged into a circos plot using Circos v0.69 software (Krzywinski et al., 2009). The chromosomal structures between *Oryzias latipes* genome (Japanese medaka with the normal XY system, ASM31367v1) and *O. fasciatus* genome (Barred knifejaw with the  $X_1X_2Y$  system) were identified using JCVI v0.7.5 (<https://github.com/tanghaibao/jcvi>) with minspan=30 -simple (Tang et al., 2017). The order of the corresponding chromosomes for the male/female *O. fasciatus* and *L. crocea* followed the standard of male *O. fasciatus* as *Chr1* to *Chr23*. The chromosomes named for the male/female *O. fasciatus* were designated machr and fechr, respectively.

### **Gene family identification for evolutionary analysis**

To identify the gene families for phylogenetic tree construction, we used the genome of male *O. fasciatus* for comparisons with those of other marine species, including *Salmo salar*, *Larimichthys crocea*, *Gadus morhua*, *Paralichthys olivaceus*, *Cynoglossus semilaevis*, *Notothenia coriiceps*, *Boleophthalmus pectinirostris*, *Branchiostoma floridae*, *Gasterosteus aculeatus*, *Callorhinchus milii*, *Danio rerio* and *Oryzias latipes*. To obtain high-confidence results, the protein sequences of the longest transcript of each gene for the abovementioned species were selected. The BLASTP program was used to search for similarities between the protein sequences in an all-to-all manner with an e-value threshold of  $1E-5$  (Altschul et al., 2012). Finally, the Hcluster\_sg software (<https://github.com/douglasgsc/official/hcluster>) was employed to obtain the clustering of gene families based on H-scores calculated from the bit-score.

### **Phylogenetic analysis and divergence time estimation**

The phylogenetic relationships between male *O. fasciatus* and the abovementioned species were constructed using orthologues. We used OrthoMCL V2.0.9 software to generate orthologous groups with a coefficient setting of 1.5 (Li et al., 2003). The

ClustalW program was applied to extract and align the protein sequences of single-copy genes from the orthogroups with a length filter (retention of proteins with length  $\geq 100$  aa) (Thompson et al., 2002). Then, the MUSCLE software was used to conduct multiple sequence alignments for each of single-copy genes from male *O. fasciatus* and other marine species (Edgar, 2004). All of the coding sequence alignments were concatenated to form a super-alignment data set for each species. RAxML software was implemented to estimate phylogenetic relationships among species using the maximum-likelihood method (Stamatakis, 2014). According to the eight calibration times obtained from the TimeTree database (<http://www.timetree.org/>), MCMCtree as implemented in the PAML package was employed to predict the divergence times among species (Yang, 2007).

### **Supplemental References**

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (2002). Basic local alignment search tool (BLAST). *J. Mol. Biol.* 215, 403-410.

Belaghal, H., Dekker, J., and Gibcus, J.H. (2017). Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. *Methods* 123, 56-65.

Benson, G. (1999). Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573.

Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* 268, 78-94.

Burton, J.N., Adey, A., Patwardhan, R.P., Qiu, R., Kitzman, J.O., and Shendure, J. (2013). Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* 31, 1119-1125.

Campbell, M.S., Holt, C., Moore, B., and Yandell, M. (2014). Genome Annotation and Curation Using MAKER and MAKER-P. *Curr. Protoc. Bioinformatics* 48, 4.11.11.

Chin, C.S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E.E., Turner, S.W., and Korlach, J. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10, 563.

Cioffi, M.B., and Bertollo, L.A.C. (2012). Chromosomal distribution and evolution of repetitive DNAs in fish. *Repetit. DNA* 7, 197-221.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674.

Delcher, A.L., Salzberg, S.L., and Phillippy, A.M. (2003). Using MUMmer to identify similar regions in large sequence set. *Current Protocol in Bioinformatics Chapter 10: Unit 10.3*.

Durand, N.C., Shamim, M.S., Machol, I., Rao S.S.P., Huntley, M.H., Lander, E.S., and Aiden E.L. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* 3, 95-98.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792-1797.

Flicek, P., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., et al. (2014). Ensembl 2014. *Nucleic Acids Res.* 42 (Database issue), D749–D755.

Gong, G.R., Dan, C., Xiao, S.J., Guo, W.J., Huang, P.P., Xiong, Y., Wu J.J., He, Y., Zhang, J.C., Li, X.H., et al., (2018). Chromosomal-level assembly of yellow catfish genome using third-generation DNA sequencing and Hi-C analysis. *GigaScience* 7, giy120.

Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S.R. (2003). Rfam: An RNA family database. *Nucleic Acids Res.* 31, 439-441.

Harris, M.A., Clark, J., Ireland, A, Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B, Mungall, C., et al. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* 32, 258D-261.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R. Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Geneome Res.* 27, 722-736.

Krzywinski, M., Schein1, J., Birol1, Í., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra1, M.A. (2009). Circos: An information aesthetic for comparative genomics. *Geneome Res.* 19, 1639-1645.

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

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34,



3094-3100.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754-1760.

Li, L., Stoeckert, C.J., and Roos, D.S. (2003). OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Res.* 13, 2178-2189.

Lobo, I. (2008). Basic local alignment search tool (BLAST). *Nature Educat.* 1.

Lowe, T.M., and Eddy, S.R. (1997). tRNAscan - SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955-964.

Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27, 764.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimell, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297-1303.

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 27, 29-34.

Pryszcz, L.P., and Gabaldón, T. (2016). Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res.* 44, e113-e113.

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.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312-1313.

Stanke, M., Steinkamp, R., Waack, S., and Morgenstern, B. (2004). AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res.* 32, W309-W312.

Sanyal, A., Lajoie, B.R., Jain, G. and Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature* 489, 109-113.

Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*, Chapter 4, Unit 4.10.

Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2002). Multiple sequence alignment using

ClustalW and ClustalX. *Current Protocols in Bioinformatics*, 4.10.1-4.10.14 1-14.

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

Tang, H.B., Krishnakumar, V., Li, J.P., MichelMoser, Maria, and Cheol Yim W. (2017). Jcvi v0.7.5: JCVI utility libraries. Zenodo. doi:10.5281/zenodo.31631.

Walker, B., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q.D., Wortman, J., Yong, S.K., et al. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLOS one* 9, e112963.

Xiao, Y.S., Xiao, Z.Z., Ma, D.Y., Liu, J. and Li, J. (2019). Genome sequence of the barred knifejaw *Oplegnathus fasciatus* (Temminck & Schlegel, 1884): the first chromosome-level draft genome in the family Oplegnathidae. *GigaScience* 8, giz013.

Yang, Z.H. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* 24, 1586-1591.