Journal Pre-proof

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, Angel Herrera-Ulloa

PII: S2589-0042(20)30223-6

DOI: https://doi.org/10.1016/j.isci.2020.101039

Reference: ISCI 101039

To appear in: *ISCIENCE*

Received Date: 11 December 2019

Revised Date: 5 March 2020

Accepted Date: 1 April 2020

Please cite this article as: Xiao, Y., Xiao, Z., Ma, D., Zhao, C., Liu, L., Wu, H., Nie, W., Xiao, S., Liu, J., Li, J., Herrera-Ulloa, A., Chromosome-level Genome Reveals the Origin of Neo-Y Chromosome in the Male Barred knifejaw *Oplegnathus fasciatus*, *ISCIENCE* (2020), doi: https://doi.org/10.1016/j.isci.2020.101039.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

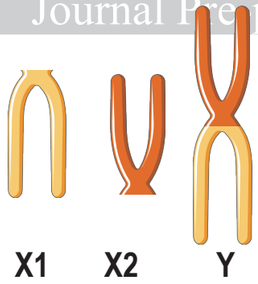
© 2020 The Author(s).

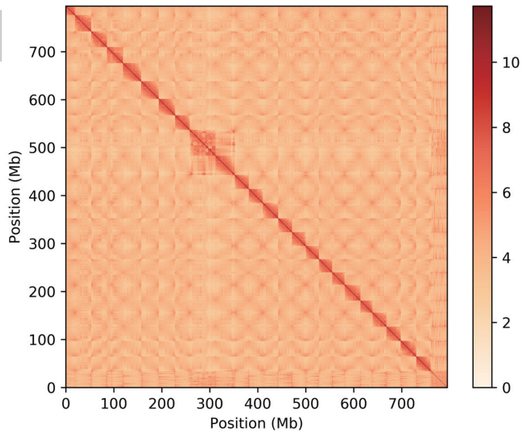
Journal Preproof

Barred knifejaw
Oplegnathus fasciatus

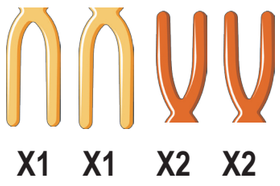
Male

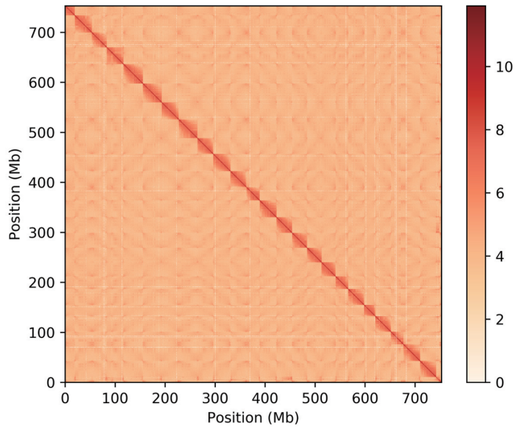
Female



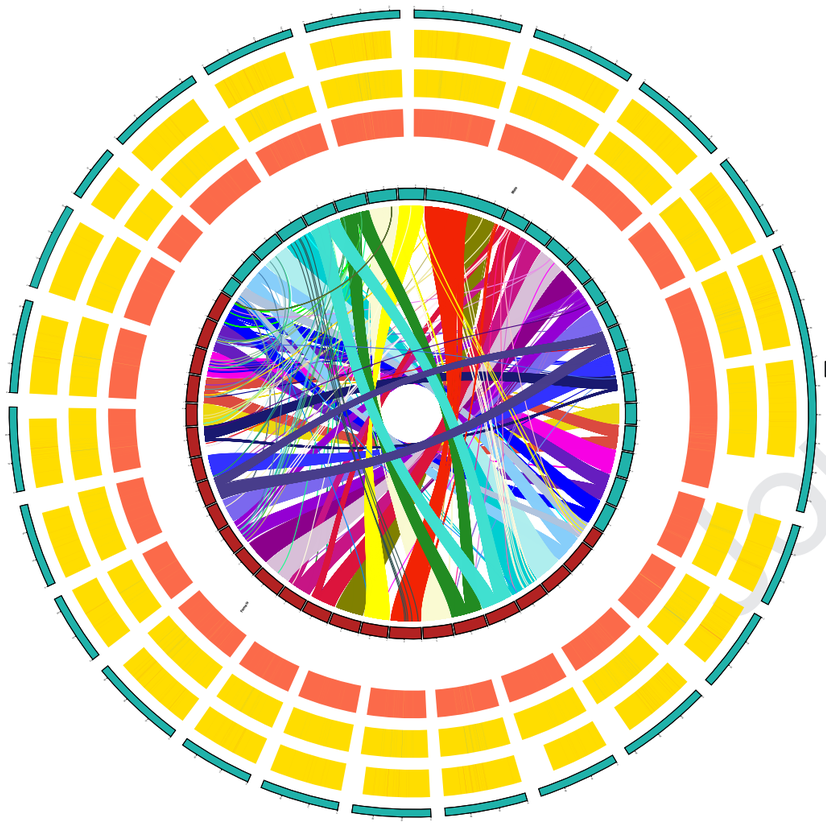


High-throughput chromosome interaction mapping for male genome



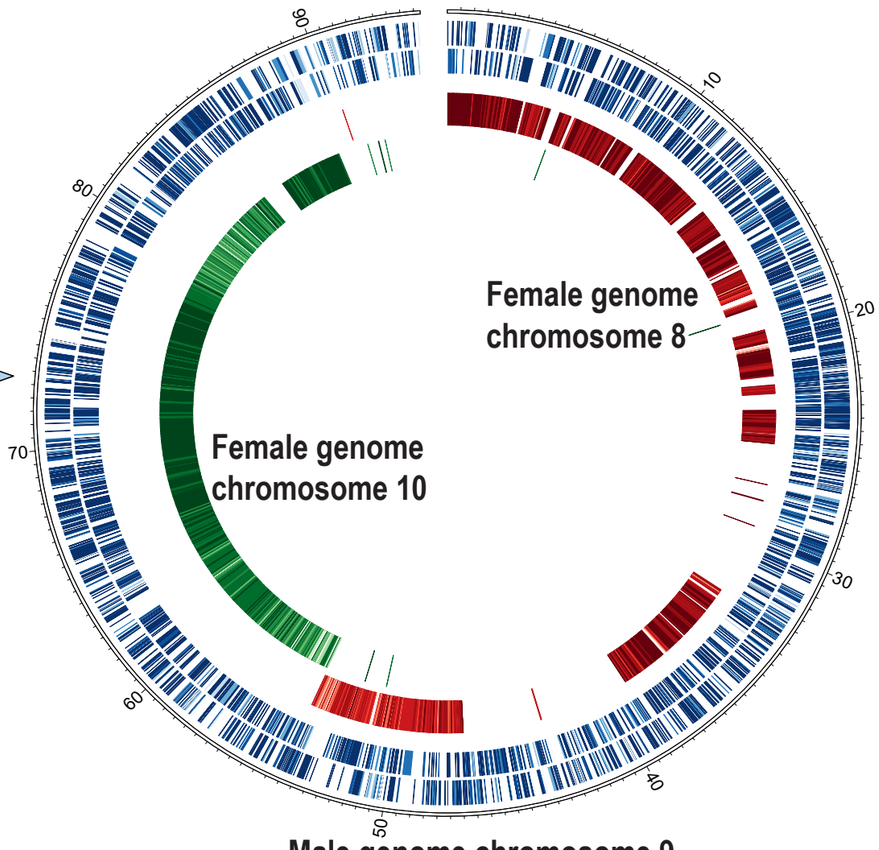


High-throughput chromosome interaction mapping for female genome

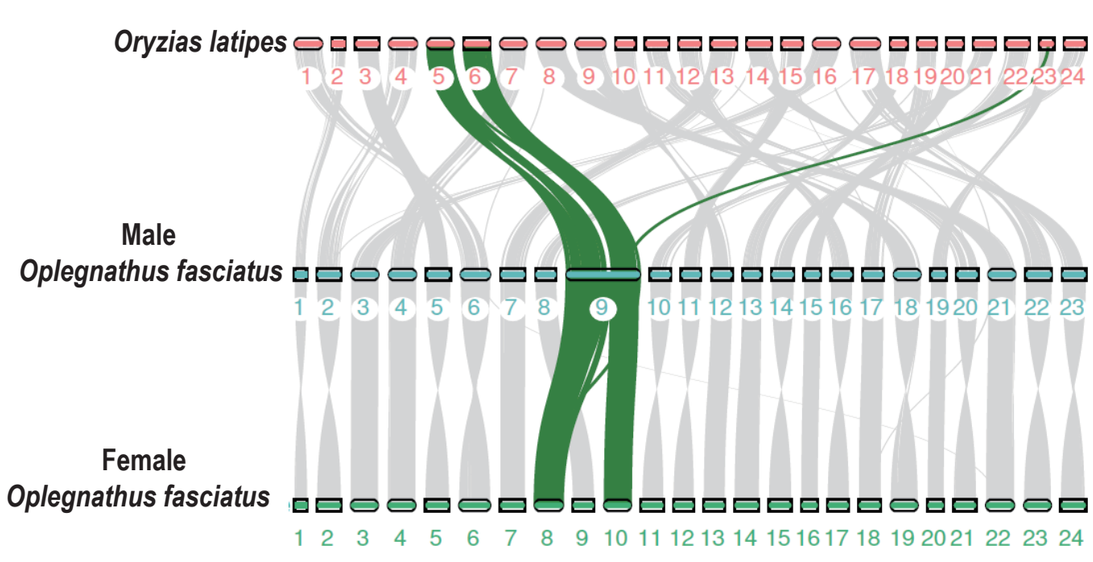


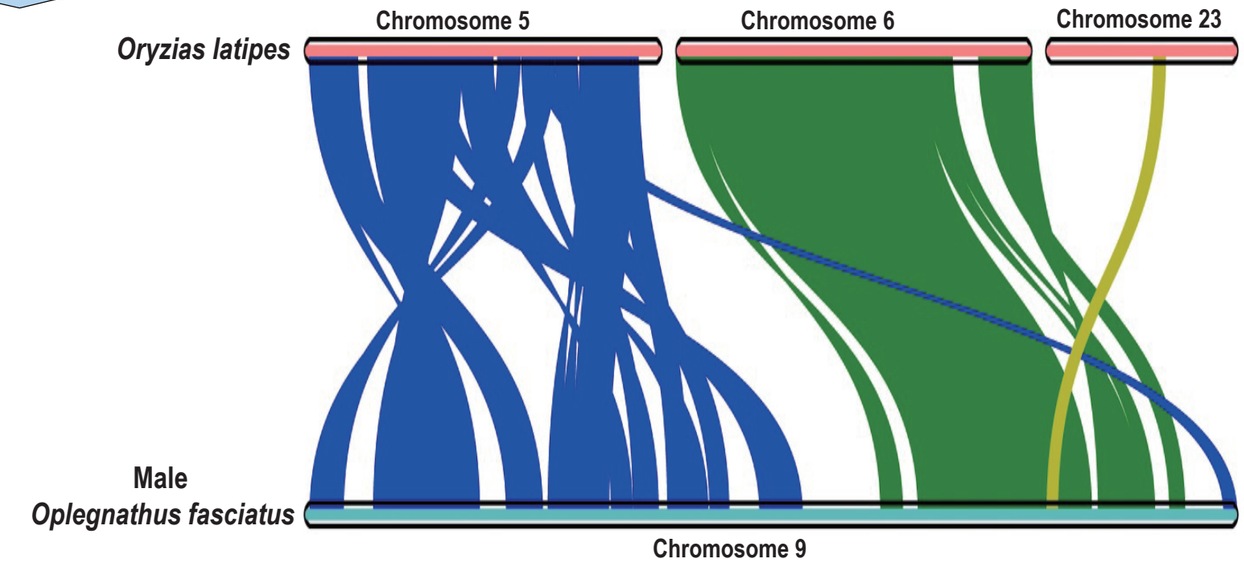
Male genome with 23 chromosomes

Male genome chromosome 9



Male genome chromosome 9





Journal Pre-proof

Chromosome-level Genome Reveals the Origin of Neo-Y Chromosome in the Male Barred knifejaw *Oplegnathus fasciatus*

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

[1]Center for Ocean Mega-Science, The Key Laboratory of Experimental Marine Biology, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China

[2]Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China

[3]BGI Education Center, University of Chinese Academy of Sciences, Shenzhen, China,

[4]Wuhan Frasergen Bioinformatics Co., Ltd. East Lake High-Tech Zone, Wuhan, China.

[5]College of Plant Protection, Jilin Agriculture University, Changchun, Jilin, China

[6]School of Computer Science and Technology, Wuhan University of Technology, Wuhan, China

[7] Escuela de Ciencias Biológicas, Universidad Nacional, San José, Costa Rica

[8]These authors contributed equally

[9]Lead Contact

∗Correspondence:

shijun_xiao@163.com (S.J.X.), jliu@qdio.ac.cn (J.L.), junli@qdio.ac.cn (J.L.)

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 & 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; Xiao et al., 2019). This fish is also a valuable species for sashimi and recreational fishing, and its the ex-factory price has reached up to 30 dollars per kilogram in China (Xiao et al., 2015; Park et al., 2018). It has been reported that the male of *O. fasciatus* has 2n=47 chromosomes (1m + 2m/sm + 44t), while females possess 2n=48 chromosomes (2m/sm + 46t) (Xu et al., 2012; Xu et al., 2019). Similar chromosome karyotypes have also been reported in male and female individuals of *O. 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; Xue et al. 2016; Xu et al., 2019). 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 et al., 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 et al., 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 & 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 need 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 high-throughput chromosome interaction mapping (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 (**Table 1, Table 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 a 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 (**Table S2, Table S3**). 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% (**Figure S1, Figure 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 (**Table S1, Table 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, 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 chromosomes (**Figure 1, Figure 2, Figure 3**). More than 99.7% of contigs that longer than 100kb were anchored on chromosomes, exhibiting the excellent anchoring rate for male *O. fasciatus* chromosome assembly (**Table S4**).

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 was used to evaluate the accuracy of the male *O. fasciatus* genome assembly, which was generated from the alignment of NGS-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 (**Table 3**). The estimation of repetitive element content for the male *O. fasciatus* genome were comparable to the result in 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%), respectively (**Table 3, Table S7, Figure S3**). The repetitive contents of the male *Ch9* and the female *Ch8*/*Ch10* were also identified for 23.79%, 26.07% and 22.70%, respectively (**Table S8**). Although the frequency of DNA transponsons, LINEs and LTRs was higher than that in *L. crocea*, *G. aculeatus*, *O. latipes*, and *D. 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 (**Table S9, Table 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 to those of six related species (*L. calcarifer*, *L. crocea*, *G. aculeatus*, *G. morhua*, *P. olivaceus* and *C. semilaevis*) and showed a similar distribution with those of other teleosts (**Figure S4, 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 functional 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% of complete BUSCO genes were successfully identified (**Table 3**). We also used tRNAscan-SE software to annotate the non-coding RNAs against the Rfam database, and 4 types of non-coding RNAs (miRNAs (0.006%), tRNAs (0.009%), rRNAs (0.007%), and snRNAs (0.015%)), including 2666 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 (Figure 2, Figure 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

Journal Pre-proof

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%), male *Ch23* / female *Ch24* (99.2%). The comparisons of chromosomal sequences of female *Ch8*/*Ch10* and the male *Ch9* were further performed (Figure 3 b). 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 a high identity (~99.0%) with those from the female *Ch8* and *Ch10* using nucmer with minimum match length of 1000 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, Figure S5**). Indeed, structure variations (SV) 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 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 (*sycp1*), caveolin 3 (*cav3*) (**Table S14**). The *ctf8* could regulate sister chromatid cohesion and fidelity of chromosome transmission (Bermudez et al., 2003). The *cenpp* involves in assembly of kinetochore proteins, mitotic progression and chromosome segregation (Okada et al., 2006). The *sycp1* is as 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_1X_2Y system) - *O. latipes* (the normal XY system) comparison using homolog gene-pairs between two species were also performed. As a result, twenty-two of female *O. fasciatus* chromosomes harbored an excellent one-to-one correspondence to those of the male *O. fasciatus* genome excepted for female C*h8* and *Ch10*. The male C*h9* showed strong conserved synteny with female *Ch8* and *Ch10*, consistent with the abovementioned results that *Ch9* might be the neo-Y chromosome (**Table S15, Figure 3, Figure 4, Figure S6**). Furthermore, we found that the synteny of chromosomes for *O. fasciatus* and *O. latipes* were largely conserved. Fourteen chromosomes of male *O. fasciatus* genome were unambiguously aligned to single chromosomes of *O. latipes* genome (**Table S15, Figure 4**). Other 9 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, 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, Figure 4, Figure 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 $\leq$ 1E-5. The results showed that 10,919 protein-coding genes 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, respectively (**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 *S. salar*, *L. crocea*, *G. morhua*, *P. olivaceu*, *C. semilaevis*, *N. coriiceps*, *B. pectinirostris*, *B. floridae*, *G. aculeatus*, *C. milii*, *D. 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 ≥100 aa, 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 data set 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 *Larimichthys crocea* in the order Perciformes (Eupercaria), consistent with the new 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 *Larimichthys 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; Xiao et al. 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; Yang et al., 2018; Sun et al., 2019).

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 the chromosomal studies, the origination of the neo-Y chromosome of *O. fasciatus* in previous studies were 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 et al., 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, 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, respectively (White et al., 1983; Kitano & Peichel, 2012). All those 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 & 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 & Miller, 1971; Bertollo et al. 2000; Bertollo et al., 2004; Ueno & Takai, 2008; Kitano & 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 et al., 1983; Kitano & Peichel, 2012). Our genome assembly for males and females of *O. fasciatus* lead to 23 and 24 chromosomes, directly corresponding to the male and female karyotypes (2n=47/48), respectively (**Figure 1, Figure 2**) (Xu et al., 2012; Xiao et al., 2019). The comparative analysis showed excellent chromosomal synteny between the male and female *O. fasciatus* genomes (**Figure** 1, **Figure** 2, **Figure** 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 & Miller, 1971; Bertollo et al., 2000; Bertollo et

al., 2004; Ueno & Takai, 2008; Kitano & 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 were 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 & Takai 2008; Shao et al., 2012). Previous studies showed that neo-chromosomes might harbor important genes or regulatory elements responsible to 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 for the genetic composition 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 on the growth has been detected in *O. fasciatus* that male fish grow faster than females (Chen et al., 2014; Xiao et al. 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 study

The chromosomes 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* has been deposited in the SRA under Bioproject number PRJNA486572. The whole genome sequencing data are available in the NBCI with the accession number SRP220007.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at http://dx.doi.org/10.17632/5xtt3d9btm.3

ACKNOWLEDGEMENTS

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-STS-QYZX-020), Qingdao National Laboratory for Marine Science and Technology (2018SDKJ0502-2, 2015ASKJ02), China Agriculture 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, Key Laboratory of Mariculture and Enhancement of Zhejiang Province (2016KF002).

AUTHOR CONTRIBUTIONS

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

DECLARATION OF INTERESTS

The authors declare no competing interests.

FIGURE LEGENDS

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 was cited from the reference (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 kb 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.

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 (SV) 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 lines precisely joined pair of genes originated from the male and female *O. fasciatus* chromosomes. (b) and (c) From outer to inner circles: a, the 9th 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 *Ch*9 aligned to *Ch8* of female *O. fasciatus* with red color. e, the chromosomal region of male *Ch*9 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 *Ch*9 genes with the female genes in the b, c track of figure (c).

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.

Figure 5. Phylogenetic analysis of male *O. fasciatus* and other related 12 species. 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.

REFERENCE

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Liponan, D.J. (2012). Basic local alignment search tool (BLAST). J. Mol. Biol. *215*, 403-410.

Belaghzal, 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.

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 cytotaxonomic 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., Lecointre, G., and Ortí, G. (2017). Phylogenetic classification of bony fishes. BMC Evol. Biol. *17*, 162.

Birney, E., Clamp, M., and Durbin, R.J. (2004). GeneWise and genomewise. Genome Res. *14*, 988.

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 $X_1X_1X_2X_2/X_1X_2Y$ 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/XY_1Y_2 sex chromosome

Journal Pre-proof

509 system and in the chromosomal differentiation in *Harttia* species (Siluriformes, Loricariidae). Rev
510 Fish Biol Fisher *23*, 127-134.
511 Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA.
512 J. Mol. Biol. *268*, 78-94.
513 Burton, J.N., Adey, A., Patwardhan, R.P., Qiu, R., Kitzman, J.O., and Shendure, J. (2013).
514 Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions.
515 Nat. Biotechnol. *31*, 1119-1125.
516 Bermudez, V.P., Maniwa, Y., Tappin, I., Ozato, K., Yokomori, K., Hurwitz, J. (2003). The
517 alternative Ctf18-Dcc1-Ctf8-replication factor C complex required for sister chromatid cohesion
518 loads proliferating cell nuclear antigen onto DNA. Proc. Natl. Acad. Sci. USA *100*, 10237-10242.
519 Bisig, C.G., Guiraldelli, M, F., Kouznetsova, A., Scherthan, H., Höög, C., Dawson, D.S., and
520 Pezza, R. J. (2012). Synaptonemal complex components persist at centromeres and are required
521 for homologous centromere pairing in mouse spermatocytes. Plos Genet. *8*, e1002701.
522 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).
523 Chromosome assembly of *Collichthys lucidus*, a fish of Sciaenidae with a multiple sex
524 chromosome system. Sci. Data *6*, 132.
525 Campbell, M.S., Holt, C., Moore, B., and Yandell, M. (2014). Genome Annotation and Curation
526 Using MAKER and MAKER-P. Curr. Protoc. Bioinformatics *48*, 4.11.11.
527 Chen, S.L., Zhang, G.J., Shao, C.W., Huang, Q.F., Liu, G., Zhang, P., Song, W.T., An, N.,
528 Chalopin, D., Volff, J.N., et al. (2014). Whole-genome sequence of a flatfish provides insights into
529 ZW sex chromosome evolution and adaptation to a benthic lifestyle. Nature *46*, 253-260.
530 Chin, C.S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland,
531 A., Huddleston, J., Eichler, E.E., Turner, S.W., and Korlach, J. (2013). Nonhybrid, finished
532 microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods *10*, 563.
533 Cioffi, M.B., and Bertollo, L.A.C. (2012). Chromosomal distribution and evolution of repetitive
534 DNAs in fish. Repetit. DNA *7*, 197-221.
535 Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO:
536 universal tool for annotation, visualization and analysis in functional genomics research.
537 Bioinformatics *21*, 3674.
538 Delcher, A.L., Salzberg, S.L., and Phillippy, A.M. (2003). Using MUMmer to identify similar

Journal Pre-proof

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.

Ferreira, M., Garcia, C., Matoso, D.A., de Jesus, I.S., and Feldberg, E. (2016). A new multiple sex chromosome system X1X1X2X2/X1Y1X2Y2 in Siluriformes: Cytogenetic characterization of *Bunocephalus coracoideus* (Aspredinidae). Genetica *144*, 591-599.

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

Griffithsjones, 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., Irenland, A, Lomax, J., Ashiburner, 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.

Jaillon, O., Aury, J.M., Brunet, F., Petit, J.L., Stange-Thomann, N., Mauceli, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A., et al. (2004). Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate protokaryotype. Nature *431*, 946.

Kasahara, M., Naruse, K., Sasaki, S., Nakatani, Y., Wei, Q., Ahsan, B., Yamada, T, Nagayasu, Y, Doi, K., Kasai, Y., et al. (2007). The medaka draft genome and insights into vertebrate genome evolution. Nature *447*, 714-719.

Kitano, J., and Peichel, C. (2012). Turnover of sex chromosomes and speciation in fishes. Environ. Biol. Fish. *94*, 549-558.

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.

Krysanov, E. and Demidova, T. (2018). Extensive karyotype variability of African fish genus Nothobranchius (Cyprinodontiformes). Comp. Cytogenet. *12*, 387-402.

Journal Pre-proof

569 Krzywinski, M., Schein1, J., Birol1, İ., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and
570 Marra1, M.A. (2009). Circos: An information aesthetic for comparative genomics. Geneome Res.
571 *19*, 1639-1645.
572 Kurtz, S., Phillippy, A., Delcher, A., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L.
573 (2004). Versatile and open software for comparing large genomes. Genome Biol. *5*, R12.
574 Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*,
575 3094-3100.

576 Li, H., and Durbin. R. (2009). Fast and accurate short read alignment with Burrows–Wheeler
577 transform. Bioinformatics *25*, 1754-1760.
578 Li, H., Sun, Z.P., Li, Q., and Jiang, Y.L. (2011). Characterization of an Iridovirus Detected in Rock
579 Bream (*Oplegnathus fasciatus*; Temminck and Schlegel). Chin. J. Virol. *27*, 158-164.
580 Li, L., Stoeckert, C.J., and Roos, D.S. (2003). OrthoMCL: Identification of Ortholog Groups for
581 Eukaryotic Genomes. Genome Res. *13*, 2178-2189.
582 Lobo, I. (2008). Basic local alignment search tool (BLAST). Nature Educat. 1.
583 Lowe, T.M., and Eddy, S.R. (1997). tRNAscan‐SE: A program for improved detection of transfer
584 RNA genes in genomic sequence. Nucleic Acids Res. *25*, 955-964.
585 Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of
586 occurrences of k-mers. Bioinformatics *27*, 764.
587 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimell, K.,
588 Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: A MapReduce
589 framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297-1303.
590 Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (2000). KEGG: kyoto
591 encyclopedia of genes and genomes. Nucleic Acids Res. *27*, 29-34.
592 Okada, M., Cheeseman, I.M., Hori, T., Okawa, K., McLeod, I.X., Yates, J. R., Desai, A., and
593 Fukagawa, T. (2006). The CENP-H-I complex is required for the efficient incorporation of newly
594 synthesized CENP-A into centromeres. Nat Cell Biol. *8*, 446-457.
595 Parise-Maltempi, P.P., Martins, C., Oliveira, C., and Foresti, F. (2007). Identification of a new
596 repetitive element in the sex chromosomes of *Leporinus elongatus* (Teleostei: Characiformes:
597 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.

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

Schembri, P.J., Bodilis, P., Evans, J., and Francour, P. (2010). Occurrence of barred kinfejaw, *Oplegnathuf 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.

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.

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.

Journal Pre-proof

628 Sun, J., Mu, H.W., CH Ip, J., Li, R.S., Xu, T., Accorsi, A., Sánchez Alvarado, A., Ross, E., Lan, Y.,
629 Sun, Y.N., et al. (2019). Signatures of Divergence, Invasiveness, and Terrestrialization Revealed
630 by Four Apple Snail Genomes. Mol. Biol. Evol. *36*, 1507-1520.
631 Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in
632 genomic sequences. Current Protocols in Bioinformatics, Chapter 4, Unit 4.10.
633 Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2002). Multiple sequence alignment using
634 ClustalW and ClustalX. Current Protocols in Bioinformatics, 4.10.1-4.10.14 1-14.
635 Trapnell, C, Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with
636 RNA-Seq. Bioinformatics *25*, 1105-1111.
637 Tang, H.B., Krishnakumar, V., Li, J.P., MicchelMoser, Maria, and Cheol Yim W. (2017). Jcvi
638 v0.7.5: JCVI utility libraries. Zenodo. doi:10.5281/zenodo.31631.
639 Ueno, K, and Takai, A. (2008). Multiple sex chromosome system of $X_1X_1X_2X_2/X_1X_2Y$ type in
640 lutjanid fish, *Lutjanus quinquelineatus* (Perciformes). Genetica *132*, 35-41.
641 Uyeno, T., and Miller, R.R. (1971). Multiple sex chromosomes in a Mexican Cyprinodontid fish.
642 Nature 231, 452-453.
643 Uyeno, T., and Miller, R.R. (1971). Multiple sex chromosomes in a Mexican cyprinodontid fish.
644 Nature *231*, 452-453.
645 Walker, B., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng,
646 Q.D., Wortman, J., Yong, S.K., et al. (2014). Pilon: An integrated tool for comprehensive
647 microbial variant detection and genome assembly improvement. PLOS one *9*, e112963.
648 White, M.J.D. (1983). Animal Cytology and Evolution. Cambridge University Press, Cambridge.
649 Xiao, Y.S., Li, J., Ren, G.J., Ma, G.Y., Wang, Y.F., Xiao, Z.Z., and Xu, S.H. (2016). Pronounced
650 population genetic differentiation in the rock bream *Oplegnathus fasciatus* inferred from
651 mitochondrial DNA sequences. Mitochondrial DNA A *27*, 2045–2052.
652 Xiao, Y.S., Xiao, Z.Z., Ma, D.Y., Liu, J. and Li, J. (2019). Genome sequence of the barred
653 knifejaw Oplegnathus fasciatus (Temminck & Schlegel, 1884): the first chromosome-level draft
654 genome in the family Oplegnathidae. GigaScience *8*, giz013.
655 Xiao, Z.Z. (2015).Study on population genetics and culture biology of *Oplegnathus fasciatus*.
656 Doctor thesis. P, 162-176.
657 Xu, D.D., Sember, A., Zhu, Q.H., de Oliveira E.A., Liehr, T., AI-Rikabi A.B.H., Xiao, Z.Z., Song,

Journal Pre-proof

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.

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

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.

Zheng, Q.Y. (2019). The role of 11β-hydroxylase (Cyp11b2) on gametogenesis in tilapia. Master dissertation. P, 10-30.

Journal Pre-proof

Table 1. Summary of male *O. fasciatus* genome assembly and annotation

Genome assembly			
	Draft scaffolds for male *O. fasciatus*	Chromosome-length scaffolds based on Hi-C for male *O. fasciatus*	**Chromosome-length scaffolds based on Hi-C for female *O. fasciatus*
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		1527.4

** The data was cited from the reference (Yongshuang Xiao, Zhizhong Xiao, Daoyuan Ma, Jing Liu, Jun Li. Genome sequence of the barred knifejaw *Oplegnathus fasciatus* (Temminck & 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)

Journal Pre-proof

Journal Pre-proof

Table 2. Genome quality of *O. fasciatus* based on the BUSCO assessment

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

Journal Pre-proof

Journal Pre-proof

Table 3. The detailed classification of repeat sequences for male *O. fasciatus*

Type	Repbase 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

Journal Pre-proof

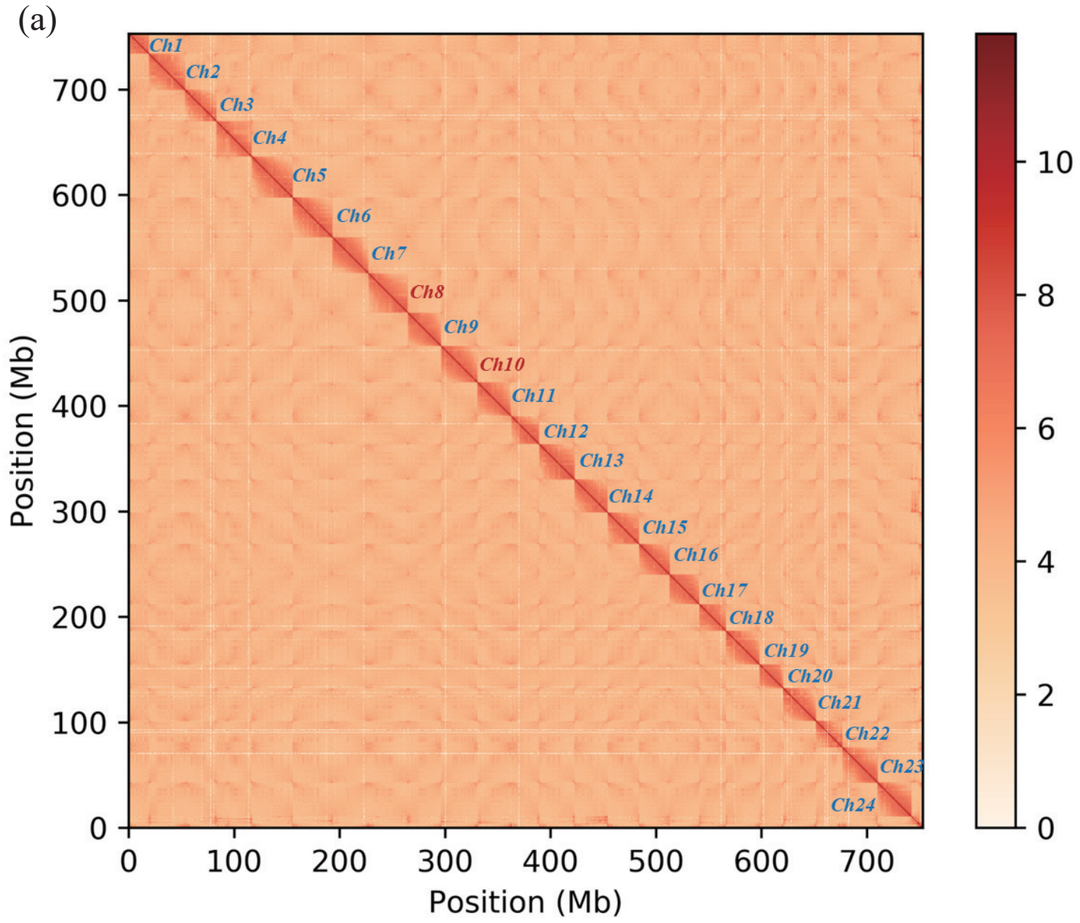
Table 4. Functional annotation of the protein-coding genes in male *O. fasciatus* genome

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

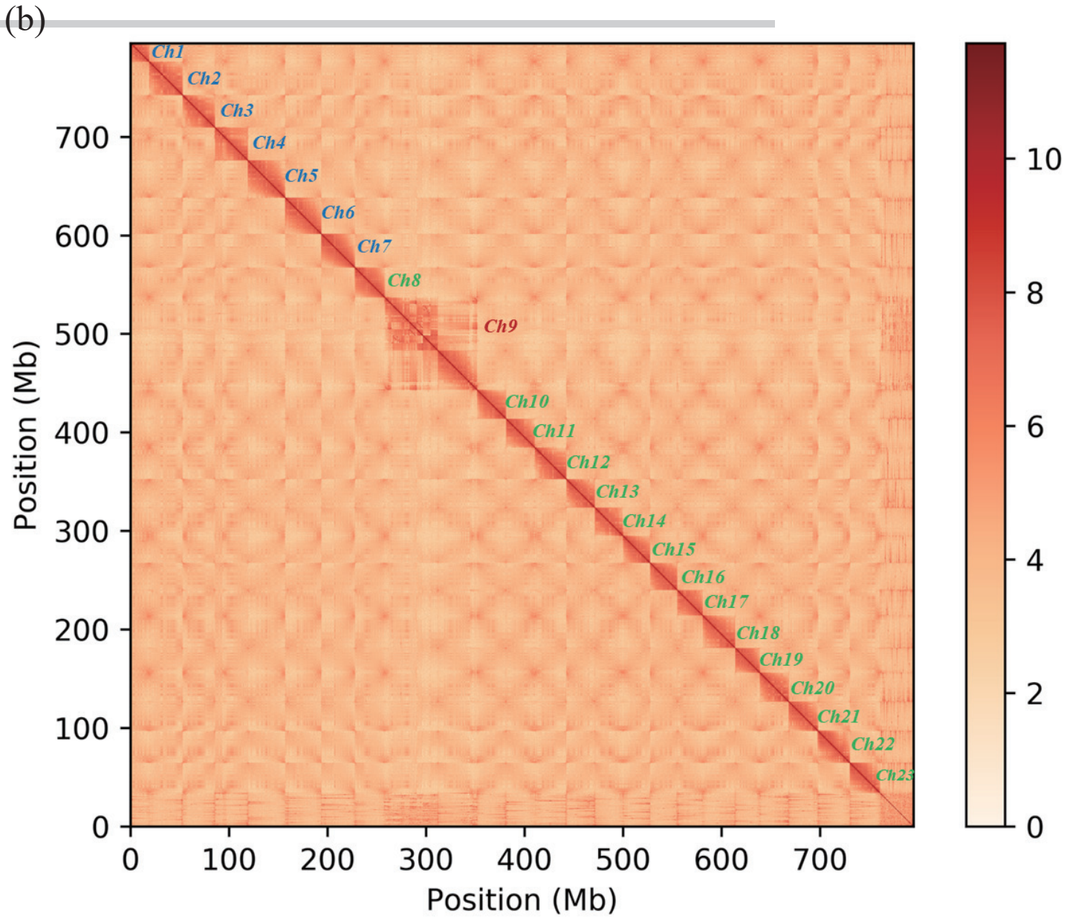
Journal Pre-proof

nal Pre-

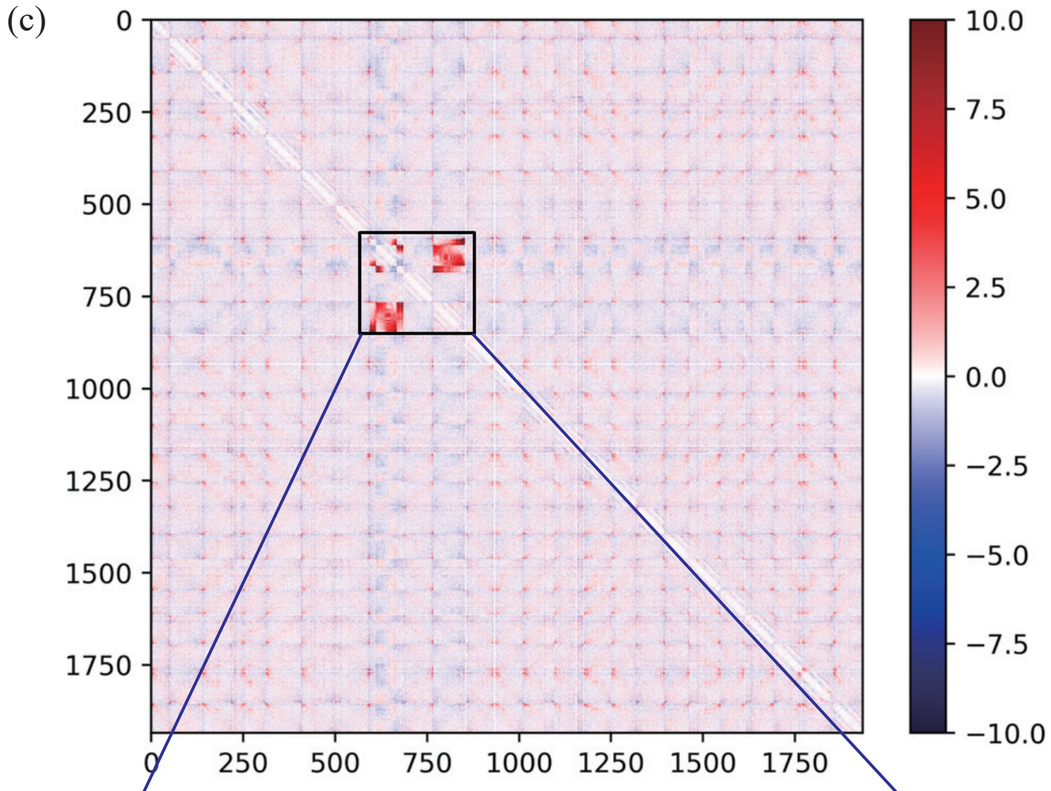
(a)

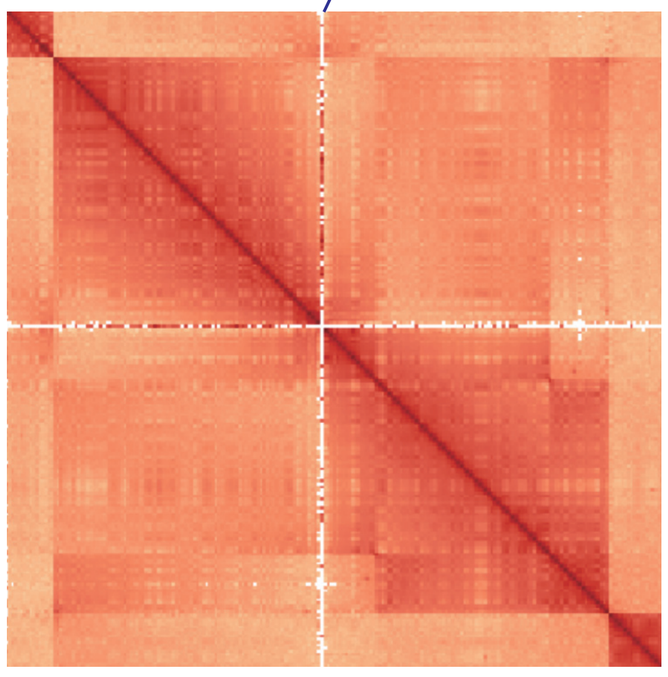


(b)



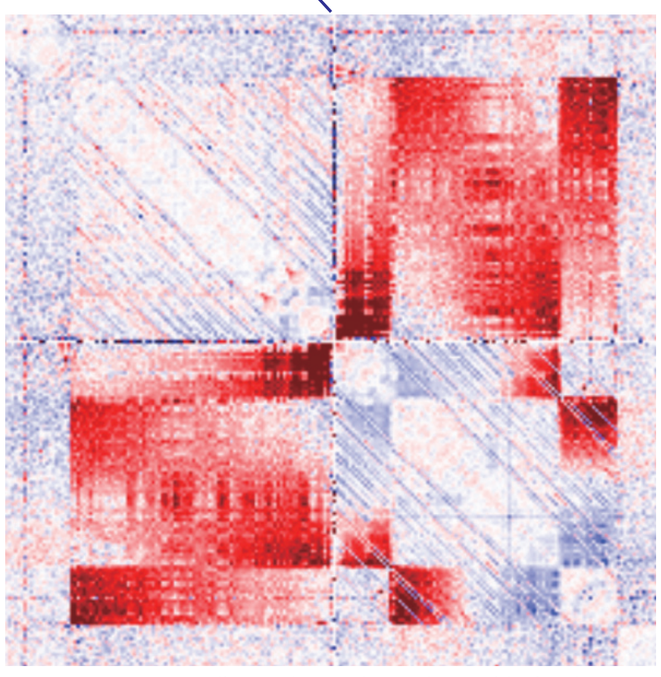
(c)





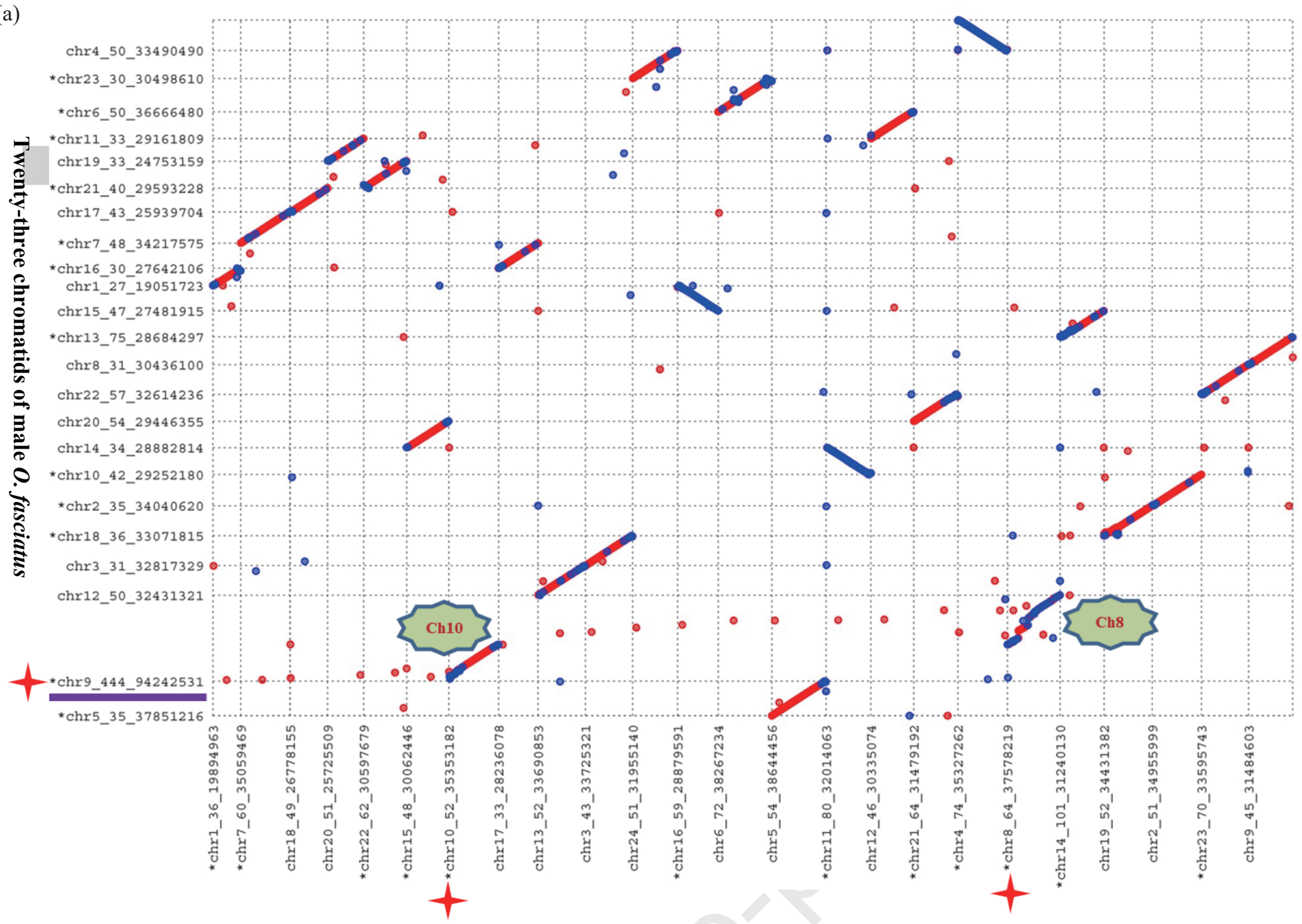
The district corresponding to the number 10 chromatid of female *O. fasciatus*

The district corresponding to the number 8 chromatid of female *O. fasciatus*



Neo-Y corresponding to the number 9 chromatid of male *O. fasciatus*

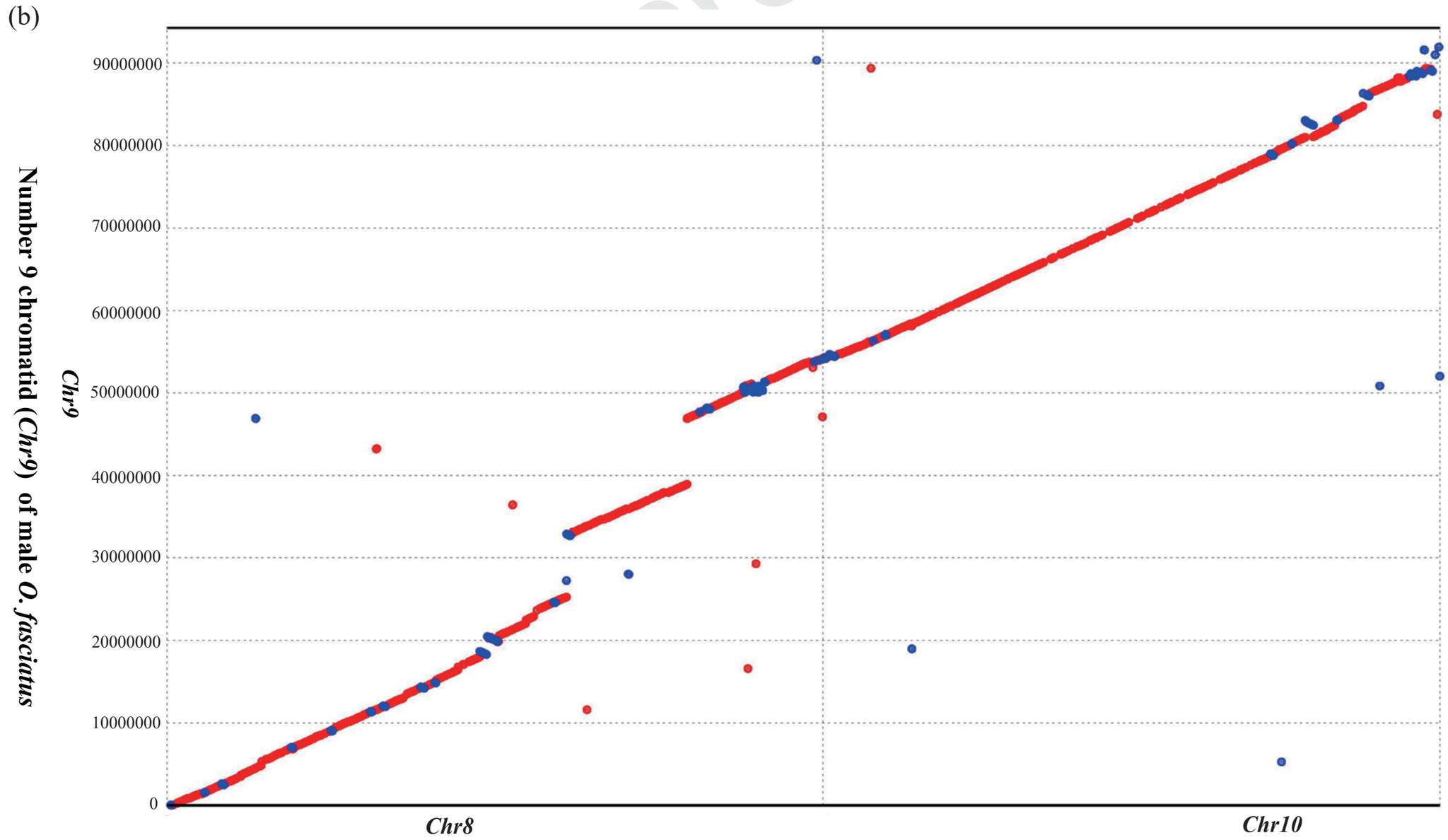
(a)



Twenty-three chromatids of male *O. fasciatus*

Twenty-four chromatids of female *O. fasciatus*

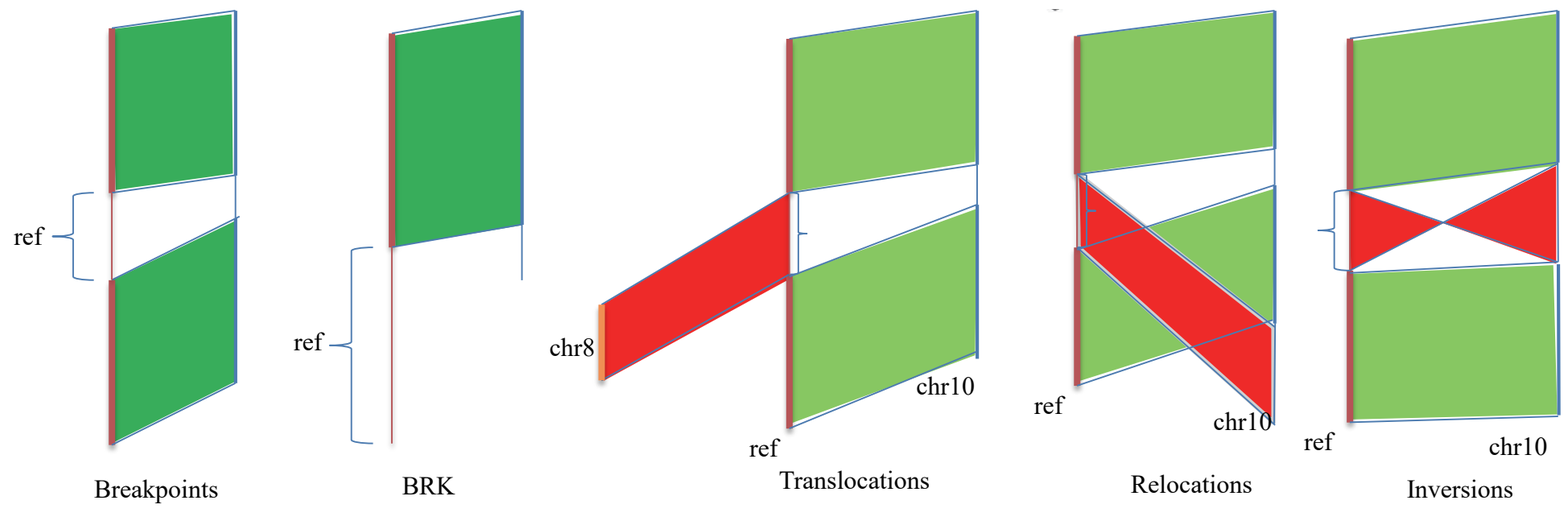
(b)



Number 9 chromatid (*Chr9*) of male *O. fasciatus*

Number 8 and 10 chromatids (*Chr8* and *Chr10*) of female O. fasciatus

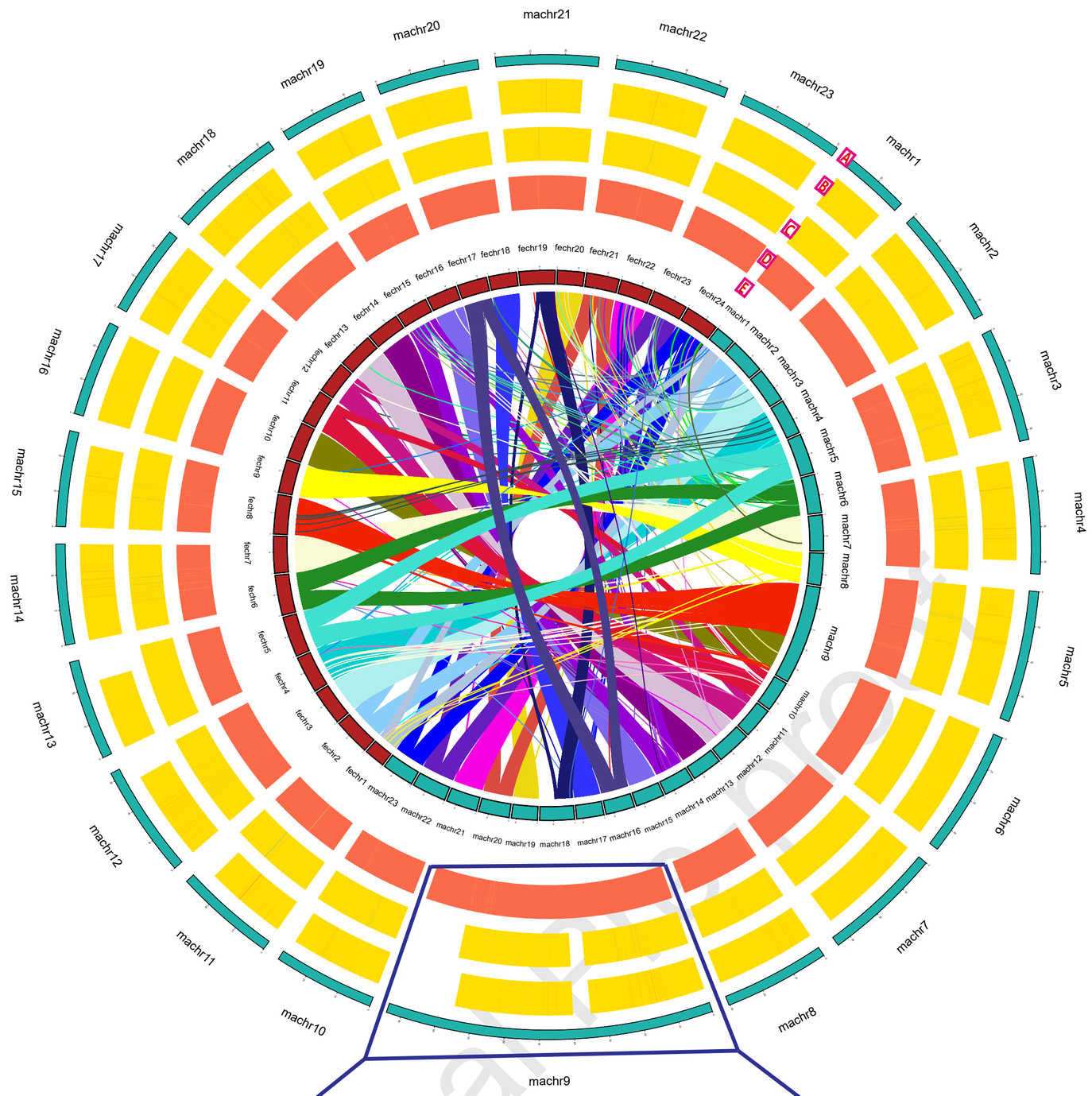
(c)



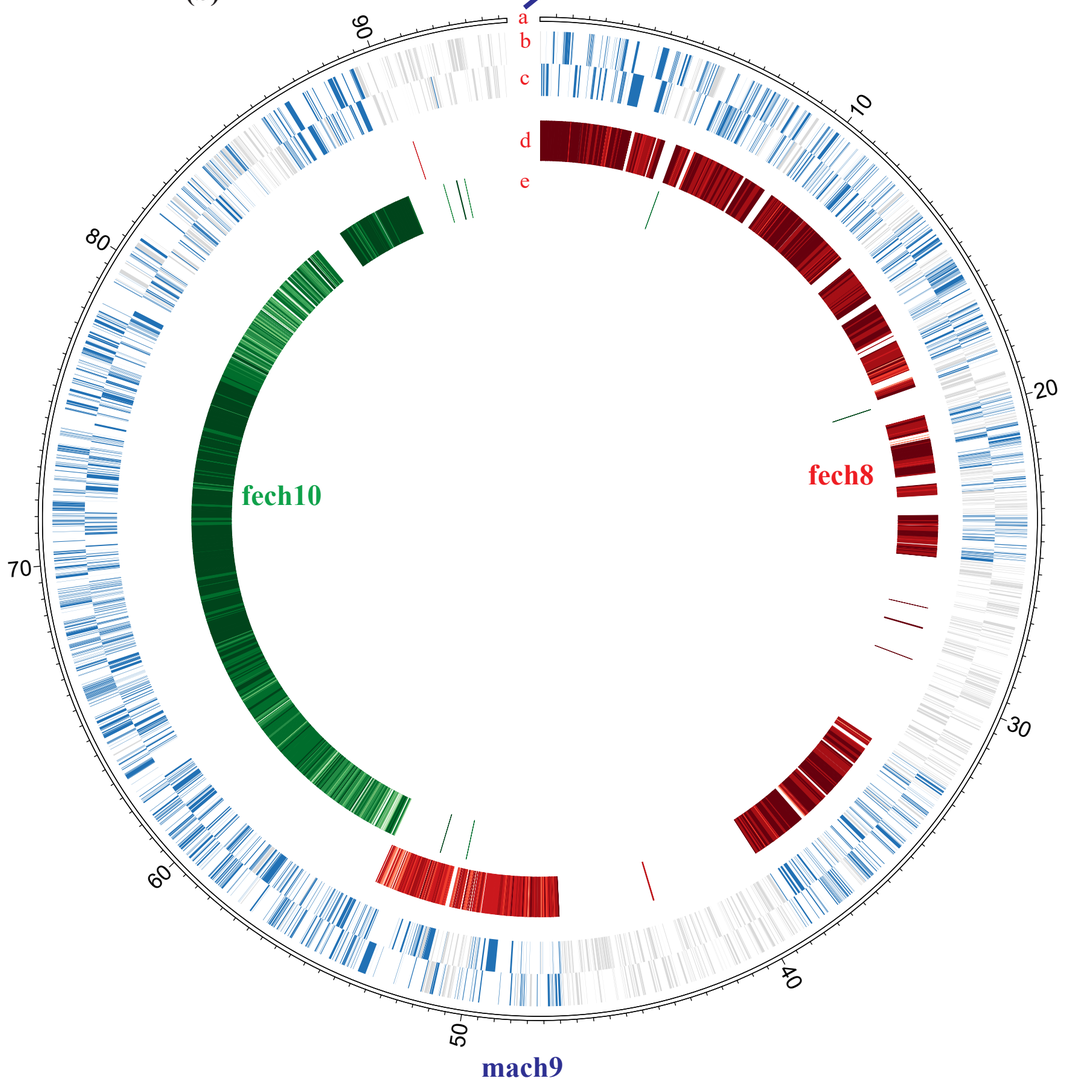
70	2	7	26	23

Journal Pre-proof

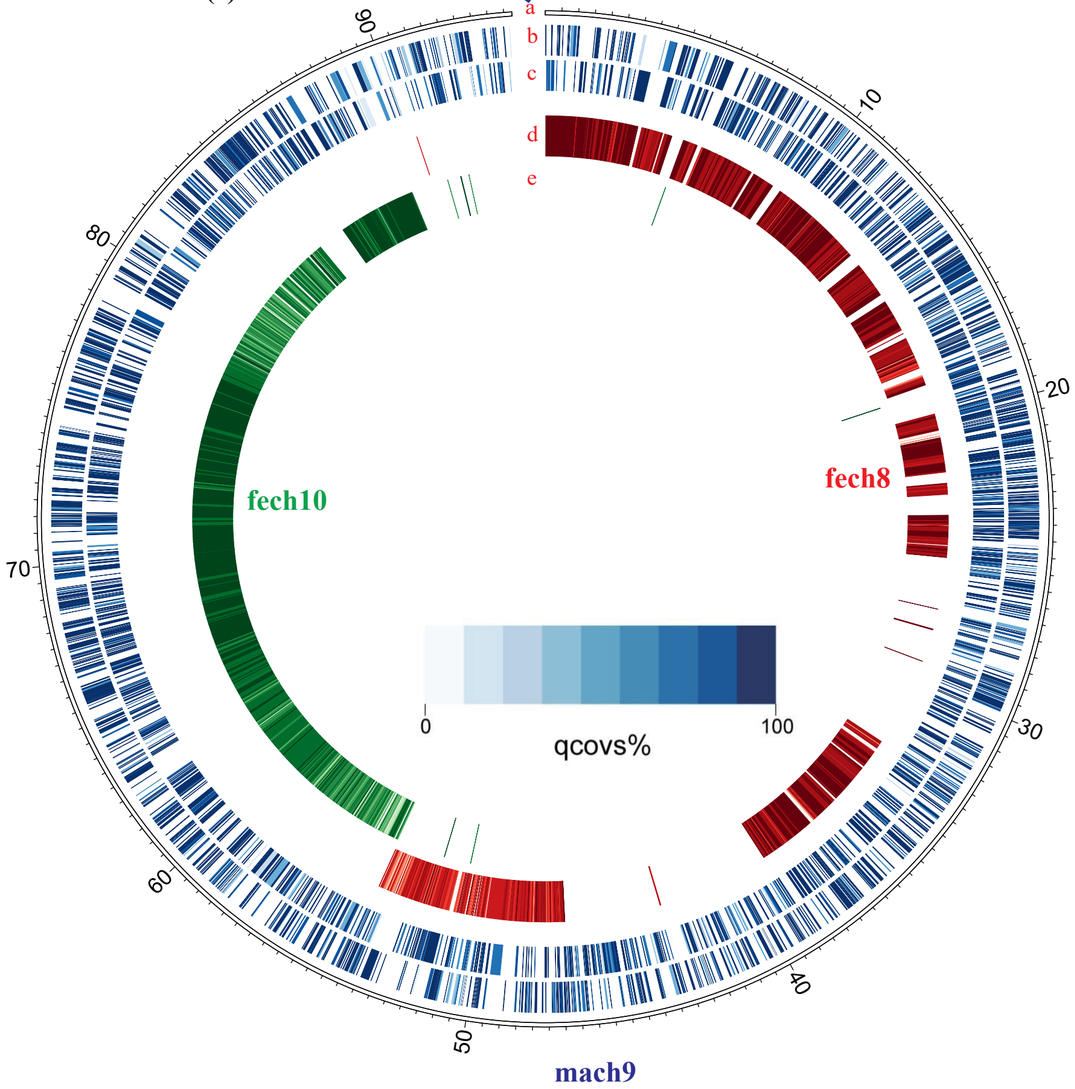
(a)

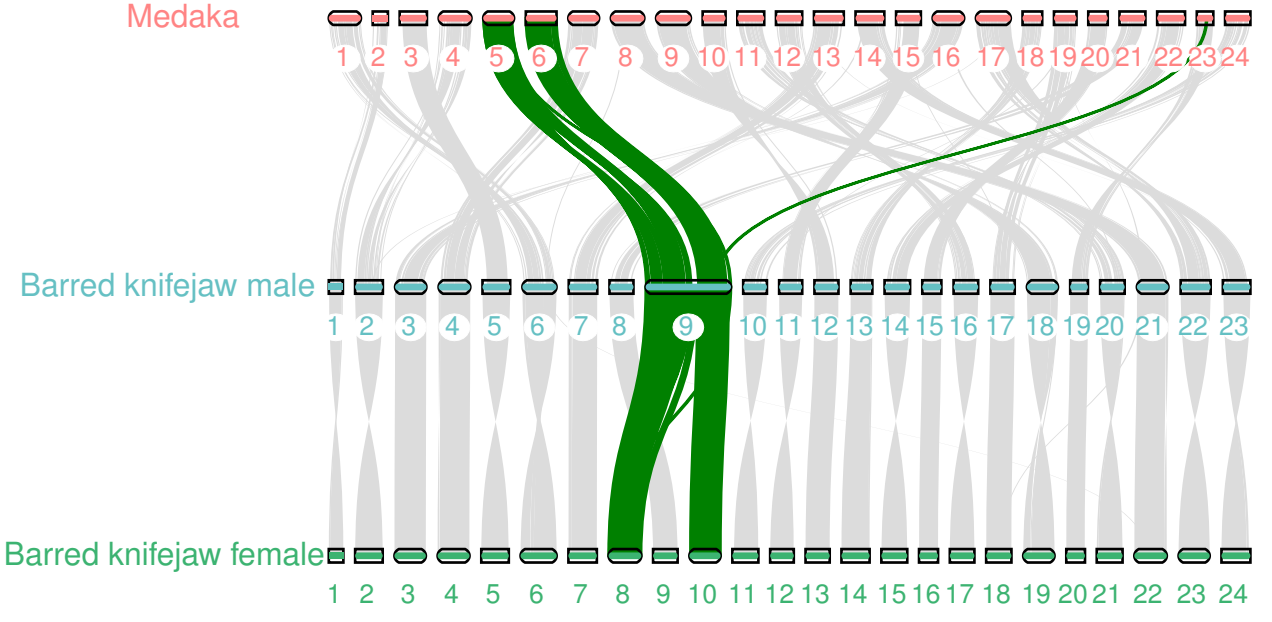


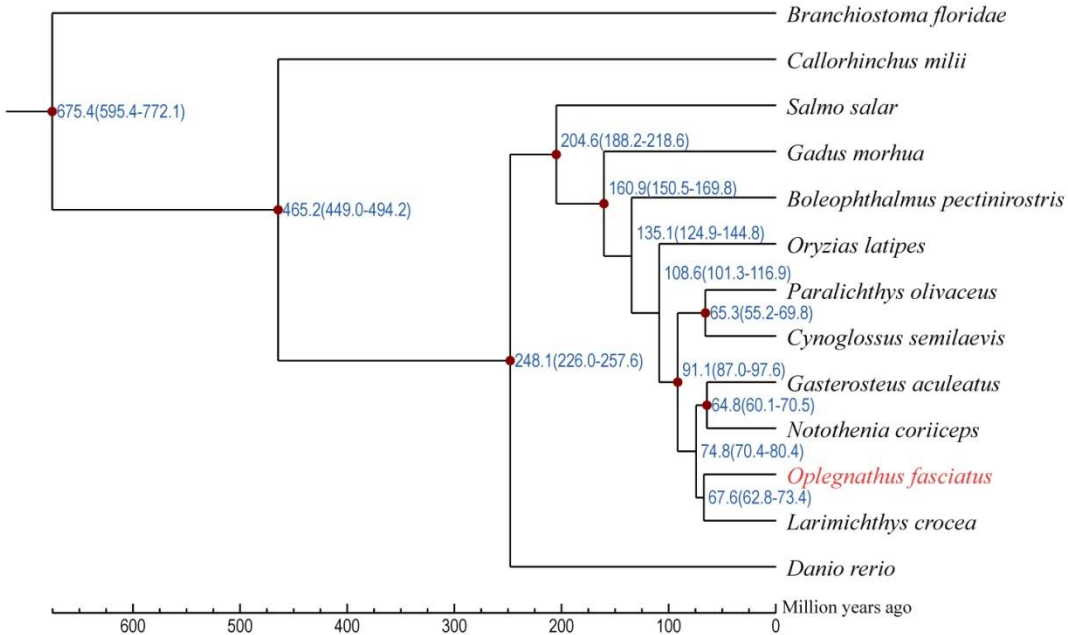
(b)



(c)




Medaka
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Barred knifejaw male
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Barred knifejaw female
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24


Branchiostoma floridae
Callorhinchus milii
Salmo salar
Gadus morhua
Boleophthalmus pectinirostris
Oryzias latipes
Paralichthys olivaceus
Cynoglossus semilaevis
Gasterosteus aculeatus
Notothenia coriiceps
Oplegnathus fasciatus
Larimichthys crocea
Danio rerio
675.4(595.4-772.1)
465.2(449.0-494.2)
248.1(226.0-257.6)
204.6(188.2-218.6)
160.9(150.5-169.8)
135.1(124.9-144.8)
108.6(101.3-116.9)
65.3(55.2-69.8)
91.1(87.0-97.6)
64.8(60.1-70.5)
74.8(70.4-80.4)
67.6(62.8-73.4)
Million years ago
600
500
400
300
200
100
0

Journal Pre-proof

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, Angel Herrera-Ulloa

HIGHLIGHTS

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

2、Identification of the origin of neo-Y chromosome to the X_1X_2Y system

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

Journal Pre-proof