

Research Article

The Assembly of the Y Chromosome Reveals Amplification of Genes Regulating Male Fertility in *Bactrocera dorsalis*

Shuangxiong Wu¹, Jiahong Wu², Quan Lei¹, Donghai He¹, Xinrui Jiang², Chao Ye¹, Dong Wei¹, Jinjun Wang¹, Luohao Xu², Hongbo Jiang²

1. Key Laboratory of Entomology and Pest Control Engineering, College of Plant Protection, Southwest University, China; 2. MOE Key Laboratory of Freshwater Fish Reproduction and Development, Key Laboratory of Aquatic Science of Chongqing, School of Life Sciences, Southwest University, China

Background: The oriental fruit fly *Bactrocera dorsalis* is an invasive pest causing considerable ecological and economic damage worldwide. The Y chromosome is an important target for the sterile insect technique (SIT) but its sequence and structure has been poorly explored.

Results: We assembled the genome of *B. dorsalis* at the chromosome level with a total size of 562.6 Mb. The assembly includes a ~7.6 Mb Y chromosome sequence, being the first reported Y chromosome in Tephritidae. The assembled Y chromosome is highly repetitive, and contains 61 genes, including 9 multi-copy genes. We surprisingly found that the M factor (*MoY*) in Tephritidae has multiple copies, verified by our droplet digital PCR (ddPCR) analysis. Besides, we identified 32 copies of *gyf-like on the Y chromosome* (*gyfY*) that were highly expressed in testis. RNAi targeting the *gyfY* resulted in depletion of live sperms, suggesting that the amplification of *gyfY* is essential for male fertility, which facilitated the understanding of high fecundity of this invasive insect.

Conclusion: We reported firstly the Y chromosome of *Bactrocera dorsalis*. Our results will also provide target genes for CRISPR/Cas9 based SIT, leading to the development of novel control strategies against tephritid flies.

Shuangxiong Wu and Jiahong Wu contributed equally to this work.

Corresponding authors: Luohao Xu, luohaox@swu.edu.cn; Hongbo Jiang, jhb8342@swu.edu.cn

1. Introduction

Tephritidae is a dipteran family comprising more than 5000 species^[1], dozens of which are invasive and extremely harmful to fruit and vegetable industries^{[2][3]}. The oriental fruit fly, *Bactrocera dorsalis* (Hendel) attacks more than 600 fruit and vegetable crops^[4], which causes great economic losses^[5]. Meanwhile, it is classified as the top member in the competitive hierarchy of tephritid flies^[6], since it could replace and drive various fruit fly species to extinction^[7]. The female flies lay eggs inside of the fruits and larvae feed on the flesh, which makes them hard to control^[8]. Current management of this fly depends on luring and killing by insecticides^[9], while very much hope has been given to the sterile insect technique (SIT)^{[10][11]}. The central principle of SIT is to produce a large number of sterile males^{[12][13][14]}. Male tephritid flies have a unique Y chromosome based on the reported karyotype as $2n = 12 (5A \times 2 + X + Y)$ ^[15]. According to previous studies, several genes on the Y chromosome are crucial for insect sex determination and male fertility^{[16][17][18][19]}, which makes them the most suitable molecular targets for the application of transgenic SIT^{[20][21][22]}.

Although the genome of female *B. dorsalis* has been reported by several groups^{[23][24][25]} including our own group^[26], the Y chromosome of *B. dorsalis* is not available yet. There are several reasons that the Y chromosomes of insects are extremely difficult to assemble. Most highly degenerated Y chromosome has experienced several rounds of recombination suppression^{[27][28][29]}, leading to repeat accumulation and heterochromatinization^{[30][31][32][33][34]}. Earlier attempts in analyzing of insect Y chromosomes have shown an extremely large proportion of satellite DNA and transposons elements^{[32][35][36][37][38]}. Sex chromosome differentiation also leads the Y chromosome to be hemizygous, which reduces its sequencing coverage^{[39][40]}. The lack of Y chromosome assembly prevents in-depth understanding the contribution of Y chromosome to the sex determination and male fertility, as well as the development of Y chromosome based transgenic SIT strategy to control this notorious fly.

In this study, we assembled the Y chromosome sequence of *B. dorsalis*, which is the first reported Y chromosome in Tephritidae. The assembled Y chromosome contains 61 genes. We surprisingly found that the M factor (*MoY*), widely conserved in Tephritidae^[41], is a multi-copy gene. In addition, *gyf-like on the Y chromosome* (*gyfY*) has experienced dramatic gene amplification in several *Bactrocera* species. By RNAi mediated gene knockdown, we found that the amplification of *gyfY* is essential for male fertility in *B. dorsalis*. This facilitated the understanding of high fecundity of this invasive insect. Our results will also

provide target genes for CRISPR/Cas9 based SIT, which may lead to the development of novel control strategies against the tephritid flies.

2. Materials and methods

2.1. Flies

The *Bactrocera dorsalis* population sequenced in the present study was maintained since a wild population sampled in Haikou, Hainan province, China, in 2008. Adults were reared under standard conditions (27 ± 1 °C, $70 \% \pm 5 \%$ relative humidity, and a 14-h photoperiod), on an artificial diet (containing honey, sugar, yeast powder, vitamin C, and water). A male fly from the inbred line at the 7th generation was collected for genomic DNA extraction. To avoid contamination by food and the intestinal microorganism, proboscis and the entire abdomens of the adults were removed and the surfaces were treated with 75 % ethanol.

2.2. Genome sequencing

Genomic DNA used for PacBio sequencing was extracted via the sodium dodecyl sulfate (SDS) method. We used the PacBio Sequel sequencing platform by Beijing Novogene Bioinformatics Technology Co., Ltd. (Novogene, Beijing, China) platform to achieve deep coverage of sequencing reads. Libraries were constructed by SMRTbell[®] Express Template Preparation Kit 2.0 (Pacific Biosciences, Delaware, USA) according to the manufacturer's instructions^[42]. For short-read sequencing, an Illumina library with an average insert size of 350 bp was constructed using NEB Next Ultra DNA Library Prep Kit (NEB, Beijing, China) with PE 150 bp mode.

Hi-C sequencing data was used to improve draft genome assemblies and assist the chromosome-level assembly^[43]. DNA from 25 male flies without the entire abdomens and proboscis was extracted and randomly broken into segments with 350 bp for Hi-C library preparation. Hi-C sequencing was conducted on the Illumina Novaseq 6000 platform with the PE 150 bp mode.

DNA used for Bisulfite-seq was extracted from eggs of *B. dorsalis* via Magnetic Universal Genomic DNA Kit (TIANGEN, Beijing, China) following the manufacturer's recommendations. After bisulfite treatment, sequencing Illumina library construction of the size selected samples was performed by Accel-NGS[®] Methyl-Seq DNA Library Kit (Swift, MI, USA) with PE 150 bp mode. Bismark (v0.22.1)^[44] was used for mapping, deduplication and methylation extraction.

2.3. Chromosome-level assembly and genome assessment

Firstly, to estimate genome size and heterozygosity, a *k*-mer analysis (jellyfish v2.2.10)^[45] was carried out by clean HiFi data and visualized using GenomeScope2 (v2.0)^[46]. Hifiasm (v0.17.7)^[47] with Hi-C integration mode was used for genome assembly. The heterozygosity of the assembly was removed by Purge_dups (v1.2.5) with default parameters^[48]. The Hi-C sequencing reads were then mapped to the contigs using the Juicer (v1.6) pipeline^[49]. Following this, the 3D-DNA (v201013) pipeline^[50] was used to generate the “.hic” file. Subsequently, we visualized the Hi-C heatmap using Juicebox Assembly Tools^[51] for manual operation. According to the Hi-C interaction heatmap in Juicebox. We demarcated chromosome boundaries and adjusted the contig order to acquire a high-quality chromosome-level genome assembly (Figure S1). Finally, Benchmarking Universal Single-Copy Orthologues (BUSCO v5.5.0)^[52] with the *insecta_odb10* lineage (n = 1367) was used to assess genome assembly completeness.

2.4. Identifying the sex chromosomes and autosomes with re-sequencing data

The chromosome quotient method^[53] was used to identify the sex chromosomes. In detail, the female and male re-sequencing reads were mapped to the assembled genome using BWA-mem (v0.7.17-r1188)^[54], and then samtools (v 1.10)^[55] and bedtools (v2.25.0)^[56] is used to calculate the sequencing coverage and depth in 100k windows. The ratio of female to male depth values is expected to cluster around 0, 1 or 2 for Y, autosomal or X chromosome respectively.

2.5. Repeat and genome annotation

Insect homology repetitive elements were obtained from RepeatMasker (<http://www.repeatmasker.org>) database (RepeatMaskerLib.h5) and Repbase (RepeatmaskerEdition-20181026)^[57]. Satellite Repeat Finder (SRF)^[58], RepeatModeler (v2.0.3) and EDTA (v2.2.0)^[59] were used for de novo prediction of repetitive elements. RepeatMasker (v4.1.5) was utilized to annotate the repeat elements. For tandem repeats, TideHunter (1.4.2)^[60] was used to predict the repeat units and StainedGlass (v0.1)^[61] was used for visualization.

We subjected Hisat2 (v2.2.1)^[62] to align RNA-seq data from 11 tissues (Table S3) and Trinity (v2.8.5)^[63] to assemble transcripts. Maker (v3.01.03)^[64] was used to predicted gene models based on assembled transcripts and protein homology based on the *Zeugodacus cucurbitae*, *Drosophila melanogaster*^[65] and

female *B. dorsalis*^{[24][26]} generated by liftoff (v1.6.3)^[66]. Finally, we used PASA (v2.5.2)^[67] pipeline to polish gene models using Trinity assembly.

For genes on the Y chromosome, we manually analyzed and removed some fake genes meeting the following four criteria that 1, the cds length of the putative gene is shorter than 250 bp; 2, the FPKM value is almost zero; 3, the number of exons not more than two; 4, lacking orthologous gene with the proximal species.

2.6. Determination of *gyfY* and *MoY* copy number using ddPCR analysis

Genomic DNA was extracted from the individual fly using TIANamp Genomic DNA kit (TIANGEN, Beijing, China) and was digested by the restriction enzymes MboI. Copy number of Y-link genes *gyfY* and *MoY* were assessed via ddPCR using a QX200 Droplet Digital™ PCR System (Bio-Rad, CA, USA) in accordance with the manufacturer's protocol of ddPCR Supermix for Probes (No dUTP) (Bio-Rad, CA, USA). Reference gene *IR25a*^[68] was fluorescently labelled with VIC, while target genes were labelled with FAM (Table S7). In brief, each 22 μ L reaction contained 11 μ L of 2 \times ddPCR Supermix, 0.99 μ L each of 10 μ M primers, 0.55 μ L each of 10 μ M probes and 0.55 ng genomic DNA. About 40 μ L of droplet reaction was generated in the QX200 Droplet Generator with the assistance of droplet generation oil. The PCR were performed following cycling conditions: 1 \times (95 $^{\circ}$ C for 10 min), 40 \times (94 $^{\circ}$ C for 30 s, 57 $^{\circ}$ C for 1 min), 1 \times (4 $^{\circ}$ C for 30 min). Following amplification, PCR plate 96 (Eppendorf, Hamburg, Germany) was transferred to the QX200 Droplet Reader and analyzed using QuantaSoft to acquire the absolute quantification (Table S8, Figure S5).

2.7. RNAi of *gyfY* and male sperm viability

A long dsRNA (1056 bp) covering all the copies of the *gyfY2* was designed to knockdown the expressions of all *gyfY2* copies, while a specific dsRNA was designed targeting the *gyf* on the X chromosome (*gyfX*) (The primers are listed in Table S7). The dsRNAs were synthesized with the TranscriptAid T7 High Yield Transcription Kit (Thermo Scientific, DE, USA). Two micrograms of dsRNA (dsGFP was used as a negative control) was injected to the immature male twice (in 1-day-old and 5-day-old, respectively) to obtain better silence efficiency with a standard protocol^[69]. Then, the mature males were dissected and the testes were collected for the viability assay in 9-day-old.

The spermatozoa were diluted in a buffer containing bovine serum albumin (10 mM HEPES, 2 % BSA, 150 mM NaCl, pH 7.4), and three pairs of testes were dissected in 100 mL buffer. According to the

manufacturer's instructions, the LIVE/DEAD™ Sperm Viability Kit (Thermo Scientific, MA, USA) was used to dye sperms. In detail, two luminous dyes, SYBR-14 and PI, were applied to distinguish live and dead sperms, respectively^[70]. The number of sperms was counted by photographing under the stereomicroscope (Leica, Wetzlar, Germany). Following the process above, the specific *dsgyfX* was used for RNA interference *gyfX*.

3. Results

3.1. A reference genome assembly of a male *Bactrocera dorsalis*

About 19.3 GB PacBio HiFi data was obtained from an individual male fly. With the assistance of about 45.4 GB Hi-C data (Table S2), we generated a raw assembly comprising 631 MB. Purge_dups (v1.2.5)^[48] was used to remove the 67.9 Mb redundant contigs. Compared with the *k*-mer based genome size estimation (495 Mb) (Figure 1a), we uncovered 67.6 Mb more sequences, in part due to successful assembly of some highly repetitive satellite DNA. The final BUSCO completeness score is 98.3 %, with 97.4 % single, 0.9 % duplicated, 0.3 % fragmented, and 1.4 % missing (Figure S2), suggesting the genome is highly completed.

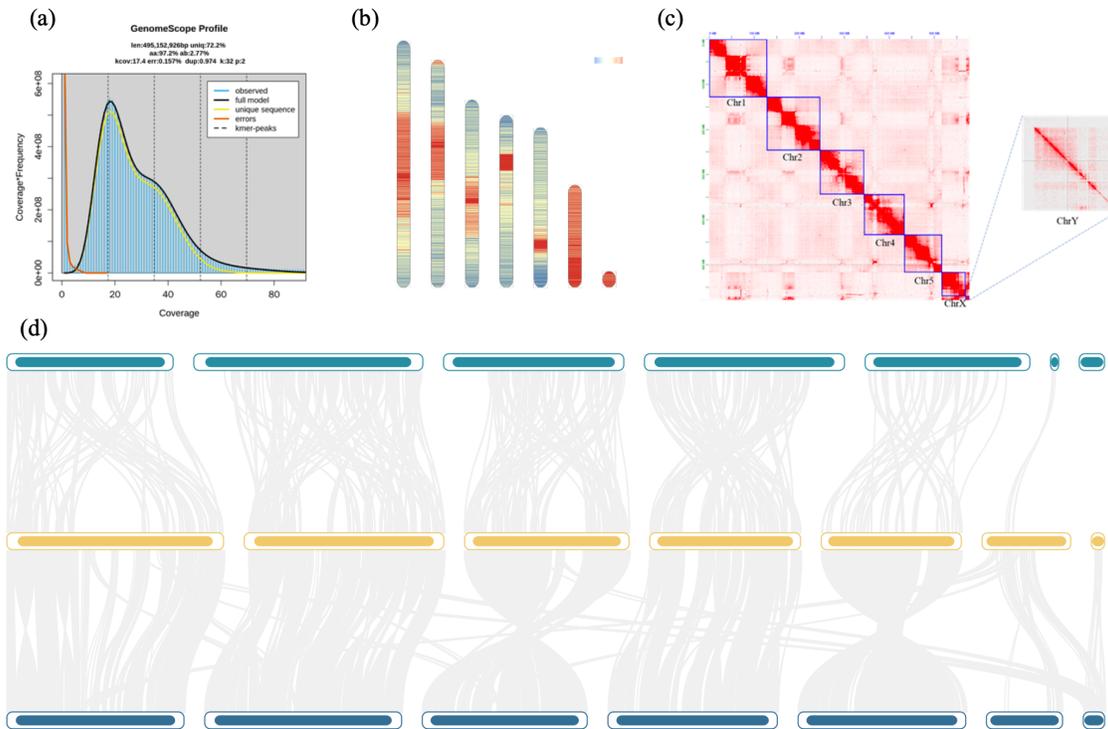


Figure 1. Chromosome-level genome assembly features of *Bactrocera dorsalis*. (a) Estimation of the genome size based on the distribution of k -mers ($k = 32$) of HiFi reads. (b) Chromosome-level genome repeat sequence characteristics. (c) Hi-C heatmap for the whole genome and zoom-in view for Y chromosomes visualised in Juicebox. (d) Whole-genome synteny between *Z. cucurbitae* (Zcuc), *D. melanogaster* (Dmel) and *B. dorsalis* (Bdor).

Chromosome-level genome assembly features of *Bactrocera dorsalis*. (a) Estimation of the genome size based on the distribution of k -mers ($k = 32$) of HiFi reads. (b) Chromosome-level genome repeat sequence characteristics. (c) Hi-C heatmap for the whole genome and zoom-in view for Y chromosomes visualised in Juicebox. (d) Whole-genome synteny between *Z. cucurbitae* (Zcuc), *D. melanogaster* (Dmel) and *B. dorsalis* (Bdor).

We further anchored the contigs to chromosome level with the assistance of Hi-C data, generating the final assembly comprising 562.6 MB, with a scaffold N50 of 93.2 MB and contig N50 of 2.6 MB (Table S4). The Hi-C heatmap revealed 7 chromosome models (Figure 1c), including 5 autosomes, X chromosome and Y chromosome, accounting for 98.1 % of the assembled sequence. Compared with other published Tephritidae genomes (Table S1), the genome of *B. dorsalis* is the most completed and the best assembled with the highest scaffold N50.

Our high-quality assembly contains large amounts of highly repetitive regions, including pericentromeric regions, putative centromeric repeats and the repeat-rich sex chromosomes (X and Y chromosomes). Overall, about 310 Mb of the assembled 563 Mb *B. dorsalis* genome are repetitive (Table S6). It is noted that SINEs (short interspersed elements) are almost absent from of *B. dorsalis*, despite the overall high repeat content. After masking the repeat sequences, a total of 16,773 protein-coding genes were annotated.

3.2. Centromere

Our Hi-C interaction map shows strong inter-chromosomal interactions among centromeres and pericentromeric regions (Figure 1c), similar to that in *D. melanogaster*^[71]. The centromeres and pericentric regions also exhibit the highest repeat content and methylation and lowest gene density across the chromosomes (Figure 1b, S4). The pericentric regions appear to be larger on larger chromosomes. For instance, in Chr1, the pericentric regions occupies almost one-third of the entire chromosome (Figure 1c). Our syntenic analysis shows that those pericentric regions are not present in *D. melanogaster*, but are partially present in *Z. cucurbitae* (Figure 1d), suggesting graduate expansion of the pericentric repeats. Satellite DNA is the major repeat category in those regions, and is dominated by five different satellite DNA families. Those five families exclusively distribute in the centromeric regions, and are present in all chromosomes (Figure S3).

3.3. Characterizations of the Y chromosome

We carried out genome resequencing of male and female flies, producing clean data of about 128.2 Gb and 56.6 Gb, respectively (Table S3). Using these data, we identified X and Y chromosomes in *B. dorsalis* with the chromosome quotient method^[53]. The second smallest chromosome (Chr6) has a nearly two-fold greater sequencing depth compared to females, consistent with the expected for the X chromosome. The smallest chromosome (Chr7) has a male-biased coverage and depth ratio and was consistent with the expectation for the Y chromosome (Figure 2a, d). The other 5 chromosomes exhibit no significant differences between females and males, with female/male ratios close to 1 (Figure 2a). According to the repeat annotation, repeat content of the total sequences is various among autosomes, X and Y chromosome, which is 47.86 %, 80.62 % and 74.47 %, respectively (Table S5).

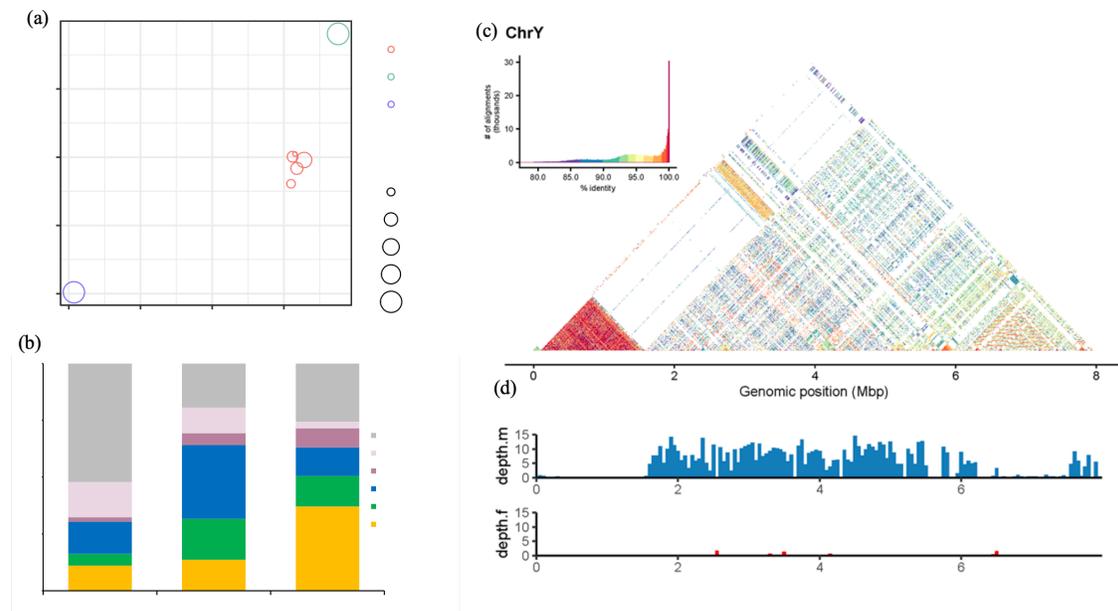


Figure 2. Identification and repeat sequence characteristics of Y chromosome. (a) Female to male depth ratio and coverage ratio for Y chromosome are both close to zero, for X chromosome are close to two and one, respectively. (b) Color-coded bar plot illustrating the proportions of TE compositions in the autosomal, X, and Y chromosomal sequences. (c) Identification of satellite DNA on the Y chromosome by StainedGlass. (d) The bar charts show male (depth.m) and female (depth.f) re-sequencing coverage and depth for Y chromosome.

Identification and repeat sequence characteristics of Y chromosome. (a) Female to male depth ratio and coverage ratio for Y chromosome are both close to zero, for X chromosome are close to two and one, respectively. (b) Color-coded bar plot illustrating the proportions of TE compositions in the autosomal, X, and Y chromosomal sequences. (c) Identification of satellite DNA on the Y chromosome by StainedGlass. (d) The bar charts show male (depth.m) and female (depth.f) re-sequencing coverage and depth for Y chromosome.

Most of these repeat sequences in Y chromosome were LINES (37.07 %), while LTRs and other DNA elements accounted for 13.46 % and 12.56 %, respectively. Compared with autosomes, the rates of LTR and LINES of Y chromosome are close to 3 (Figure 2b). And the satellite DNA of the Y chromosome is mainly concentrated at the beginning (Figure 2c).

Due to the repetitive nature of the Y chromosome, some bioinformatically predicted gene models may be incomplete or erroneous. To increase the quality of gene annotations, we manually curated gene models, leading to the identification of 61 Y-linked genes (Figure 3). Those Y-linked genes include 9 multi-copy

genes, among which 5 were tandemly amplified on the Y chromosome. Besides, we uncovered that the M factor (*MoY*) in Tephritidae has two copies due to tandem duplication, verified by our droplet digital PCR (ddPCP) analysis (Table S8).

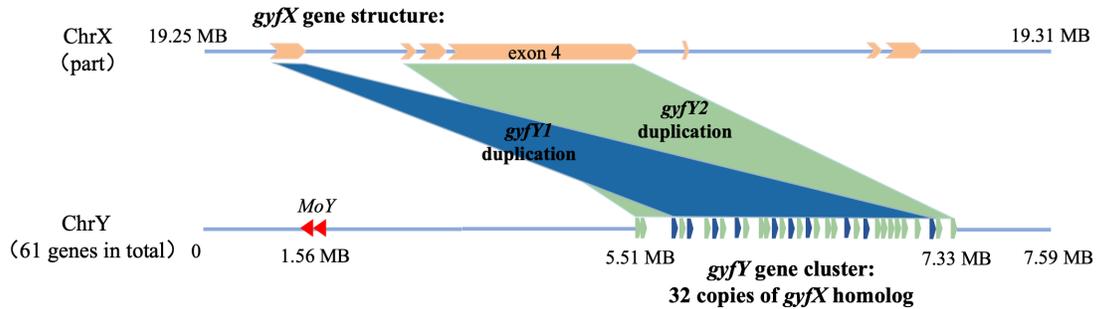


Figure 3. Schematic diagram of *gyfY* and *MoY* gene distribution on the Y chromosome. *MoY* (Red) has two adjacent copies. *gyfY1* (Blue) is homologous to exon 1 of *gyfX*, and *gyfY2* (Green) is homologous to exon 2, 3 and part of 4 of *gyfX*. The *gyfY* genes are all on the positive strand of the Y chromosome. For more details, please see Figure S6, the multiple sequence alignments of the putative *gyfY* and *gyfX* proteins.

Schematic diagram of *gyfY* and *MoY* gene distribution on the Y chromosome. *MoY* (Red) has two adjacent copies. *gyfY1* (Blue) is homologous to exon 1 of *gyfX*, and *gyfY2* (Green) is homologous to exon 2, 3 and part of 4 of *gyfX*. The *gyfY* genes are all on the positive strand of the Y chromosome. For more details, please see Figure S6, the multiple sequence alignments of the putative *gyfY* and *gyfX* proteins.

3.4. Male sperm viability after *gyfY* RNAi

gyf (glycine-tyrosine-phenylalanine) protein is ubiquitous in eukaryotes^{[72][73]}. We found that *gyfY* was amplified dramatically on the Y chromosome and was identified as a truncated paralog of *gyf* on the X chromosome (*gyfX*) in *B. dorsalis* (Figure 3). The copy number of *gyfY* is 27 (9 copies of *gyfY1* and 18 copies of *gyfY2*) confirmed by ddPCR, that was similar with the prediction by genome annotation (32 copies) (Figure 3, Table S8).

In order to explore the function of the *gyfY*, we designed a dsRNA with a length of 1056 bp, theoretically capable of broadly interfering with all *gyfY2* genes. Male flies were injected twice in 1-day-old and 5-day-old, and then 30 male flies were detected in 9-day-old. We found that the treated males had almost no live sperm compare to controls that had about 20 mature sperms per square millimeter (Figure 4b, d),

indicating that the amplification of *gyfY* is essential for male fertility. In addition, RNA interference with *gyfX* also resulted in an about 30% reduction in the number of live sperm (Figure 4c, d).

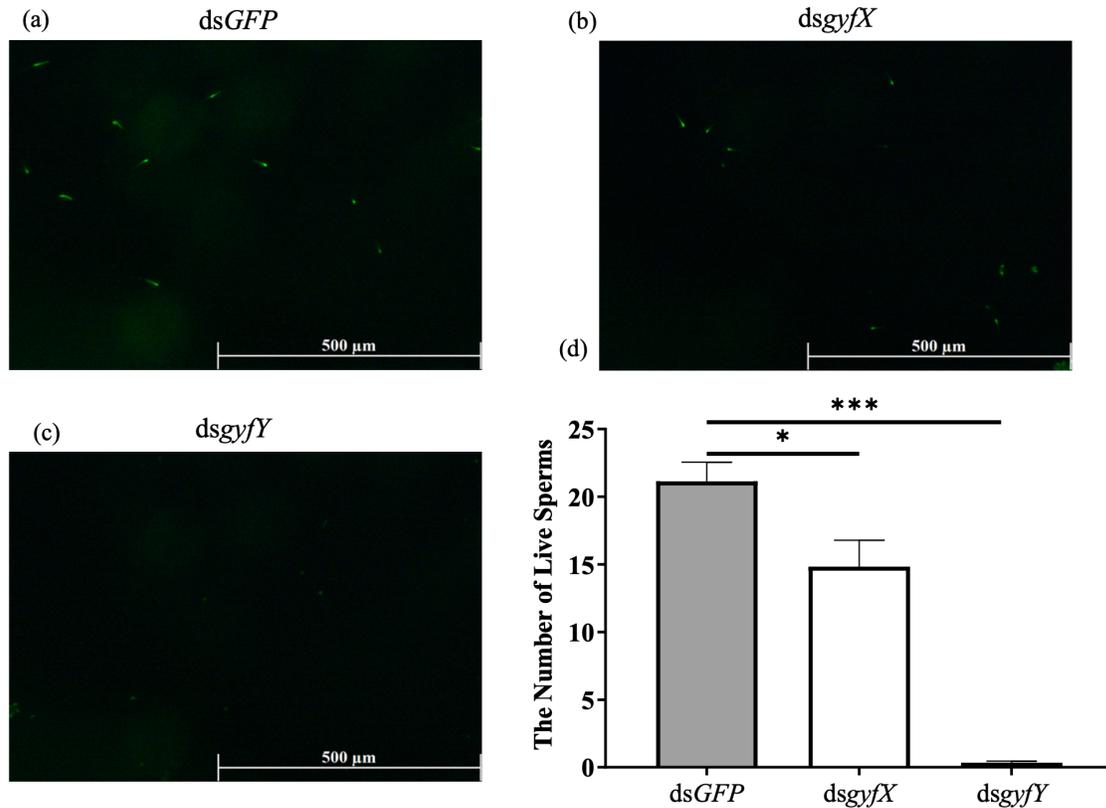


Figure 4. Effect of *gyfX* and *gyfY* RNA interference (RNAi) on sperm viability. (a-c) Live sperm after injection with dsGFP, dsGyfX and dsGyfY. Green indicates live sperm, and scale bars are indicated in the bottom left corner. (d) Number of live sperm after injected twice dsGFP, dsGyfX and dsGyfY. Data were presented as mean \pm SE ($n = 50$). Asterisks represent a significant difference determined by Student's *t*-test (* $P < 0.05$; *** $P < 0.001$)

Effect of *gyfX* and *gyfY* RNA interference (RNAi) on sperm viability. (a-c) Live sperm after injection with dsGFP, dsGyfX and dsGyfY. Green indicates live sperm, and scale bars are indicated in the bottom left corner. (d) Number of live sperm after injected twice dsGFP, dsGyfX and dsGyfY. Data were presented as mean \pm SE ($n = 50$). Asterisks represent a significant difference determined by Student's *t*-test (* $P < 0.05$; *** $P < 0.001$)

4. Discussion

Male tephritid flies own a unique Y chromosome, which is crucial for insect sex determination and male fertility^{[18][41][74]}. However, the Y chromosome is the most difficult part of the genome, because it is highly heterochromatic and repetitive^{[75][76]}, and has low sequence coverage^[40]. Using the PacBio HiFi, Hi-C and re-sequencing data, we generated the chromosome-level genome assembly of male *B. dorsalis*, including a characterized Y chromosome. This coordinated Y chromosome of *B. dorsalis* is the first assembled Y chromosome reported in tephritid flies. The completion of the Y chromosome assembly in tephritid flies is significant for understanding the intricacies of genome function and evolution.

The overall TE content of tephritid flies range from 21.87 % to 55.03 %, according to the available genomic information (Table S1, Table S9). This likely contributes to the genome size variation of the flies. Based on our current assembly, we found that the overall TE content of *B. dorsalis* (55.03 %) is much higher than *D. melanogaster* (21.53 %)^[65]. However, no SINE family has been detected in *B. dorsalis*, which also occurs in most insects^[77]. The repeat sequence content of X and Y chromosomes is 80.62 % and 74.47 % respectively, which is significantly higher than that of autosomes (Figure 2b, Table S5). While the expansion of repeats on the Y chromosome is consistent with prediction due to the loss of recombination^{[38][78]}, it remains a puzzle about the repeat expansion on the X chromosome. It is possible that the X chromosome has a reduced selection efficacy to purge repeats due to its smaller effective population size compared with autosomes^[79].

The high-quality genome assembly allowed us to identify the centromere and the repetitive pericentric regions of *B. dorsalis* which we found were much larger than those of *Drosophila*^{[76][80][81]}. The cause and consequence of pericentric expansion in *B. dorsalis* warrants future epigenetic investigations. Moreover, we identified the putative centromere tandem duplication sequences, which primarily consist of 5 sequences ranging from 165 bp to 180 bp (Figure S3), but to determine the core centromeric repeat, CENP-A ChIP-seq data is required. To the best of our knowledge, this is the first exploration of centromeric sequences in an agricultural insect. Our study provides a working protocol to identify centromeric repeats in insects, which will be useful to illuminate the origin and evolution of centromeric sequences.

Interestingly, we first confirmed the *M factor* (*MoY*) in Tephritidae are multi-copy gene by ddPCP (Table S8). This is in accordance with the prediction of 5 copies of *MoY* on the Y chromosome in *Z. cucurbitae*. However, there is only one copy of M factor in other Dipteran insects, such as mosquitos *Anopheles stephensi*, *Aedes aegypti* and *Anopheles gambiae* and housefly *Musca domestica* (Table S9)^{[41][82][83][84][85]}.

To masculinize *B. dorsalis* and the *Z. cucurbitae*, multiple M factors are required to work together or each M factor can work separately. Further study on the function of different copies of the *MoY* will provide a clue why multiple copy number of *MoY* is needed in the tephritid flies.

gyf (glycine-tyrosine-phenylalanine) proteins are characterized by the presence of a conserved domain (Pfam domain PF02213), which is ubiquitous in eukaryotes^{[72][73]}. In *D. melanogaster*, *gyf* is on the dot chromosome (homologous to the X chromosome of Tephritidae)^[86], which functions in regulation of autophagy and mRNA regulatory element binding as a translation repressor^{[87][88]}. In the current study, a copy of *gyf* (*gyfX*) was predicted on the X chromosome, while 32 copies of *gyf-like on the Y chromosome* (*gyfY*) were predicted on the Y chromosome of *B. dorsalis* (Figure 3). This dramatic amplification of *gyfY* may be caused by non-allelic homologous recombination to enhance the expression^[89]. The functional domains of *gyfX* were intact and well conserved comparing to other eukaryotes species. However, multicopy *gyfY* is identified as the truncated paralog of *gyfX*.

We identified two types of *gyfY* tandem repeats namely *gyfY1* (derived from the first exon of *gyfX*) and *gyfY2* (derived from the second, third and part of fourth exons of *gyfX*), instead of only *gyfY2* in a previous study^[18]. Furthermore, we confirmed 9 out of 10 predicted copies of *gyfY1*, while confirmed 18 out of the 22 predicted copies of *gyfY2*. This is in accordance with the previous report, while the few unconfirmed copies may be due to the complexity and repetition of the Y chromosome. Based on the analysis of the genome information available, similar gene amplification of *gyfY* occurs on the Y chromosome of *Bactrocera tryoni*^[90], but surprisingly not in *Bactrocera oleae*^[91] or *Z. cucurbitae*. This might be due to the rapid differential evolution of the Y chromosome^[38]. It is noticeable that *gyfY* has the domain of disorder protein but no PF02213 domain which was also verified by the PONDR (Predictor of Natural Disordered Regions; <http://pondr.com/>) (Figure S7)^[92].

A long *dsgyfY* (1056 bp) was designed from the region of the *gyfX* that is homologous to the *gyfY2*. The *dsgyfY* was decomposed into many siRNAs to knockdown all the *gyfY2*. With the 73% silence efficiency (Figure S8), the flies treated with *dsgyfY* almost had no live sperm (Figure 4), which indicated that the explosive amplification of *gyfY* enhanced the male fitness in *B. dorsalis*. Since the long dsRNA covers all the copies of *gyfY2*, our RNAi showed more significant reduction on the number of the live sperms in the male flies than a former study, that live sperms decreased by ~ 50% due to knocking down a *gyfY* with a 237 bp dsRNA^[18]. Moreover, similar gene amplification on Y chromosome occurs and is crucial for male fertility in *D. melanogaster*. Y-link *Mst77Y* undergoes amplification from the autosomal gene *Mst77F* to the

Y chromosome, collectively participating in the process of regulating sperm nuclear compression^[93]. In addition, when knocking down the *gyfX* resulted in an about 30% reduction of the live sperms (Figure 4). This suggested that *gyfX* also have an effect on sperm viability in *B. dorsalis* but less significant than *gyfY*. In conclusion, we successfully assembled the Y chromosome of *B. dorsalis*, which is the first reported Y chromosome in Tephritidae. The assembled Y chromosome contains 61 genes, including 9 multi-copy genes. We surprisingly found that the M factor (*MoY*), widely conserved in Tephritidae, is a multi-copy gene. In addition, amplification of *gyf* occurred and plays important roles in the male fertility in this species. Our data revealed male specific genes involved in male fertility which are the ideal molecular targets to develop novel transgenic control methods against this notorious fly, such as the genetic sexing strains targeting Y chromosome. Furthermore, CRISPR/Cas9 targeting Y-link multi-copy *gyfY* may lead to “Y-shredding”, resulting in female-only offspring.

Statements and Declarations

Author Contributions

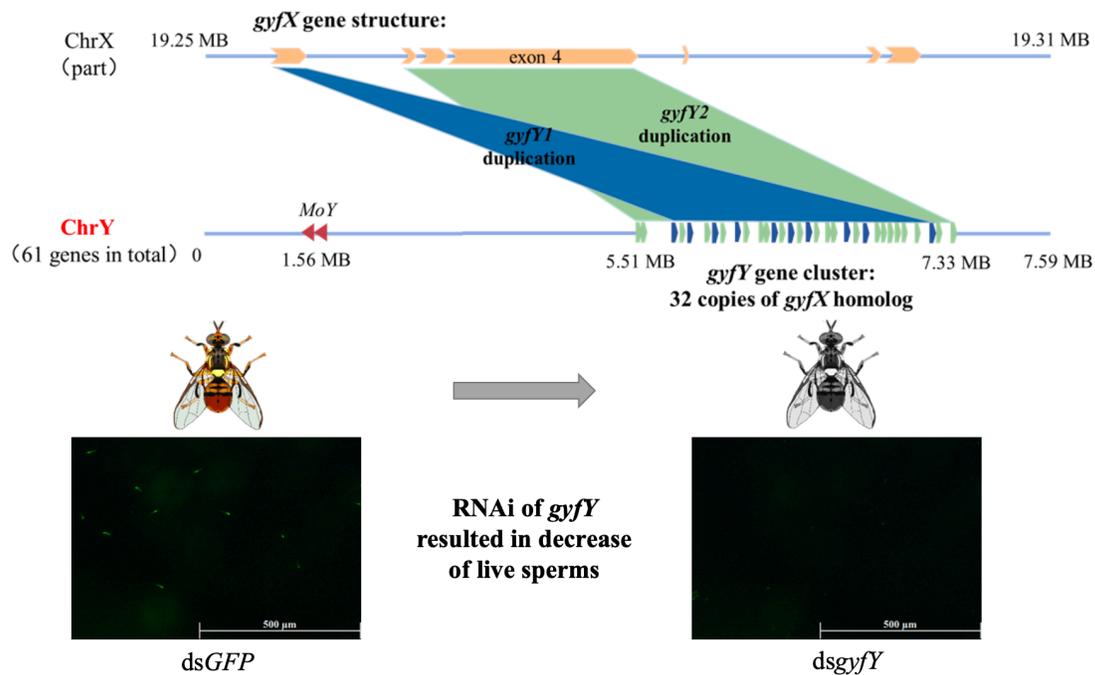
Hongbo Jiang and Luohao Xu designed research; Shuangxiong Wu, Jiahong Wu, Xinrui Jiang and Luohao Xu worked on analysis of the genome assembly and annotation; Shuangxiong Wu, Quan Lei, Donghai He, Chao Ye, and Dong Wei worked on the functional verification of *gyfY*; Shuangxiong Wu and Hongbo Jiang wrote the paper; Luohao Xu and Hongbo Jiang revised the manuscript.

Conflict of Interests

The authors declare no conflict of interests.

Data Availability

All raw data from this study has been deposited into the China National Center for Bioinformation (CNCB, <https://www.cncb.ac.cn/>) with the BioProject accession PRJCA026299.



Acknowledgements

We thank Dr. Yang Yang and Mr. Changhao Liang from Southwest University and Miss Xing Li from Yangzhou University for assistance on bioinformatics analysis. We also thank Prof. Marc F. Schetelig and Prof. Ying Yan from Justus-Liebig-University Giessen for their comments on an earlier version of this manuscript. This research was supported by funding from the National Key R&D Program of China (2022YFC2601000), National Natural Science Foundation of China (U21A20222, 32072491, 32370445), 111 Project (B18044) and the China Agriculture Research System of MOF and MARA.

References

1. [△]Bragard C, Dehnen-Schmutz K, Di Serio F, Gonthier P, Jacques MA, Jaques Miret JA, et al. Pest categorisation of non-EU Tephritidae. *EFSA Journal*. 18: 5931 (2020).
2. [△]Paini DR, Sheppard AW, Cook DC, De Barro PJ, Worner SP, Thomas MB. Global threat to agriculture from invasive species. *Proceedings of the National Academy of Sciences of the United States of America*. 113: 7575–7579 (2016).
3. [△]Berenbaum MR, Schutze MK, Virgilio M, Norrbom A, Clarke AR. Tephritid Integrative Taxonomy: Where We Are Now, with a Focus on the Resolution of Three Tropical Fruit Fly Species Complexes. In *Annual Review of Entomology*, ed. by Berenbaum MR, pp. 147–164 (2017).

4. [△]Liquido NJ, Mcquate GT, Suiter KA. Compendium of fruit fly host information (CoFFHI), edition 3.0. <https://coffhi.cphst.org/> [accessed October 24 2017].
5. [△]*Bactrocera dorsalis* (Oriental fruit fly). doi:10.1079/cabicompendium.17685 [accessed 25 September 2022].
6. [△]Zhao Z, Carey JR, Li Z. The Global Epidemic of *Bactrocera* Pests: Mixed-Species Invasions and Risk Assessment. *Annual Review of Entomology*. 69: 219–237 (2024).
7. [△]Duyck PF, David P, Quilici S. A review of relationships between interspecific competition and invasions in fruit flies (Diptera: Tephritidae). *Ecological Entomology*. 29: 511–520 (2004).
8. [△]Papadopoulos NT, De Meyer M, Terblanche JS, Kriticos DJ. Fruit Flies: Challenges and Opportunities to Stem the Tide of Global Invasions. *Annual Review of Entomology*. 69: 355–373 (2024).
9. [△]Xu L, Jiang H-B, Yu J-L, Pan D, Tao Y, Lei Q, et al. Two odorant receptors regulate 1-octen-3-ol induced oviposition behavior in the oriental fruit fly. *Communications Biology*. 6: 176 (2023).
10. [△]Koyama J, Kakinohana H, Miyatake T. Eradication of the melon fly, *Bactrocera cucurbitae*, in Japan: Importance of behavior, ecology, genetics, and evolution. *Annual Review of Entomology*. 49: 331–349 (2004).
11. [△]Hendrichs J, Robinson AS, Cayol JP, Enkerlin W. Medfly area wide sterile insect technique programmes for prevention, suppression or eradication: The importance of mating behavior studies. *Florida Entomologist*. 85: 1–13 (2002).
12. [△]Knipling EF. Possibilities of insect control or eradication through the use of sexually sterile males. *Journal of Economic Entomology*. 48: 459–462 (1955).
13. [△]Bushland RC, Lindquist AW, Knipling EF. Eradication of screw-worms through release of sterilized males. *Science*. 122: 287–288 (1955).
14. [△]Vanderplank FL. Hybridization between *Glossina* species and suggested new method for control of certain species of Tsetse. *Nature*. 154: 607–608 (1944).
15. [△]Carraretto D, Aketarawong N, Di Cosimo A, Manni M, Scolari F, Valerio F, et al. Transcribed sex-specific markers on the Y chromosome of the oriental fruit fly, *Bactrocera dorsalis*. *BMC Genetics*. 21: 125 (2020).
16. [△]Hafezi Y, Sruba SR, Tarrash SR, Wolfner ME, Clark AG. Dissecting Fertility Functions of *Drosophila* Y Chromosome Genes with CRISPR. *Genetics*. 214: 977–990 (2020).
17. [△]Krsticevic FJ, Santos HL, Januário S, Schrago CG, Carvalho AB. Functional Copies of the Mst77F Gene on the Y chromosome of *Drosophila melanogaster*. *Genetics*. 184: 295–307 (2010).
18. [△][♂][♀][♂]Liu P, Yu S, Zheng W, Zhang Q, Qiao J, Li Z, et al. Identification and functional verification of Y-chromosome-specific gene typo-gyf in *Bactrocera dorsalis*. *Insect Science* (2024).

19. [△]Hafezi Y, Omurzakov A, Carlisle JA, Caldas IV, Wolfner MF, Clark AG. The *Drosophila melanogaster* Y-linked gene, *WDY*, is required for sperm to swim in the female reproductive tract. *Communications Biology*. 7: 90 (2024).
20. [△]Carrami EM, Eckermann KN, Ahmed HMM, Sánchez C HM, Dippel S, Marshall JM, et al. Consequences of resistance evolution in a Cas9-based sex conversion-suppression gene drive for insect pest management. *Proceedings of the National Academy of Sciences*. 115: 6189–6194 (2018).
21. [△]Gamez S, Chaverra-Rodriguez D, Buchman A, Kandul NP, Mendez-Sanchez SC, Bennett JB, et al. Exploiting a Y chromosome-linked Cas9 for sex selection and gene drive. *Nature Communications*. 12: 7202 (2021).
22. [△]Champer J, Geci R, Willis K, Burt A. Gene drive designs for efficient and localisable population suppression using Y-linked editors. *PLOS Genetics*. 18 (2022).
23. [△]Jiang F, Liang L, Wang J, Zhu S. Chromosome-level genome assembly of *Bactrocera dorsalis* reveals its adaptation and invasion mechanisms. *Communications Biology*. 5: 25 (2022).
24. ^{a, b}Wang YH, Fang GQ, Xu PH, Gao BL, Liu XJ, Qi XW, et al. Behavioral and genomic divergence between a generalist and a specialist fly. *Cell Reports*. 41: 111654 (2022).
25. [△]Zhang Y, Liu S, De Meyer M, Liao Z, Zhao Y, Virgilio M, et al. Genomes of the cosmopolitan fruit pest *Bactrocera dorsalis* (Diptera: Tephritidae) reveal its global invasion history and thermal adaptation. *Journal of advanced research*. 53: 61–74 (2022).
26. ^{a, b}Yang Y, Jiang H-B, Liang C-H, Ma Y-P, Dou W, Wang J-J. Chromosome-level genome assembly reveals potential epigenetic mechanisms of the thermal tolerance in the oriental fruit fly, *Bactrocera dorsalis*. *International journal of biological macromolecules*. 225: 430–441 (2022).
27. [△]Charlesworth B. Model for evolution of Y-chromosomes and dosage compensation. *Proceedings of the National Academy of Sciences of the United States of America*. 75: 5618–5622 (1978).
28. [△]Bachtrog D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nature Reviews Genetics*. 14: 113–124 (2013).
29. [△]Cortez D, Marin R, Toledo-Flores D, Froidevaux L, Liechti A, Waters PD, et al. Origins and functional evolution of Y chromosomes across mammals. *Nature*. 508: 488–493 (2014).
30. [△]Xu L, Auer G, Peona V, Suh A, Deng Y, Feng S, et al. Dynamic evolutionary history and gene content of sex chromosomes across diverse songbirds. *Nature Ecology & Evolution*. 3: 834–844 (2019).
31. [△]Rhie A, Nurk S, Cechova M, Hoyt SJ, Taylor DJ, Altemose N, et al. The complete sequence of a human Y chromosome. *Nature*. 621: 344–354 (2023).

32. ^{a, b}Chakraborty M, Ramaiah A, Adolfi A, Halas P, Kaduskar B, Ngo LT, et al. Hidden genomic features of an invasive malaria vector, *Anopheles stephensi*, revealed by a chromosome-level genome assembly. *BMC Biology*. 19: 28 (2021).
33. ^ΔMahajan S, Bachtrog D. Convergent evolution of Y chromosome gene content in flies. *Nature Communications*. 8: 785 (2017).
34. ^ΔChang C-H, Larracuente AM. Heterochromatin-Enriched Assemblies Reveal the Sequence and Organization of the *Drosophila melanogaster* Y Chromosome. *Genetics*. 211: 333–348 (2019).
35. ^ΔMérel V, Gibert P, Buch I, Rodriguez Rada V, Estoup A, Gautier M, et al. The Worldwide Invasion of *Drosophila suzukii* Is Accompanied by a Large Increase of Transposable Element Load and a Small Number of Putatively Adaptive Insertions. *Molecular Biology and Evolution*. 38: 4252–4267 (2021).
36. ^ΔHall AB, Papathanos PA, Sharma A, Cheng CD, Akbari OS, Assour L, et al. Radical remodeling of the Y chromosome in a recent radiation of malaria mosquitoes. *Proceedings of the National Academy of Sciences of the United States of America*. 113: E2114–E2123 (2016).
37. ^ΔMeisel RP, Gonzales CA, Luu H. The house fly Y Chromosome is young and minimally differentiated from its ancient X Chromosome partner. *Genome Research*. 27: 1417–1426 (2017).
38. ^{a, b, c}Chang C-H, Gregory LE, Gordon KE, Meiklejohn CD, Larracuente AM. Unique structure and positive selection promote the rapid divergence of *Drosophila* Y chromosomes. *eLife*. 11: e75795 (2022).
39. ^ΔBracewell RR, Bentz BJ, Sullivan BT, Good JM. Rapid neo-sex chromosome evolution and incipient speciation in a major forest pest. *Nature Communications*. 8: 1593 (2017).
40. ^{a, b}Xue L, Gao Y, Wu M, Tian T, Fan H, Huang Y, et al. Telomere-to-telomere assembly of a fish Y chromosome reveals the origin of a young sex chromosome pair. *Genome Biol*. 22: 203 (2021).
41. ^{a, b, c}Meccariello A, Salvemini M, Primo P, Hall B, Koskiniot P, Dalikova M, et al. Maleness-on-the-Y (MoY) orchestrates male sex determination in major agricultural fruit fly pests. *Science*. 365: 1457–1460 (2019).
42. ^ΔPendleton M, Sebra R, Pang AWC, Ummat A, Franzen O, Rausch T, et al. Assembly and diploid architecture of an individual human genome via single-molecule technologies. *Nature Methods*. 12: 780–786 (2015).
43. ^ΔBelton J-M, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods*. 58: 268–276 (2012).
44. ^ΔKrueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*. 27: 1571–1572 (2011).
45. ^ΔMarçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*. 27: 764–770 (2011).

46. [△]Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications*. 11: 1432 (2020).
47. [△]Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods*. 18: 170–175 (2021).
48. ^a ^bGuan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*. 36: 2896–2898 (2020).
49. [△]Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, et al. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Systems*. 3: 95–98 (2016).
50. [△]Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*. 356: 92–95 (2017).
51. [△]Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, et al. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Systems*. 3: 99–101 (2016).
52. [△]Manni M, Berkeley MR, Seppely M, Simao FA, Zdobnov EM. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution*. 38: 4647–4654 (2021).
53. ^a ^bHall AB, Qi Y, Timoshevskiy V, Sharakhova MV, Sharakhov IV, Tu Z. Six novel Y chromosome genes in *Anopheles* mosquitoes discovered by independently sequencing males and females. *BMC Genomics*. 14: 273 (2013).
54. [△]Li H. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics*. 30: 2843–2851 (2014).
55. [△]Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25: 2078–2079 (2009).
56. [△]Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 26: 841–842 (2010).
57. [△]Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*. 6: 11 (2015).
58. [△]Zhang Y, Chu J, Cheng H, Li H. De novo reconstruction of satellite repeat units from sequence data. *Genome research*. 33: 1994–2001 (2023).
59. [△]Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol*. 20: 275 (2019).

60. [△]Gao Y, Liu B, Wang Y, Xing Y. TideHunter: efficient and sensitive tandem repeat detection from noisy long-reads using seed-and-chain. *Bioinformatics*. 35: I200–I207 (2019).
61. [△]Vollger MR, Kerpedjiev P, Phillippy AM, Eichler EE. "StainedGlass: interactive visualization of massive tandem repeat structures with identity heatmaps". *Bioinformatics*. 38: 2049–2051 (2022).
62. [△]Kim D, Langmead B, Salzberg SL. "HISAT: a fast spliced aligner with low memory requirements". *Nature Methods*. 12: 357–U121 (2015).
63. [△]Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. "Full-length transcriptome assembly from RNA-Seq data without a reference genome". *Nature Biotechnology*. 29: 644–U130 (2011).
64. [△]Holt C, Yandell M. "MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects". *BMC Bioinformatics*. 12 (2011).
65. [△][‡]Hoskins RA, Carlson JW, Wan KH, Park S, Mendez I, Galle SE, et al. "The Release 6 reference sequence of the *Drosophila melanogaster* genome". *Genome Research*. 25: 445–458 (2015).
66. [△]Shumate A, Salzberg SL. "Liftoff: accurate mapping of gene annotations". *Bioinformatics*. 37: 1639–1643 (2021).
67. [△]Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK, Hannick LI, et al. "Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies". *Nucleic Acids Research*. 31: 5654–5666 (2003).
68. [△]Wu S-X, Chen Y, Lei Q, Peng Y-Y, Jiang H-B. "Sublethal Dose of β -Cypermethrin Impairs the Olfaction of *Bacotropera dorsalis* by Suppressing the Expression of Chemosensory Genes". *Insects*. 13: 721 (2022).
69. [△]Meng L-W, Yuan G-R, Chen M-L, Zheng L-S, Dou W, Peng Y, et al. "Cuticular competing endogenous RNAs regulate insecticide penetration and resistance in a major agricultural pest". *BMC Biology*. 21: 187 (2023).
70. [△]Chalah T, Brillard JP. "Comparison of assessment of fowl sperm viability by eosin-nigrosin and dual fluorescence (SYBR-14/PI)". *Theriogenology*. 50: 487–493 (1998).
71. [△]Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, et al. "Three-Dimensional Folding and Functional Organization Principles of the *Drosophila* Genome". *Cell*. 148: 458–472 (2012).
72. [△][‡]Nishizawa K, Freund C, Li J, Wagner G, Reinherz EL. "Identification of a proline-binding motif regulating CD2-triggered T lymphocyte activation". *Proceedings of the National Academy of Sciences of the United States of America*. 95: 14897–14902 (1998).
73. [△][‡]Freund C, Dötsch V, Nishizawa K, Reinherz EL, Wagner G. "The GYF domain is a novel structural fold that is involved in lymphoid signaling through proline-rich sequences". *Nature Structural Biology*. 6: 656–660 (1999).

74. [△]Willhoeft U, Franz G. "Identification of the Sex-Determining Region of the *Ceratitidis capitata* Y chromosome by Deletion Mapping". *Genetics*. 144: 737–745 (1996).
75. [△]Brown EJ, Nguyen AH, Bachtrog D. "The Y chromosome may contribute to sex-specific ageing in *Drosophila*". *Nature Ecology & Evolution*. 4: 853–862 (2020).
76. ^a [△]Nguyen AH, Bachtrog D. "Toxic Y chromosome: Increased repeat expression and age-associated heterochromatin loss in male *Drosophila* with a young Y chromosome". *PLOS Genetics*. 17: e1009438 (2021).
77. [△]Gilbert C, Peccoud J, Cordaux R. "Transposable Elements and the Evolution of Insects". *Annual Review of Entomology*. 66: 355–372 (2021).
78. [△]Martí E, Larracuente AM. "Genetic conflict and the origin of multigene families: implications for sex chromosome evolution". *Proceedings of the Royal Society B: Biological Sciences*. 290: 20231823 (2023).
79. [△]Xu L, Wa Sin SY, Grayson P, Edwards SV, Sackton TB, Mank J. "Evolutionary Dynamics of Sex Chromosomes of Paleognathous Birds". *Genome Biology and Evolution*. 11: 2376–2390 (2019).
80. [△]Bracewell R, Chatla K, Nalley MJ, Bachtrog D. "Dynamic turnover of centromeres drives karyotype evolution in *Drosophila*". *eLife*. 8: e49002 (2019).
81. [△]Becker PB, Chang C-H, Chavan A, Palladino J, Wei X, Martins NMC, et al. "Islands of retroelements are major components of *Drosophila* centromeres". *PLOS Biology*. 17: e3000241 (2019).
82. [△]Sharma A, Heinze SD, Wu YL, Kohlbrenner T, Morilla I, Brunner C, et al. "Male sex in houseflies is determined by *Mdmd*, a paralog of the generic splice factor gene *CWC22*". *Science*. 356: 642–645 (2017).
83. [△]Krzywinska E, Dennison NJ, Lycett GJ, Krzywinski J. "A maleness gene in the malaria mosquito *Anopheles gambiae*". *Science*. 353: 67–69 (2016).
84. [△]Hall AB, Basu S, Jiang XF, Qi YM, Timoshevskiy VA, Biedler JK, et al. "A male-determining factor in the mosquito *Aedes aegypti*". *Science*. 348: 1268–1270 (2015).
85. [△]Criscione F, Qi Y, Tu Z. "GUY1 confers complete female lethality and is a strong candidate for a male-determining factor in *Anopheles stephensi*". *eLife*. 5: e19281 (2016).
86. [△]Vicoso B, Bachtrog D. "Reversal of an ancient sex chromosome to an autosome in *Drosophila*". *Nature*. 499: 332–335 (2013).
87. [△]Kim M, Semple I, Kim B, Kiers A, Nam S, Park H-W, et al. "*Drosophila* Gyf/GRB10 interacting GYF protein is an autophagy regulator that controls neuron and muscle homeostasis". *Autophagy*. 11: 1358–1372 (2015).
88. [△]Ruscica V, Bawankar P, Peter D, Helms S, Igreja C, Izaurralde E. "Direct role for the *Drosophila* GIGYF protein in 4EHP-mediated mRNA repression". *Nucleic Acids Res*. 47: 7035–7048 (2019).

89. [△]Bachtrog D, Mahajan S, Bracewell R. "Massive gene amplification on a recently formed *Drosophila* Y chromosome". *Nature Ecology & Evolution*. 3: 1587–1597 (2019).
90. [△]Choo A, Nguyen TNM, Ward CM, Chen IY, Sved J, Shearman D, et al. "Identification of Y-chromosome scaffolds of the Queensland fruit fly reveals a duplicated *gyf* gene paralogue common to many *Bactrocera* pest species". *Insect Molecular Biology*. 28: 873–886 (2019).
91. [△]Bayega A, Djambazian H, Tsoumani KT, Gregoriou ME, Sagri E, Drosopoulou E, et al. De novo assembly of the olive fruit fly (*Bactrocera oleae*) genome with linked-reads and long-read technologies minimizes gaps and provides exceptional Y chromosome assembly. *BMC Genomics*. 21: 259 (2020).
92. [△]Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN. PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochimica Et Biophysica Acta-Proteins and Proteomics*. 1804: 996–1010 (2010).
93. [△]Park JI, Bell GW, Yamashita YM. Derepression of Y-linked multicopy protamine-like genes interferes with sperm nuclear compaction in *D. melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America*. 120: e2220576120 (2023).

Supplementary data: available at <https://doi.org/10.32388/L6SNAM>

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.