Chromosome-level Assembly, Dosage Compensation and Sex-biased Gene Expression in the Small Brown Planthopper, *Laodelphax striatellus*

Qing-Ling Hu1,2, Yu-Xuan Ye1, Ji-Chong Zhuo2, Hai-Jian Huang2, Jun-Min Li2, and Chuan-Xi Zhang*,1,2

1Institute of Insect Science, Zhejiang University, Hangzhou 310058, China
2State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-Products, Key Laboratory of Biotechnology in Plant Protection of Ministry of Agriculture and Zhejiang Province, Institute of Plant Virology, Ningbo University, Ningbo 315211, China

*Corresponding author: E-mail: chxzhang@zju.edu.cn.

Accepted: 22 October 2022

**Abstract**

In insects, sex chromosome differentiation often results in unequal gene dosages between sexes. Dosage compensation mechanisms evolve to balance gene expression, but the degree and mechanism of regulation often vary by insect species. In hemipteran species, the small brown planthopper (SBPH), *Laodelphax striatellus*, is an injurious crop pest, with a sex chromosome type XX in females and XO in males. This species offers the opportunity to study dosage compensation and sex-biased gene expression. In this study, we generated a chromosome-level genome of SBPH using Oxford Nanopore Technologies and high-throughput chromatin conformation capture (Hi-C) technology. We also sequenced RNA-seq data from 16 tissue samples to annotate the genome and analyze gene dosage compensation. We finally obtained a 510.2 megabases (Mb) genome with 99.12% of the scaffolds anchored on 15 chromosomes (14 autosomes and 1 X chromosome) and annotated 16,160 protein-coding genes based on full-length cDNA sequencing data. Furthermore, we found complete dosage compensation in all *L. striatellus* somatic tissues, but lack of dosage compensation in gonad tissue testis. We also found that female-biased genes were significantly enriched on the X chromosome in all tissues, whereas male-biased genes in gonad tissues were enriched on autosomes. This study not only provides a high-quality genome assembly but also lays a foundation for a better understanding of the sexual regulatory network in hemipteran insects.

**Key words:** small brown planthopper, chromosome-level assembly, ONT, dosage compensation, sex-biased gene expression.

**Significance**

The previous chromosome-level genome of the *Laodelphax striatellus* had relatively low percentage (79.6%) of scaffolds anchored to chromosomes. Moreover, in Hemipteran insects, the dosage compensation level for X chromosomes in different tissues is not well known. We constructed a new genome with 99.13% of the scaffolds anchored on 15 chromosomes. And we found dosage compensation was complete in somatic tissues, while lacking in testis tissue of *L. striatellus*. Third-generation sequencing technology ONT was applied to assemble the draft genome, and Hi-C technology was further applied for chromosomes construction. The transcriptomes from somatic and gonad tissues were used to analyze the dosage compensation extent and sex-biased genes.

© The Author(s) 2022. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
Introduction

Sex chromosome differentiation is a common way to determine sexes in nature (e.g., two basic sex chromosome types: XX females and XY males or ZW females and ZZ males). The sex chromosomes evolve from a normal chromosome pair that carries sex-determining genes (Ohno 1967). The region containing these sex-determining genes will become recombination suppressed, which contributes to subsequent genetic degeneration of sex chromosome regions (e.g., massive gene loss or even complete loss of the Y or W chromosome (Bergero and Charlesworth 2009). This degeneration triggers the problem of unequal gene dosage on sex chromosomes, which could be lethal or deleterious. To mitigate the impact, the mechanism of dosage compensation evolves to balance gene expression on sex chromosomes between females and males and between sex chromosomes and autosomes (Disteche 2012; Graves 2016). Studies involving different lineages have demonstrated that the extent of dosage compensation and the precise molecular mechanisms vary among species. In therian mammals, one entire X chromosome is randomly inactivated in somatic cells of females by epigenetic silencing to balance the dosage difference (Lyon 1961; Gartler and Riggs 1983). In Drosophila, dosage compensation is regulated by upregulating the X chromosome in males (Larschan et al. 2011). In birds, reptiles and fish, dosage compensation is partial and gene specific (Ellegren et al. 2007; Vicoso et al. 2013; Shao et al. 2014).

Insects are the most speciose group in the animal kingdom. At present, limited studies regarding dosage compensation have been conducted in certain species of Diptera (Nozawa et al. 2014; Vensko and Stone 2015), Lepidoptera (Catalán et al. 2018; Ranz et al. 2021), Hemiptera (Pal and Vicoso 2015; Richard et al. 2017), Orthoptera (Chauhan et al. 2021), Strepsiptera (Mahajan and Bachtrog 2015), and Coleoptera (Prince et al. 2010), whereas it remains unclear in a tremendous number of other species. Based on existing results, the degree of dosage compensation and regulation mechanisms vary among insect species and even different tissues (Nozawa et al. 2014; Ranz et al. 2021), developmental stages (Rose et al. 2016), and time periods of segment added to X chromosomes (Mahajan and Bachtrog 2015; Gu et al. 2019). Hemiptera is an ideal clade to study the evolution of sex chromosomes. In this group, many species show degenerated Y chromosomes with XY or XO sex determination systems. The dosage compensation in pea aphid, Acyrthosiphon pisum (Richard et al. 2017), and three true bugs (Pal and Vicoso 2015) have been analyzed, all show the existence of dosage compensation, but these studies only analyze whole body samples, which could not represent the dosage compensation state among different tissues. In addition, X-chromosome masculinization has been found in pea aphid, whereas feminization has been found in other three true bugs. We still know little about dosage compensation in rice planthoppers. The small brown planthopper, Laodelphax striatellus (Fallén) (Delphacidae, Hemiptera), is an important agricultural pest that damages a variety of cereal crops, including rice, wheat, maize, oats, tall oat grass, and barley, as well as transmits several important rice viruses. Its distribution widely covers Asia, Europe and some areas of Oceania and Africa. From 2006 to 2015, rice planthoppers have caused an average of ~13.1 million tons of economic and yield losses each year in China (Liu et al. 2016).

In recent years, genome assembly combined with RNA-seq data has revealed the state of dosage compensation in many species. The previous draft genome version of SBPH has been assembled by high-depth of Illumina short reads and use low-depth of Pacific Biosciences (PacBio) data to fill the gaps. The scaffold-level genome has a contig N50 of 118 kilobases (kb) and a scaffold N50 of 1.08 Mb (Zhu et al. 2017). Later, this draft genome has been assembled to chromosomal level with Hi-C technology (Ma et al. 2021). However, nearly 20.4% of the genome regions have failed to locate on any chromosome, which may be due to the short length of contig N50 and the lack of restriction sites in many short contigs. The third-generation sequence representatives by PacBio and Oxford Nanopore Technologies (ONT) MinION generate longer reads, reaching tens of kilobases (Laver et al. 2015), greatly improving genome assembly quality. For example, application of ONT and Hi-C data has increased scaffold N50 of Illumina sequence assembly from 356.6 kb to 69.96 Mb for Nilaparvata lugens (Ye et al. 2021).

In this study, we assembled the chromosome-level genome of SBPH by combining ONT sequences and Hi-C technology to improve genome quality. By analyzing the transcriptome gene expression, we tested the level of dosage compensation and characterized how sex-biased genes were distributed across chromosomes in different tissues of SBPH.

Results

Chromosome-level Assembly of the L. striatellus Genome

We sequenced the male genome of L. striatellus using four 1D flow cells by ONT 1D. After quality control, a total of 106.89 gigabases (Gb; ~200X) of clean data was retained, including 16,381,943 subreads with a read N50 length of 21,754 bases and an average read length of 6,525 bases (supplementary table S1, Supplementary Material online). We also sequenced 34 Gb (~68X) of clean Illumina data of male L. striatellus to estimate genome size and polish the genome assembly. The genome length was estimated to be 498,082,955 bp by Genomescan v2.0 with a
heterozygosity of 2.9% when kmer was set to 19 (supplementary fig. S1, Supplementary Material online). The subreads were primarily assembled and further polished to obtain the raw contig-level assembly. The raw-contig-level genome was 762,184,311 bp in length, containing 1,731 contigs and the contig N50 was 909,801 bp (supplementary table S2, Supplementary Material online). After removing the haplotigs, a haploid contig-level assembly was obtained in length of 509,864,764 bp, including 603 contigs with a contig N50 of 1,613,830 bp (supplementary table S2, Supplementary Material online).

We applied Hi-C data (supplementary table S3, Supplementary Material online) to construct haploid contigs for chromosome-level assembly. The length of the final chromosome-level genome was 510,210,580 bp and the scaffold N50 was 34,959,101 bp (table 1). The genome assembly included 15 chromosome-level scaffolds (lengths ranged from 13,928,342 to 59,587,101 bases, accounting for 99.13% of the total genome), and additional 297 small scaffolds (4,460,860 bases in total, accounting for 0.87% of the genome; fig. 1A). The GC content of the male L. striatellus genome was 34.52% (fig. 2B), which was very close to the previous version’s 34.54% (Zhu et al. 2017). To assess the completeness of the genome assembly, benchmarking universal single-copy orthologs (BUSCOs) of Eukaryota (n = 255), Arthropoda (n = 1,013), Hemiptera (n = 2,510), and Insecta (n = 1,367) were used to align against the genome, and the completeness rate ranged from 92.2% to 95.9% (fig. 3A). These results demonstrated the high quality of our genome assembly. Furthermore, we identified a total of 207,334 simple sequence repeats (SSRs) from the whole genome (supplementary table S4, Supplementary Material online). We also characterized a total of 157,801,173 bp of repetitive elements, accounting for 30.93% of the whole genome (supplementary table S5, Supplementary Material online).

### Table 1

| Assembly and Annotation Statistics of Three Laodelphax striatellus Genome Version |
|--------------------------------------------|
| L. striatellus genome version               |
|                                          |
| Illumina + PacBio                        | Hi-C ONT + Illumina + Hi-C |
| Base pairs (Mb)                           | 541                        | 541.1                      | 510.20                    |
| % Ns                                      | 1.9                        | 2                          | 0.04                      |
| Contig numbers                            | 50,020                     | 50,020                     | 1,073                     |
| ContigN50 (kb)                            | 71                         | 71                         | 1,362                     |
| Scaffolds numbers                         | 20,450                     | 1,908                      | 312                       |
| Scaffold N50 (kb)                         | 1,185                      | 29,238                     | 34,959                    |
| % of scaffolds anchored in chromosomes    | —                          | 79.60%                     | 99.13%                    |
| Protein-coding genes                      | 21,254                     | 16,412                     | 16,160                    |
| Reference                                  | Zhu et al. (2017)          | Ma et al. (2021)           | This study                |

**Fig. 1.**—Chromosome genome assembly of Laodelphax striatellus. (A) Heatmap representing the frequency of Hi-C contacts along the L. striatellus genome assembly. Each chromosome was named in order of chromosome length from longest to shortest, as shown on the right and bottom axes. The heatmap was created by Juicebox software using Hi-C data. (B) Illumina-sequenced male (M) to female (F) read coverage ratio in a 100-kb fixed window of the L. striatellus genome. The lines represent the smooth fitted curve using the LOESS method.
**X Chromosome Identification**

The karyotype of *L. striatellus* consists of 15 haploid chromosomes (14A + 1X; Zhu et al. 2017), which is the same number as our Hi-C chromosome-level scaffolds (figs. 1A and 2A). To identify the X chromosome, ~34 Gb of males and 33 Gb of females Illumina paired-end DNA short sequences were mapped against the genome to estimate the coverage of the mapped reads. We then calculated male to female coverage ratio in a 100-kb sliding window across chromosomes. The method of sex determination in *L. striatellus* is that females have two X chromosomes (XX) and males have only one X chromosome (XO), so the sequence read coverage of the X chromosome in males should be half that of females. In our results, only one chromosome, the X chromosome, was observed with a male to female coverage of about 0.5, while the ratio of remaining autosomes fluctuated close to 1 (fig. 1B).

**Genome Annotation-based on ONT Long-read and Illumina Short-read cDNA Sequencing**

To obtain accurate gene structure annotation, we sequenced both Illumina short-read and ONT long-read cDNA data to perform genome annotation. Illumina data were obtained from six independent libraries of *L. striatellus*...
Dosage Compensation and Sex-biased Gene Expression in *L. striatellus*

After quality control, ∼97.64 Gb of clean data was retained, containing 650,918,608 sequences of 150 bp. Then, these short reads were assembled into 154,426 transcripts by Trinity. ONT data were generated from four independent libraries of *L. striatellus* at different developmental stages, including eggs, nymphs, adult males, and adult females. A total of 8,322,589,622 bases were obtained, consisting of 6,821,102 reads, with an average read length of 1,220 bp. The reads were further classified into 155,475 full-length consensus isoforms with a read length N50 of 3,003 bp (supplementary table S6, Supplementary Material online).

We finally annotated 16,160 protein-coding genes across the genome of *L. striatellus* (supplementary table S7, Supplementary Material online), and 99.34% of genes were anchored on 15 chromosome-level scaffolds. The gene set contained 93% of the conserved Hemiptera (*n* = 2,510) and Insecta (*n* = 1,367) BUSCO genes, respectively (fig. 3B). The genes were evenly distributed across the genome (fig. 2C). For functional annotation, we aligned 16,160 protein-coding genes against commonly used databases. In summary, 13,399 genes had a hit in the non-redundant protein sequence database (NR), 8,503 genes had a hit in the nucleotide sequence (NT) database, 8,964 genes in the UniProtKB/Swiss-Prot database, 6,907 genes in the Orthologous Groups of proteins (KOG) database, 55 ribosomal RNAs (rRNAs), 147 small nuclear RNAs (snRNAs), and 1,860 transfer RNAs (tRNAs; supplementary table S9, Supplementary Material online). Furthermore, 10,551 genes were annotated with Gene Ontology (GO) terms (supplementary fig. S2, Supplementary Material online), and 8,192 genes were mapped to Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (supplementary fig. S3, Supplementary Material online). A total of 14,599 genes were hit at least once in these databases, accounting for 90.3% of the 16,160 annotated genes (supplementary table S8, Supplementary Material online). Overall, the BUSCO completeness and the high annotation rate indicated that our gene annotation was accurate and reliable.

In addition to protein-coding RNAs, we also identified non-coding RNAs (ncRNAs), which are essential components of transcriptional and epigenetic regulation of gene expression. In total, 2,116 ncRNAs were identified across the genome, including 54 microRNAs (miRNAs), 55 ribosomal RNAs (rRNAs), 147 small nuclear RNAs (snRNAs), and 1,860 transfer RNAs (tRNAs; supplementary table S9, Supplementary Material online).

### Dosage Compensation of *L. striatellus*

To investigate whether the degree of dosage compensation differs in different *L. striatellus* body tissues, we obtained Illumina RNA-sequencing data from gonad tissues (female ovaries and male testes) and used publicly available somatic transcriptome data from legs, antennae, and heads (Li et al. 2021). To analyze dosage compensation, we compared gene expression levels between females (XX) and males (X) and between the X chromosome and autosomes within each sex.

We first examined the expression levels of the X chromosome versus autosomes in males (X:AA) and females (XX: AA; fig. 4A). We found that expression levels of autosomes were more strongly expressed than X chromosomes in all male tissues (fig. 4A; supplementary table S10, Supplementary Material online) (Wilcoxon P-values: head, 0.042; antennae, 0.0056; leg, 0.0066; gonad, 2E-16). Testis showed 81%,
and somatic tissues showed 3–5% higher median expression in autosomes compared with that of the X chromosome (supplementary table S11, Supplementary Material online). In contrast, there were no significant differences between X chromosomes and autosomes in female tissues (except legs; fig. 4A, supplementary table S10, Supplementary Material online; Wilcoxon P-values: head, 0.086; antenna, 0.1328; leg, 0.0442; gonad, 0.19). We also investigated the X:A expression ratio of females and males median chromosome expression. We found that in males, the X:A ratio ranged from 0.952 to 0.976, and in females, it ranged from 0.965 to 1.002 in somatic tissues. For gonad tissues, X:A ratio was 0.552 in testis and 1.031 in ovary (fig. 5A, supplementary table S10, Supplementary Material online).

In species with dosage compensation, expression should be balanced between one X (X) in males and two X (XX) in females. Therefore, we further compared differences in the X-chromosome expression levels between males and females. We found no significant differences in the X-chromosome expression levels between males and females in all somatic tissues (fig. 4B; supplementary table S10, Supplementary Material online; Wilcoxon P-values: head, 0.35; antenna, 0.57; leg, 0.86). Furthermore, the median F:M expression ratios for X chromosome ranged from 0.972 to 1.0 in all somatic tissues (fig. 5B; supplementary table S10, Supplementary Material online). However, in gonad tissue, we found significantly lower expression levels of X chromosome in male testis than in female ovary (figs. 2D, E and 4B; supplementary table S10, Supplementary Material online; Wilcoxon P-value <2E-16). The median F:M expression ratio was 1.76 in the X chromosome (fig. 5B; supplementary table S10, Supplementary Material online). When testing for difference in F:M ratio in X chromosomes and autosomes by the Wilcoxon rank-sum test, we found that X chromosomes and autosomes F:M ratios differed significantly in gonad tissue, but not in somatic tissues, with the exception of antenna (fig. 5B; Wilcoxon P-values: head, 0.7574; antenna, 0.0009; leg, 0.0526; gonad, 2E-16).

Taken together, these results suggest the existence of complete dosage compensation in the somatic tissue of L. striatellus. In gonad tissue, however, the male testis lacks dosage compensation.

Genomic Distribution of sex-biased Genes

Sex-biased genes are often nonrandomly located on chromosomes in many species, with sex chromosomes usually showing a pattern of masculinization or feminization. To analyze whether sex-biased genes have skewed distribution in L. striatellus, we used our gene structure annotation
results to infer the chromosomal locations of sex-biased and unbiased genes in somatic and gonad tissues. We observed a significantly higher proportion of female-biased genes on X chromosomes than autosomes in all tissues of *L. striatellus*, suggesting an enrichment of female-biased genes on the X chromosome (fig. 6A; supplementary table S12, Supplementary Material online). On the other hand, we found a significantly higher proportion of male-biased genes in autosomes compared with X chromosome in gonad tissue. In somatic tissues, the proportion of male-biased genes was higher in autosomes than on the X chromosome, although this difference was not significant (fig. 6B; supplementary table S12, Supplementary Material online). Sexual antagonism and possible lack of dosage compensation in the testis may drive X-chromosome enrichment of male-biased genes and autosome enrichment of female-biased genes and autosome enrichment of male-biased genes (see Discussion).

**Discussion**

The previous two versions of the *L. striatellus* genome either have a relatively short scaffold N50 (Zhu et al. 2017) or leave ~20% of scaffolds unplaced on chromosomes (Ma et al. 2021). Here, we improved the genome quality of *L. striatellus*. The new assembly demonstrated higher contiguity than previous versions at both contig and scaffold levels, with contig N50 increased from 71 to 1,362 kb and scaffold N50 increased from 29.24 to 34.96 Mb (table 1). In addition, most scaffolds (99.12%) were anchored to chromosomes, with a lower proportion of ambiguous nucleotides
Argued that sex chromosome dosage compensation in males was complete. Gu and Walters (2017) argued that sex chromosome dosage compensation should be limited to that heterogametic X-linked gene expression equal to ancestral level of the proto-X, and the equality of sex-linked expression between the sexes was termed as dosage balance. In *L. striatellus*, we found equal level of X chromosome expression between males and females, and no difference in expression between X chromosome and autosomes in homogametic females. Thus, complete compensation with balance existed in *L. striatellus* somatic tissues. Although we found that the X-chromosome expression levels were significantly lower than autosomes (fig. 4A), and the median X:A ratios were close to but still <1 in male somatic tissues (fig. 5A). First of all, some genes may escape dosage compensation. It has been reported that *Drosophila*'s ability to bear a deficiency in gene dose varied depending on the gene (Lee et al. 2016). Thus, lack of dosage compensation in some genes may cause subtle or even no effect on phenotypes, which may explain the escape of dosage compensation in some genes. In addition, the enrichment of female-biased genes and the absence of male-biased genes on the X chromosome may also contribute to decreased X-chromosome expression in males (fig. 6). Moreover, we also observed significantly higher levels of autosomal expression in males compared with females (fig. 4B; supplementary table S10, Supplementary Material online; Wilcoxon P-values: head, 0.000017; antenna, 0.18; leg, 0.0061; gonad, 0.02). Therefore, higher autosomal expression may also lead to unbalanced expression of autosomes and X chromosome in males.

For gonad tissue, gene expression levels on the X chromosome were significantly lower in male testes than in female ovaries, reflecting the absence of dosage compensation in male testes. A lack of dosage compensation also has been found in *Drosophila melanogaster* testes, suggesting most of the dosage compensation complex genes are not expressed in the *Drosophila* germline; different testis cells also have different degrees of dosage compensation (Witt et al. 2021). Details of dosage compensation in testis cells and the associated molecular mechanisms require further analysis in *L. striatellus*. Thus, our results show that the degree of dosage compensation differs between somatic and gonad tissues of *L. striatellus*.

Two modes of dosage compensation have been observed in many other species, by downregulating gene expression of two X/Z copies in homogametic sex or upregulating gene expression of one X/Z copy in hemi-gametic sex. Downregulation of two copies of X results in unequal gene expression between autosomes and X chromosomes in homogametic sex; however, this phenomenon was not observed in female *L. striatellus* (fig. 4A). Therefore, the dosage compensation mechanism of *L. striatellus* may be the upregulation of the male X chromosome.

In many species, sex-biased genes often exhibit a nonrandom genomic distribution. We found that the autosomes were rich in male-biased genes, while X chromosomes were rich in female-biased genes (fig. 6A and B). Similar patterns have been found in *Drosophila* genus, and three possible mechanisms have been proposed to explain these patterns, including dosage compensation, sexual antagonism, and meiotic sex chromosome inactivation (MSCI). Lack of dosage compensation in the *Drosophila* germline is thought to be an important cause of male-biased gene defects in the X chromosome (Bachtrog et al. 2010; Meiklejohn and Presgraves 2012). For *L. striatellus*, a similar scenario is plausible, where the lack of dosage compensation in male testis results in an enrichment of female-biased genes on the X chromosome. However, this mechanism alone does not explain the observations in somatic tissues where female-biased genes were enriched on the X chromosome (fig. 6A). In this context, it is possible that another evolutionary force sexual antagonism is at work, namely the opposing selection pressure between females and males may explain the deviations of gene distribution in somatic tissues. In the sexual antagonism theory, dominant mutations with beneficial fitness for females and detrimental fitness for males will be more likely to be fixed on the X chromosome in female homogametic organisms (Rice 1984). MSCI refers to the inactivation of X-linked genes during male meiosis. Since the X chromosome is inactivated in spermatoocytes, genes required for sperm maturation are expected to escape to autosomes (Lu and Wu 2011). We observed significantly higher levels of autosomal gene expression in testis than in ovary (fig. 4B), and that testis-biased genes were enriched in autosomes (fig. 6B). Thus, genes critical for testis development appear to escape to the autosomes, which may indicate the presence of MSCI in SBPH testis; however, the existence of this process in testis remains to be verified.

### Conclusion

This study provided an improved chromosome-level genome assembly of *L. striatellus*. We also confirmed that the degree of dosage compensation varied by tissues in *L. striatellus*. The complete dosage compensation regulation was present in somatic tissues but not in testis tissue. We also characterized the X-chromosome feminization of *L.


**Materials and Methods**

**Insect Strain and Sample Preparation**

The *L. striatellus* strain used in this study was initially collected from rice fields (in Hangzhou, China) in 2018 and reared in the laboratory for >10 generations. The strain was kept at 27 ± 0.5 °C with relative humidity >80% on rice seedlings (Xiushui 134) under a 16-h photoperiod.

**Genome Sequencing**

We used the long reads generated by Oxford Nanopore platform to do de novo genome assembly of *L. striatellus*. Genomic DNA from ~500 male adult was extracted by QIAamp DNA MiniKit (Qiagen) following the manufacturer’s instructions. The ONT sequence library was constructed by the ONT 1D Ligation Sequencing Kit (LSK109) according to the manufacturer’s procedure and then sequenced on Oxford Nanopore PromethION sequencers. To correct the sequencing error of ONT data, and estimate the genome size, we further sequenced the DNA genome extracted from 100 unmated male adults on Illumina platform HiSeq X. The Illumina sequencing library was constructed by the NEBNext Ultra II DNA Library Prep Kit.

**Draft Genome Assembly**

A genome survey was carried out to estimate the genome size and heterozygosity using the males Illumina paired-end reads by GenomeScope v2.0 (Ranallo-Benavidez et al. 2020), and the kmer was set to 19. The original draft genome was assembled by NextDenovo v2.4.0 (https://github.com/Nextomics/NextDenovo), the NextCorrect module performed error correction, and the NextGraph module assembled the genome based on the ONT data, and parameters in configuration file were “rerun=3; read_type=ont; read_cutoff=1k; genome_size=500m.” The initial draft genome assembly was further polished twice to correct base errors. The first polish applied the raw Nanopore reads by NextPolish v1.3.1 with default parameters (Hu et al. 2020), and the second polish used Illumina reads by Pilon v1.2.3 with default parameters (Walker et al. 2014). The heterozygosity was reduced by HaploMerger2 with default parameters (Huang et al. 2017) and Purge Haplotigs v1.1.1 was further used to obtain the haploid contigs with parameters “-a 60” (Roach et al. 2018).

**Hi-C Sequencing**

Thirty male-fifth instar nymphs were collected for Hi-C library construction. To construct the Hi-C library, the samples were first mechanically disrupted with a homogenizer. The cross-linking reaction was conducted by incubating the cells in 2% formaldehyde. The reaction was stopped by adding 2.5 M glycerol. Thereafter, the cells were flash-frozen in liquid nitrogen and stored at −80°C. After cross-linking, the chromatin was digested with 400 IU of Mbol restriction enzyme (NEB). The Hi-C library was then prepared by adding biotin, adding T4 DBA ligase (NEB) to ligate DNA, adding proteinase K for reverse cross-linking, DNA fragments purification, fragment shearing to 300–500 bp, and DNA ends repair. Finally, DNA fragments labeled by biotin were separated on Dynabeads® M-280 Streptavidin. The Hi-C library was sequenced on an Illumina Hiseq platform. The Hi-C library construction and sequencing were performed by Novogene Co., Ltd.

**Hi-C Assembly**

Chromosomal-level assembly was constructed based on the 3D de novo assembly (3D-DNA) pipeline (Dudchenko et al. 2017). The Hi-C sequencing data were first aligned to the draft genome by Juicer version 1.6 (https://github.com/aidenlab/juicer) to obtain linkage matrix using “-s Mbol.” Then 3D-DNA version 180922 (https://github.com/aidenlab/3d-dna) was performed to obtain the chromosomal-level genome using “-r 3.” Juicebox Assembly Tools (JBAT) v1.11.08 (Durand et al. 2018) was used to visualize and modify potential assembly errors before obtaining final high-quality genome. BUSCO version v5.1.3 (Simão et al. 2015) was used to evaluate the completeness of the genome assembly. Basic genomic features including GC content, gene density, and transcriptome reads coverage of ovary and testis were displayed by circos v0.69-8 (Krzywinski et al. 2009).

**X-Chromosome Determination**

In the *L. striatellus* sex chromosome determination system, females have two X chromosomes (XX), while males have only one X chromosome (XO). The X chromosome can be identified by calculating the coverage ratio between males and females. DNA sequencing coverage on the X chromosome in males was expected to be approximately half that in females at the same sequencing depth. Therefore, ~100 unmated males and females were prepared separately for coverage analysis. Genomic DNA extraction and library construction were performed as described above and sequenced on the Illumina HiSeq X platform. Approximately 34G of males and 33G of females clean DNA sequencing reads were generated. The sequencing data were aligned...
to the genome assembly by BWA v0.7.17-r1188 (Li 2013) with default parameters. The low-quality mapped reads were filtered out by SAMtools v1.10 using “-q 30” (Li et al. 2009). Read coverage was then calculated by sliding a 100-kb window across the genome using the command coverageBed from BEDTools version 2.25.0 (Quinlan and Hall 2010).

Transcriptome Sequencing

To better annotate the genome, we prepared the transcriptome of *L. striatellus* from representative tissues and developmental stages, and generated both long Nanopore and short Illumina RNA-seq reads. To construct ONT full-length cDNAs, RNA samples were prepared from different developmental stages of *L. striatellus*, including eggs, nymphs, adult females, and adult males. Samples from different body tissues, including carcass, fat body, gut, salivary gland, ovary, and testis, were prepared to sequence on the Illumina platform. All total RNA was extracted by the Takara RNizol Total RNA Isolation Kit according to the manufacturer’s protocols. The ONT and Illumina RNA libraries construction and sequencing were performed by Novogene Co., Ltd following the manufacturer’s instruction.

Repetitive Elements and Non-coding RNA Annotation

The repetitive sequences and transposable elements across the genome were identified by RepeatMasker v4.1.2 with default parameters (Chen 2004). Both the eukaryote RepBase library version 20181026 (https://www.girinst.org) and a de novo repeat library constructed by RepeatModeler v2.0.2 with default parameters (Flynn et al. 2020) were used as target databases. SSRs or microsatellites were identified using MISA (Beier et al. 2017). The ncRNAs including miRNAs, snRNAs, and rRNAs were detected using the Rfam database (http://rfam.xfam.org), and tRNAs were predicted by tRNAscan-SE v2.0.9 (Lowe and Eddy 1997).

Genome Annotation

Protein-coding genes were annotated by combining three lines of evidence, including RNA-seq-based, homology-based, and de novo methods. Illumina short-read cDNA sequences from different tissues of carcass, fat body, gut, salivary gland, ovary, and testis were first assembled into transcripts by Trinity v2.11.0 with default parameters (Haas et al. 2013). ONT reads from eggs, nymphs, adult females, and adult males were mapped to the genome using Minimap2 v2.17-r941 (Li 2018), andrnaSPAdes v3.14.1 (Bushmanova et al. 2019) was used to obtain full-length transcripts with default parameters. Illumina-assembled transcripts and ONT full-length transcripts were mapped to the genome by PASA v2.4.1 with default parameters (Haas et al. 2003) to predict gene structures. GeneWise v2.4.1 (Birney et al. 2004) was applied for homology prediction. Before de novo prediction, repeat sequences were soft masked. AUGUSTUS v3.4.0 (Stanke et al. 2006), SNAP version 20060728 (Korf 2004), GeneMark-ET Suite v4.59 (Lomsadze et al. 2014), BRAKER v2.1.2 (Hoff et al. 2016), and MAKER v3.01.03 (Holt and Yandell 2011) were implemented separately to perform ab initio gene prediction. EvidenceModeler v1.1.1 (EVM; Haas et al. 2008) was used to merge the above results to the final annotation results with parameters “–segmentSize 1000000 –overlapSize 10000” and weights “ab initio, 1; homologs, 5; transcripts, 10.”

Functional annotation of protein-coding genes was performed by aligning sequences to NR (Deng et al. 2006), NT, UniProtKB/Swiss-Prot (Bairoch and Boeckmann 1991), and KOG databases using BLAST (Altschul et al. 1990; E-value < 1e-5). Sequences aligned to CDD using the webserver (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi). The softwares eggnoc-mapper v2.1.5 (Huerta-Cepas et al. 2017) and Interproscan v5.53-87.0 (Jones et al. 2014) were used to obtain the GO annotations. The KEGG pathways were annotated using the KEGG Automatic Annotation Server (Moriya et al. 2007).

Dosage Compensation

Illumina RNA-seq data for gonad tissues (testis and ovary) were generated as described above, and two biological replicates were prepared for each. RNA-seq data for the somatic tissues leg, antenna, and head were downloaded from publicly available database with no biological replicate (Li et al. 2021). RSEM v1.3.1 (Li and Dewey 2011) was applied to calculate the FPKM expression matrix using default settings, and TMM normalization was applied across all samples using the run_TMM_scale_matrix.pl script in Trinity v2.11.0 (Haas et al. 2013). The relationship between biological replicates of gonad tissues was assessed by Spearman’s rank correlation, with correlation coefficients of ∼0.97 for both ovary and testis (supplementary table S13, Supplementary Material online). The average expression of all replicates for each gene was used for the following analysis. We filtered genes with an average expression level of FPKM < 1 in males and females, as this approach has been used in dosage compensation studies for other species (Pal and Vicoso 2015; Catalán et al. 2018). We then log-transformed autosomal and X-chromosomal expression to obtain log2FPKM values in males and females. The Wilcoxon rank-sum test was applied to detect bias in female and male expression. The X:A expression ratio was calculated by dividing the median gene expression on the X chromosome by the median of all genes expressed on autosomes. The deviation of the female: male expression
ratio on the X chromosome and autosomes was tested by the Wilcoxon rank-sum test.

Sex-biased Gene Distribution

The RNA-seq data were mapped to the genome using HISAT2 v2.1.0 with default parameters (Kim et al. 2019), and then featureCounts v2.02 (Liao et al. 2014) was applied to calculate the raw read count of each gene using parameters “-p -B -C -t exon -g gene_id.” Sex-biased genes were identified by the Bioconductor package edgeR v3.32.1 (Robinson et al. 2010) as conducted in R v4.0.2 (Team 2013). For gonad tissues with biological replicates, we retained genes with normalized counts per million (CPM) value >1 in both two replications. We then estimated dispersion using the tag-wise method, and applied a generalized log-linear model to fit the one-coefficient analysis. For somatic tissues, due to the downloaded data from public database had no biological replication, we retained genes with CPM normalized value >1. The square-root dispersion (bcv) was set to 0.2. All P-values were corrected with Benjamini and Hochberg methods (Benjamini and Hochberg 1995). Sex-biased genes were identified as those genes showing logFC >2 or <-2 and false discovery rate <0.05 (supplementary tables S14–S17, Supplementary Material online). Chromosome counts for sex-biased and unbiased genes on autosomes and X chromosomes formed a 2×2 contingency table, and Fisher’s exact test was carried out to test the deviation of gene proportions located on X chromosomes and autosomes.

Supplementary material

Supplementary data are available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org).

Author Contributions

Conceptualization: C.X.Z.; software: Y.X.Y. and Q.L.H.; formal analysis: Q.L.H.; resources: J.C.Z., H.J.H., and J.M.L.; writing-original draft preparation: Q.L.H.; writing-review and edition: C.X.Z., Y.X.Y., and J.C.Z.; supervision: C.X.Z.; project administration: C.X.Z.; funding acquisition: C.X.Z. All authors have read and agreed to the published version of the manuscript.

Funding

This work was funded by National Key Research and Development Program of China (2021YFD1401100), Natural Science Foundation of Zhejiang Province (LQ21C140006), National Natural Science Foundation of China (31900361) and the Fundamental Research Funds for the Central Universities (K20210106).

Data Availability

The Nanopore and Illumina sequencing data were submitted to the NCBI Sequence Read Archive (SRA) database under BioProject accession number PRJNA841412. This Whole Genome Shotgun project has been deposited at GenBank under the accession JADWYY0000000000. The version described in this paper is version JADWYY0100000000. The gene structure annotation and gene functional annotation files were deposited at figshare (https://doi.org/10.6084/m9.figshare.20751286.v2).

Literature Cited

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215:403–410.
Bachtrog D, Toda NR, Lockton S. 2010. Dosage compensation and de-masculinization of X chromosomes in Drosophila. Curr Biol. 20: 1476–1481.
Bairach A, Boeckmann B. 1991. The SWISS-PROT protein sequence data bank. Nucleic Acids Res. 19:2247.
Beier S, Thiel T, Münch T, Scholz U, Mascher M. 2017. MISA-web: a web server for microsatellite prediction. Bioinformatics 33: 2583–2585.
Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc: Ser B (Methodological). 57:289–300.
Bergero R, Charlesworth D. 2009. The evolution of restricted recombination in sex chromosomes. Trends Ecol Evol. 24:94–102.
Binney E, Clamp M, Durbin R. 2004. Genewise and genomewise. Genome Res. 14:988–995.
Bushmanova E, Antipov D, Lipidus A, Prijibelski AD. 2019. rnasPades: a de novo transcriptome assembler and its application to RNA-Seq data. GigaScience 8:gis100.
Catalán A, Macias-Munoz A, Briscoe AD. 2018. Evolution of sex-biased gene expression and dosage compensation in the eye and brain of Heliconius butterflies. Mol Biol Evol. 35:2120–2134.
Chauhan P, et al. 2021. Genome assembly, sex-biased gene expression and dosage compensation in the damselfly Ischnura elegans. Genomics 113:1828–1837.
Chen N. 2004. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics. S: 4.10.11–14.10.14.
Deng Y, et al. 2006. Integrated nr database in protein annotation system and its localization. Comput Eng. 32:71–74.
Disteche CM. 2012. Dosage compensation of the sex chromosomes. Annu Rev Genet. 46:537–560.
Dudchenko O, et al. 2017. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science 356:92–95.
Durand NC, et al. 2016. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Systems. 3:99–101.
Ellegren H, et al. 2007. Faced with inequality: chicken do not have a general dosage compensation of sex-linked genes. BMC Biol. 5:40.
Flynn JM, et al. 2020. Repeatmodeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 117: 9451–9457.
Gartler SM, Riggs AD. 1983. Mammalian X-chromosome inactivation. Annu Rev Genet. 17:155–190.
Graves JAM. 2016. Evolution of vertebrate sex chromosomes and dosage compensation. Nat Rev Genet. 17:33–46.
Gu L, et al. 2019. Dichotomy of dosage compensation along the Neo Z chromosome of the monarch butterfly. Curr Biol. 29: 4071–4077.e4073.
Gu L, Walters JR. 2017. Evolution of sex chromosome dosage compensation in animals: a beautiful theory, undermined by facts and be-deviled by details. Genome Biol Evol. 9:2461–2476.
Haas BJ, et al. 2003. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 31:5654–5666.
Haas BJ, et al. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol. 9:R7.
Haas BJ, et al. 2013. De novo transcript sequence reconstruction from RNA-Seq using the Trinity platform for reference generation and analysis. Nat Protoc. 8:1494–1512.
Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics 32:767–769.
Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinf. 12:491.
Hu J, Fan J, Sun Z, Liu S. 2020. Ntxpolish: a fast and efficient genome polishing tool for long-read assembly. Bioinformatics 36:2253–2255.
Huang S, Kang M, Xu A. 2017. Haplomerger2: rebuilding both haploid sub-assemblies from high-heterozygosity diploid genome assembly. Bioinformatics 33:2577–2579.
Huerta-Cepas J, et al. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol Biol Evol. 34:2115–2122.
Jones P, et al. 2014. Intergenoscan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240.
Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based contig reassignment for third-gen diploid genome assemblies. Bioinformatics 35:W182–W185.
Korf I. 2004. Gene finding in novel genomes. BMC Bioinf. 5:59.
Krzywinski M, et al. 2009. Circos: an information aesthetic for comparative genomics. Genome Res. 19:1639–1645.
Larschan E, et al. 2011. X chromosome dosage compensation via enhanced transcriptional elongation in *Drosophila*. Nature 471:115–118.
Laver T, et al. 2015. Assessing the performance of the Oxford Nanopore Technologies MinION. Biomol Detect Quantif 3:1–8.
Lee H, et al. 2016. Effects of gene dose, chromatin, and network topology on expression in *Drosophila melanogaster*. PLoS Genet. 12: e1006295.
Li H, et al. 2009. The sequence alignment/map format and SAM tools. Bioinformatics 25:2078–2079.
Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997.
Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100.
Li Y, et al. 2021. Comparative transcriptome analysis of chromoreception organs of *Laodelphax striatellus* in response to rice stripe virus infection. Int J Mol Sci. 22:10299.
Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinf. 12:1–16.
Liao Y, Smyth GK, Shi W. 2014. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923–930.
Liu W, et al. 2016. Statistics and analysis of crop yield losses caused by main diseases and insect pests in recent 10 years. Plant Prot. 42:1–9.
Lomsadze A, Burns PD, Borodovsky M. 2014. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. Nucleic Acids Res. 42:e119.
Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25:955–964.
Lu X, Wu CL. 2011. Sex, sex chromosomes and gene expression. BMC Biol. 9:30.
Lyon MF. 1961. Gene action in the X-chromosome of the mouse (*Mus musculus L.*). Nature 190:372–373.
Ma W, et al. 2021. Chromosomal-level genomes of three rice plant hoopers provide new insights into sex chromosome evolution. Mol Ecol Resour. 21:226–237.
Mahajan S, Bachtrog D. 2015. Partial dosage compensation in Strepsiptera, a sister group of beetles. Genome Biol Evol. 7:591–600.
Meiklejohn CD, Presgraves DC. 2012. Little evidence for demasculinization of the *Drosophila* X chromosome among genes expressed in the male germline. Genome Biol Evol. 4:1007–1016.
Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35:W182–W185.
NextDenovo [Internet]. Nxtomics. Available from: https://github.com/Nxtomics/NextDenovo. (last updated date: 13 October 2021; cited date: 6 November 2022)
Nozawa M, Fukuda N, Ikeo K, Gojobori T. 2014. Tissue-and stage-dependent dosage compensation on the neo-X chromosome in *Drosophila pseudoobscura*. Mol Biol Evol. 31:614–624.
Ohno S. 1967. Sex chromosomes and sex-linked genes. Berlin, Heidelberg, New York: Springer.
Pal A, Vicoso B. 2015. The X chromosome of hemipteran insects: conservation, dosage compensation and sex-biased expression. Genome Biol Evol. 7:3259–3268.
Prince EG, Kirkland D, Demuth JP. 2010. Hyperexpression of the X chromosome in both sexes results in extensive female bias of X-linked genes in the flour beetle. Genome Biol Evol. 2:336–346.
Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.
Team RC. 2013. R: A language and environment for statistical computing.
Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. Genomescope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 11:1–10.
Ranz JM, et al. 2021. A de novo transcriptional atlas in * Danaus plexippus* reveals variability in dosage compensation across tissues. Commun Biol 4:1–13.
Rice WR. 1984. Sex chromosomes and the evolution of sexual dimorphism. Evolution 38:735–742.
Richard G, et al. 2017. Dosage compensation and sex-specific epigenetic landscape of the X chromosome in the pea aphid. Epigenetics Chromatin 10:30.
Roach MJ, Schmidt SA, Borneman AR. 2018. Purge haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinf. 19:1–10.
Robinson MD, Mccarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140.
Rose G, et al. 2016. Dosage compensation in the African malaria mosquito anopheles gambiae. Genome Biol Evol. 8:411–425.
Shao C, et al. 2014. Epigenetic modification and inheritance in sexual reversal of fish. Genome Res. 24:604–615.
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212.
Stanke M, et al. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. 34:W435–W439.
Vensko SP, Stone EA. 2015. Recent progress and open questions in Drosophila dosage compensation. Fly (Austin). 9:29–35.
Vicoso B, Emerson JJ, Zektser Y, Mahajan S, Bachtrog D. 2013. Comparative sex chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global dosage compensation. PLoS Biol. 11:e1001643.
Walker BJ, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963.
Witt E, Shao Z, Hu C, Krause HM, Zhao L. 2021. Single-cell RNA-sequencing reveals pre-meiotic X-chromosome dosage compensation in Drosophila testis. PLoS Genet. 17:e1009728.
Ye YX, et al. 2021. Chromosome-level assembly of the brown planthopper genome with a characterized Y chromosome. Mol Ecol Resour. 21:1287–1298.
Zhu J, et al. 2017. Genome sequence of the small brown planthopper, Laodelphax striatellus. GigaScience 6:1-12.

Associate editor: Judith Mank