G3, 2024, **14(11)**, jkae223
https://doi.org/10.1093/g3journal/jkae223
Advance Access Publication Date: 18 September 2024
Genome Report

Genome sequence of the sugarcane aphid, *Melanaphis sacchari* (Hemiptera: Aphididae)

Jinshuai Zhao [1,2,†] Liqiang Xie,[1,2,†] Xinrui Zhao,[3] Luhua Li,[4] Jianghui Cui,[3,*] Jinfeng Chen [1,*]

[1]State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China
[2]College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
[3]North China Key Laboratory for Crop Germplasm Resources of Education Ministry, Hebei Agricultural University, Baoding, Hebei 07100, China
[4]College of Agriculture, Guizhou University, Guiyang 550025, China

*Corresponding author: North China Key Laboratory for Crop Germplasm Resources of Education Ministry, Hebei Agricultural University, Baoding, Hebei 07100, China. Email: 13663123545@126.com; *Corresponding author: State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China. Email: chenjinfeng@ioz.ac.cn
[†]These authors contributed equally to the article.

The sugarcane aphid, *Melanaphis sacchari*, is an agricultural pest that causes damage to plants in the Poaceae (the grasses) family, such as sorghum and sugarcane. In this study, we used nanopore long reads and a high-throughput chromosome conformation capture chromatin interaction maps to generate a chromosome-level assembly with a total length of 356.1 Mb, of which 85.5% (304.6 Mb) is contained within the 3 autosomes and the X chromosome. Repetitive sequences accounted for 16.29% of the chromosomes, and a total of 12,530 protein-coding genes were annotated, achieving 95.8% Benchmarking Universal Single-Copy Ortholog gene completeness. This offered a substantial improvement compared with previous low-quality genomic resources. A phylogenomic analysis by comparing *M. sacchari* with 24 published aphid genomes representing 3 aphid tribes revealed that *M. sacchari* belonged to the tribe Aphidini and maintained a conserved chromosome structure with other Aphidini species. The high-quality genomic resources reported in this study are useful for understanding the evolution of aphid genomes and studying pest management of *M. sacchari*.

Keywords: the sugarcane aphid; *Melanaphis sacchari*; synteny; genome assembly

Introduction

Aphids (Hemiptera: Aphididae) are one of the most destructive agricultural pests. They feed on plant phloem (phloem-feeding) and transmit more than 55% of all known insect-transmitted plant viruses (Ng and Falk 2006; Mathers *et al.* 2020). About 100 out of the ~5,000 aphid species have been recognized as important agricultural pests (Blackman and Eastop 2017). Additionally, the complex life cycle, extensive phenotypic plasticity, and rapid environmental adaptation of aphids make them an important model for the study of plant–insect interactions (Ferry *et al.* 2004; Hogenhout and Bos 2011), speciation (Hawthorne and Via 2001; Peccoud *et al.* 2009), and sex chromosome evolution (Jaquiery *et al.* 2013, 2018).

There is a lack of genomic resources on aphids, which hampers the study of these destructive pests (Mathers 2020). The first aphid genome, the pea aphid *Acyrthosiphon pisum*, was sequenced in 2010 (The International Aphid Genomics Consortium 2010). Subsequently, some important pest aphid genomes were sequenced, such as *Aphis gossypii* (Mathers *et al.* 2022), *Myzus persicae* (Mathers *et al.* 2021), and *Sitobion avenae* (Mathers *et al.* 2023). However, these aphids represent only a small fraction of all aphid species, and most published aphid genomes are not assembled at the chromosome level (Mathers *et al.* 2022). This greatly affects large-scale genome structural variation analysis and genome-wide synteny analysis (Chaisson *et al.* 2015; Chakraborty *et al.* 2018). Therefore, high-quality aphid genomic data are needed to understand the diversity, adaptation, and genome evolution of aphids.

The sugarcane aphid, *Melanaphis sacchari*, is one of the major pests on sorghum and sugarcane in many areas of Asia, Africa, Australia, the Far East, and parts of Central, North, and South America (Singh *et al.* 2004; Bowling *et al.* 2016; Pekarcik and Jacobson 2021). *M. sacchari* is difficult to control due to its incredible fertility, rapid population expansion through parthenogenetic reproduction, and its effective dispersal strategy (Brewer *et al.* 2017; Neupane *et al.* 2020). *M. sacchari* feeds on plant phloem nutrients directly harming plant growth and development and also acts as an insect vector causing serious indirect harm to crop production (Hogenhout *et al.* 2008) by transmitting plant viruses, such as sugarcane yellow leaf virus (Ahmad *et al.* 2007) and millet red leaf virus (Blackman and Eastop 1984) in a persistent, circulative, nonpropagative manner (Gray and Gildow 2003), as well as sugarcane mosaic virus (Yang 1986; Setokuchi and Muta 1993) in a nonpersistent manner (Gadhave *et al.* 2020). In addition, *M. sacchari* produces large amounts of honeydew and sooty mold, which results in a serious reduction in yields (Bowling *et al.* 2016). Apart from harming sorghum and sugarcane crops, *M. sacchari* can also damage corn, rice, and so on (Singh *et al.* 2004; Exilien *et al.* 2022). However, a high-quality reference genome is still lacking for *M. sacchari*. In this study, we report a high-quality chromosome-level genome of *M. sacchari* using nanopore long reads and a high-throughput chromosome conformation capture (Hi-C) chromatin interaction map.

Received on 28 May 2024; accepted on 11 September 2024

Methods

Genomic DNA isolation and sequencing

Female *M. sacchari* were collected from the greenhouse at the Hebei Agricultural University, located in Baoding, Hebei Province. The aphids originated from a local wild population native to the region and were named as "Hebei1." The whole body of insects was collected for short reads, long reads, Hi-C, and mRNA analysis. Genomic DNA was extracted and purified from female *M. sacchari* individuals. For short-read sequencing, we constructed a 150-bp paired-end sequencing library and performed sequencing on the MGISEQ-2000RS platform (MGI Tech Co., Ltd, Shenzhen, China). Trimmomatic (v0.39) (Bolger *et al.* 2014) was used to filter out low-quality bases and remove sequencing adaptors. For long-read sequencing, we quantified the genomic DNA using Qubit (v4.0) (Invitrogen). Then, genomic DNA libraries were prepared using the Ligation Sequencing Kit (Oxford Nanopore Technologies, Oxford, UK: SQK-LSK109) following the manufacturer's protocol and sequenced on an Oxford Nanopore PromethION flow cell.

Hi-C library preparation and sequencing

To further improve the continuity of the assembled genomes of *M. sacchari*, we generated Hi-C data using chromosome conformation capture experiments. We extracted and purified genomic DNA from whole bodies of female *M. sacchari* individuals. Subsequently, the nuclei were isolated, digested with 100 units of *Dpn*II restriction enzyme, and end-labeled via biotinylation with biotin-14-dATP. The ligated DNA was sheared into 300–600 bp fragments. These fragments underwent end repair, A-tail, and purification steps. Finally, the Hi-C libraries were quantified and sequenced on the DNBSEQ-T7 platform (MGI Tech Co., Ltd).

RNA library preparation and sequencing

Total mRNA was extracted from whole bodies of female *M. sacchari* individuals using TRIzol reagent (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's protocol. The RNA-seq libraries were prepared for 150-bp paired-end sequencing on the DNBSEQ-T7 platform (MGI Tech Co., Ltd).

Genome assembly

Before de novo assembly, we estimated the genome size of *M. sacchari* based on the short reads. Jellyfish (v1.1.10) (Marcais and Kingsford 2011) was employed to calculate the frequency of each of the K-mers ($n = 17$) with the parameters "count -m 17 -s 200000 M –C." The resulting data were then inputted into GenomeScope (v1.0) (Vurture *et al.* 2017) to estimate the genome size. The preliminary genome assembly was performed using NextDenovo (v2.4.0) (Hu *et al.* 2024) with customized parameters (read cutoff = 10k, seed depth = 45, genome size = 380 M). Subsequently, we used one round of Pilon (v1.24) (Walker *et al.* 2014) polishing with short reads to acquire a corrected assembly. The completeness of the assembly was assessed using Benchmarking Universal Single-Copy Orthologs (BUSCOs) (v4.1.4) (Manni *et al.* 2021) with the Insect_odb10 dataset genes ($n = 1{,}367$). We used ALLHiC (v0.9.8) (Zhang *et al.* 2019) to obtain a scaffold-level genome based on Hi-C data. The scaffold-level genome was imported into Juicebox (v1.11.08) (Durand *et al.* 2016) and manually corrected to obtain a chromosome-level genome based on Hi-C interaction signals. Finally, 4 chromosomes, namely MSAC_01 (1), MSAC_02 (2), MSAC_03 (3), and MSAC_04 (4 or X), and 132 scaffolds (150 contigs) made up the final primary assembly. During manual correction, scaffolds6–136 with short lengths (0.5–2.1 Mb) could not be clearly located on chromosomes, so we classified them as "unplaced assembly." Meanwhile, we observed that scaffold5, which is 22.4 Mb and contains 19 contigs, displayed no significant global Hi-C interaction with other regions of the genome except for a weak interaction with part of the chromosome 4. The average coverage of short reads on scaffold5 (~41-fold coverage) was significantly lower than that of chromosomes 1, 2, 3, and 4 (92–87-fold coverage). Despite annotating 1,321 genes on scaffold5, their expression levels were significantly lower than those of genes on the autosomes and sex chromosomes, with almost no expression. Moreover, the genome and BUSCO gene completeness assessment for scaffold5 was 0%. According to the transposable element (TE) annotation, the proportion of TE was 16.29% in chromosomes 1, 2, 3, and 4 (304.6 Mb), but it was 56.61% in scaffold5 (22.4 Mb). We also found that scaffold5 not only had homologous genes but also exhibited syntenic regions, primarily on the X chromosome. Considering these results, we were unable to place the 22.4-Mb-long scaffold5 on any of the 4 chromosomes of *M. sacchari*. So, we classified it together with 131 other scaffolds as "unplaced assembly." Subsequent evolutionary and comparative genomic analyses were performed on the chromosome assembly, which includes chromosomes 1, 2, 3, and 4.

TE and gene annotation

For TE annotation, we constructed a *M. sacchari*-specific repeat database de novo using RepeatModeler (v2.0.1) (Flynn *et al.* 2020) with default parameters. We utilized RepeatMasker (v4.0.9) (http://www.repeatmasker.org) and *M. sacchari*-specific repeat database to identify repeat sequences throughout the *M. sacchari* genome with the parameters: "-e rmblast -div 40 -xsmall -nolow -norna." Our gene prediction strategy involved a comprehensive approach that combined transcriptome-based, de novo-based, and homology-based methods. For de novo prediction, we employed the AUGUSTUS tool in BRAKER (v2.1.4) (Hoff *et al.* 2016) for gene prediction using BAM files from the RNA-seq alignments as input data. To incorporate homologous evidence into our predictions, we imported protein-coding sequences from *Rhopalosiphum padi* (Mathers *et al.* 2022), *Rhopalosiphum maidis* (Chen *et al.* 2019), *Aphis glycines* (Mathers 2020), and *Aphis fabae* (Mathers *et al.* 2022) into miniprot (v0.11) (Li 2023) using the parameters "-I30 kb –gff" to perform a gene structure analysis based on homologous evidence. In terms of transcriptome-based prediction, raw reads were filtered using Trimmomatic (v0.39). The filtered reads were then mapped to the genome assembly using HISAT2 (v2.2.1) (Kim *et al.* 2019). StringTie (v2.21) (Pertea *et al.* 2015) was used to identify the transcript position in the genome assembly, and transcript sequences were extracted. We further mapped these extracted transcript sequences back to the genomes utilizing PASA (v2.41) (Haas *et al.* 2008). Additionally, TransDecoder v5.50 (https://github.com/TransDecoder/TransDecoder) was employed for generating gene predictions based on PASA-extracted transcripts. Finally, EvidenceModeler (v2.1.0) (Haas *et al.* 2008) [weights for each: Augustus (de novo): 2; TransDecoder (de novo): 3; Miniprot (homology): 8; PASA (transcriptome): 10] was used to integrate gene predictions from all 4 tools. The following parameters were applied during integration: "–segmentSize 10000000 –overlapSize 100000."

Phylogenetic tree construction and species divergence time estimation

We estimated a phylogeny of Hemiptera using protein sequences from our new genome assembly of *M. sacchari* and 25 previously

reported Hemiptera species, including 6 Aphidini species [*R. padi* (Mathers *et al.* 2022), *R. maidis* (Chen *et al.* 2019), *A. fabae* (Mathers *et al.* 2022), *A. gossypii* (Mathers *et al.* 2022), *Aphis thalictri* (Mathers *et al.* 2022), and *A. glycines* (Mathers 2020)], 17 Macrosiphini species [*A. pisum* (Mathers *et al.* 2021), *Brachycaudus cardui* (Mathers *et al.* 2022), *Brachycaudus helichrysi* (Mathers *et al.* 2022), *Brachycaudus klugkisti* (Mathers *et al.* 2022), *Brevicoryne brassicae* (Mathers *et al.* 2022), *Diuraphis noxia* (Mathers *et al.* 2022), *Macrosiphum albifrons* (Mathers *et al.* 2022), *Metopolophium dirhodum* (GCF_019925205.1) (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_019925205.1), *Myzus ligustri* (Mathers *et al.* 2022), *Myzus lythri* (Mathers *et al.* 2022), *Myzus varians* (Mathers *et al.* 2022), *Myzus cerasi* (Mathers *et al.* 2020), *M. persicae* (Mathers *et al.* 2021), *Phorodon humuli* (Mathers *et al.* 2022), *Pentalonia nigronervosa* (Mathers *et al.* 2020), *S. avenae* (Mathers *et al.* 2023), and *Sitobion miscanthi* (Mathers *et al.* 2023)], one Eriosomatini species [*Eriosoma lanigerum* (Biello *et al.* 2021)], and one hemipteran outgroup species [*Bemisia tabaci* (GCA_918797505.1)] (https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_918797505.1). In brief, BUSCO (v4.1.4) was used to identify conserved gene orthologs using the insecta_odb10 gene set ($n = 1{,}367$) in each genome assembly. The identifiers for complete genes present in all species were extracted using python scripts (https://github.com/ypchan/GPA/blob/main/gpa/singel_copy_BUSCO_datasets.py). The protein sequences for each single-copy BUSCO gene from all species were extracted. The sequence alignment of the resulting 832 single-copy BUSCO genes was performed using MAFFT (v7.310) (Katoh and Standley 2013). We extracted conserved sites using Gblocks (v0.91b) (Castresana 2000) with "-b4 = 5 -b5 = h -t = p" parameters after the sequence alignment. The best-fitting model (JTT + F + R6) was determined by the ModelFinder program (Kalyaanamoorthy *et al.* 2017) implemented in IQ-TREE (v2.0.3) (Nguyen *et al.* 2015) based on the Bayesian information criterion and ultrafast bootstrap approximation with 1,000 replicates (-bb 1000). We estimated a divergence time using MCMCTree, a tool within the PAML (v4.9) (Yang 2007) package. Calibration information for fossil nodes was obtained from the TimeTree website (Kumar *et al.* 2022).

TE divergence distribution

To estimate the relative age of TE and its transposition history in *M. sacchari*, we performed a Kimura distance–based pair divergence analysis of TE superfamilies based on Kimura 2-parameter distances (*K*-values) (Kimura 1980). In brief, we utilized the TE annotation output (alignment files) as input to calculate Kimura distances using calcDivergenceFromAlign.pl and createRepeatLandscape.pl (Perl scripts in the RepeatMasker util directory). Finally, transition and transversion rates were calculated for alignments and transformed into Kimura distances (Kimura 1980) using the following equation: $K = -1/2 \ln(1-2p-q) - 1/4 \ln(1-2q)$, where p represents the proportion of sites with transitions and q represents the proportion of sites with transversions.

Comparative analysis of orthologous gene families and synteny analysis

To perform a synteny analysis, we employed the following 2 methods: (1) We selected 9 aphid genomes at the chromosome level, encompassing 3 tribes: Aphidini (*R. padi*, *A. fabae*, *A. gossypii*, and *M. sacchari*), Macrosiphini (*A. pisum*, *S. miscanthi*, *B. brassicae*, and *M. persicae*), and Eriosomatini (*E. lanigerum*). OrthoFinder was used to identify strictly single-copy genes for each species. These strictly single-copy genes served as the input for JCVI (v0.7.5) (Tang *et al.* 2024) to plot the chromosomal synteny among species. (2) Syntenic blocks were pairwise identified between species using MCScanX (Wang *et al.* 2012), with a minimum requirement of 5 genes to call a syntenic block (-s 5). MCScanX results were visualized using SynVisio (https://synvisio.github.io/#).

Results

Assembly of the *M. sacchari* genome

To assemble a high-quality, chromosome-scale, reference genome of *M. sacchari*, we generated 54.6 million Oxford nanopore ultra-long reads (325 Gb), which is equivalent to ~915-fold genome coverage (Supplementary Table 1). These reads were assembled using NextDenovo, resulting in an assembly with a contig N50 of 12.5 Mb and a genome size of 355.7 Mb comparable to the estimated genome size of 359.4 Mb (Supplementary Fig. 1). The assembly was polished using paired-end short reads, and contigs were anchored onto chromosome using Hi-C reads (Supplementary Figs. 2 and 3), resulting in a final assembly of 356.1 Mb. The chromosome assembly consists of 4 chromosomes covering 304.6 Mb sequences and containing 85 contigs (Fig. 1a and Supplementary Table 2) (Blackman 1980). In the final assembly, additional 132 contigs or scaffolds, which were 51.5 Mb in size, were classified as "unplaced assembly" (see *Methods* for details). We reported data analysis based on the 304.6 Mb genome sequences composed of 4 chromosomes in the rest of the study. The completeness of the genome was 97.3% (Complete and single-copy or S: 95.8%, Complete and duplicated or D: 1.5%) assessed by BUSCO genes ($n = 1{,}367$). Compared with a previous genome assembly available at the National Center for Biotechnology Information (NCBI) (GCF_002803265.2), which was highly fragmented, containing numerous gaps, the BUSCO score has increased from 95.2 to 97.3%, and the fragmented BUSCO gene has decreased from 1.3 to 0.4%. Additionally, the number of contigs reduced from 1,347 to 85 and the contig N50 increased from 0.275 to 12.5 Mb (Fig. 1b and Supplementary Table 2). Therefore, the assembled genome of *M. sacchari* was highly contiguous and of high quality.

Annotation of TEs and protein-coding genes in the *M. sacchari* genome

TEs in the *M. sacchari* genome were annotated using RepeatModeler and RepeatMasker. TEs made up 16.29% of the assembled *M. sacchari* genome (Fig. 1c and d). Most TEs were unknown repetitive elements (10.57%), followed by DNA transposons (3.54%), long interspersed nuclear elements (LINEs) (1.39%), and long terminal repeat (LTR) retrotransposons (0.61%) (Fig. 1d and Supplementary Table 3). Short interspersed nuclear elements (SINEs) were absent in the *M. sacchari* genome, which is consistent with the TE distribution reported in aphid genomes (Baril *et al.* 2023). We observed that chromosome 4 (MSAC_4) exhibited a higher density of TEs compared with other chromosomes, with DNA transposons contributing an average of 5.7% of the sequences and primarily concentrated at the ends of the chromosome (Fig. 1c). In contrast, DNA transposons accounted for ~2.8–3.1% of other chromosomes in the *M. sacchari* genome. The distributions of sequence divergence for TE superfamilies estimated by Kimura 2-parameter distance indicated that *M. sacchari* experienced a recent transposition peak for both DNA transposons and LTR retrotransposons (Fig. 1d and Supplementary Fig. 4; Kimura 1980; Chalopin *et al.* 2015).

Annotation of protein-coding genes was performed using a combination of homology-based methodology, ab initio predictions, and transcriptome data. A total of 12,530 protein-coding genes were predicted in the *M. sacchari* genome, with 80.2% of

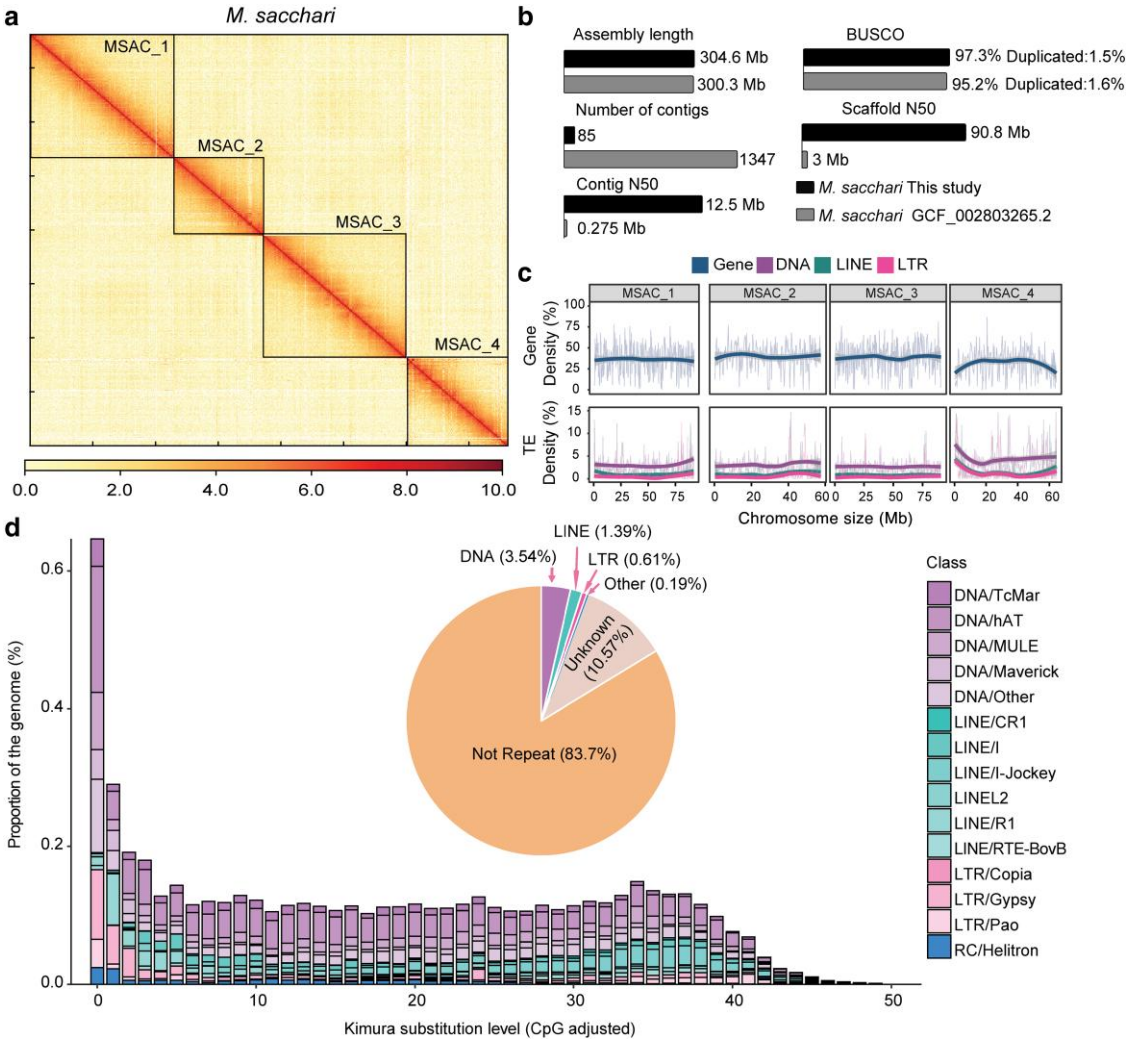

Fig. 1. A genome assembly, comparison, and annotation of *M. sacchari*. a) Hi-C matrix of *M. sacchari* genome. A heatmap showing the frequency of Hi-C contacts along our *M. sacchari* genome assembly. The different chromosomes are separated by a black frame. b) A comparison of genome assembly parameters with previously published genomes for *M. sacchari*. c) A density map of gene and TE distribution on chromosomes. Gene and TE density were calculated as the percentage of the length of genes or TEs within a 300-kb window along the chromosomes. d) A repeat landscape of the *M. sacchari* genome. The x-axis shows the Kimura 2-parameter distance between repeat copies and their respective consensus sequence, with low score indicating that the repeat copy is more recent. The y-axis shows the cumulative percentage of repeats in the genome. TE superfamilies are shown in different colors. A pie chart shows the percentage of TEs of the whole genome. Unknown refers to TEs that were unable to be classified into known superfamilies. A barplot shows the percentage of TEs of known superfamilies. A barplot with unknown repeats is shown in Supplementary Fig. 4.

them annotated to function using Kyoto Encyclopedia of Genes and Genomes, Gene Ontology, Pfam, or Orthologous Matrix (Supplementary Table 4). The average transcript length was 1,533 bp, which is similar to that of other aphid genomes. The completeness of the protein-coding genes of the *M. sacchari* genome was 95.8% assessed by BUSCO using the insecta_odb10 database ($n = 1{,}367$), indicating an improvement in single-copy BUSCO genes from 64.1 to 93.3% and a decrease in duplicated BUSCO genes from 31.2 to 2.5% (Supplementary Table 2), indicating the high quality of the *M. sacchari* gene annotation.

Comparative analysis of genomes among aphids

To better understand the evolutionary position of *M. sacchari* within the Aphididae, we incorporated 24 additional Aphididae genomes and used *B. tabaci* as an outgroup to construct the phylogenetic tree (Fig. 2a). Among these 26 species, a total of 832 strictly single-copy genes were identified and used for a phylogenetic analysis. The analysis revealed 3 well-supported clades of Aphididae, namely Macrosiphini, Aphidini, and Eriosomatini (Fig. 2a). The estimated divergence time was ~76 million years ago (MYA) for these 3 clades, whereas Macrosiphini and Aphidini separated ~41 (MYA). *M. sacchari* belongs to Aphidini and is diverged from the genus *Aphis* ~24 MYA. We performed a gene family analysis using 10 representative species with chromosome-level genomes available (Macrosiphini: *M. persicae*, *B. brassicae*, *S. miscanthi*, and *A. pisum*; Aphidini: *M. sacchari*, *R. padi*, *A. fabae*, and *A. gossypii*; Eriosomatini: *E. lanigerum*; and Aleyrodidae: *B. tabaci*). A total of 2,178 single-copy genes and 6,425 genes/gene families were found to be shared in all 10 species; however, Aphididae shared 6,253 single-copy genes and 8,154 genes/gene families (Fig. 2b). This suggests that multiple-copy gene families evolved rapidly in Aphididae. Many gene families were present in a specific lineage or species (Fig. 2b). Additionally, the TE content of the aphid genomes varies from 9.67% in *A. gossypii* to 40.32% in *M. albifrons* (Fig. 2a and Supplementary Table 5). There is a significant positive correlation between the genome size and TE content ($R = 0.92$, $P = 4.27e{-}11$; Supplementary Fig. 5).

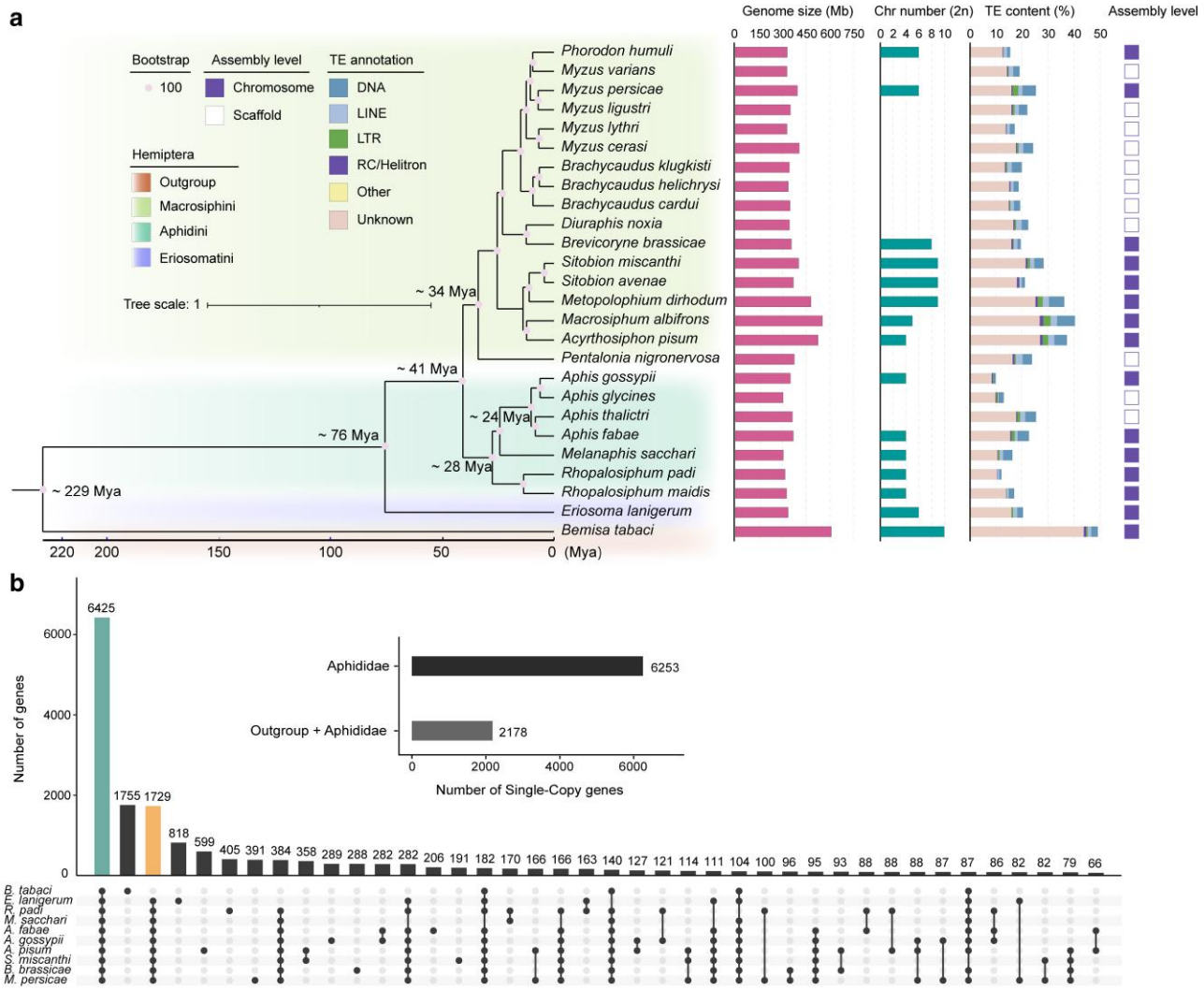

Fig. 2. Comparative genomics of the tribes Macrosiphini and Aphidini. a) A phylogenetic tree and characteristic comparison of 25 aphid species. The branch length represents the divergence time, with *B. tabaci* as an outgroup, while each node is supported by a bootstrap value of 100. The tree is annotated with genome size, chromosome number, TE content, and assembly level. b) A comparative analysis of homologous gene families from the 9 aphid species and the *B. tabaci*. The solid black dots in each column represent the set of species corresponding to the number of genes. The first bar on the left indicates the number of genes shared by outgroup and Aphididae. The first and third bars on the left indicate the number of genes shared by the Aphididae. The horizontal bar indicates the number of single-copy genes. Outgroup indicates *B. tabaci*, and Aphididae indicates *M. persicae*, *B. brassicae*, *S. miscanthi*, *A. pisum*, *M. sacchari*, *R. padi*, *A. fabae*, *A. gossypii*, and *E. lanigerum*.

Synteny analysis of aphid genomes

To investigate chromosome evolution in *M. sacchari* and other aphids, we performed a synteny analysis between 4 Macrosiphini species (*M. persicae*, *B. brassicae*, *S. miscanthi*, and *A. pisum*), 4 Aphidini species (*R. padi*, *A. fabae*, *A. gossypii*, and *M. sacchari*), and 1 Eriosomatini species (*E. lanigerum*). We first analyzed synteny using single-copy genes and found that *M. sacchari* showed conserved chromosome structure compared to Aphidini species (Fig. 3). However, extensive intra- and inter-chromosomal rearrangements were observed when compared to Eriosomatini and Macrosiphini species, which were consistent with earlier findings reported in a comparative analysis of aphid genomes (Supplementary Fig. 6; Mathers *et al.* 2021, 2023). The analysis also showed that chromosome 4 of *M. sacchari* (MSAC_4) is homologous to X chromosome of *A. pisum*. Aphidini does not exhibit strong rearrangement between X chromosome and autosomal chromosome as in Macrosiphini (Mathers *et al.* 2021, 2023). There is a lack of correspondence between the single-copy genes on chromosome 4 of *M. sacchari* and those located on the arms of the X chromosome within the Macrosiphini, *R. padi*, and *A. fabae*, suggesting that some additional chromosome material appears to have been introduced into the latter species. To further elucidate this phenomenon, we then performed synteny analysis using MCScanX to identify syntenic genome regions based on all protein-coding genes (Fig. 4a). We observed that the protein-coding genes on the arms of X chromosome of *A. pisum*, *S. miscanthi*, *M. persicae*, *B. brassicae*, and *A. fabae* have homologous genes on chromosome 4 of *M. sacchari*, indicating that these genes may arise from gene duplications and accumulate in these regions on X chromosomes of these species. The synteny analysis revealed that homologous genes present on each autosome of *M. sacchari* were identified on the 3 autosomes of *A. pisum*, with over 75% of these genes concentrated on chromosomes 1 and 2 in *A. pisum*. This is also the case when comparing *M. sacchari* to *S. miscanthi*, *B. brassicae*, and *M. persicae* (Fig. 4). Additionally, we observed chromosomal fission and fusion events in *S. miscanthi*, *B. brassicae*, and *M. persicae* (Mathers *et al.* 2023; Li *et al.* 2024).

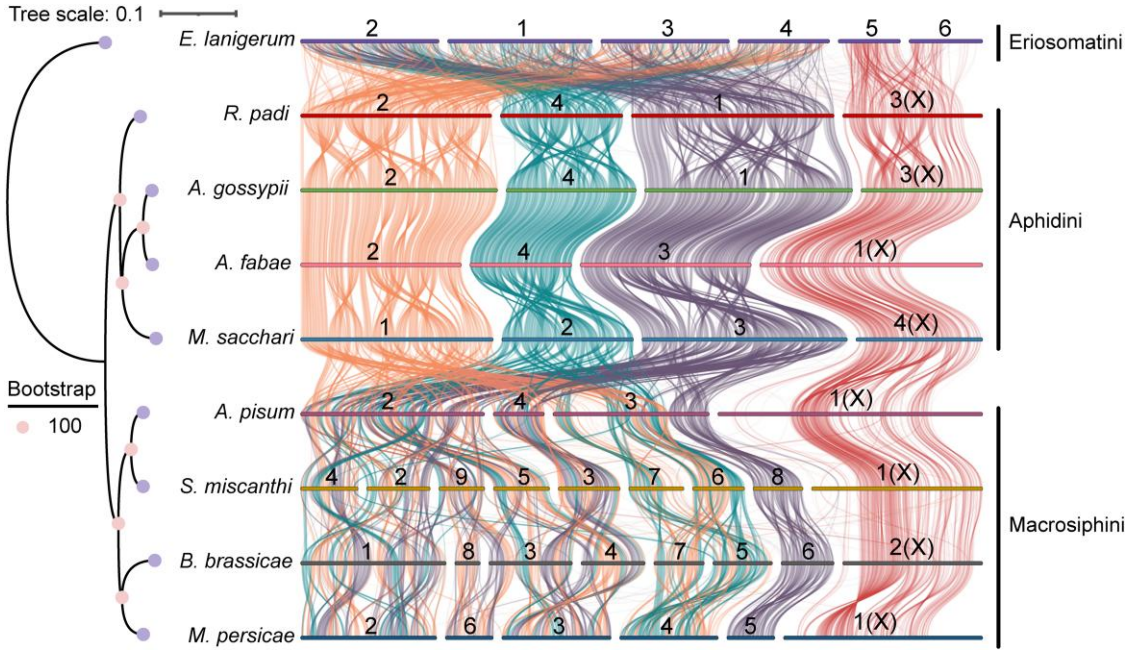

Fig. 3. A phylogenetic and chromosomal synteny of single-copy orthologous genes among the aphid tribes Macrosiphini, Aphidini, and Eriosomatini. The left side shows a phylogenetic tree of 9 aphid species. Each line represents a single-copy gene ($n = 6{,}253$), and the line color is referenced by *M. sacchari*. The number indicates the chromosome numbering of the aphid, with "X" used to denote the sex chromosome.

These findings suggest that chromosome fission and fusion events occur less frequently in tribe Aphidini compared to the tribe Macrosiphini, although they diverged within a similar divergence time.

Discussion

In this study, we assembled a chromosome-level reference genome for the diploid *M. sacchari* with a genome size of 304.6 Mb and a contig N50 of 12.5 Mb. The primary haplotype assembly represents high continuity, accuracy, and integrity compared to the previously released *M. sacchari* genome in the NCBI. Phylogenomic analyses have indicated that *M. sacchari* is a member of the tribe Aphidini and is closely related to the genus Aphis, which includes several agricultural pests, such as *A. gossypii*, *A. fabae*, *A. thalictri*, and *A. glycines*. Therefore, the availability of a high-quality genome of *M. sacchari* will facilitate studying of these important agricultural pests.

In the assembly, we identified a scaffold5, which lacks significant global interaction signals with other chromosomes on the Hi-C interaction map, suggesting that it may be an extra, supernumerary chromosome segment. This is reminiscent of the B chromosomes first discovered in the somatic cells of Hemiptera (Wilson 1907). Scaffold5 exhibits homologous genes and syntenic regions with the other 4 *M. sacchari* chromosomes and the genomes of 8 other aphid species, with these genes and regions predominantly localized to the arms of X chromosome. This observation is analogous to the findings in cichlid fish, where B chromosome sequences were found to be homologous to A chromosome sequences (Ramos *et al.* 2017; Clark *et al.* 2018). In the grasshopper *Eyprepocnemis plorans* and the rodent group Oryzomyini, sex chromosomes appear to be involved in the origin of B chromosomes (Lopez-Leon *et al.* 1994; Ventura *et al.* 2015). The proportion of TEs in scaffold5 is significantly higher than the average levels observed in other scaffolds and chromosomes, yet the types of TEs are fundamentally the same as those found in the A chromosomes. Based on cytological staining patterns, similar phenomena have been reported in animals, plants, and fungi, where B chromosomes exhibit heterochromatic characteristics (Ma *et al.* 2010; Ventura *et al.* 2015). In maize, B chromosome protein-coding gene homologs are widely dispersed across the 10 A chromosomes, without detectable syntenic gene regions of the B chromosome, indicating a high degree of heterogeneity in this region (Blavet *et al.* 2021). However, the absence of clear Hi-C signals presents a challenge in determining whether scaffold5 is indeed a B chromosome. Despite this, we believe that the unique characteristics of scaffold5 could provide valuable material for the study of B chromosomes. In the future, the use of FISH may help us directly observe the specific location of scaffold5 within cells, offering direct evidence for its chromosomal status.

High rates of autosomal chromosome rearrangement have been reported in aphids, such as the formation of *A. pisum* chromosome 3 through a fusion event involving homologues of *M. persicae* chromosomes 4 and 5 (Mathers *et al.* 2021). *A. pisum* chromosome 2 is homologous to 4 small chromosomes in *S. miscanthi* (Mathers *et al.* 2023). In this study, we observed a lower occurrence of inter-chromosome rearrangement events in tribe Aphidini within the same divergence time. A karyotype analysis indicated that most Aphidini species have 4 chromosomes ($2n = 8$), with very few species reaching the highest chromosome number within this tribe of 6 ($2n = 12$) (Blackman 1980). By contrast, the karyotypes of Macrosiphini show a broader range, from $2n = 4$ to $2n = 72$. The extensive variability of karyotypes in aphids suggests that different evolutionary forces act on genome evolution in these species, which warrant further investigations. The assembly of the *M. sacchari* chromosome will provide further insights into the intricate evolutionary history of aphid genomes.

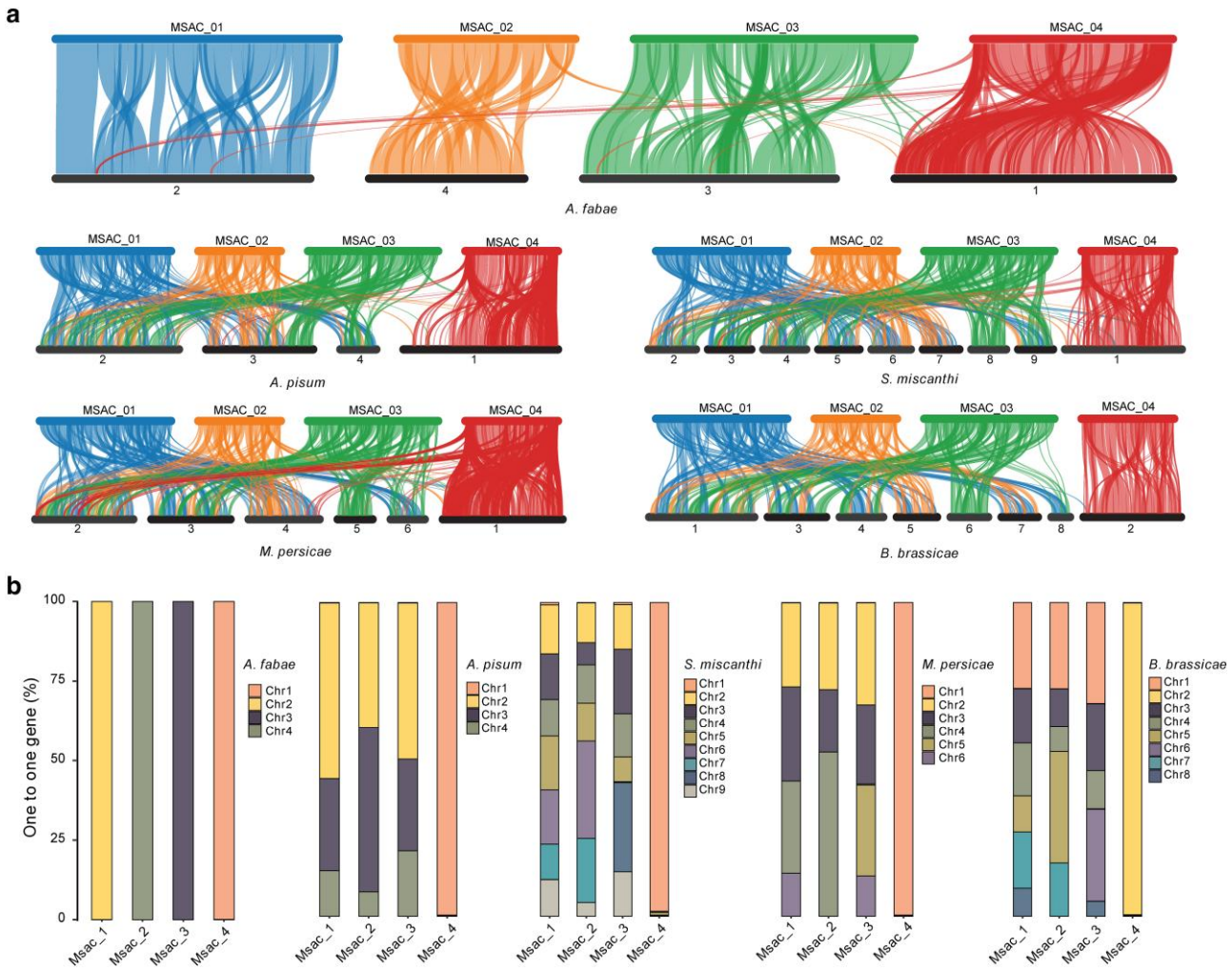

Fig. 4. A pairwise synteny analysis of *M. sacchari* and 5 aphids of the tribe Aphidini and Macrosiphini. a) A synteny analysis of *M. sacchari* and other 5 species of aphids (Aphidini: *A. fabae* and Macrosiphini: *A. pisum*, *S. miscanthi*, *B. brassicae*, and *M. persicae*). The plot shows blocks of syntenic genes. b) A homology of single-copy genes from *M. sacchari* and other 5 species of the tribe Aphidini and Macrosiphini of aphids at the chromosome level. The x-axis represents the chromosome on which the *M. sacchari* gene is located. The y-axis represents the proportion of genes located on homologous chromosomes that exhibit synteny with another aphid.

Data availability

The raw sequences of nanopore ultra-long reads (SRR17399617; SRR17399618), whole-genome sequence short reads (SRR17399616), RNA-seq reads (SRR22746183), and Hi-C reads (SRR21203420; SRR21203421) have been deposited in the NCBI SRA (BioProject accession no. PRJNA792680). The genome assembly has been deposited in the NCBI Genome (JBCITD000000000). The assembled genome sequences and gene and TE annotations and the scripts used for analyses are available on Zenodo (https://doi.org/10.5281/zenodo.13283872). All study data are included in the main article and Supplementary Materials.

Supplemental material available at G3 online.

Acknowledgments

The authors thank Dr Yang Liu and Jianfei Shi from the Institute of Zoology, Chinese Academy of Sciences, for their efforts in processing the sequencing data.

Funding

This work was supported by the National Key Research and Development Program of China (grant no. 2023YFD1400800), the Open Research Fund Program of State Key Laboratory of Integrated Management of Pest Insects and Rodents (grant no. IPM2108), and the Initiative Scientific Research Program of Institute of Zoology, Chinese Academy of Sciences (grant no. 2023IOZ0203).

Conflicts of interest

The author(s) declare no conflicts of interest.

Literature cited

Ahmad YA, Costet L, Daugrois JH, Nibouche S, Letourmy P, Girard JC, Rott P. 2007. Variation in infection capacity and in virulence exists between genotypes of sugarcane yellow leaf virus. Plant Dis. 91(3):253–259. doi:10.1094/PDIS-91-3-0253.

Baril T, Pym A, Bass C, Hayward A. 2023. Transposon accumulation at xenobiotic gene family loci in aphids. Genome Res. 33(10): 1718–1733. doi:10.1101/gr.277820.123.

Biello R, Singh A, Godfrey CJ, Fernández FF, Mugford ST, Powell G, Hogenhout SA, Mathers TC. 2021. A chromosome-level genome assembly of the woolly apple aphid, Eriosoma lanigerum

Hausmann (Hemiptera: Aphididae). Mol Ecol Resour. 21(1): 316–326. doi:10.1111/1755-0998.13258.
Blackman RL. 1980. Chromosome numbers in the Aphididae and its taxonomic significance. Syst Entomol. 5(1):7–25. doi:10.1111/j.1365-3113.1980.tb00393.x.
Blackman RL, Eastop VF. 1984. Aphids on the World's Crops: An Identification and Information Guide. London: Wiley. p. 476.
Blackman R, Eastop V. 2017. Taxonomic issues. In: van Emden HF, Harrington R, editors. Aphids as Crop Pests. 2nd ed. Wallingford: CAB International. p. 1–36.
Blavet N, Yang H, Su H, Solansky P, Douglas RN, Karafiatova M, Simkova L, Zhang J, Liu Y, Hou J, *et al.* 2021. Sequence of the supernumerary B chromosome of maize provides insight into its drive mechanism and evolution. Proc Natl Acad Sci U S A. 118(23): e2104254118. doi:10.1073/pnas.2104254118.
Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30(15):2114–2120. doi:10.1093/bioinformatics/btu170.
Bowling RD, Brewer MJ, Kerns DL, Gordy J, Seiter N, Elliott NE, Buntin GD, Way MO, Royer TA, Biles S, *et al.* 2016. Sugarcane aphid (Hemiptera: Aphididae): a new pest on sorghum in North America. J Integr Pest Manag. 7(1):12. doi:10.1093/jipm/pmw011.
Brewer MJ, Gordy JW, Kerns DL, Woolley JB, Rooney WL, Bowling RD. 2017. Sugarcane aphid population growth, plant injury, and natural enemies on selected grain sorghum hybrids in Texas and Louisiana. J Econ Entomol. 110(5):2109–2118. doi:10.1093/jee/tox204.
Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 17(4):540–552. doi:10.1093/oxfordjournals.molbev.a026334.
Chaisson MJ, Huddleston J, Dennis MY, Sudmant PH, Malig M, Hormozdiari F, Antonacci F, Surti U, Sandstrom R, Boitano M, *et al.* 2015. Resolving the complexity of the human genome using single-molecule sequencing. Nature. 517(7536):608–611. doi:10.1038/nature13907.
Chakraborty M, VanKuren NW, Zhao R, Zhang X, Kalsow S, Emerson JJ. 2018. Hidden genetic variation shapes the structure of functional elements in Drosophila. Nat Genet. 50(1):20–25. doi:10.1038/s41588-017-0010-y.
Chalopin D, Naville M, Plard F, Galiana D, Volff J-N. 2015. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. Genome Biol Evol. 7(2):567–580. doi:10.1093/gbe/evv005.
Chen W, Shakir S, Bigham M, Richter A, Fei Z, Jander G. 2019. Genome sequence of the corn leaf aphid (Rhopalosiphum maidis Fitch). Gigascience. 8(4):giz033. doi:10.1093/gigascience/giz033.
Clark FE, Conte MA, Kocher TD. 2018. Genomic characterization of a B chromosome in Lake Malawi Cichlid fishes. Genes (Basel). 9(12): 610. doi:10.3390/genes9120610.
Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. 2016. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. 3(1):99–101. doi:10.1016/j.cels.2015.07.012.
Exilien R, Brodeur J, Fournier V, Martini X. 2022. Host range and phenology of sugarcane aphid (Hemiptera: Aphididae) and natural enemy community in Sorghum in Haiti. J Econ Entomol. 115(6):1956–1963. doi:10.1093/jee/toac173.
Ferry N, Edwards MG, Gatehouse JA, Gatehouse AM. 2004. Plant-insect interactions: molecular approaches to insect resistance. Curr Opin Biotechnol. 15(2):155–161. doi:10.1016/j.copbio.2004.01.008.
Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 117(17):9451–9457. doi:10.1073/pnas.1921046117.
Gadhave KR, Gautam S, Rasmussen DA, Srinivasan R. 2020. Aphid transmission of *Potyvirus*: the largest plant-infecting RNA virus genus. Viruses. 12(7):773. doi:10.3390/v12070773.
Gray S, Gildow FE. 2003. Luteovirus-aphid interactions. Annu Rev Phytopathol. 41(1):539–566. doi:10.1146/annurev.phyto.41.012203.105815.
Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol. 9(1):R7. doi:10.1186/gb-2008-9-1-r7.
Hawthorne DJ, Via S. 2001. Genetic linkage of ecological specialization and reproductive isolation in pea aphids. Nature. 412(6850):904–907. doi:10.1038/35091062.
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(5): 767–769. doi:10.1093/bioinformatics/btv661.
Hogenhout SA, Ammar E-D, Whitfield AE, Redinbaugh MG. 2008. Insect vector interactions with persistently transmitted viruses. Annu Rev Phytopathol. 46(1):327–359. doi:10.1146/annurev.phyto.022508.092135.
Hogenhout SA, Bos JI. 2011. Effector proteins that modulate plant-insect interactions. Curr Opin Plant Biol. 14(4):422–428. doi:10.1016/j.pbi.2011.05.003.
Hu J, Wang Z, Sun Z, Hu B, Ayoola AO, Liang F, Li J, Sandoval JR, Cooper DN, Ye K, *et al.* 2024. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. Genome Biol. 25(1):107. doi:10.1186/s13059-024-03252-4.
Jaquiery J, Peccoud J, Ouisse T, Legeai F, Prunier-Leterme N, Gouin A, Nouhaud P, Brisson JA, Bickel R, Purandare S, *et al.* 2018. Disentangling the causes for faster-X evolution in aphids. Genome Biol Evol. 10(2):507–520. doi:10.1093/gbe/evy015.
Jaquiery J, Rispe C, Roze D, Legeai F, Le Trionnaire G, Stoeckel S, Mieuzet L, Da Silva C, Poulain J, Prunier-Leterme N, *et al.* 2013. Masculinization of the X chromosome in the pea aphid. PLoS Genet. 9(8):e1003690. doi:10.1371/journal.pgen.1003690.
Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 14(6):587–589. doi:10.1038/nmeth.4285.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780. doi:10.1093/molbev/mst010.
Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 37(8):907–915. doi:10.1038/s41587-019-0201-4.
Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 16(2):111–120. doi:10.1007/bf01731581.
Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, Stecher G, Hedges SB. 2022. TimeTree 5: an expanded resource for species divergence times. Mol Biol Evol. 39(8):msac174. doi:10.1093/molbev/msac174.
Li H. 2023. Protein-to-genome alignment with miniprot. Bioinformatics. 39(1):btad014. doi:10.1093/bioinformatics/btad014.
Li G, Huang C, Ji B, Shi Z, Wang J, Yuan J, Yang P, Xu X, Jing H, Xu L, *et al.* 2024. A comparative genomic analysis at the chromosomal-level reveals evolutionary patterns of aphid chromosomes. PREPRINT (Version 1) available at Research Square. doi:10.21203/rs.3.rs-4737612/v1.

Lopez-Leon MD, Neves N, Schwarzacher T, Heslop-Harrison JS, Hewitt GM, Camacho JPM. 1994. Possible origin of a B chromosome deduced from its DNA composition using double FISH technique. Chromosome Res. 2(2):87–92. doi:10.1007/bf01553487.

Ma LJ, van der Does HC, Borkovich KA, Coleman JJ, Daboussi MJ, Di Pietro A, Dufresne M, Freitag M, Grabherr M, Henrissat B, *et al.* 2010. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. Nature. 464(7287):367–373. doi:10.1038/nature08850.

Manni M, Berkeley MR, Seppey M, Zdobnov EM. 2021. BUSCO: assessing genomic data quality and beyond. Curr Protoc. 1(12):e323. doi:10.1002/cpz1.323.

Marcais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. Bioinformatics. 27(6):764–770. doi:10.1093/bioinformatics/btr011.

Mathers TC. 2020. Improved genome assembly and annotation of the soybean aphid (*Aphis glycines* Matsumura). G3 (Bethesda). 10(3):899–906. doi:10.1534/g3.119.400954.

Mathers TC, Mugford ST, Hogenhout SA, Tripathi L. 2020. Genome sequence of the banana aphid, *Pentalonia nigronervosa* Coquerel (Hemiptera: Aphididae) and its symbionts. G3 (Bethesda). 10(12):4315–4321. doi:10.1534/g3.120.401358.

Mathers TC, Mugford ST, Wouters RHM, Heavens D, Botha A-M, Swarbreck D, Van Oosterhout C, Hogenhout SA. 2022. Zenodo. doi:10.5281/zenodo.5908005.

Mathers TC, Wouters RHM, Mugford ST, Biello R, van Oosterhout C, Hogenhout SA. 2023. Hybridisation has shaped a recent radiation of grass-feeding aphids. BMC Biol. 21(1):157. doi:10.1186/s12915-023-01649-4.

Mathers TC, Wouters RHM, Mugford ST, Swarbreck D, van Oosterhout C, Hogenhout SA. 2021. Chromosome-scale genome assemblies of aphids reveal extensively rearranged autosomes and long-term conservation of the X chromosome. Mol Biol Evol. 38(3):856–875. doi:10.1093/molbev/msaa246.

Neupane SB, Kerns DL, Szczepaniec A. 2020. The impact of sorghum growth stage and resistance on life history of sugarcane aphids (Hemiptera: Aphididae). J Econ Entomol. 113(2):787–792. doi:10.1093/jee/toz310.

Ng JCK, Falk BW. 2006. Virus-vector interactions mediating nonpersistent and semipersistent transmission of plant viruses. Annu Rev Phytopathol. 44(1):183–212. doi:10.1146/annurev.phyto.44.070505.143325.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 32(1):268–274. doi:10.1093/molbev/msu300.

Peccoud J, Ollivier A, Plantegenest M, Simon JC. 2009. A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. Proc Natl Acad Sci U S A. 106(18):7495–7500. doi:10.1073/pnas.0811117106.

Pekarcik AJ, Jacobson AL. 2021. Evaluating sugarcane aphid, *Melanaphis sacchari* (Hemiptera: Aphididae), population dynamics, feeding injury, and grain yield among commercial sorghum varieties in Alabama. J Econ Entomol. 114(2):757–768. doi:10.1093/jee/toab013.

Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 33(3):290–295. doi:10.1038/nbt.3122.

Ramos E, Cardoso AL, Brown J, Marques DF, Fantinatti BE, Cabral-de-Mello DC, Oliveira RA, O'Neill RJ, Martins C. 2017. The repetitive DNA element BncDNA, enriched in the B chromosome of the cichlid fish *Astatotilapia latifasciata*, transcribes a potentially noncoding RNA. Chromosoma. 126(2):313–323. doi:10.1007/s00412-016-0601-x.

Setokuchi O, Muta T. 1993. Ecology of aphids on sugarcane III. Relationship between alighting of aphid vectors of sugarcane mosaic virus and infecting in fields. Jpn J Appl Entomol Zool. 37(1):11–16. doi:10.1303/jjaez.37.11.

Singh BU, Padmaja PG, Seetharama N. 2004. Biology and management of the sugarcane aphid, *Melanaphis sacchari* (Zehntner) (Homoptera: Aphididae), in sorghum: a review. Crop Prot. 23(9):739–755. doi:10.1016/j.cropro.2004.01.004.

Tang H, Krishnakumar V, Zeng X, Xu Z, Taranto A, Lomas JS, Zhang Y, Huang Y, Wang Y, Yim WC, *et al.* 2024. JCVI: a versatile toolkit for comparative genomics analysis. iMeta. 3(4):e211. doi:10.1002/imt2.211.

The International Aphid Genomics Consortium. 2010. Genome sequence of the pea aphid Acyrthosiphon pisum. PLoS Biol. 8(2):e1000313. doi:10.1371/journal.pbio.1000313.

Ventura K, O'Brien PC, do Nascimento Moreira C, Yonenaga-Yassuda Y, Ferguson-Smith MA. 2015. On the origin and evolution of the extant system of B chromosomes in Oryzomyini radiation (Rodentia, Sigmodontinae). PLoS One. 10(8):e0136663. doi:10.1371/journal.pone.0136663.

Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC. 2017. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics. 33(14):2202–2204. doi:10.1093/bioinformatics/btx153.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, *et al.* 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 9(11):e112963. doi:10.1371/journal.pone.0112963.

Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, *et al.* 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 40(7):e49. doi:10.1093/nar/gkr1293.

Wilson EB. 1907. The supernumerary chromosomes of Hemiptera. Science. 26:870–871. https://cir.nii.ac.jp/crid/1572261550039895808.

Yang CR. 1986. Transmission of sugarcane mosaic virus by three kinds of aphids. Chin J Entomol. 6:43–49.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24(8):1586–1591. doi:10.1093/molbev/msm088.

Zhang X, Zhang S, Zhao Q, Ming R, Tang H. 2019. Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. Nat Plants. 5(8):833–845. doi:10.1038/s41477-019-0487-8.

Editor: M. Sachs