A high-quality chromosome-level genome assembly provides insights into wing dimorphism in *Metopolophium dirhodum* (Walker)

Bin Zhu[1], Rui Wei[1], WenJuan Hua[1], Wenlin Zhang[2], Pei Liang[3], and Xiwu Gao[1]

[1]China Agricultural University College of Plant Protection
[2]Berry Genomics Corporation
[3]China Agricultural University

April 05, 2024

Abstract

The rose-grain aphid *Metopolophium dirhodum* (Walker) (Hemiptera: Aphididae) is one of the most important aphid pests for cereals worldwide. Some studies have examined the biological and ecological characteristics of *M. dirhodum*. However, the lack of genomic data limits in-depth studies of this organism. Here, we present a chromosome-level genome assembly of *M. dirhodum* using PacBio long HiFi reads and Hi-C technology. The final genome assembly is 447.8 Mb, with 98.50% of the assembled sequences anchored to nine chromosomes. The contig and scaffold N50 values are 7.82 Mb and 37.54 Mb, respectively. A total of 18,003 protein-coding genes were predicted, of which 92.05% were functionally annotated. Comparative transcriptomic analyses identified a number of genes that might be related to wing dimorphism, including the insulin receptor, insulin receptor substrate, forkhead box protein O (Foxo), and ecdysone receptor. These results may provide an important reference for understanding the ecology, genetics, and evolution of this organism or even other aphid insects.

A high-quality chromosome-level genome assembly provides insights into wing dimorphism in *Metopolophium dirhodum* (Walker)

Bin Zhu[1], Rui Wei[1], Wenjuan Hua[1], Wenlin Zhang[2], Pei Liang[1*], Xiwu Gao[1*]

[1]Department of Entomology, China Agricultural University, Beijing, 100193, P. R. China

[2]Berry Genomics Corporation, Beijing, 100015, P. R. China

Correspondence:

Prof. Pei Liang

Department of Entomology

China Agricultural University

2 YuanmingyuanWest Road

Beijing 100193, China

Phone: (86) 10-6273 1306

Fax: (86) 10-6273 1306

Email: *liangcau@cau.edu.cn*

Prof. Xiwu Gao

Department of Entomology

China Agricultural University

2 YuanmingyuanWest Road

Beijing 100193, China

Phone: (86) 10-6273 2974

Fax: (86) 10-6273 2974

Email: *gaoxiwu@263.net.cn*

ABSTRACT

The rose-grain aphid *Metopolophium dirhodum* (Walker) (Hemiptera: Aphididae) is one of the most important aphid pests for cereals worldwide. Some studies have examined the biological and ecological characteristics of *M. dirhodum* . However, the lack of genomic data limits in-depth studies of this organism. Here, we present a chromosome-level genome assembly of *M. dirhodum* using PacBio long HiFi reads and Hi-C technology. The final genome assembly is 447.8 Mb, with 98.50% of the assembled sequences anchored to nine chromosomes. The contig and scaffold N50 values are 7.82 Mb and 37.54 Mb, respectively. A total of 18,003 protein-coding genes were predicted, of which 92.05% were functionally annotated. Comparative transcriptomic analyses identified a number of genes that might be related to wing dimorphism, including the insulin receptor, insulin receptor substrate, forkhead box protein O (Foxo), and ecdysone receptor. These results may provide an important reference for understanding the ecology, genetics, and evolution of this organism or even other aphid insects.

KEYWORDS : chromosome-level genome, *Metopolophium dirhodum* , wing dimorphism, Aphidinae

1 INTRODUCTION

The rose-grain aphid *Metopolophium dirhodum* (Walker) (Hemiptera: Aphididae) is one of the most common and economically important aphid pests for cereals, including wheat, barley, rye and oat, worldwide (Cannon, 1986; Honek, 1991; Ma et al., 2004; Li et al., 2020). *M. dirhodum* is native in the Holarctic and was then introduced to North America, South America, South Africa, Europe, Oceania and Eastern Asia (Honek et al., 2018; Blackman and Eastop, 2000). In the continental climate of central Europe, *M. dirhodum* is usually the most abundant aphid species on cereals (Honek et al., 1987, 2018; Praslicka et al., 1996).

Damage caused to cereals by *M. dirhodum* takes several forms. These include sucking the juice from wheat leaves, stems and young ears, which further results in the deterioration of plant nutrition (Holt et al., 1984), defecating sticky honeydew that further obstructs photosynthesis and reduces wheat quality (Jiang et al., 2019), and transmission of a number of pathogenic plant viruses, including the barley yellow dwarf virus (Kennedy et al., 2005). Both nymphs and adults of this aphid may cause yield losses of 27–30% during the latter part of flowering stages of wheat (Holt et al., 1984; Chopa et al., 2012).

Recently, several studies have focused on the population dynamics (Honek et al., 2018), control methods (Cambier et al., 2001; Chopa et al., 2012) and symbiotic microorganisms (Telesnicki et al., 2012) of *M. dirhodum* . However, the challenge to developing a further understanding is the lack of genomic resources for this pest. To date, chromosome-level genomes with annotation information for several aphid species, including *Acyrthosiphon pisum* (International Aphid Genomics Consortium., 2010), *Sitobion miscanthi* (Jiang et al., 2019), *Eriosoma lanigerum* (Mathers et al., 2021) and *Rhopalosiphum maidis* (Chen et al., 2019), are available, which will be very helpful for the further study of these aphids. Here, we report the genome sequence of *M. dirhodum*assembled by incorporating Pacific Biosciences (PacBio) long HiFi reads and Hi-C technology. Subsequently, gene prediction, functional annotation and phylogenetic analysis were also performed. The

genomic resource developed here for *M. dirhodum* is valuable for understanding its genetics, development and evolution and will provide an important reference for the study of other insect genomes.

M. dirhodum , like most aphids, can produce wing morphs when experiencing crowding, poor nutrition and temperature or photoperiod changes (Müller et al., 2001; Braendle et al., 2006; Zhang et al., 2019). Wing dimorphism in insects is an adaptive switch to environmental changes. Specifically, wingless morphs allocate more resources to reproduction, enabling rapid colony growth. Winged morphs are devoted to dispersal, which enables them to look for new habitats and food resources. Moreover, winged morphs are better at long-distance migration and host alternation, thus causing more serious host damage and virus transmission (Zhang et al., 2019). Therefore, understanding the molecular mechanisms of *M. dirhodum*in triggering wing dimorphism is important for its effective control. In the present study, we constructed and sequenced RNA libraries of winged and wingless *M. dirhodum* from third- and fourth-instar nymphs and adults. Differentially expressed genes between the winged and wingless populations were analyzed and identified according to the current assembly of the *M. dirhodum* genome. The results lay a solid foundation for further study of the molecular mechanisms of wing dimorphism in *M. dirhodum* or other insects.

2 MATERIALS AND METHODS

2.1 Insects

The *Metopolophium dirhodum* population used in the present study was originally collected from Langfang, Hebei Province, China in 2018 and then reared on wheat seedlings in our laboratory, which was maintained at 22 +- 2 degC and 60% relative humidity with a 16 h light:8 h dark cycle, for more than 2 years.

2.2 Sample preparation, library construction and sequencing

An isogenic colony was started from a single parthenogenetic female of*M. dirhodum* and was maintained alone on wheat seedlings prior to the collection of insects for sequencing. Two hundred milligrams of fresh mixed *M. dirhodum*(including first- to fourth-instar nymphs and winged and wingless adults) were collected for DNA extraction and genome sequencing. Total genomic DNA was extracted using a Blood & Cell Culture DNA Mini Kit according to the manufacturer's protocol (Qiagen, Hilden, Germany). For short-read sequencing, a paired-end library (2x150 bp) with short insert sizes of approximately 500 bp was constructed using the VAHTSTM Universal DNA Library Prep Kit for Illumina V2 (Vazyme, Nanning, China) and then sequenced on an Illumina NovaSeq 6000 platform (San Diego, CA, USA). For long-read genomic sequencing, the PacBio SMRTbell 15 kb library was constructed using a SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, CA, USA) and then sequenced on the PacBio Sequel II platform for circular consensus sequencing (CCS) (Pacific Biosciences, CA, USA).

To assist chromosome-level assembly, the Hi-C (high-throughput chromosome conformation capture) technique was applied to capture genome-wide chromatin interactions. Approximately 200 mg of fresh mixed*M. dirhodum* (including first- to fourth-instar nymphs and winged and wingless adults) was ground in 2% formaldehyde to allow cross-linking of cellular protein, and approximately 100 μg of DNA was extracted. Subsequently, chromosome integrity and cross-linked protein residues were assessed. Chromatin digestion was performed with the restriction enzyme *Mbo* I. Biotinylated residues were added during repair of the sticky ends, and the resulting blunt-end fragments were ligated under dilute conditions (Lieberman-Aiden et al., 2009; Belton et al., 2012; Rao et al., 2014; Belaghzal et al., 2017; Pan et al., 2021). The DNA was extracted and randomly sheared to fragments of 300–500 bp. The biotin-labeled fragments were isolated with magnetic beads. The next four steps, including end repair, dA tailing, adapter ligation and DNA purification, were accomplished by adding the corresponding reaction components sequentially. The library quantity was estimated by Qubit 2.0, an Agilent 2100 instrument (Agilent Technologies, Santa Clara, CA, USA), and quantitative PCR. The Hi-C library was then sequenced using the Illumina NovaSeq 6000 platform with paired-end 150 bp reads.

For PacBio full-length transcriptome sequencing, total RNA was isolated from fresh mixed*M. dirhodum* (including first- to fourth-instar nymphs and winged and wingless adults of equal quality) using an EASYspin

Plus Cell/Tissue RNA Isolation Kit (Aidlab Biotechnologies, Beijing, China) and quantified using a NanoDrop ND-2000 spectrophotometer (NanoDrop products, Wilmington, DE, USA). Ten micrograms of total RNA wee reverse transcribed into cDNA using a SMARTer PCR cDNA Synthesis Kit (Takara, Dalian, China) following the manufacturer's protocols. The SMRT library was constructed using the SMRTbell template prep kit (Takara) following the manufacturer's protocols. The library was sequenced on the PacBio Sequel II platform, and SMRTlink was used to obtain full-length consensus isoform sequences.

For Illumina transcriptome sequencing, total RNA was isolated from winged or wingless *M. dirhodum* of third- and fourth-instar nymphs and adults of equal quality using an EASYspin Plus Cell/Tissue RNA Isolation Kit (Aidlab Biotechnologies, Beijing, China) and then quantified using a NanoDrop ND-2000 spectrophotometer. cDNA libraries were constructed using a VAHTSTM mRNA-seq V3 Library Prep Kit (Vazyme, Nanjing, China). A total of 18 libraries were constructed with three biological replicates per sample. Sequencing was performed on an Illumina NovaSeq instrument (Illumina, San Diego, CA, USA), and 150 bp paired-end reads were generated.

2.3 Genome survey and assembly

The K-mer distribution was analyzed to estimate the genome size, heterozygosity, and repeat content using Illumina paired-end reads. The K-mer distribution was analyzed using the Jellyfish and GenomeScope tools based on a k value of 17 (Vurture et al., 2017).

PacBio subreads were obtained from the raw polymerase reads after removal of short and low-quality reads and the adaptor sequences, which were then filtered and corrected using the pbccs pipeline with default parameters (*https://github.com/PacificBiosciences/ccs*). The resulting HiFi reads (high-fidelity reads) were subjected to hifiasm for de novo assembly (*https://github.com/chhylp123/hifiasm*). BWA v0.7.15 (*https://sourceforge.net/projects/bio-bwa/files/*) (Li, 2013) and SAMtools v1.4 (*https://sourceforge.net/projects/samtools/files/samtools/*) (Li et al., 2009) were used for read alignment and SAM/BAM format conversion. Genome assembly and completeness were assessed using the conserved genes in BUSCO v3.0.2 (*https://busco.ezlab.org/*) (Simao et al., 2015).

2.4 Chromosome assembly using Hi-C

The Hi-C sequence data were aligned against the draft genome using JUICER v1.6.2 (*https://github.com/aidenlab/juicer*) (Durand et al., 2016). The uniquely mapped sequences were analyzed using 3D-DNA software (*https://github.com/theaidenlab/3d-dna*) to assist genomic assembly (Dudchenko et al., 2017). The algorithms "misjoin" and "scaffolding" were used to remove the misjoins and obtain scaffolds at the chromosomal level. The algorithm "seal" was employed to find the scaffolds that had been incorrectly removed by the "misjoin". A heatmap of chromosome interactions was constructed to visualize the contact intensity among chromosomes using JUICER v1.6.2.

2.5 Genome Annotation

Tandem repeats and interspersed repeats were identified using Tandem Repeats Finder (TRF) v4.09 (*http://tandem.bu.edu/trf/trf407b.linux64.download.html*) (Benson, 1999) and RepeatModeler v2.0 (*http://www.repeatmasker.RepeatModeler/*) (Flynn et al., 2020), respectively. RepeatMasker v4.1.0 (*http://www.repeatmasker.org/RMDownload*) was used to mask the predicted and known repeated sequences (Tarailo-Graovac and Chen, 2009). tRNAscan-SE v1.4alpha (Chan et al., 2019) was used to predict tRNAs, and Infernal v1.1.3 (*http://eddylab.org/*) was used to search the Rfam database v11.0 with an E-value cutoff of 10^{-5} to identify other types of noncoding RNAs (ncRNAs). (Nawrocki et al., 2013).

Protein-coding genes were predicted through the combination of homology-based, RNA sequencing-based, and ab initio predictions. For the homolog-based approach, the protein sequences of several related species, including *A. pisum* (International Aphid Genomics Consortium., 2010), *R. maidis* (Chen et al., 2019),*Diuraphis noxia* (Nicholson et al., 2015), *Aphis gossypii*(Quan et al., 2019), *Aphis glycines* (Wenger et al., 2020) and*Myzus persicae* (Mathers et al., 2017), were downloaded from NCBI and aligned against the assembled genome using Gene Model Mapper (GeMoMa) v1.6.1.jar

(*http://www.jstacs.de/index.php/GeMoMa*) (Keilwagen et al., 2016) to refine the blast hits to define exact intron/exon positions. For the RNA sequencing-based method, the PacBio full-length transcriptome, which was obtained from the pooled sample of*M. dirhodum* , was used to predict the open reading frames (ORFs) with PASA (*https://sourceforge.net/projects/pasa/files/stats/timeline*) (Campbell et al., 2006) using default settings. For the ab initio method, two de novo programs, Augustus v3.2.2 (*http://augustus.gobics.de/binaries/*) (Stanke and Waack, 2003) and SNAP (*http://snap.stanford.edu/snap/download.html*) (Korf, 2004), were employed with default parameters to predict genes in the repeat-masked genome sequences. All predicted genes from the three approaches were integrated with EVidenceModeler (EVM) (*https://sourceforge.net/projects/evidencemodeler/*) (Haas et al., 2008) to generate high-confidence gene sets, and the untranslated regions and alternative splicing were predicted using PASA.

The gene set was annotated by aligning protein sequences to functional databases, including NR (nonredundant sequence database) (Deng et al., 2006), Swiss-Prot (Bairoch & Boeckmann, 1991), eggNOG (evolutionary genealogy of genes: Nonsupervised Orthologous Groups) (Huerta-Cepas et al., 2019), GO (Gene Ontology) (Dimmer et al., 2012) and KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa and Goto, 2000), using BLAST with a threshold e-value [?] 1e-5.

2.6 Phylogeny and comparative genomics

Orthologous groups were identified using the OrthoMCL pipeline (*https://orthomcl.org/orthomcl/*) (Li et al., 2003) with default parameters for *M. dirhodum* and nine other species, including*A. pisum* (International Aphid Genomics Consortium., 2010), *R. maidis* (Chen et al., 2019),*D. noxia* (Nicholson et al., 2015), *M. persicae* (Mathers et al., 2017), *Melanaphis sacchari* (GCA_002803265.2),*Nilaparvata lugens* (Ma et al., 2021), *Bemisia tabaci* (Xie et al., 2017; Chen et al., 2019) and *Apolygus lucorum* (Liu et al., 2021). *Drosophila melanogaster* (Hoskins et al., 2015) was used as an outgroup. MAFFT (*https://mafft.cbrc.jp/alignment/software/*) (Katoh & Standley, 2013) was used to align each orthologous gene sequence with default parameters. RAxML (Stamatakis, 2014) was used to infer the maximum-likelihood tree with the best-fit substitution model and 1000 bootstrap replicates. Mummer (*https://github.com/mummer4/mummer*) was applied for a detailed collinearity analysis between the *M. dirhodum* and *A. pisum* genomes.

2.7 Transcriptome analysis

Raw data (raw reads) in FASTQ format were first processed through primary quality control. In this step, clean data (clean reads) were obtained by removing reads containing adapters, reads containing poly-N, low-quality reads (lower than 5) and contaminants from the raw data. All downstream analyses were based on clean data of a high quality. Paired-end clean reads were aligned to the assembled genome of *M. dirhodum* using TopHat with default parameters. Differential expression analysis was performed using the DEGSeq R package (1.20.0) (Wang et al., 2010). We used the adjusted *P value* (padj) [?] 0.001 and a |log (fold-change)| [?] 2 as the criteria for the significant difference in expression.

3 Results

3.1 Chromosomal-level de novo genome of *M. dirhodum*

A total of 41.17 Gb of high-quality paired-end reads were obtained by Illumina genomic sequencing (˜92.22X coverage, Table S1). The genome size of *M. dirhodum* was estimated to be 457.2 M based on k -mer counting. The k -mer distribution analysis revealed a peak at 79.8× of the sequencing depth, suggesting a moderate level of heterozygosity (0.445%) and highly repetitive sequence content (59.20%) in the genome (Fig. 1A). To obtain a reference genome for*M. dirhodum* , we generated 161.53 Gb of PacBio long reads using the CCS model (Table S1), which were subsequently corrected to 10.34 Gb HiFi reads. The genome was initially assembled using hifiasm, resulting in 296 contigs with a contig N50 of 7.82 Mb and the longest contig of 23.64 Mb (Table 1). A total of 41.17 Gb of short reads generated by the Illumina NovaSeq 6000 platform was then mapped against our assembly, resulting in a mapping rate of 92.18%. BUSCO analysis showed that

96.9% (single-copied genes: 92.5%; duplicated genes: 4.4%) of 1,367 single-copy genes in the insecta_odb9 database were identified as complete, 0.4% of genes were fragmented, and 2.7% of genes were missing in the assembled genome.

For the chromosome-level assembly, 38.09 Gb of clean reads (150 bp paired-end) were obtained from the Hi-C library (coverage: 85.31X, Table S1). In total, 118,367,396 (86.83%) reads were mapped to the draft genome, and 96,331,684 (70.67%) of them were uniquely mapped. The uniquely mapped sequences were analyzed with 3D-DNA software to assist genomic assembly. As a result, 68 scaffolds were assembled with an N50 length of 37.54 Mb (Table 1). Finally, 447.8 Mb genomic sequences were located on 9 chromosomes, accounting for 98.50% of the whole assembled length (Fig. 1B, Table 1, Table S2). The contig N50 and scaffold N50 of*M. dirhodum* were much higher than those of previously reported aphid genome assemblies (Table 1).

3.2 Genome Annotation

Repeatmasker and Repbase were used to annotate the repeat sequences. In total, 34.97% of the *M. dirhodum* genome was annotated as repeat sequences. Long terminal repeats (LTRs), long interspersed nuclear elements (LINEs) and DNA transposons accounted for 9.23%, 2.25% and 10.33% of the whole genome, respectively, and 13.16% of repeat sequences were annotated as unclassified (Table S5). A total of 286 tRNAs were predicted by trnascan-SE. Using infernal, we also identified 51 small nucleolar RNAs (snoRNAs), 586 ribosomal RNAs (rRNAs), 73 small nuclear RNAs (snRNAs), 59 microRNAs (miRNAs), 286 tRNAs and 639 other types of ncRNAs.

After masking repeat sequences, 18,003 protein-coding genes with a mean CDS length of 1,776 bp were identified from the *M. dirhodum*genome using de novo, homology- and RNA sequencing-based methods. The number of genes in the *M. dirhodum* genome is comparable to that in other insect species (Table 1). Functional annotation found that 16,548 (91.92%), 9,030 (50.16%), and 12,836 (71.30%) genes had significant hits with proteins cataloged in NR, SwissProt and eggNOG, respectively. There were 9,260 (51.44%) and 6,254 (34.74%) genes annotated to GO terms and KEGG pathways, respectively (Fig. S1).

3.3 Genome synteny and Phylogeny analysis

To gain insights into an evolutionary perspective for *M. dirhodum* , a whole genome-based phylogenetic analysis was performed with eight other hemipteran insect species, including *M. persicae* ,*D. noxia* , *A. pisum* ,*R. maidis* , *M. sacchari* , *N. lugens* , *B. tabaci*and *A. lucorum* . *D. melanogaster* was used as the outgroup. A total of 209,881 genes to 22,945 orthogroups for the 10 species comparison were assigned (Fig. 2). A phylogenetic tree was constructed using the single-copy orthologous genes (Table S3). As a result,*M. dirhodum* and the five other Aphididae insects formed an Aphididae cluster, which showed that *M. dirhodum* is close to*A. pisum* and separated from *M. sacchari* and *R. maidis* . Three other Hemipteran insects, including *B. tabaci* ,*N. lugens* and *A. lucorum,* formed another cluster (Fig. 2).

Syntenic relationships between the *M. dirhodum* and *A. pisum* genomes were compared. The results reveal high levels of genome rearrangement between chromosomes of *M. dirhodum* and *A. pisum* , and a number of fission and fusion events were observed. Chr1 in*M. dirhodum* shares 81.9% of the syntenic blocks of chr X in*A. pisum* (Fig. 3). Considering the conservation of the X chromosome in Aphidini insects (Biello et al., 2021), we inferred that chr 1 might be the sex chromosome in *M. dirhodum* . In addition, chrA1 in *A. pisum* is mainly syntenic to chr2, chr4, chr5 and chr8 of *M. dirhodum* . ChrA2 in *A. pisum* is mainly syntenic to chr6, chr7 and chr9 of *M. dirhodum* . ChrA3 in *A. pisum* is mainly syntenic to chr3 in *M. dirhodum* . However, many fusion events covering small regions occurred in all chromosomes between these two insect species (Fig. 3).

3.4 Identification of DEGs between winged and wingless *M. dirhodum*

All aphids are born through viviparous reproduction with wing primordia, but they degenerate by the second instar in the unwinged morph (Ding et al., 2017). In the winged individuals, the wing primordia continue to slowly develop from first- to third-instar nymphs and then rapidly grow in the fourth instar. In *M. dirhodum* , winged and wingless individuals can be distinguished from third-instar nymphs to adults

under a microscope (Ishikawa et al., 2008, 2013). To identify potential genes related to wing dimorphism and development, RNA-seq was performed between populations of winged and wingless *M. dirhodum* in third- and fourth-instar nymphs and adults using our assembled genome as a reference.

As a result, 4641 DEGs were detected between the wingless (WW) and winged (WY) third-instar nymphs (WW3/WY3), including 2114 upregulated and 2527 downregulated genes in the wingless nymphs, compared to the winged individuals (Fig. 4A, Table S4). A total of 4349 DEGs were detected between the wingless and winged fourth-instar nymphs (WW4/WY4), including 1936 upregulated and 2413 downregulated genes in the wingless nymphs, compared to the winged individuals (Fig. 4B, Table S4). In addition, 4375 DEGs were detected between the wingless and winged adults (WWA/WYA), including 1877 upregulated and 2498 downregulated genes in the wingless adults, compared to the winged individuals (Fig. 4C, Table S4). Among them, 3638 genes were differentially expressed in both the WW3/WY3 and WW4/WY4 groups, 3362 genes were differentially expressed in both the WW3/WY3 and WWA/WYA groups, 3433 genes were differentially expressed in both the WW4/WY4 and WYA/WWA groups, and 3050 genes were differentially expressed in all three comparison groups (WW3/WY3, WW4/WY4 and WWA/WYA) (Fig. 4D, Table S4).

Juvenile hormone or ecdysone signals, c-Jun NH2-terminal kinases (JNK), and insulin/insulin-like growth factor signaling (IIS) pathways have been reported to contribute to the regulation of wing dimorphism in many insects (Xu et al., 2015, 2017; Zhang et al., 2019). Moreover, Wnt2, Fng (fringe), Uba1 (ubiquitin-activating enzyme E1), Hh (hedgehog), Foxo, Dpp (decapentaplegic), Brk (Brinker), Ap (alar process), Dll (distal-less), Hth (helix-turn-helix), Tsh (thyroid-stimulating hormone), Nub (nubbin), Scr (sex combs reduced), Antp (antennapedia), Ubx (ultrathorax), Asc, Srf (serum response factor) and Fl (flugellos) have been also reported to play important roles in regulating wing polyphenism or development (Zhang et al., 2019). All the DEGs annotated as these genes were selected and are listed in Tables 2, 3 and 4, including insulin receptor substrate, insulin-like receptor, ecdysone receptor, broad-complex core protein, and Foxo. In addition, many genes related to muscle composition, energy metabolism and reproduction were also identified between the winged and wingless aphids.

3.5 Identification of DEGs among different developmental stages from winged or wingless *M. dirhodum*

For the transcriptome analysis of the different developmental stages of wingless *M. dirhodum* , 55 DEGs were detected between third- and fourth-instar nymphs (WW3/WW4), including 12 upregulated and 43 downregulated genes in third-instar nymphs compared to fourth-instar individuals. Thirty-three DEGs were detected between fourth-instar nymphs and adults (WW4/WWC), of which 30 genes were upregulated and 3 were downregulated in fourth-instar nymphs compared to adults. A total of 231 DEGs were detected between the wingless third-instar nymphs and adults (WW3/WWC), including 129 upregulated and 102 downregulated genes in the third-instar nymphs, compared to the adults (Fig. 5A, Table S5). Among them, 39 genes were differentially expressed in both the WW3/WW4 and WW4/WWC groups, 28 genes were differentially expressed in both the WW3/WWC and WW4/WWC groups, and no differentially expressed genes were found in either the WW3/WWC or WW3/WW4 groups (Fig. 5A, Table S5).

For the transcriptome analysis of the different developmental stages of winged *M. dirhodum* , 134 DEGs were detected between the third- and fourth-instar nymphs (WY3/WY4), including 22 upregulated and 112 downregulated genes in the third-instar nymphs compared to the fourth-instar individuals. A total of 334 DEGs were detected between fourth-instar nymphs and adults (WY4/WYC), including 289 upregulated and 45 downregulated genes in the fourth-instar nymphs, compared to adults. A total of 692 DEGs were detected between the third-instar nymphs and adults (WY3/WYC), including 320 upregulated and 372 downregulated genes in the third-instar nymphs, compared to winged adults (Fig. 5B, Table S6). Among them, 66 genes were differentially expressed in both the WY3/WY4 and WY3/WYC groups, 61 genes were differentially expressed in both the WY3/WY4 and WY4/WYC groups, 141 genes were differentially expressed in both the WY3/WYC and WY4/WYC groups, and 13 genes were differentially expressed in all three comparison groups (WY3/WY4, WY4/WYC and WY3/WYC) (Fig. 5B, Table S6). All these DEGs might be related to the development and metamorphosis of *M. dirhodum* .

4 DISCUSSION

M. dirhodum is an important pest of wheat and other cereals worldwide. Previous studies of this species have mainly focused on occurrence regularity (Honek et al., 2018), control methods (Cambier et al., 2001; Chopa et al., 2012) and reproductive strategies (Wratten, 1977), whereas more in-depth studies are limited due to the lack of high-quality genome data. Therefore, we used PacBio long HiFi reads and Hi-C technology to produce a chromosomal-level genome for *M. dirhodum* . This is the first high-quality chromosome-level genome of*M. dirhodum* , which will be very helpful for cloning, functional verification and evolutionary analysis of genes in this important species or even in other hemipteran insects.

The k -mer analysis showed that the *M. dirhodum* genome harbored a moderate level of heterozygosity and a high level of repetitive sequence content, which is similar to other Aphidinae insects with low or moderate level heterozygosity (Jiang et al., 2019; Chen et al., 2019). BUSCO assessment revealed that 96.9% of the complete BUSCOs could be found in the current assembled *M. dirhodum* genome. This percentage is higher than those in the genomes of some other species, such as *S. miscanthi* (90.2%) (Jiang et al., 2019), *R. maidis* (94.5%) (Chen et al., 2019), *A. pisum* (93.5%) (International Aphid Genomics Consortium., 2010) and *E. lanigerum*(96.8%) (Mathers et al., 2021). Considering the moderate level of heterozygosity and the high-level repetitiveness of the genome, the current result represents a high-quality genome assembly of *M. dirhodum* . Nine chromosomes were finally obtained after Hi-C-assisted assembly, supporting a 2n = 18 karyotype for *M. dirhodum,* which is identical to *S. miscanthi* (Jiang et al., 2019). Furthermore, a total of 18,003 protein-coding genes were predicted in the genome of*M. dirhodum* , which was comparable to that of several other Aphidinae species, such as 16,006 protein-coding genes in *S. miscanthi* (Jiang et al., 2019) and 19,097 protein-coding genes in *D. noxia* (Nicholson et al., 2015), but less than *R. maidis* (Chen et al., 2019), *A. pisum*(International Aphid Genomics Consortium., 2010), *M. persicae*(Jiang et al., 2013) and *E. lanigerum* (Mathers et al., 2021), which have 26,286, 36,195, 23,910 and 28,186 protein-coding genes, respectively.

Wing polymorphism is an evolutionarily successful feature in a wide variety of insect species, including Hemiptera, Coleoptera, Hymenoptera, Orthoptera, Diptera, Lepidoptera, Isoptera, Psocoptera and Dermaptera (Xu et al., 2017; Zhang et al., 2019). Insect wing polymorphisms have been widely studied, and some progress has been made, especially in the rice plant hopper *Nilaparvata lugens* , in which two putative insulin receptors (InRs), InR1 and InR2, were identified. Interestingly, InR1 and InR2 play fully opposite roles in wing morph determination by regulating the activity of Foxo. In detail, knockdown of InR2 at the nymph stage led to a strong trend toward long-winged adults, whereas dysfunction of InR1 resulted in the development of short-winged adults (Xu et al., 2015). In the present study, two transcripts annotated as insulin receptors (insulin-like receptors and insulin-like peptide receptors) were identified in *M. dirhodum* , which is consistent with the annotation of the *A. pisum* genome (International Aphid Genomics Consortium., 2010). Comparative transcriptome analysis showed that the insulin-like receptor was downregulated in third-instar*M. dirhodum* wingless nymphs and that the insulin-like peptide receptor was upregulated in fourth-instar wingless nymphs compared to winged individuals. In addition, one transcript annotated as insulin receptor substrate 1 was downregulated in both third- and fourth-instar wingless nymphs, and one transcript annotated as insulin gene enhancer protein isl-1 was upregulated in both third-instar and fourth-instar wingless nymphs compared to the winged individuals. No DEGs associated with the insulin signaling pathway were identified between the wingless and winged *M. dirhodum* adults. These results suggest that insulin signaling may play an important role in the early wing dimorphism of*M. dirhodum* . Foxo was identified to be upregulated in wingless*M. dirhodum* in third- and fourth-instar nymphs and adults, which is another important factor in the regulation of wing dimorphism. Vellichirammal et al. (2017) revealed that ecdysone signaling plays a critical role in transgenerational plasticity in wing dimorphism in*A. pisum* . Their results showed that more winged offspring would be produced by injection of ecdysone or its analog, while fewer winged offspring were produced by knockdown of EcR or treatment with an EcR antagonist. Our results showed that EcR is downregulated in wingless*M. dirhodum* from third-instar nymphs to adults compared to winged nymphs, suggesting that the decrease in ecdysone is not conducive to the production of winged individuals. Moreover, DEGs annotated as juvenile hormone epoxide hydrolase, brc, FTZ-F1, Wnt, Dpp, Ubx, Srf and Fl were also detected between the wingless and winged *M. dirhodum* . These findings provide

a valuable reference for revealing the mechanism of wing dimorphism.

5 CONCLUSION

In this study, we presented a high-quality chromosome-level genome assembly of *M. dirhodum* with the aid of PacBio long HiFi reads and Hi-C technology. The 447.8 M *M. dirhodum* genome encoded 18,003 protein-coding genes. This is the first assembled genome for*M. dirhodum* and will facilitate further research related to pesticide resistance, virus transmission and other aspects of *M. dirhodum* biology. Comparative transcriptomic analyses identified a number of DEGs that might be related to wing dimorphism, including ILR, Foxo, ECR, brc, and FTZ-F1. These results lay the foundation for future in-depth research on wing dimorphism in *M. dirhodum* .

ACKNOWLEDGEMENTS

This work was supported by the China Agriculture Research System (Grant Number: CARS-05-03A-02). We thank Berry Genomics Corporation for technical support in Illumina, PacBio and Hi-C.

CONFLICTS OF INTEREST

The authors declare no competing interests.

AUTHOR CONTRIBUTIONS

X. W. Gao, P. Liang and B. Zhu planned and coordinated the project. R. Wei and B. Zhu prepared the samples for PacBio and Hi-C. W. J. Hua and B. Zhu prepared the samples for Illumina sequencing. W. L. Zhnag and B. Zhu performed the experiments and analyzed the data. B. Zhu wrote the manuscript. P. Liang revised the manuscript, and all authors approved the final manuscript.

DATA AVAILABILITY STATEMENT

Raw genome sequencing reads and RNA-seq reads were deposited in the National Center for Biotechnology Information using BioProject Accession no. PRJNA751716 and PRJNA751719. The whole genome shotgun sequencing project has been deposited at NCBI GenBank under accession no. JAIOUA000000000.

REFERENCES

Bairoch, A., & Boeckmann, B. (1991). The SWISS-PROT protein sequence data bank. *Nucleic Acids Research* , *19* Suppl, 2247-2249.

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

Belton, J. M., McCord, R. P., Gibcus, J. H., Naumova, N., Zhan, Y., & Dekker, J. (2012). Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods* , *58* (3), 268–276.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* , *27* (2), 573-80.

Biello, R., Singh, A., Godfrey, C.J., Fernández, F.F., Mugford, S.T., Powell, G., Hogenhout, S.A., Mathers TC. (2021). A chromosome-level genome assembly of the woolly apple aphid,*Eriosoma lanigerum* Hausmann (Hemiptera: Aphididae).*Molecular Ecology Resources* , *21* (1), 316-326.

Blackman, R. L., and Eastop V. F. (2000). Aphids on the World's crops. John Wiley and Sons, Chichester, New York, Brisbane, Toronto and Singapore.

Braendle, C., Davis, G.K., Brisson, J.A., Stern, D.L. (2006). Wing dimorphism in aphids.*Heredity* , *97* (3), 192–99.

Cambier, V., Hance, T., De Hoffmann, E. (2001). Effects of 1,4-benzoxazin-3-one derivatives from maize on survival and fecundity of *Metopolophium dirhodum*(Walker) on artificial diet. *Journal of Chemical Ecology* ,*27* (2), 359-70.

Campbell, M.A., Haas, B.J., Hamilton, J.P., Mount, S. M., Buell, C. R. (2006). Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis. *BMC Genomics* , *7* , 327.

Cannon, R.J.C. (1986). Summer populations of the cereal aphid*Metopolophium dirhodum* (Walker) on winter wheat: three contrasting years. *Journal of Applied Ecology* , *23* (1), 101–114.

Chan, P.P., Lowe, T.M. (2019). tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods in Molecular Biology* , 1962, 1–14.

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.

Chen W, Wosula EN, Hasegawa DK, Casinga C, Shirima RR, Fiaboe KKM, Hanna R, Fosto, A., Goergen, G., Tamò, M., Mahuku, G., Murithi, H.M., Tripathi, L., Mware, B., Kumar, L.P., Ntawuruhunga, P., Moyo, C., Yomeni, M., Boahen, S., Edet, M., Awoyale, W., Wintermantel, W.M., Ling, K.S., Legg, J.P., Fei, Z. (2019). Genome of the African cassava whitefly*Bemisia tabaci* and distribution and genetic diversity of cassava-colonizing whiteflies in Africa. *Insect Biochemistry and Molecular Biology* , 110, 112-120.

Chopa, C.S., Descamps, L.R. (2012). Composition and biological activity of essential oils against *Metopolophium dirhodum* (Hemiptera: Aphididae) cereal crop pest. *Pest Managent Science* ,*68* (11),1492-500.

Deng, Y. Y., Li, J. Q., Wu, S. F., Zhu, Y., Chen, Y. W., & He, F. C. (2006). Integrated nr database in protein annotation system and its localization. *Computer Engineering* , *32* , 71-72.

Dimmer, E.C., Huntley, R.P., Alam-Faruque, Y., Sawford, T., O'Donovan, C., Martin, M.J., Bely, B., Browne, P., Mun Chan, W., Eberhardt, R., Gardner, M., Laiho, K., Legge, D., Magrane, M., Pichler, K., Poggioli, D., Sehra, H., Auchincloss, A., Axelsen, K., Blatter, M.C., Boutet, E., Braconi-Quintaje, S., Breuza, L., Bridge, A., Coudert, E., Estreicher, A., Famiglietti, L., Ferro-Rojas, S., Feuermann, M., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., James, J., Jimenez, S., Jungo, F., Keller, G., Lemercier, P., Lieberherr, D., Masson, P., Moinat, M., Pedruzzi, I., Poux, S., Rivoire, C., Roechert, B., Schneider, M., Stutz, A., Sundaram, S., Tognolli, M., Bougueleret, L., Argoud-Puy, G., Cusin, I., Duek-Roggli, P., Xenarios, I., Apweiler, R. (2012). The UniProt-GO Annotation database in 2011. *Nucleic Acids Research* , *40*(Database issue), D565-70.

Ding, B.Y., Shang, F., Zhang, Q., Xiong, Y., Yang, Q., Niu, J.Z., Smagghe, G., Wang, J.J. (2017). Silencing of Two Insulin Receptor Genes Disrupts Nymph-Adult Transition of Alate Brown Citrus Aphid.*International Journal of Molecular Sciences* , *18* (2), 357.

Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S., Huntley, M. H., Lander, E. S., & Aiden, E. L. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Systems* , *3* (1), 95-98.

Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., . . . Aiden, A. P. (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds.*Science* , *356* (6333), 92-95.

Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., Smit, A.F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America* , 117(17), 9451–7.

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biololy* , *9* (1), R7.

Holt, J., Griffiths, E. & Wratten, S. D. (1984). The influence of wheat growth stage on yield reductions caused by the rose-grain aphid, *Metopolophium dirhodum* . *Annals of Applied Biology* , *105* (1), 7–14.

Honek, A. (1987). Effect of plant quality and microclimate on population growth and maximum abundances of cereal aphids, *Metopolophium dirhodum* (Walker) and *Sitobion avenae* (F.) (Hom., Aphididae). *Journal*

of Applied Entomology ,*104* (3): 304–314.

Honek, A. (1991). Factors determining the peak abundance of *Metopolophium dirhodum*(Homoptera: Aphididae) on cereals. *Bulletin Of Entomological Research* , *81* (1), 57–64.

Honek, A., Martinkova, Z., Saska, P. & Dixon, A. F. G. (2018). Aphids (Homoptera: Aphididae) on winter wheat: Predicting maximum abundance of *Metopolophium dirhodum* .*Journal of Economic Entomology* , *111* (4), 1751–1759.

Hoskins, R.A., Carlson, J.W., Wan, K.H., Park, S., Mendez, I., Galle, S.E., Booth, B.W., Pfeiffer, B.D., George, R.A., Svirskas, R., Krzywinski, M., Schein, J., Accardo, M.C., Damia, E., Messina, G., Méndez-Lago, M., de Pablos, B., Demakova, O.V., Andreyeva, E.N., Boldyreva, L.V., Marra, M., Carvalho, A.B., Dimitri, P., Villasante, A., Zhimulev, I.F., Rubin, G.M., Karpen, G.H., Celniker, S.E. (2015). The Release 6 reference sequence of the *Drosophila melanogaster*genome. *Genome Research* , *25* (3), 445-58.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., von Mering, C., Bork, P. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research .47* (D1), D309-D314.

International Aphid Genomics Consortium. (2010). Genome sequence of the pea aphid *Acyrthosiphon pisum* . *PLoS Biology* , *8* (2), e1000313.

Ishikawa, A., Hongo, S., Miura, T. (2008). Morphological and histological examination of polyphenic wing formation in the pea aphid *Acyrthosiphon pisum* (Hemiptera, hexapoda). *Zoomorphology* , *127* (2), 121–133.

Ishikawa, A., Gotoh, H., Abe, T., Miura, T. (2013). Juvenile hormone titer and wing-morph differentiation in the vetch aphid *Megoura crassicauda* . *Journal of Insect Physiology* , 59(4), 444–449.

Jiang, X., Zhang, Q., Qin, Y., Yin, H., Zhang, S., Li, Q., Zhang, Y., Fan, J., Chen, J. A. (2019). Chromosome-level draft genome of the grain aphid *Sitobion miscanthi* . *Gigascience* . *8* (8): giz101.

Jiang, Z., Jones, D.H., Khuri, S., Tsinoremas, N.F., Wyss, T., Jander, G., Wilson, A.C. (2013). Comparative analysis of genome sequences from four strains of the Buchnera aphidicola Mp endosymbion of the green peach aphid, *Myzus persicae* . *BMC Genomics* , *4* , 917.

Keilwagen, J., Wenk, M., Erickson, J. L., Schattat, M. H., Grau, J., & Hartung, F. (2016). Using intron position conservation for homology-based gene prediction. *Nucleic Acids Research* ,*44* (9), e89-e89.

Kennedy, T. F. & Connery, J. (2005). Grain yield reductions in spring barley due to barley yellow dwarf virus and aphid feeding. *Irish Journal of Agricultural And Food Research* , *44* (1), 111–128.

Kanehisa, M., Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* , *28* , 27–30.

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics* . *5* , 59.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* , *25* (16), 2078–2079.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* , *1303* (3997).

Li, L., Stoeckert, C. J., Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research* ,*13* (9), 2178-2189.

Li, X., Gong, P., Wang, B., Wang, C., Li, M., Zhang, Y., Li, X., Gao, H., Ju, J., Zhu, X. (2020). Selection and validation of experimental condition-specific reference genes for qRT-PCR in *Metopolophium dirhodum* (Walker) (Hemiptera: Aphididae). *Scientific Reports .10* (1), 21951.

Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., & Dorschner, M. O. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* , *326* (5950), 289–293.

Liu, Y., Liu, H., Wang, H., Huang, T., Liu, B., Yang, B., Yin, L., Li, B., Zhang, Y., Zhang, S., Jiang, F., Zhang, X., Ren, Y., Wang, B., Wang, S., Lu, Y., Wu, K., Fan, W., Wang, G. (2021). *Apolygus lucorum*genome provides insights into omnivorousness and mesophyll feeding.*Molecular Ecology Resources* . *21* (1), 287-300.

Ma, C.S., Hau, B. and Poehling, H.M. (2004). Effects of pattern and timing of high temperature exposure on reproduction of the rose grain aphid, *Metopolophium dirhodum* . *Entomologia Experimentalis Et Applicata* , *110* (1): 65–71.

Ma, W.H., Xu, L., Hua, H.X., Chen, M.Y., Guo, M.J., He, K., Zhao, J., Li, F., (2021). Chromosomal-level genomes of three rice planthoppers provide new insights into sex chromosome evolution. *Molecular Ecology Resources* . *21* (1), 226-237.

Mathers, T.C., Chen, Y., Kaithakottil, G., Legeai, F., Mugford, S.T., Baa-Puyoulet, P., Bretaudeau, A., Clavijo, B., Colella, S., Collin, O., Dalmay, T., Derrien, T., Feng, H., Gabaldón, T., Jordan, A., Julca, I., Kettles, G.J., Kowitwanich, K., Lavenier, D., Lenzi, P., Lopez-Gomollon, S., Loska, D., Mapleson, D., Maumus, F., Moxon, S., Price, D.R., Sugio, A., van Munster, M., Uzest, M., Waite, D., Jander, G., Tagu, D., Wilson, A.C., van Oosterhout, C., Swarbreck, D., Hogenhout, S.A. (2017). Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species.*Genome Biology* , *18* (1), 27.

Muller, C.B., Williams, I.S., Hardie, J. (2001). The role of nutrition, crowding, and interspecfic interactions in the development of winged aphids. *Ecology Entomology* , *26* (3), 330–40.

Nawrocki, E.P., Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches.*Bioinformatics* , *29* (22), 2933–5.

Nicholson, S.J., Nickerson, M.L., Dean, M., Song, Y., Hoyt, P.R., Rhee, H., Kim, C., Puterka, G.J. (2015). The genome of *Diuraphis noxia* , a global aphid pest of small grains. *BMC Genomics* . *16* , 429.

Pan, Y., Fang, G., Wang, Z., Cao, Y., Liu, Y., Li, G., Liu, X., Xiao, Q., Zhan, S. (2021) Chromosome-level genome reference and genome editing of the tea geometrid. *Molecular Ecology Resources* , doi: 10.1111/1755-0998.13385.

Praslicka, J. (1996). Influence of some growing factors on the occurrence of cereal aphids associated with winter wheat. Rostlinna Vyroba, *42* (11): 499–502.

Quan, Q., Hu, X., Pan, B., Zeng, B., Wu, N., Fang, G., Cao, Y., Chen, X., Li, X., Huang, Y., Zhan, S. (2019). Draft genome of the cotton aphid*Aphis gossypii* . *Insect Biochemistry and Molecular Biology* ,*105* , 25-32.

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., & Aiden, E. L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* ,*159* (7), 1665–1680.

Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* ,*31* (19), 3210–3212.

Stanke, M., Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel.*Bioinformatics* . *19* (Suppl 2), ii215–25.

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

Telesnicki, M.C., Ghersa, C.M., Martínez-Ghersa, M.A., Arneodo, J.D. (2012). Molecular identification of the secondary endosymbiont Hamiltonella defensa in the rose-grain aphid*Metopolophium dirhodum* . *Revista Argentina De Microbiologia* , *44* (4), 255-8.

Vellichirammal, N.N., Gupta, P., Hall, T.A., Brisson, J.A. (2017). Ecdysone signaling underlies the pea aphid transgenerational wing polyphenism. *Proceedings of the National Academy of Sciences of the United States of America* , *114* , 1419–23.

Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., Schatz, M.C. (2017). GenomeScope: fast reference-free genome profling from short reads. *Bioinformatics* , *33* (14), 2202–4.

Wang, L., Feng, Z., Wang, X., Wang, X., Zhang, X. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* . *26* (1), 136-8.

Wenger, J.A., Cassone, B.J., Legeai, F., Johnston, J.S., Bansal, R., Yates, A.D., Coates, B.S., Pavinato, V.A.C., Michel, A. (2020). Whole genome sequence of the soybean aphid, *Aphis glycines* . *Insect Biochemistry and Molecular Biology* , *123* , 102917.

Wratten, S.D. (1977). Reproductive strategy of winged and wingless morphs of the aphids *Sitobion avenae* and *Metopolophium dirhodum* . *Annals of Applied Biology* , 85(3), 319-31.

Xie, W., Chen, C., Yang, Z., Guo, L., Yang, X., Wang, D., Chen, M., Huang, J., Wen, Y., Zeng, Y., Liu, Y., Xia, J., Tian, L., Cui, H., Wu, Q., Wang, S., Xu, B., Li, X., Tan, X., Ghanim, M., Qiu, B., Pan, H., Chu, D., Delatte, H., Maruthi, M.N., Ge, F., Zhou, X., Wang, X., Wan, F., Du, Y., Luo, C., Yan, F., Preisser, E.L., Jiao, X., Coates, B.S., Zhao, J., Gao, Q., Xia, J., Yin, Y., Liu, Y., Brown, J.K., Zhou, X.J., Zhang, Y. (2019). Genome sequencing of the sweet potato whitefly*Bemisia tabaci* MED/Q. *Gigascience* . 6(5), 1-7.

Xu, H.J., Xue, J., Lu, B., Zhang, X.C., Zhuo, J.C., He, S.F., Ma, X.F., Jiang, Y.Q., Fan, H.W., Xu, J.Y., Ye, Y.X., Pan, P.L., Li, Q., Bao, Y.Y., Nijhout, H.F., Zhang, C.X. (2015). Two insulin receptors determine alternative wing morphs in planthoppers. *Nature* , *519* , 464-467.

Xu, H.J., and Zhang, C.X. (2017). Insulin receptors and wing dimorphism in rice planthoppers. *Philosophical Transactions of The Royal Society of London. Series B, Biological Sciences* , *372* , 1713.

Zhang, C.X., Brisson, J.A., Xu, H.J. (2019). Molecular mechanisms of wing polymorphism in insects. *Annual Review of Entomology* ,*64* , 297-314.

Zhang, R.J., Chen, J., Jiang, L.Y., Qiao, G.X. (2019). The genes expression difference between winged and wingless bird cherry-oat aphid*Rhopalosiphum padi* based on transcriptomic data. Scientific Reports. 9(1), 4754.

Figure legends

Fig. 1: *k*-mer (K=17) distribution of Illumina genome sequencing reads of *Metopolophium dirhodum* (A) and Hi-C contact heat map of the assembled genome (B)

Fig. 2: Maximum likelihood phylogeny of *Metopolophium dirhodum* and nine other aphid species based on a concatenated alignment of the conserved single copy orthologues. The histograms are subdivided to represent different categories of orthology: 1:1:1 (single copy orthologous genes in communal gene families); N:N:N (multiple copy orthologous genes in communal gene families); specific (genes from unique gene families from each species); other (genes that do not belong to any of the abovementioned orthologous categories); uncluster (genes that do not cluster to any families)

Fig. 3: Chromosome-level synteny analysis between*Metopolophium dirhodum* and *Acyrthosiphon pisum*

Fig. 4: Scatter plot of differentially expressed genes between winged (WY) and wingless (WW) *Metopolophium dirhodum*

Fig. 5: Scatter plot of differentially expressed genes among third- and fourth-instar nymphs and adults of winged (WY) and wingless (WW) *Metopolophium dirhodum*

Supplementary Information

Fig. S1: Venn diagram of functional annotation based on five databases.

Table S1: Statistics for sequencing data for *Metopolophium dirhodum* genome assembly.

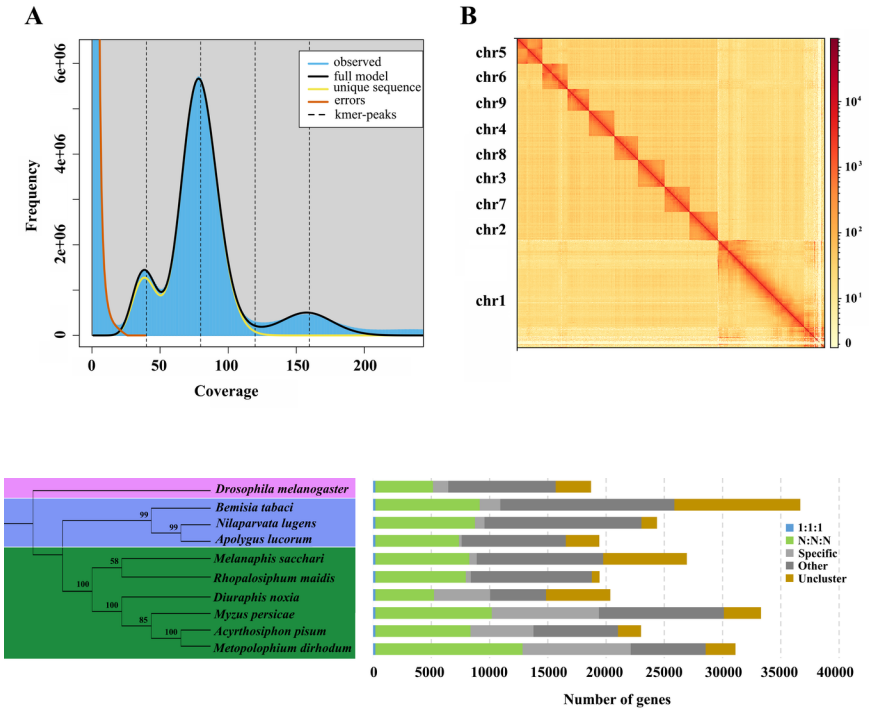
Table S2: Overview of chromosome length of *Metopolophium dirhodum*assembly.

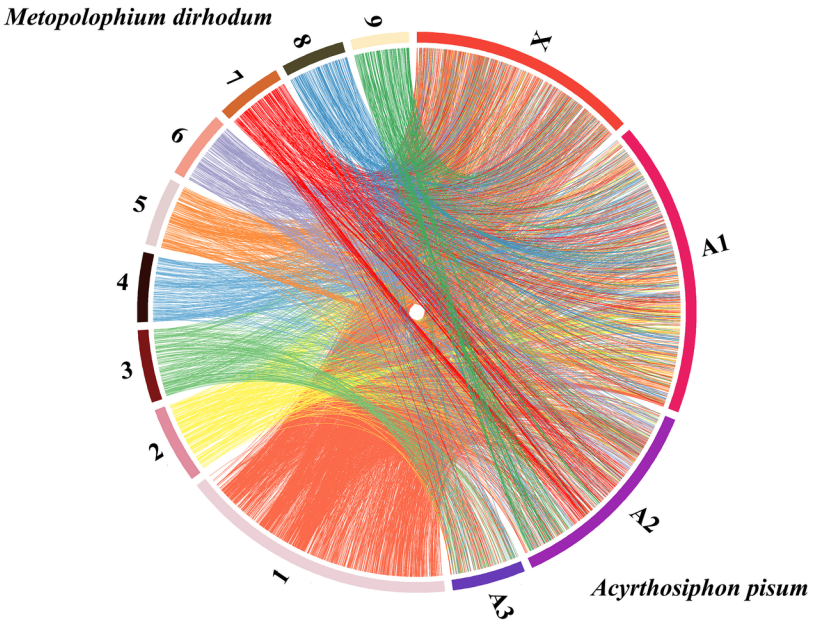
Table S3: Gene family clusters of ten species.

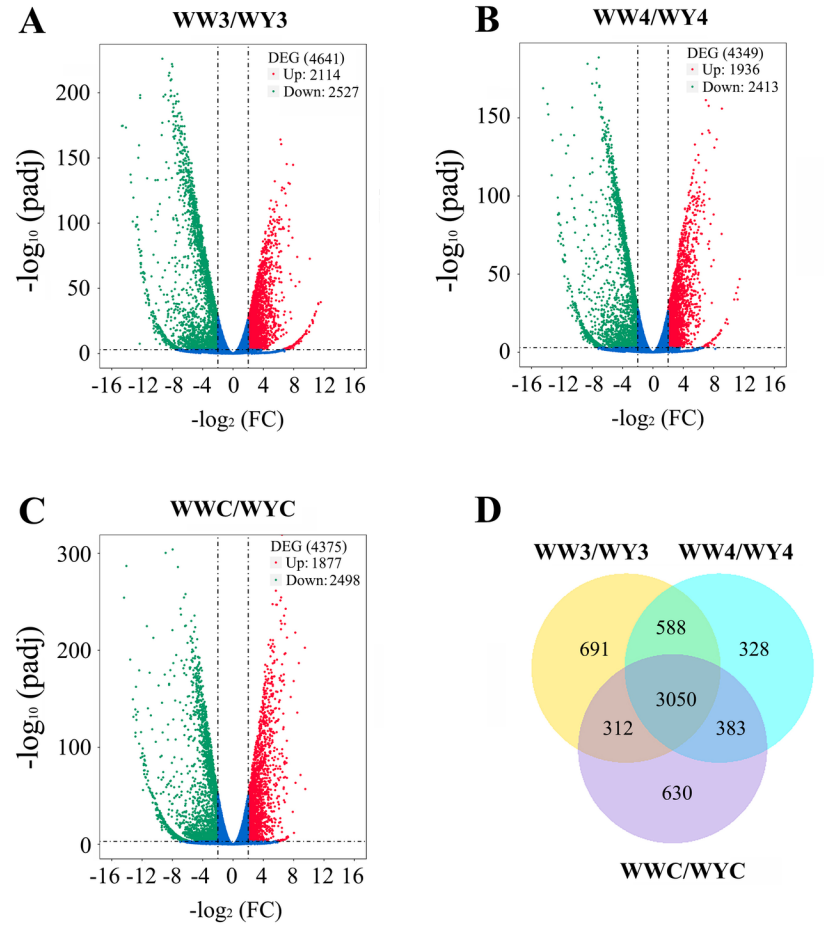
Table S4: Differentially expressed genes between winged (WY) and wingless (WW) *Metopolophium dirhodum* .

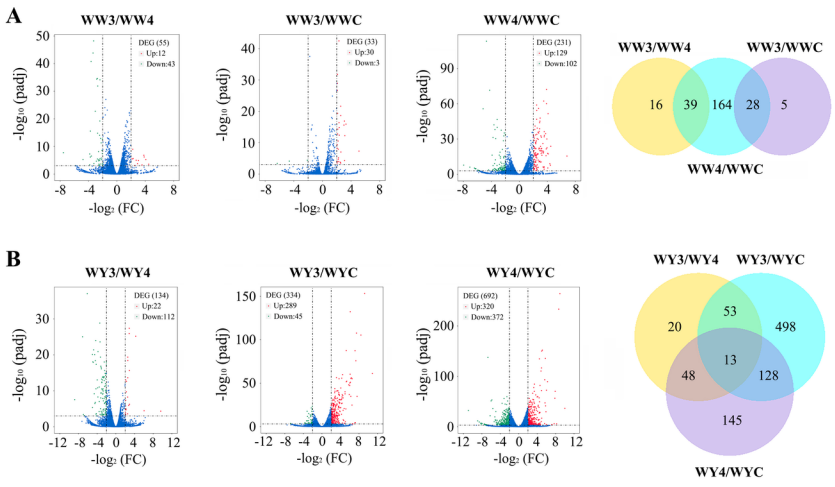
Table S5: Differentially expressed genes among third- and fourth-instar nymphs and adults of wingless (WW) *Metopolophium dirhodum* .

Table S6: Differentially expressed genes among third- and fourth-instar nymphs and adults of winged (WY) *Metopolophium dirhodum* .









Hosted file

Table 1-4.docx available at https://authorea.com/users/370395/articles/711893-a-high-quality-chromosome-level-genome-assembly-provides-insights-into-wing-dimorphism-in-metopolophium-dirhodum-walker