Chromosome-level genome assembly of the aphid parasitoid *Aphidius gifuensis* using Oxford Nanopore sequencing and Hi-C technology

Bingyan Li1 | Zhenyong Du1 | Li Tian1 | Limeng Zhang2 | Zhihua Huang2 | Shujun Wei3 | Fan Song1 | Wanzhi Cai1 | Yanbi Yu4 | Hailin Yang2 | Hu Li1

1Department of Entomology and MOA Key Lab of Pest Monitoring and Green Management, College of Plant Protection, China Agricultural University, Beijing, China
2Tobacco Company, Yuxi, China
3Institute of Plant and Environmental Protection, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China
4Yunnan Tobacco Company of China National Tobacco Corporation, Kunming, China

**Correspondence**
Hu Li, Department of Entomology and MOA Key Lab of Pest Monitoring and Green Management, College of Plant Protection, China Agricultural University, Beijing 100193, China. Email: tigerleecau@hotmail.com (H.L.)
Hailin Yang, Tobacco Company, Yuxi 653100, China. Email: 3641775@qq.com (H.Y.)
Yanbi Yu, Yunnan Tobacco Company of China National Tobacco Corporation, Kunming 650011, China. Email: yuyanbi@sina.com (Y.Y.)

**Funding information**
China National Tobacco Corporation of Science and Technology Major Projects

**ABSTRACT**
*Aphidius gifuensis* is a parasitoid wasp that has been commercially bred and released in large scale as a biocontrol agent for the management of aphid pests. As a highly efficient endoparasitoid, it is also an important model for exploring mechanisms of parasitism. Currently, artificially bred populations of this wasp are facing rapid decline with undetermined cause, and mechanisms underlying its parasitoid strategy remain poorly understood. Exploring the mechanism behind its population decline and the host–parasitoid relationship is impeded partly due to the lack of a comprehensive genome data for this species. In this study, we constructed a high-quality reference genome of *A. gifuensis* using Oxford Nanopore sequencing and Hi-C (proximity ligation chromatin conformation capture) technology. The final genomic assembly was 156.9 Mb, with a contig N50 length of 3.93 Mb, the longest contig length of 10.4 Mb and 28.89% repetitive sequences. 99.8% of genome sequences were anchored onto six linkage groups. A total of 11,535 genes were predicted, of which 90.53% were functionally annotated. Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis showed the completeness of assembled genome is 98.3%. We found significantly expanded gene families involved in metabolic processes, transmembrane transport, cell signal communication and oxidoreductase activity, in particular ATP-binding cassette (ABC) transporter, Cytochrome P450 and venom proteins. The olfactory receptors (ORs) showed significant contraction, which may be associated with the decrease in host recognition. Our study provides a solid foundation for future studies on the molecular mechanisms of population decline as well as host–parasitoid relationship for parasitoid wasps.

**KEYWORDS**
*Aphidius gifuensis*, chromosome-level genome, Hi-C, nanopore, parasitoid wasp
INTRODUCTION

Parasitoid wasps have been widely used as effective biological control agents against numerous destructive pests in the agricultural industry, including species of Lepidoptera, Diptera, Coleoptera and Hemiptera (Beckage & Gelman, 2004; Wang et al., 2019). The parasitoid wasps lay eggs inside or on the surface of pests, and the parasitoid larvae absorb nutrients from their hosts until emergence. To increase the efficiency of parasitism, the wasps develop a variety of strategies, producing and transmitting bioactive substances into hosts, such as venom (Asgari & Rivers, 2011; Mrinalini & Werren, 2015), polydnavirus (Belle et al., 2002), teratocytes (Dahlman et al., 2003; Strand, 2014), ovarian proteins (Luckhart & Webb, 1996) and larval secretions. Multiple parasitoid wasps have been bred successfully on a commercial scale and widely applied to control specific agricultural pests (Wang et al., 2019), such as Aphidius ervi (Dennis et al., 2020), Cotesia flavipes (Muirhead et al., 2012), Encarsia formosa (Hoddle et al., 1998) and Trichogramma galloi (Postali Parra & Coelho, 2019).

Aphidius gifuensis Ashmead (Hymenoptera: Braconidae) is a common endoparasitoid of many destructive aphid species (Figure 1), including Myzus persicae, Sitobion avenae and Schizaphis graminum (Pan & Liu, 2014). In China, this wasp has been widely bred and used as a biocontrol agent for more than four decades to control the green peach aphid, Myzus persicae, a severe pest on vegetables and tobaccos (Li et al., 2018; Yang et al., 2011). The life history, physiological characteristics, parasitic ability and breeding technology of A. gifuensis have been extensively studied to improve its captive breeding behaviour and aphid control efficiency (Khan et al., 2016; Li et al., 2013; Pan et al., 2018; Yang et al., 2009). For instance, using comparative transcriptome and proteome analysis, Zhang et al. (2018) identified genes associated with diapause behaviour. Several chemosensory receptors and candidate genes associated with environmental adaptation were also identified, including chemosensory: olfactory receptors (ORs), gustatory receptors (GRs) and ionotropic receptors (IRs); detoxification systems: cytochrome P450, peroxidase and glutathione S-transferase; and heat shock proteins (Fan et al., 2018; Gao et al., 2017; Kang et al., 2017, 2018). Nevertheless, in recent years, commercially bred A. gifuensis has shown rapid population decline and a significant decrease in aphid control efficiency, which can be reflected by a marked decrease in its viability, fecundity and capability of recognizing and parasitizing its aphid host (Xie et al., 2020). These issues have severely compromised the efficiency of A. gifuensis as a biocontrol agent in the field. However, knowledge about the molecular mechanisms of parasitism and population degradation of this wasp remains limited, in part due to the absence of the whole-genome information of A. gifuensis.

Whole-genome information plays an important role in genomic studies of the parasitoid wasps. Genome data can provide important insights into molecular mechanisms behind physiology and parasitic biology of parasitoid wasps. The unique haplodiploid mating system of the Hymenoptera is also beneficial to whole-genome sequencing. To date, whole genomes have been obtained for more than 40 parasitoid wasp species (Branstetter et al., 2018), seven of which were braconid wasps (Burke et al., 2018; Geib et al., 2017; Tvedte et al., 2019; Yin et al., 2018), including Aphidius ervi (Dennis et al., 2020), Cotesia vestalis (Shi, Wang, et al., 2019), Diachasma alloeum (Tvedte et al., 2019), Fopius arisanus (Geib et al., 2017), Lysiphlebus fabarum (Dennis et al., 2020), Macrocentrus cingulum (Yin et al., 2018) and Micropplitis demolitor (Burke et al., 2018). Most braconid genomes were sequenced primarily on Illumina platforms, which precluded the assembly of target genome into chromosome-level scaffolds. The application of single molecule real-time sequencing technology can improve the quality, particularly in terms of contiguity, of the genome assembly.

In this study, we report a high-quality genome assembly of A. gifuensis, the first Aphidiinae genome reported at chromosome level, using Nanopore sequencing and Hi-C (proximity ligation chromatin conformation capture) technology. Comparative analyses with available wasp genomes were conducted to understand the genomic evolution of A. gifuensis. This reference genome will provide a foundation for future studies on the molecular mechanism of host recognition, parasitizing behaviour and population degradation of A. gifuensis, key features of this important biocontrol agent.

FIGURE 1 Aphidius gifuensis. (a) Adult and (b) Adult emerging from mummified aphid [Colour figure can be viewed at wileyonlinelibrary.com]
2 MATERIALS AND METHODS

2.1 Sampling and genome sequencing

All A. gifuensis samples for genome sequencing were collected from a laboratory-bred population reared on Myzus persicae in the Biological Control for Tobacco Diseases and Insect Pests Engineering Research Center of China Tobacco, Yunnan, China. About 17.4 μg genomic DNA was extracted from a pooled sample of 800 adult males using the Blood & Cell Culture DNA Kit (Qiagen). DNA was quantified by 0.75% agarose gel electrophoresis, Nanodrop spectrophotometry (Thermo Fisher) and Qubit 3.0 fluorometry (Invitrogen).

Using the TrueSeq Nano DNA HT Sample preparation Kit (Illumina), a paired-end (PE) library, with an insert size of 350 bp, was constructed for genomic survey and sequenced with PE 150 bp on an Hiseq X Ten platform (Illumina). Libraries with long fragments were prepared and sequenced on a GridION X5 sequencer (Nanopore) at GrandOmics. The quality of these libraries was measured using 0.75% agarose gel electrophoresis.

The Hi-C sequencing library was prepared following Belton et al. (2012) with minor modifications. In brief, the cells of approximately 200 male samples were fixed with 2% formaldehyde for crosslinking. The fixed tissue was frozen in liquid nitrogen and ground to a powder before re-suspending in nuclei isolation buffer, following the protocol described by Shi, Ma, et al. (2019). The cross-linked DNA was digested with DpnII restriction endonuclease and marked by biotin-14-dCTP to remove nonligated DNA fragments. The ligated DNA was then sheared to 300–600 bp followed by a standard Illumina library preparation protocol, described in Meyer and Kircher (2010). The library was sequenced on the Illumina NovaSeq platform (Illumina) with paired-end (PE) 150 bp reads.

2.2 Genome size estimation and assembly

Raw data of short reads from the Illumina platform were filtered using strict quality controls by fastp v0.12.6 (Chen et al., 2018), to remove duplications, reads containing adapters and low-quality reads (-q 5 -u 50). These were then used to estimate the genome size based on the K-mer distribution analysed by kmerfreq v2.0 (Marçais & Kingsford, 2011). To assess the degree of microbial contamination, we checked 20,000 randomly selected Illumina sequencing reads by searching against the online NCBI-NR database (https://blast.ncbi.nlm.nih.gov/Blast.cgi). We applied an e-value threshold of <1e-5, and according to this criterion, no microbial contaminant sequences were found in the sequencing reads. For genomic contig assembly, we first used canu v1.3 (Koren et al., 2017) to correct and trim the long sequencing reads with default parameters. The trimmed reads were assembled by using the software smartdenovo v1.0 (https://github.com/ruanjue/smartdenovo, -k 19, -j 5000, -e dmo). To correct for errors of assembly, the short reads were mapped to the genome assembly using bwa-mem v0.7.12-r1039 (Li & Durbin, 2009) with default parameters, followed by three iterations of polishing with Illumina short reads using nextpolish v1.0 (genome_size =auto, -min_read_len 10k, -max_read_len 150k, -max_depth 60) (Hu et al., 2019). The assembled contigs were also assessed against NCBI-NR database using the same approach as described for raw sequence reads and were again found to contain no microbial sequences. We calculated genome and assembly statistics, such as GC content and contig N50, using a custom python script (Supplemental text).

To further scaffold the assembly into chromosome-scale linkage groups, we performed Hi-C analysis. The raw Hi-C data were primarily filtered using the default parameters in Hi-C-Pro (Burton et al., 2013). Then, all the cleaned reads were mapped to the draft assembly contigs using bowtie2 v2.3.2 with end-to-end alignment mode (--very-sensitive -L 30) (Langmead & Salzberg, 2012), after quality control with Fastp as described previously. The uniquely aligned paired reads were retained for analysis. Based on the matrix of valid interaction reads, the contigs were anchored to chromosomal linkage groups using lachesis (https://github.com/shendurelab/LACHESIS) (Burton et al., 2013). The completeness of the final draft genome was assessed using BUSCO v3.0.1 (Benchmarking Universal Single-Copy Orthologs) at nucleotide level based on 1658 genes in the insecta_odb9 database (Simão et al., 2015).

2.3 Transcriptome sequencing

Total RNA was extracted from 75 adult males using TRizol Universal (Tiangen) (Rio et al., 2010). The NanoDrop spectrophotometer (Thermo Fisher) and Agilent 2100 Bioanalyzer were used to evaluate the quality of extracted RNA (OD260/280 = 2.0–2.2, OD260/230 = 1.8–2.1, 28 s:18 s ≥ 1.5, RIN ≥ 8). A total of 25.54 μg RNA was eluted with nuclease-free water. The PacBio Iso-Seq protocol was followed for library preparation: total RNA was reverse transcribed into cDNA using the Clontech SMARTer PCR cDNA Synthesis Kit. The cDNA was amplified for library construction with an insert size between 0.5 and 6 kb after size-selection using the BluePippin (Sage). The sequencing was performed on one SMRT cell on the PacBio Sequel platform (Pacific Biosciences) in circular consensus sequencing (CCS) mode. Raw sequence data were filtered with the standard IsoSeq3 protocol before subsequent analysis, including circular consensus calling through ccs, clustering and polishing through isoseq3 (https://github.com/ylipacbio/IsoSeq3).

2.4 Genome structure and functional annotation

RepeatMasker version 1.331 (https://github.com/rmhubley/RepeatMasker, Smit et al., 2013–2015), a de novo repeat library builder by repeatmodeler version-open-1.0.11 (http://www.repeatmasker.org/RepeatModeler, Smit & Hubley, 2008–2015) and repbase (http://www.girinst.org/repbase) (Bao et al., 2015), was used to identify repetitive sequences and transposable elements. LTR_Finder v1.06 (-C -w 0) (Xu & Wang, 2007) and MITE-Hunter (-n
Gene structure prediction was performed using three strategies: homology-based prediction, ab initio prediction and transcriptome-based prediction. For homology-based prediction, GEMOMA v1.6.1 (Keilwagen et al., 2016, 2019) was used with default parameters and the published genomes from GenBank, including two model insects (Bombbyx mori and Drosophila melanogaster) (Adams et al., 2000; Xia et al., 2004) and five hymenopteran species (Apis mellifera, Microplitis demolitor, Fopius arisanus, Nasonia vitripennis and Ceratosolen solmsi) (Table S1). We used each of the seven species as reference for homology-based prediction. The prediction based on Apis mellifera generated the highest BUSCO values and was used for the final integration with EVidenceModeler (EVM). Augustus v3.3.1 (--gff3=on --hintsfile=hints.gff --extrinsicCfgFile=extrinsic.cfg --min_intron_len=30 --softmasking=1 --allow_hinted_splicesites=g-cag,atac) (Stanke & Waack, 2003) was used to predict genes in ab initio prediction, training parameters with the transcriptome data of A. gifuensis. Combining TransDecoder (http://transdecoder.github.io) with default parameters and RASA v2.3.3 (-c alignAssembly.config -C -R -g genome.fasta -T -u trans.fasta -t trans.clean.fasta -f fl.acc --ALIGNERS gmap) (Haas et al., 2003), the intact open reading frame (ORF) and retained de novo gene prediction were identified to improve the accuracy of prediction by using transcriptome data. Finally, we integrated the results of these three strategies using EVidenceModeler (EVM) v1.1.1 (--segmentSize 1000000 --overlapSize 1000000) (Haas et al., 2008).

Gene functional annotation was performed based on homologous searches and the best match to the databases of KEGG (Kanehisa & Goto, 2000; Ogata et al., 1999), KOG (Tatusov et al., 2001), Swiss-Prot (Bairocha & Apweiler, 2000) and NCBI-NR using BLAST and HMMER analyses. To identify collinear gene blocks between A. gifuensis, Apis mellifera and Nasonia vitripennis, we exported coding sequences (CDs), searched syntenic genes and visualized the high-quality blocks using MCscan (Multiple Collinearity Scan Toolkit) in JCVI (https://github.com/tanghaibao/jcvl) with default parameters.

2.6 Gene family expansion and contraction analysis

To identify the gene family expansion and contraction, protein sequences for A. gifuensis and 13 hymenopteran species were obtained from GenBank and other databases (Table S1). CAFE (Computational Analysis of Family Evolution) v4.2.1 (-p 0.05 -t 10 -r 10000) (Hahn et al., 2007; Han et al., 2013) was used to compare generated gene family clusters, with a birth and death rate model estimated over a phylogeny. For gene families exhibiting significant expansion and contraction with Viterbi p-values < 0.05 (De Bie et al., 2006) in A. gifuensis genome, KEGG pathway enrichment and GO analysis were executed by r package clusterProfiler (Yu et al., 2012).

2.7 Gene family annotation

We manually annotated four gene families of cytochrome P450 monooxygenase (P450s), ATP-binding cassette (ABC) transporter, venom proteins and olfactory receptors (ORs). A set of orthologous protein sequences of related species (such as Apis mellifera, Aphidius ervi, Fopius arisanus and Lysiphlebus fabarum) from NCBI GenBank and Hidden Markov models (HMMs) were used as references for gene identification. BLAST v2.10.0 (E-value < 1e-5) and HMMER v3.3 were used to search the candidate genes in the A. gifuensis genome. Then, the results of BLAST and HMMER analyses were integrated through the bioinformatic pipeline BITACORA (full mode) (Vizuet al., 2020). The annotated candidate genes sequences were aligned using MAFFT v7.471 with G-INS-I strategy (Katoh & Standley, 2013).
We constructed the phylogenetic tree using IQ-TREE v1.6.11 (-m TEST -bb 1000 -alrt 1000).

3 | RESULTS AND DISCUSSION

3.1 | Genome sequencing and assembly

A total of 18.6 Gb clean reads were generated by the Illumina platform. The genome size is approximately 154.9 Mb based on K-mer frequency distribution analysis (Figure S1). For long-read genome sequencing, we obtained 10.9 Gb Nanopore reads after removing low-quality sequences, which corresponds to approximately 69.5-fold coverage of the A. gifuensis genome (Table 1). The mean and N50 length of filtered reads were 16.7 kb and 23 kb, respectively.

Under the correction and trimming of Canu and contig assembly of Smartdenovo, a 156.9 Mb draft genome was generated by de novo assembly, consisting of 136 contigs with a contig N50 of 3.93 Mb and longest contig of 10.4 Mb (Table 2). The average GC content was 26.5%. The resulting genome was slightly larger than the estimated size but smaller than most published genomes of parasitoid wasps. Our assembly of A. gifuensis genome using long-read Nanopore sequencing obtained a larger contig N50 of 3.93 Mb and longest contig of 10.4 Mb (Table 2). The average GC content was 26.5%. The resulting genome was slightly larger than the estimated size but smaller than most published genomes of parasitoid wasps. Our assembly of A. gifuensis genome using long-read Nanopore sequencing obtained a larger contig N50 compared to the other seven Braconidae genomes previously assembled using combined Illumina and PacBio sequencing methods (Table 3; Burke et al., 2018), in which only Microplitis demolitor and Cotesia vestalis had a scaffold N50 longer than 1 Mb (Burke et al., 2018). The other five genomes had contig N50 s ranging from 192.4 to 980 kb. Our results also highlight the importance of using long-read sequencing technology which largely increased the quality of assembly compared with solely using Illumina sequencing. Along with the other two Aphidinae species, Aphidius ervi and Lysiphlebus fabarum, the genomic GC content was also particularly low (<27%). There may be biological factors which further contribute to the low GC content. For some bacteria (Almpanis et al., 2018; Barahimipour et al., 2015; McCutcheon et al., 2009) and plant species (Smarda et al., 2014; Veleba et al., 2017), environmental conditions, such as limited nitrogen availability, could contribute to the low GC level. Other factors that may influence GC content include genome size and the extent of DNA methylation, which has been shown in vertebrate and hymenopteran insect like Polistes dominula (Mugal et al., 2015; Standage et al., 2016). The genomes of hymenopteran insect are in general characterized by relatively low GC levels and the gene conversion process and high recombination rates are hypothesized to contribute to GC bias (Branstetter et al., 2018). The smaller genome size and limited host species have been hypothesized to underline the extreme low GC content of the previously characterized aphid parasitoids (Dennis et al., 2020). We generated 80 Gb of Hi-C filtered data (543,717,626 paired-end reads) to construct a chromosome-level assembly (Figure S2). After mapping these reads onto the draft genome, 138,972,005 PE reads were retained, including 113,641,396 valid interaction PE reads. Of the 136 contigs, 99.8% of the sequence length could be anchored onto six linkage groups, with the length ranging from 18.38 to 32.14 Mb, resulting in a scaffold N50 of 27.48 Mb (Table S2).

We assessed the completeness of genome at chromosome level using BUSCO. We identified 1641 (98.97%) complete (single-copy genes: 94.75%, duplicated genes: 4.22%, 6 (0.36%) partial genes and 11 (0.66%) missing genes in the 1658 highly conserved Insecta data set (insecta_odb9) (Table S3). These results supported the high level of accuracy and completeness in the genome assembly.

3.2 | Gene annotation and phylogenetic analysis

A total of 45.34 Mb of repeat sequences was identified, constituting 28.89% of the A. gifuensis genome (Table S4; Fig S3). DNA transposons and retroelements accounted for 4.85% and 3.37% of the genome, respectively. For retroelements, 2.15% were classified as long terminal repeats (LTRs), 1.07% as long interspersed elements (LINEs), and 0.14% as short interspersed elements (SINEs). Genome size usually appears to be positively associated with abundance of its repeat content (Bosco et al., 2007; Hartl, 2000; Maumus et al., 2015; Yin et al., 2018). Among all published genomes of parasitoid wasps, the genome of Macrocentrus cingulum, which is the smallest (132.4 Mb), contains 24.9% repeat elements (Yin et al., 2018), whereas larger genomes, such as those of Nasonia vitripennis

| TABLE 1 | Sequencing statistics generated by different platforms for Aphidius gifuensis genome assembly |
| Platform | Library size (Gb) | Raw data (Gb) | Clean data (Gb) | Coverage (X) |
|---|---|---|---|---|
| Illumina | 350 bp | 25.2 | 18.6 | 118.5 |
| Nanopore | >10 kb | 11.8 | 10.9 | 69.5 |
| Hi-C | 400 bp | 82.7 | 80.0 | 509.9 |

| TABLE 2 | Genome assembly statistics of Aphidius gifuensis |
| Methods | Statistic | Value |
|---|---|---|
| Nanopore assembly | Genome size (Mb) | 156.9 |
| | Number of contigs | 136 |
| | Maximum contig size (bp) | 10,405,750 |
| | Contig N50 (bp)/L50 | 3,929,594/13 |
| | Contig N90 (bp)/L90 | 694,987/45 |
| | BUSCO (%) | 98.97 |
| Hi-C assembly | Number of pseudo-chromosomes/scaffolds | 6/118 |
| | Maximum scaffold size (bp) | 32,149,644 |
| | Scaffold N50 (bp)/L50 | 27,481,583/3 |
| | Total size (bp) | 155,454,520 |
| | BUSCO (%) | 98.90 |
Gene prediction results based on three strategies

TABLE 4

| Features                              | 
|---------------------------------------|
| Assembly level                        | Chromosome | Scaffold | Contig | Scaffold | Scaffold | Scaffold | Scaffold | Scaffold |
| Total length (Mb)                     | 156.9      | 138.8    | 140.7  | 178.5    | 384.4    | 153.6    | 132.3    | 241.2    |
| Number of scaffolds                   | 6          | 5743     | NA     | 1437     | 3313     | 1042     | 5696     | 1794     |
| Scaffold N50 (kb)                     | 27,480     | 581.4    | NA     | 2609.6   | 657.0    | 980.0    | 192.4    | 1,140    |
| Number of contigs                     | 13         | 12,948   | 1698   | 6820     | 24,824   | 8510     | 13,289   | 27,508   |
| Contig N50 (kb)                       | 3929       | 25.2     | 216.1  | 51.3     | 45.5     | 51.9     | 64.9     | 14.12    |
| Completeness (%)                      | 98.97      | 93.7     | 95.9   | 96.7     | 99.0     | 97.0     | 99.4     | 97.0     |
| GC content (%)                        | 26.5       | 25.8     | 23.8   | 29.9     | 38.3     | 39.4     | 35.6     | 33.1     |
| Repeat content (%)                    | 28.9       | 29.3     | 49.1   | 24.0     | 49.0     | NA       | 24.9     | 36.2     |
| Number of protein-coding genes        | 10,443     | 20,226   | 15,170 | 11,278   | 19,064   | 18,906   | 11,993   | 18,586   |

(295.8 Mb), *Diadromus collaris* (399.1 Mb) and *Diaclasma alloeum* (384.3 Mb), contain 42.1%, 37.0% and 49.0% repeat elements, respectively (Tvedte et al., 2019; Werren et al., 2010). However, an exception to this association can be found in *Lysiphlebus fabarum*, which has a much smaller genome size (140.7 Mb) but contains a high proportion of repeat sequences (49.1%) (Dennis et al., 2020).

Using three methods of gene prediction, 11,535 genes were annotated in the *A. gifuensis* genome (Table 4). The transcriptome-based method predicted fewer genes (6660 genes) compared to other wasps; for example, there are 11,278 genes in *Cotesia vestalis*, 15,328 genes in *Diaadromus collaris* (Shi, Ma, et al., 2019; Shi, Wang, et al., 2019) and 19,597 genes in *Microplitis demolitor* (Burke et al., 2018), which were all predicted from RNA-seq data sets. Fewer putative genes with RNA-seq prediction may be affected by the limited life stage tissues because only male adults were used for the transcriptome sequencing in our study. The average gene length was 5921.89 bp, and the average CDS length was 1694.13 bp. The average exon number per gene, average exon length and average intron length were 5.36, 347.58 bp and 926.29 bp, respectively. In addition, approximately 90.53% of the genes (10,443 genes) could be functionally annotated (Table 5). We identified 98.90% of the BUSCO Insecta database (insecta_odb9) genes (single-copy genes: 96.30%, duplicated genes: 2.60%; Table S3) at protein level, further underlining the accuracy and completeness of gene predictions. Different types of noncoding RNAs were also annotated, yielding 846 tRNA, 130 miRNA, 100 rRNA and 77 snRNA (Table S6). The annotated gene set of *A. gifuensis* was compared with other parasitoid wasps (Table S5). The number of annotated genes was relatively low despite the high assembly quality and completeness of the *A. gifuensis* genome. In general, the existence of gap with shorter assembly scaffold or contig could increase the pseudogenes or false-positive annotations (Li et al., 2019).

3.3 | Orthology, synteny and phylogenetic relationship

Along with 13 other hymenopteran species, we identified 13,139 gene family groups in total and assigned 10,302 genes to 8261 gene families in the *A. gifuensis* genome using OrthoMCL. 3504 single-copy genes identified were employed to reveal the phylogenetic relationships among these species (Figure 2a). As shown in the phylogenetic tree, *A. gifuensis* was more closely clustered with *Fopius arisanus* than with *Macrocentrus cingulum* and *Microplitis demolitor*. These four species formed a clade of Ichneumonidae, which diverged from Chalcidoidea at approximately 194.6 Mya. The inter-Parasitica phylogenetic relationships were consistent with previous studies (Peters et al., 2017; Tang et al., 2019).

We compared the orthologous analysis between *A. gifuensis* and other three parasitoid wasps, *Ceratosolen solmsi*, *Fopius arisanus* and *Macrocentrus cingulum*. 5754 homology gene families were shared by the four species. *A. gifuensis* shared 7536 gene families with *Fopius arisanus*, higher than 6874 with *Ceratosolen solmsi* and 6646 with *Macrocentrus cingulum* (Figure 2b), that showed more homology between *A. gifuensis* and *Fopius arisanus*.

Syntenic relationships between *A. gifuensis*, *Apis mellifera* and *Nasonia vitripennis* showed a high level of collinearity among the three
chromosome-level genomes from the Hymenoptera (Figure 2c). We defined a syntenic block as including at least three orthologous genes. In total, 542 syntenic blocks were found between A. gifuensis and Apis mellifera, and the gene number in these blocks ranged 4–23 with a mean of 5.97. 428 blocks were found between A. gifuensis and Nasonia vitripennis, with the same gene number range of 4–23 and a mean of 5.60. Syntenic analysis usually identifies putatively homologous genome regions by anchoring neighbouring gene pairs, which may influence differences in gene density, tandem duplication, gene transpositions and chromosomal rearrangements (Tang et al., 2008; Wang et al., 2012). In our analysis, A. gifuensis showed slightly higher synteny with Apis mellifera than Nasonia vitripennis, despite the closer phylogenetic relationship of A. gifuensis and Nasonia vitripennis, an observation which may be related to the above factors or to differing annotation qualities among these species.

3.4 | Gene family expansion and contraction

Using café, we estimated the gene family expansion and contraction in A. gifuensis genome, compared with 13 hymenopteran species (Table S1) in the phylogenetic analyses. Significant expansions and contractions of gene families are usually related to the adaptive evolution of the species (Wu, Zhang, et al., 2019; Zhang et al., 2020). In the genome of A. gifuensis, a total of 405 and 663 orthologous groups were significantly expanded and contracted, respectively (Viterbi p < .05), compared to the most recent common ancestor (MRCA) of A. gifuensis and Fopius arisanus (Figure 3a).

As shown by KEGG enrichment analysis among the expanded groups, several GO-terms were significantly over-represented: ABC transporters (02010, 17 genes, p = 7.76e-10), fatty acid biosynthesis (00061, 15 genes, p = 2.03e-7), AMPK signalling pathway (04152, 29 genes, p = 8.11e-7) and cell cycle (04112, 21 genes, p = 8.11e-7) (Figure 3b, Table S7). Based on GO analysis, some significantly enriched in metabolic process (GO:0008152, 316 genes, p = 1.11e-11), transmembrane transport (GO:0055085, 59 genes, p = 6.32e-6), chromosome organization (GO:0051276, 14 genes, p = .000112), signal transduction (GO:0007165, 18

### Table 5

| Database   | Number | Per cent (%) |
|------------|--------|--------------|
| KOG        | 7763   | 67.30        |
| KEGG       | 6185   | 53.62        |
| NR         | 10,332 | 89.57        |
| Swiss-Prot | 8877   | 76.96        |
| GO         | 5746   | 49.81        |
| **Total**  | 10,443 | 90.53        |

**FIGURE 2**  Phylogenetic tree, gene orthology and synteny blocks. (a) The phylogenetic tree was constructed based on 3504 single-copy gene families with 13 hymenopteran insects (shown in Table S1), using RAxML maximum-likelihood methods. Bootstrap values were 100 in all nodes based on 100 replicates. Bars are subdivided to represent different types of orthology with different colours. The red nodes indicate calibration times. (b) Venn diagram of the orthologous gene families from four parasitoid wasps. (c) Synteny blocks between Aphidius gifuensis, Apis mellifera and Nasonia vitripennis [Colour figure can be viewed at wileyonlinelibrary.com]
genes, $p = .00476$), cellular macromolecule biosynthesis process (GO:0034645, 80 genes, $p = .00679$), ATP binding (GO:0005524, 114 genes, $p = 1.22 \times 10^{-14}$) and transporter activity (GO:0005215, 63 genes, $p = 3.13 \times 10^{-5}$) (Table S8). Conversely, among the contracted gene groups, KEGG and GO analyses showed enrichments of olfactory receptor activity (GO:0004984, 5 genes, $p = .00773$), inorganic anion transport (GO:0015698, 3 genes, $p = .00138$), single-organism process (GO:0044765, 16 genes, $p = .0006$) and chromatin (GO:0000785, 2 genes, $p = .00334$) (Table S9, Table S10). In A. gifuensis, nutrients including proteins, carbohydrates and inorganic salts for larvae development are mainly absorbed from their hosts; therefore, cell transport and digestion are important.

According to the results of 

**FIGURE 3** Gene family evolution between genomes of *Aphidius gifuensis* and 13 other hymenopteran species. (a) Green indicates gene family expansions and red indicates gene family contractions. The length of branch indicated the divergence time. MRCA: Most Recent Common Ancestor. Mya, million years ago. (b) Significant results of KEGG enrichments analysis among expanded gene families. The value around each bar meant the number involved in each KEGG pathway. (c) GO classification of expanded gene families, including the top 20 significant GO categories ($p < .05$). The detail GO classification can be viewed in the Table S10. BP, Biological process; CC, Cellular component; MF, Molecular function [Colour figure can be viewed at wileyonlinelibrary.com]

**FIGURE 4** Phylogenetic relationships of cytochrome P450 (a) and ATP-binding cassette (ABC) transporter (b) in the *Aphidius gifuensis* genome. (a) Four main clades of P450 genes were indicated. The gene names in red indicated the expansion of CYP6 subfamily. (b) The gene names in red indicated the expansion of ABCA, ABCC and ABCG subfamily. The results of Café are shown in Table S11 [Colour figure can be viewed at wileyonlinelibrary.com]
**CONCLUSIONS**

In this study, we report a chromosome-level genome assembly of *A. gifuensis*, an important parasitoid biocontrol agent for multiple aphids. Genome assembly and annotation showed high completeness and continuity, with a contig N50 longer than most published genomes of parasitoid wasps. In addition, we identified genes putatively involved in insecticide resistance, parasitoid venom protein and chemosensing. The high-quality genome will provide a solid base for future studies on mechanism underlying parasitic biology, parasitoid-host interactions and currently population decline of this wasp in the artificial breeding, and will help to improve the commercial rearing and control efficiency of this wasp as an aphid control agent.

**ACKNOWLEDGEMENTS**

This work was supported by the China National Tobacco Corporation of Science and Technology Major Projects (Nos. 110202001036[LS-05] and 110201801023[LS-02]). We thank Prof. Xin Zhou, Dr. Shiqi Luo and Min Tang for their constructive comments on the genome assembly and annotation.

**AUTHOR CONTRIBUTIONS**

H. L., H. Y. and Y. Y designed the research. B. L., Z. D., S. W. and H. Y. designed the research. B. L., Z. D., L. T. and H. L. wrote the manuscript. All authors reviewed the manuscript.

**DATA AVAILABILITY STATEMENT**

All raw data have been deposited into the NCBI Sequence Read Archive (SRA) database with a BioProject accession PRJNA615161 and a BioSample accession SAMN14446980. The whole-genome sequencing and transcriptome sequencing data are also available under Accession no. SRR11568632, SRR11568634, SRR11578829, SRR11578830 and SRR11577123–SRR11577219. In addition, the genome assembly and annotation information are available under the Accession nos of JACMRX000000000 and GCA_014905175.1. The version described in this paper is version JACMRX010000000.

**ORCID**

Zhenyong Du https://orcid.org/0000-0002-4569-6713
Shujun Wei https://orcid.org/0000-0001-7398-0968
Hu Li https://orcid.org/0000-0001-8590-1753

**REFERENCES**

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Gemanitis, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., Sutton, G. G., Woram, J. R., Yandell, M. D., Zhang, Q. ... Venter, J. C. (2000). The genome sequence of *Drosophila melanogaster*. *Science*, 287(5461), 2185–2195. https://doi.org/10.1126/science.287.5461.2185

Ahmed, T., Zhang, T., Wang, Z., He, K., & Bai, S. (2016). Gene set of chemosensory receptors in the polyembryonic endoparasitoid *Macrocentrus cingulum*. *Scientific Reports*, 6(1), 24078. https://doi.org/10.1038/srep24078

Alpannis, A., Swain, M., Gatherer, D., & McEwan, N. (2018). Correlation between bacterial G+C content, genome size and the G+C content
of associated plasmids and bacteriophages. *Microbial Genomics*, 4(4), e000168. https://doi.org/10.1099/mgen.0.000168

Asgary, S., & Rivers, D. B. (2011). Venom proteins from endoparasitoid wasps and their role in host-parasite interactions. *Annual Review of Entomology*, 56, 313–335. https://doi.org/10.1146/annurev-ento-120709-144849

Bairocha, A., & Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Research*, 28(1), 45–48. https://doi.org/10.1093/nar/28.1.45

Bao, W., Kojima, K., & Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, 6, 11. https://doi.org/10.1186/s13100-015-0041-9

Barahimipour, R., Strenkert, D., Neupert, J., Schroda, M., Merchant, S. S., Asgari, S., & Rivers, D. B. (2011). Venom proteins from endoparasitoid wasps. *Insect Molecular Biology*, 12(5), 527–534. https://doi.org/10.1046/j.1365-2583.2003.00439.x

Danneels, E. L., Rivers, D. B., & de Graaf, D. C. (2010). Venom proteins of the parasitoid wasp *Nasonia vitripennis*: recent discovery of an untapped pharmacopee. Toxins, 2(4), 494–516. https://doi.org/10.3390/toxins2040494

De Bie, T., Cristianini, N., Demuth, J. P., & Hahn, M. W. (2006). CAFE: a computational tool for the study of gene family evolution. *Bioinformatics*, 22(10), 1269–1271. https://doi.org/10.1093/bioinformatics/btl097

Dennis, A. B., Ballesteros, G. I., Robin, S., Schrader, L., Bast, J., Berghöfer, J., Beukeboom, L. W., Belghazi, M., Bretaudeau, A., Buellesbach, J., Cash, E., Colinet, D., Dumas, Z., Erbii, M., Falabella, P., Gatti, J.-L., Geuerink, E., Gibson, J. D., Hertaeg, C., ... Gadau, J. (2020). Functional insights from the GC-poor genomes of two aphid parasitoids, *Aphidius ervi* and *Lysiphlebus fabarum*. BMC Genomics, 21(1), 376. https://doi.org/10.1186/s12864-020-6764-0

Fan, J., Zhang, Q., Xu, Q., Xue, W., Han, Z., Sun, J., & Chen, J. (2018). Differential expression analysis of olfactory genes based on a combination of sequencing platforms and behavioral investigations in *Aphidius gifuensis*. *Frontiers in Physiology*, 9, 1679. https://doi.org/10.3389/fphys.2018.01679

Feyereisen, R. (2006). Evolution of insect P450. *Biochemical Journal*, 39(3), 417–422. https://doi.org/10.1042/BJ20060017

Han, M. V., Thomas, G. W. C., Lugo-Martinez, J., & Hahn, M. W. (2013). Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Molecular Biology and Evolution*, 30(8), 1987–1997. https://doi.org/10.1093/molbev/mst100

Han, Y., & Wessler, S. R. (2010). **miteHUNTER**: A program for discovering miniature inverted-repeat transposable elements from genomic
Wu, C., Chakrabarty, S., Jin, M., Liu, K., & Xiao, Y. (2019). Insect ATP-Binding Cassette (ABC) transporters: roles in xenobiotic detoxification and Bt insecticidal activity. *International Journal of Molecular Sciences, 20*(11), 2829. https://doi.org/10.3390/ijms20112829

Wu, N., Zhang, S., Li, X., Cao, Y., Liu, X., Wang, Q., Liu, Q., Liu, H., Hu, X., Zhou, X. J., James, A. A., Zhang, Z., Huang, Y., & Zhan, S. (2019). Fall webworm genomes yield insights into rapid adaptation of invasive species. *Nature Ecology and Evolution, 3*(1), 105–115. https://doi.org/10.1038/s41559-018-0746-5

Xia, Q., Zhou, Z., Lu, C., Cheng, D., Dai, F., Li, B., Zhao, P., Zha, X., Cheng, T., Chai, C., Pan, G., Jinshan, X., Liu, C., Lin, Y., Qian, J., Hou, Y., Zhengli, W., Li, G., Pan, M., ... Yang, H. (2004). A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science, 306*(5703), 1937–1940. https://doi.org/10.1126/science.1102210

Xie, Y., Zhang, H., Li, Y., Kong, L., Xiang, M., Yang, H., & Zhang, L. (2020). Population degradation rule of *Aphidius gifuensis* (Hymenoptera:Aphidiidae). *Chinese Journal of Biological Control, 36*(2), 163–168. https://doi.org/10.16409/j.cnki.2095-039x.2020.02.002

Xu, Z., & Wang, H. (2007). LTR_FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research, 35*(Web Server), W265–W268. https://doi.org/10.1093/nar/gkm286

Yang, Z. (2007). *paml* 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution, 24*(8), 1586–1591. https://doi.org/10.1093/molbev/msm088

Yang, S., Wei, J., Yang, S., & Kuang, R. (2011). Current status and future trends of augmentative release of *Aphidius gifuensis* for control of *Myzus persicae* in China’s Yunnan Province. *Journal of the Entomological Research Society, 13*(3), 87–99.

Yang, S., Yang, S., Zhang, C., Wei, J., & Kuang, R. (2009). Population dynamics of *Myzus persicae* on tobacco in Yunnan Province, China, before and after augmentative releases of *Aphidius gifuensis*. *Biocontrol Science and Technology, 19*(2), 219–228. https://doi.org/10.1080/0958315080269625

Ye, X., Yan, Z., Yang, Y., Xiao, S., Chen, L., Wang, J., Wang, F., Xiong, S., Mei, Y., Wang, F., Yao, H., Song, Q., Li, F., Fang, Q., Werren, J. H., ... Ye, G. (2020). A chromosome-level genome assembly of the parasitoid wasp *Pteromalus puparum*. *Molecular Ecology Resources, 20*, 1384–1402. https://doi.org/10.1111/1755-0998.13206

Yin, C., Li, M., Hu, J., Lang, K., Chen, Q., Liu, J., Guo, D., He, K., Dong, Y., Luo, J., Song, Z., Walters, J. R., Zhang, W., Li, F., & Chen, X. (2018). The genomic features of parasitism, Polyembryony and immune evasion in the endoparasitic wasp *Macrocentrus cingulat*. *BMC Genomics, 19*(1), 420. https://doi.org/10.1186/s12864-018-4783-x

Yu, G., Wang, L., Han, Y., & He, Q. (2012). *clusterProfiler*: An R package for comparing biological themes among gene clusters. *OMICS: A Journal of Integrative Biology, 16*(5), 284–287. https://doi.org/10.1089/omi.2011.0118

Zdobnov, E. M., & Apweiler, R. (2001). InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics, 17*(9), 847–848. https://doi.org/10.1093/bioinformatics/17.9.847

Zhang, H.-Z., Li, Y.-Y., An, T., Huang, F.-X., Wang, M.-Q., Liu, C.-X., Mao, J.-J., & Zhang, L.-S. (2018). Comparative transcriptome and iTRAQ proteome analyses reveal the mechanisms of diapause in *Aphidius gifuensis* Ashmead (Hymenoptera: Aphidiidae). *Frontiers in Physiology, 9*, 1697. https://doi.org/10.3389/fphys.2018.01697

Zhang, S., Shen, S., Peng, J., Zhou, X., Kong, X., Ren, P., Liu, F., Han, L., Zhan, S., Huang, Y., Zhang, A., & Zhang, Z. (2020). Chromosome-level genome assembly of an important pine defoliator, *Dendrolimus punctatus* (Lepidoptera; Lasiocampidae). *Molecular Ecology Resources, 20*(4), 1023–1037. https://doi.org/10.1111/1755-0998.13169

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Li B, Du Z, Tian L, et al. Chromosome-level genome assembly of the aphid parasitoid *Aphidius gifuensis* using Oxford Nanopore sequencing and Hi-C technology. *Mol Ecol Resour*. 2021;21:941–954. https://doi.org/10.1111/1755-0998.13308