Chromosome-level genome assembly of the horned-gall aphid, Schlechtendalia chinensis (Hemiptera: Aphididae: Erisomatinae)

Hongyuan Wei[1], Yu-Xuan Ye[2], Hai-Jian Huang[3], Ming-Shun Chen[4], Zi-Xiang Yang[5], Xiaoming Chen[5], and Chuan-Xi Zhang[2]

[1]Chinese Academy of Forestry
[2]Zhejiang University
[3]Ningbo University
[4]Kansas State University
[5]Research Institute of Resource Insects, Chinese Academy of Forestry

February 22, 2024

Abstract

The horned gall aphid Schlechtendalia chinensis, is an economically important insect that induces galls valuable for medicinal and chemical industries. S. chinensis manipulates its host plant to form well-organized horned galls during feeding. So far, more than twenty aphid genomes have been reported; however, all of those are derived from free-living aphids. Here we generated a high-quality genome assembly of S. chinensis, representing the first genome sequence of a galling aphid. The final genome assembly was 280.43 Mb, with 97% of the assembled sequences anchored into thirteen chromosomes. S. chinensis presents the smallest aphid genome size among available aphid genomes to date. The contig and scaffold N50 values were 3.39 Mb and 20.58 Mb, respectively. The assembly included 96.4% of conserved arthropod and 97.8% of conserved Hemiptera single-copy orthologous genes based on BUSCO analysis. A total of 13,437 protein-coding genes were predicted. Phylogenomic analysis showed that S. chinensis formed a single clade between the Eriosoma lanigerum clade and the Aphidini+Macrosiphini aphid clades. In addition, salivary proteins were found to be differentially expressed when S. chinensis underwent host alternation, indicating their potential roles in gall formation and plant defense suppression. A total of 36 cytochrome P450 genes were identified in S. chinensis, considerably fewer compared to other aphids, probably due to its small host plant range. The high-quality S. chinensis genome assembly and annotation provide an essential genetic background for future studies to reveal the mechanism of gall formation and to explore the interaction between aphids and their host plants.

Chromosome-level genome assembly of the horned-gall aphid, *Schlechtendalia chinensis*(Hemiptera: Aphididae: Eriosomatinae)

Hong-Yuan Wei[1], Yu-Xuan Ye[2], Hai-Jian Huang[4], Ming-Shun Chen[3], Zi-Xiang Yang[1*], Xiao-Ming Chen[1*], Chuan-Xi Zhang[2,4*]

[1]Research Institute of Resource Insects, Chinese Academy of Forestry, Kunming, China

[2]Institute of Insect Sciences, Zhejiang University, Hangzhou, China

[3]Department of Entomology, Kansas State University, Manhattan, KS, USA

[4]State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products; Key Laboratory of Biotechnology in Plant Protection of MOA of China and Zhejiang Province, Institute of Plant Virology, Ningbo University, Ningbo, China

Contributed equally.

*Correspondence

Zi-Xiang Yang, Research Institute of Resource Insects, Chinese Academy of Forestry, Kunming, China.

E-mail: yzx1019@163.com

Xiao-Ming Chen, Research Institute of Resource Insects, Chinese Academy of Forestry, Kunming, China.

E-mail: cafcxm@139.com

Chuan-Xi Zhang, Institute of Insect Sciences, Zhejiang University, Hangzhou, China.

E-mail: chxzhang@zju.edu.cn

Funding information

National natural science foundation of China (31872305); the basic research program of Yunnan Province (202001AT070016); the grant for innovative team of Yunnan Province (202005AE160011).

Abstract

The aphid *Schlechtendalia chinensis* is an economically important insect that induces horned galls, which are valuable for medicinal and chemical industries. So far, more than twenty aphid genomes have been reported. However, most of those sequenced genomes are derived from free-living aphids. Here we generated a high-quality genome assembly form a galling aphid. The final genome assembly is 271.52 Mb, with 97% of the assembled sequences anchored into thirteen chromosomes. *S. chinensis* represents one of the smallest aphid genomes sequenced to date. The contig and scaffold N50 values were 3.77 Mb and 20.41 Mb, respectively. The assembly included 96.9% of conserved arthropod and 98.5% of conserved Hemiptera single-copy orthologous genes based on BUSCO analysis. A total of 14,089 protein-coding genes were predicted. Phylogenomic analysis showed that*S. chinensis* diverged from the common ancestor of *Eriosoma lanigerum* at approximately 57.16 million years ago. In addition, some genes encoding salivary gland proteins were found expressed differentially when *S. chinensis* forms a gall, indicating their potential roles in gall formation and plant defense suppression. The high-quality *S. chinensis* genome assembly and annotation provide a solid genetic background for future studies to reveal the mechanism of gall formation and to explore the interaction between aphids and their host plants.

Key words

Schlechtendalia chinensis ; PacBio sequencing; Chromosome-level genome assembly; Comparative genomics; Gall formation; Host alternation

1 Introduction

Many aphid species are economically important plant pests that feed on plant sap and also caused damage by transmitting plant viruses. Around 100 aphid species have been identified as significant agricultural pests among the approximately 5,000 known species (Blackman & Eastop, 2020). Up to this point, studies on aphid genomes have mostly focused on the subfamily Aphidinae (International Aphid Genomics Consortium, 2010; Li et al., 2019; Mathers, 2020; Mathers et al., 2017; Mathers, Mugford, et al., 2020; Mathers, Wouters, et al., 2020; Nicholson et al., 2015; Thorpe et al., 2018; Wenger et al., 2016). Genome sequencing efforts on other subfamilies that are distantly related to Aphidinae are very limited (Julca et al., 2020; Biallo et al., 2020). Unlike most free-living aphids, galling aphids induce the formation of galls on their primary host plants and live in the galls. Galling aphids may serve as good models for the study of unique ecological and behavioral phenomena for insect-plant interaction and coevolution (Moran, 1989; Wool, 2004). Up to now, galling aphids whose genomes have been sequenced include *Eriosoma lanigerum* and *Hormaphis corn* . The aphid*E. lanigerum* often causes bark deformation and cancer-like swellings on roots, trunk or brunches of apple, and sometimes induces leaf-rosette galls on American elm (*Ulmus americana*) (Blackman and Eastop, 2020). Another aphid *H. corn* induces a gall which has an ostiole on the underside of the leaf (Kurosu et al., 1992). The galls induced by *E. lanigerum* and *H. corn* are quite different to the completely closed gall

since its aphids have peculiar strategies to adapt the specific characteristics such as high CO_2concentration, honeydew treatment and nutrition exchange of the closed internal environment (Chen et al., 2020).

The horned gall aphid,*Schlechtendalia chinensis* (Hemiptera: Aphididae: Eriosomatinae: Fordini), is one of the most economically valuable insects. Its gallnuts are valuable for medicinal purposes and in chemical industries. Some of the gallnuts components such as tannins are useful in producing inks, wine, food, cosmetic antioxidants, and animal feed. High levels of tannins ranged from 50 to 70% have been found in horned galls (Zhang, Tang, & Cheng, 2008). The annual yield of gallnuts in China is 8,000-10,000 tons, accounting for >90% of the total yield worldwide (Zhang, Tang, & Cheng, 2008).

S. chinensis has a complex life cycle involving both sexual and asexual reproduction stages with a host alternation between the Chinese sumac (*Rhus chinensis* , Anacardiaceae) and mosses (*Plagiomnium spp.*Mniaceae). In this holocyclic life cycle, a fundatrix produced by a mated female crawls along the trunk and feeds on a new leaf, where it induces the formation of a horned-gall. The fundatrix produces wingless fundatrigeniae via parthenogenesis in gall. In autumn, wingless fundatrigeniae produce winged fundatrigeniae named autumn migrants. When galls become mature and burst open, alate autumn migrants fly to nearby mosses and produce nymphs for overwintering. In the following spring, nymphs on mosses develop into spring migrants that fly back to the primary host, *R. chinensis* and produce female and male offspring (sexuales). After mating, each female reproduces only a fundatrix, starting the cycle again (Figure 1) (Zhang, Qiao, Zhong & Zhang, 1999; Blackman and Eastop, 2020). This unusual life cycle with various morphologically distant aphids at different stages is likely driven by adaptation to different environmental conditions. Unlike most free-living aphids from the Aphidinae taxon, galling aphids have many distinct biological characteristics. The most striking characteristic is that the feeding of most galling aphids does not seriously affect the health of their host plants. In fact, the formation of galls can provide some benefits to the primary host plant (Chen et al., 2020).

The complexity both in its developmental process and in the structure of its induced galls implies that *S. chinensis* may have unique gene sets that regulate its development and manipulate its host plants (Takeda et al., 2019; Hirano et al., 2020). The underlying molecular mechanisms for its complex life cycle remain largely unknown. Galls result from dramatic reprogramming of plant cell biology driven by insect molecules. Previous studies have shown that gall induction is highly, species-specific and different galling insects deliver unique sets of effectors into plant tissues, resulting in gall formation (Zhao et al., 2015; Alibory et al., 2018). The underlying mechanisms for the ability of the galling aphid to parasitize on host plants via apparently harmless galls remained unknown either. To understand the genetic basis of the complex lifestyle, we generated a high-quality chromosome-level genome assembly of *S. chinensis* , representing the first genome sequence of aphids that induces the formation of completely closed galls. In addition, we analyzed the phylogenetic relationship between*S. chinensis* and closely related species to give a better understanding of the unique biological characteristics of *S. chinensis* , such as suppressing plant defense and inducing gall formation.

2 Materials and Methods

2.1 Sample collection

S. chinensis samples were collected in October, 2019, from fresh mature galls on *R. chinensis* , in Wufeng county (30°10' N,110deg52' E,960 m above sea level), Hubei Province, China. *S. chinensis* individual for sequencing were from a single gall for PacBio sequencing, Illumina sequencing, RNA-seq and Hi-C analysis. All aphids within the gall were presumed to be clonal offspring of a single fundatrix, because all*S. chinensis* galls have contained only a single fundatrix and the fundatrix produces offspring in the gall via parthenogenesis. Fundatrigeniae (female) contained within a gall were moved to a petri dish after dissecting the gall. Impurities such as waxes were removed manually. Samples were separated into randomly groups with each group containing about 20 individuals. The samples were immediately frozen in liquid nitrogen for 2 hours and then stored at -80 f for later analysis.

In addition to the original samples, a colony was established through artificial cultivation for other genetic

studies. To establish a colony, autumn migrants of *S. chinensis* were collected from mature galls and transferred to a nursery of the moss *Plagiomnium maximoviczii* , which were maintained in a greenhouse. In the following year, spring migrants (sexuparae) were collected from the mosses and cultivated in lab. Aphids were transferred to host trees after fundatrix emergence for gall induction. Aphid samples of different stages including fundatrix, fundatrigeniae, autumn migrants, overwinter nymphs, and spring migrants as well as male and female sexuales were collected separately. All aphid samples were immediately frozen in liquid nitrogen for two hours and then stored at -80 until further analysis.

2.2 Genomic and transcriptomic sequencing

Genomic DNA was isolated from aphid samples using a DNeasy Blood & Tissue Extraction Kit (Qiagen Inc., Valencia, CA, USA) by following the manufacturer's instructions. After the measurement of quality and quantity, the DNA samples were used for making a paired-end sequencing library (150 bp in length). The library was sequenced using the Illumina NovaSeq6000 platform for genome size assessment. In addition, a 20 kb library was constructed and sequenced using the PacBio RSII platform at Annoroad Gene Technology Co., Ltd. (Beijing, China). Hi-C libraries were also constructed from aphid samples according to the Proximo Hi-C procedure and its quality and concentration were determined via an Agilent 2100 Bioanalyzer and qPCR. Briefly, aphid samples were mixed with 1% formaldehyde for 10 min at room temperature and the nuclei were extracted and permeabilized. After the quality of the library was ascertained, different libraries were pooled to achieve required concentrations for Illumina sequencing (Rao et al., 2014).

Transcriptomes were generated from RNA samples extracted from the fundatrix, the fundatrigeniae, autumn migrants, nymphs, spring migrants (sexuparae), male and female sexuales, respectively, following the standard protocols provided by the manufacturer. RNA quantity, purity and integrity were determined by a NanoPhotometer and the Agilent 2100 Bioanalyzer. cDNA libraries were built following the chain specific method. The libraries were initially quantified with qubit 2.0 fluorometer and diluted to 1.5 ng/ul. Different libraries were pooled according to the requirements of effective concentration and target data volume for Illumina sequencing. Low-quality bases in the RNA-Seq raw reads were first filtered using Trimmomatic (version 0.36) (Bolger, Lohse, & Usadel, 2014). Then, the clean reads were mapped to the genome assembly using Hisat2 (version2.1.0.5) (Kim et al., 2015) to obtain putative transcripts. The transcripts gene expression was analyzed by using cufflinks (version2.2.1) (Ghosh, & Chan, 2016).

2.3 Genome assembly

We first estimated the genome size using Illumina data. We selected a k-mer length of 17 bases and used Illumina paired end reads for k-mer analysis to estimate the genome size and heterozygosity. The k-mer number and distribution were calculated by Jellyfish (version 1.1.10, parameters set to -C, -m 17, -s 10G, -t 80) and GenomeScope (version 2.0, parameters set to 12, 150) counted and visualized genomic information (Ranallo-Benavidez, Jaron, & Schatz, 2020, Marcais & Kingsford, 2011). Then, Pacbio sequencing data were used to assemble the draft genome using Wtdbg2 (version 2.5, parameters set to -t 8, -p 21, -S 4, -s 0.05, -g 274m, -L 5000) (Ruan & Li, 2020). Long reads were used to correct sequencing errors using NextPolish (Hu, Fang, Su, & Liu, 2019). In addition, Illumina sequencing data was mapped to draft genome assembly using bowtie2 (version 2.4.4, parameters set to score-min L, -0.3, -0.3 -p 8 -I 0 -X 1000) and was used for error correction in Pilon (version 1.23, with default parameters) (Walker et al., 2014). Finally, HaploMerger2 (set default parameters) and purge_haplotigs (parameters set to -m 4G; -t 60; -l value1, -m value2, -h value3; -t 60, -a 70) was used to remove the heterozygous regions in the genome (Huang, Kang, & Xu, 2017, Roach, Schmidt, & Borneman, 2018).

To construct the chromosome-level genome assembly, Hi-C sequences were aligned with the draft genome assembly using Juicer (version 1.5, with default parameters). An initial assembly was generated via a 3D de novo assembly (3D-DNA) (version 180114) analysis with parameter "-r 3" (Dudchenko et al., 2017). The initial assembly was reviewed using Juicebox Assembly Tools (JBAT, version 1.11.0, with default parameters) (Dudchenko et al., 2018), resulting in a finally chromosome-level genome assembly. The completeness of genome assembly was assessed using BUSCO (v5.1.3) (Waterhouse et al., 2018) to scan universal single-copy

orthologous genes selected from Eukaryota, Arthropoda, Insecta and Hemiptera datasets (odb_10). The final assembly was validated using the Illumina short reads and RNA sequencing (RNA-seq) reads. The reads were aligned against the assembled genome sequence using Hisat2 (version2.1.0.5) (Kim et al., 2015).

2.4 Locating the sex chromosomes and autosomes with short reads

Male and female DNA reads were mapped separately to the genomic scaffolds using Bowtie2 with default parameters (Langmead & Salzberg, 2012). The resulting alignments were filtered to remove the low-quality mapped reads by SAMtools (view -b -q 30). The read counts of each chromosome were calculated by using SAMtools idxstats (Li et al., 2009). The mapped reads per million (MRPM) of each chromosome was calculated for both female reads and male reads. Syntenic blocks of genes were identified between the chromosome-level genome assemblies of *S. chinensis* , *Acyrthosiphon pisum* , *Rhopalosiphum maidis* ,*E. lanigerum* use TBtools (version 1.09, Chen et al., 2020).

2.5 Gene annotation

To predict repetitive regions, RepeatMasker (version 4.1.1) (Tarailo-Graovac & Chen, 2009) was used to screen the *S. chinensis* genome against the Hemiptera ch repeat database, set this parameter to RepeatMasker -pa 4 -e ncbi -species Hemiptera ch -dir. To predict transposons and repetitive regions, an aphid-specific database was generated using RepeatModeler (version 2.0.1, set to default parameters) (Flynn et al., 2020). Statistical results of RepeatMasker and Repeatmodeler analyses were combined. For the RNA-seq assisted method, RNA-seq data generated from Illumina were aligned to the *S. chinensis* genome using Hisat2 (version2.1.0.5) (Kim et al., 2015). RNA-seq evidence was used for gene structure predictions using GETA (version 2.4.2). Gene structures were also predicted based on homology to those from *E. lanigerum* , *Ac. pisum* , *Myzus persicae* , *Aphis glycines* , *R. maidis* by genewise (version 2.4.1) (Birney, Michele, & Durbin, 2004). Comprehensive gene prediction results of RNA-seq and homologous proteins were used to generate accurate and complete gene models for Augustus (version 2.5.5) (Stanke et al., 2006) training, Augustus was used to perform gene prediction (Stanke et al., 2006; Blanco, Parra & Guigo, 2007). Finally, gene prediction results were integrated and sifted by PFAM database. We aligned the genes to seven functional databases to annotate genes in the*S. chinensis* genome using BLASTP with an E-value cutoff of $1x10^{-5}$. The databases used in the study were NCBI Non-Redundant Protein Sequence (Nr), Non-Redundant Nucleotide Sequence Database (Nt), SwissProt, Cluster of Orthologous Groups for eukaryotic complete genomes (KOG), The Integrated Resource of Protein Domains and Functional Sites (InterPro), Conserved Domain Database (CDD), Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes, Orthology database (KEGG) and evolutionary genealogy of genes: Non-supervised Orthologous Groups (eggNOG). Potential sequences form bacteria, fungi and other microorganisms were removed by aligning the genome sequences to the Nt database. A localBlast2GO database was built for GO annotation and was processed via Blast2GO (version 2.5). KAAS of KEGG was used to annotate *S. chinensis* genome sequence quickly, and the pattern of BBH was chosen.

2.6 Non-coding RNA identification

Transfer RNAs (tRNAs) were identified using the tRNAscan-SE (version 1.3.1, with default parameters for eukaryotes) program (Chan & Lowe, 2019). RNAmmer (version 1.2) with parameter "-s euk -m tsu, ssu, lsu" was used to identify 5S/ 8S, 16S/ 18S and 23S/ 28S rRNA (Karin et al., 2007). Ribosomal RNAs (rRNAs), microRNAs (miRNAs) and small nuclear RNAs (snRNAs) were identified based on the Rfam database (version 12.2) using BLASTN (E-value [?]$1x10^{-5}$) (Kalvari et al., 2017).

2.7 Phylogenetic analysis

We constructed phylogenetic trees using whole-genome sequences of*S. chinensis* and eight other aphid species including*Daktulosphaira vitifoliae* ,*Sipha flava* , *Aphis glycines* , *R. maidis* , *A. pisum* , *Myzus persicae* , *Diuraphis noxia* , *E. lanigerum* . The whitefly, *Bemisia tabaci* was used as the outgroup. The aphid genome sequence and gene structure annotation files were downloaded from the NCBI genome database, and genes containing mRNA information were retained and the CDS was modified. The longest sequence is

selected as the representative sequence. Finally, the protein and CDS sequences of all genes were obtained. Orthologous groups were assigned using OrthMCL (v2.0.9) (Li, Stoeckert & Roos, 2003) based on all-versus-all BLASTP results (E-value [?]1x10^{-5}). Single-copy orthologs OrthoMCL results were used to extract single copy ortholog groups according to some standards: as long as it appears in 50% of species, it is considered to be a single copy gene. If the shortest sequence of the single copy ortholog group is greater than 6000, the gene family is filtered out. Multi-sequence alignments of single copy ortholog genes were performed using MAFFT (version 7.221, Katoh, Misawa, Kuma, & Miyata, 2002; Katoh & Standley, 2013) and conserved amino-acid sites were identified by Gblocks (version 0.91, Clore, 2014). RAxML (version 8.1.24) (Stamatakis 2014) were used to construct the phylogenetic tree under the GTRGAMMA model with 1000 bootstrap replicates (Castresana, 2000). The branch length of homologous genes was analyzed using PAML (Yang, 2007), and then compared with the standard tree to eliminate abnormal genes. Then the tree was built using RAxML again (Stamatakis, 2014). By providing the root number and multiple sequence alignment results with calibration point information, species divergence time was calculated using mcmctree (a part of the PAML software, version 14.9). Divergence time within the evolutionary tree was obtained with 95% confidence (Yang, 2007). Divergence times and ages of fossil records were derived from TimeTree (http://www.timetree.org/) and applied as calibration point. The nodal dates of *Ac. pisum* and *Ap. glycines* were 28-61 MYA, *D. vitifoliae* and *S. flava* were 87-162 MYA and *B. tabaci* and*D. vitifoliae* were 245- 351 MYA according to the divergence times from TimeTree (Johnson et al., 2018). **2.8 Gene family expansion and contraction**

We used CAFE (version 3.1) (Hahn et al., 2007) to analyze the gene family expansion and contraction by comparing our genome with those from 9 other aphids (*D. vitifoliae* , *S. flava, E. lanigerum, Ap. glycines, R. maidis, Ac. pisum, D. noxia* and *M. persicae*). Briefly, the quantitative information of gene families of 10 insects was obtained according to OrthoMCL results. The number of gene families in each species and the trees with divergence time were used as the input information of CAFE (with parameter "lambda -s, -t"), best rates for gene birth and death were decided by CAFE and all branches have the same rate of gene birth and death. Expansion and contraction of gene families were measured by CAFE (Hahn, Demuth & Han, 2007). The GO and KEGG enrichment analyses were conducted using Omicshare CloudTools under this tool's default instructions (http://www.omicshare.com/).

2.9 Identification of genes potentially involved in gall formation and host manipulation

A total of 141 proteins have been identified from saliva of *S. chinensis* in a previous study (Yang et al., 2018). BLASTX was used to search the corresponding genes in *S. chinensis* genome with the 141 salivary proteins as queries (E-value [?]1x10-5, identify [?] 50). The expression of salivary protein-encoding genes was analyzed in three stages based on RNA-seq data. The GO and KEGG enrichment analyses highly expressed genes of fundatrix were conducted using Omicshare CloudTools under this tool's default instructions (http://www.omicshare.com/).

3 Results and Discussion

3.1 Genome sequencing and *de novo* assembly

Sequencing of the fundatrigenia genome using the PacBio PS II platform generated 130 Gb of raw data with N50 21,033. The raw contig-level assembly was comprised of 304,774,269 bases with 1,409 contigs and N50 2,961,835 (Table 1). The k-mer (K=17) analysis indicated that the heterozygosity of *S. chinensis* was 0.79% and the estimated genome size was 273,985,190 bp (Figure S2). The contig-level assembly comprised 271,416,320 bp with 378 contigs, and N50 4,333,385 after removing the heterozygosity (Table 1).

The chromosome-level genome was assembled into a total length of 271,524,833 bp, with a scaffold N50 20,405,002 using PacBio and Hi-C data (Table 1, S1). More than 97% of the total genome bases were successfully anchored to 13 chromosomes containing 97.2% of the total sequences. The chromosomic lengths ranged from 14,859,000 bp to 10,104,278 bp. Three hundred and forty-one small scaffolds make up the 2.8% of the total genome (Table 1; Figure 2, 3A). BUSCO analyses against Eukaryota, Arthropoda, Insecta and Hemiptera datasets were performed. *S. chinensis* genome assembly contains the highest number of conserved

single-copy Arthropoda genes of any published aphid genome, suggesting the completeness and high quality of our genome assembly (Figure 4A). The reads mapped to the assembled genome sequences with 97.70% mapping rate and 20 G average sequence depth (Table S2), and more than 86% of the assembled RNA-seq transcripts mapped to the genome (Table S3).

3.2 Sex chromosomes and autosomes

The male and female Illumina paired-end DNA reads were mapped separately to the genomic scaffolds to estimate the MRPM. In chrX1, chrX2, chrX3, the MRPM of female reads length is 1,439,092, 1,333,387 and 1,051,602 and the MRPM of female reads length is 781,901, 726,210 and 576,946 respectively. The MRPM of female reads was nearly twice as high as that of male reads in chrX1, chrX2 and chrX3. The other 10 chromosomes exhibited no significant differences between females and males, with female/male ratios ranging from 0.90 to 1.00 (Table S5). It has been shown that the X chromosome is conserved in aphids while chromosomal rearrangements are common on autosomes (Li et al. 2021, Mathers et al. 2021). The syntenic block between the *S. chinensis* assembly and*Ac. pisum* from Macrosiphini (Li, et al., 2020), *R. maidis*from Aphidini (Chen et al. 2019), and *E. lanigerum* (Figure 3C) from Eriosomatinae were identified. All comparisons reveal high levels of genome rearrangements between autosomes. Interestingly, three*S. chinensis* chromosomes were mapped the conserved X chromosome of Macrosiphini and Aphidini, and two X chromosomes of *E. lanigerum* . The observed multiple X chromosomes are consistent with previous reports (Biello et al., 2020). We speculate that it may be fragmentation of the X chromosome in *S. chinensis* and *E. lanigerum* or the result of an ancient fusion event of the large X chromosome in Aphidinae (Macrosiphini + Aphidini). The above evidence strongly suggests that chrX1, chrX2 and chrX3 is sex chromosomes and the karyotype of *S. chinensis* is XX+X.

3.3 Genome annotation

A total of 124.22 Gb raw data was generated by the Illumina platform. A total of 261 transcripts (280,520,495 bp in total) were generated by Trinity (Table S4). A total of 79,136,004 bp repetitive sequences were obtained in the *S. chinensis* genome and the proportion of repeats was 27.14% (Table S6). In total, the number of predicted protein-coding genes was 14,089 (15,987 transcripts). A total of 97.37% of the annotated genes were located on the 13 chromosome-level scaffolds. The average CDS length, exons number per gene, exon length and intron length were 1,536 bp, 7.3, 212 bp and 910 bp, respectively, similar to most of those reported aphid species (Table S7, Figure S1). The results that 96.9%, 97.7%, 97.8% and 96.7% of BUSCO genome/gene set could be identified in the *S. chinensis* genome in comparison with the Eukaryota, Arthropod, Hemiptera and Insecta datasets showed completeness of the gene set (Figure 4B). The percentage of RNA-Seq reads assigned to a gene feature up to 90%.

Among the 14,078 predicted genes, 12,584 (89.31%) genes were annotated functionally. This was based on the combination of 8,739 (65.81%) genes found via GO database and 6,866 (51.71%) genes present in the KEGG database (Table 2). Additionally, the non-coding RNAs in the *S. chinensis* genome, including 128 tRNAs, 32 rRNAs, 29 miRNAs, and 81 snRNAs were identified (Table S8).

3.4 Phylogenomic analysis

To explore this new genome assembly in a phylogenetic context and to investigate gene family evolution among aphids, the proteins of *S. chinensis* derived from the complete set of annotated protein coding genes were compared to the proteins form nine other insect species with fully sequenced genome. The corresponding proteins from the *B. tabaci* genome were used to root the tree. A total of 3479 single copy ortholog groups extracted by OrthoMCL were used to construct the phylogenetic tree. The results showed that *S. chinensis* is a sister taxon to the wooly apple aphid *E. lanigerum* . The two Eriosomatinae species diverged from their common ancestor at approximately 57.16 million years ago (Figure 5). Eriosomatinae and Aphidinae (including *Ap. glycines* , *R. maidis* , *Ac. pisum* ,*M. persicae* or *D. noxia*) may have diverged from a common ancestor about 63.22 Mya ago. The results are similar to the previous reports (Mather et al., 2020). The subfamily Eriosomatinae has a closer relationship with the subfamily Aphidinae, than the subfamily Chaitophorinae (including *S. flava*) in the family Aphididae. Significant expansion or contraction of gene

families is often related to adaptive divergence of species. To elucidate key genomic changes associated with adaptation, significantly expanded and contracted of gene families were analyzed in all the nine aphids and *B. tabaci* . Eriosomatinae showed 40 expanded and 986 contracted gene families compared to those of the common ancestor of Aphidinae and Eriosomatinae (Figure S3A). KEGG and GO annotations suggest that most of the expanded genes were involved in the detoxification of natural xenobiotics from plants (Figure S3B, S3C). Gene family evolution analysis indicated that the *S. chinensis* genome displayed 235 expanded and 1,037 contracted gene families compared with gene families of the common ancestor of *S. chinensis* and *E. lanigerum* . The KEGG annotations suggest that most of the expanded genes were involved in IL-17 signaling pathway, arachidonic acid metabolism, NF-kappa B signaling pathway, ovarian steroidogenesis, VEGF signaling pathway, necroptosis, regulation of lipolysis in adipocyte, TNF signaling pathway, and c-type lectin receptor signaling pathway (Figure S3E). The GO annotations suggest that most of the expanded genes were involved in prostaglandin-endoperoxide synthase activity, arachidonate 15-lipoxygenase activity, nuclear nucleosome, ovarian cumulus expansion, intrinsic apoptotic signaling pathway in response to osmotic stress, regulation of fever generation, regulation of platelet-derived growth factor production, response to lead ion, chromatin assembly or disassembly (Figure S3D, Table S9). *S. chinensis* expanded gene families were enriched not only for detoxification but also in immune system.

3.4 Salivary protein-encoding genes and other gall formation associated genes

S. chinensis can induce closed galls on host plants. Previous studies have shown that gall induction is highly species-specific and galling insects deliver effectors into plant tissues, resulting in gall formation. The gall midge *Mayetiola destructor* can inject effector proteins into tissues through its saliva during feeding, resulting in converting a whole wheat seedling into a gall (Wang et al., 2018; Aljbory et al., 2020). A novel family of insect secreted proteins named BICYCLE was identified in *Hormaphis cornu* , which induces galls on the leaves of witch hazel, *Hamamelis virginiana*(Korgaonkar et al., 2021). *BICYCLE* may regulate many aspects of gall development because they are expressed very abundantly in salivary glands specifically in gall aphids.*S. chinensis* feeds on host leaves where it injects saliva into host leaf cells, resulting in gall formation. A total of 141 proteins have been identified from its salivary glands by LC-MS/MS analysis (Yang et al., 2018). In comparison to salivary proteins from 10 other free-living Hemipterans, the presence of a high proportion of proteins with binding activity was noticeable, including DNA-, protein-, ATP-, and iron-binding proteins, which may all be involved in gall formation. BICYCLE proteins were not identified in salivary glands suggested different mechanisms of gall induction between*S. chinensis* and *H. cornu.* From the RNA-Seq analysis, transcripts corresponding to 35 genes (Sc.chr03.1184- Sc.chr10.506) encoding salivary gland proteins have shown high expression levels at the gall forming fundatrix stage (Figure S4). These salivary proteins are potentially may be related to interaction between insects and host plants. According to their functions, these genes can be divided into detoxification, signal transduction, secreted protein metabolism, energy metabolism, basic biological processes, movement and function unknow (Table S10). The largest number of genes related to detoxification may be related to host plant defense inhibition. Gene belonging to movement and energy metabolism may be related to the contraction of salivary gland muscle and provide energy to salivate.

Conclusion:

We provided a high-quality chromosome-level genome assembly of the galling aphid *S. chinensis* . Phylogenomic analysis indicated that*S. chinensis* diverged from the common ancestor of *Eriosoma lanigerum* at approximately 57.16 million years ago. Transcriptome comparison showed 35 genes encoding salivary gland proteins highly expressed at the gall forming fundatrix stage. Some of these salivary proteins may involve in gall formation. Our results may benefit future research on elucidating the molecular basis for the unique biology associated with galling aphids, especially for that which inducing completely closed galls and on revealing the molecular mechanisms for gall induction and host interactions.

Acknowledgements

We thank to Prof. Kirst King-Jones (University of Alberta), and Dr. Yi-Yuan Li (University of Texas at

Austin), for their helpful comments.

Author contributions

Z.C.X., Y.Z.X. and C.X.M conceived and designed the study. Z.C.X., Y.Z.X and W.H.Y. collected samples. W.H.Y. and Y.Y.X. performed the genome assembly, gene model prediction, gene annotation and comparative analyses. H.H.J. performed the chromosome analyses. W.H.Y., Y.Z.X. performed the transcriptome analyses. W.H.Y., Y.Z.X. and Y.Y.X. wrote the manuscript with input from all authors. Z.C.X., Y.Z.X. and C.M.S. analyzed the data and discussed the results. All authors reviewed the manuscript.

Data availability statement

All data mentioned in this paper have been deposited in the National Center for Biotechnology Information with the BioProject accession number PRJNA700780 (genomic sequencing) and PRJNA702264 (transcriptome sequencing). The Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JAFHKX000000000. The genome assembly and annotation, orthogroup clustering results and salivary gland genes are available for download from Zenodo (10.5281/zenodo.3797131).

References

Al-Jbory, Z., El-Bouhssini, M., Chen, M.S. (2018) Conserved and unique putative effectors expressed in the salivary glands of three related gall midges species. Journal of Insect Science 18(5): 15. https://doi.org/10.1093/jisesa/iey094 Al Jibory, Z., Micheal, J. A., Park, Y., Reeck, G. R., & Chen M. S. (2020). Differential localization of Hessian fly candidate effectors in resistant and susceptible wheat plants. *Plant direct, 00*: 1-15. https://doi.org/10.1002/pld3.246 Biello, R., Singh, A., Godfrey, C. J., Fernandez, F. F., Mugford, S. T., & Powell, & G., Hogenhout, S. A., | Mathers, T. C. (2021). A chromosome level genome assembly of the woolly apple aphid,*Eriosoma lanigerum* Hausmann (Hemiptera: Aphididae).*Molecular Ecology Resources, 21*(1), 316-326. https://doi.org/10.1111/1755-0998.13258 Birney, Ewan, Clamp, Michele, Durbin, & Richard. (2004). GeneWise and Genomewise. *Genome Research, 14*(5), 988-995. https://doi.org/10.1101/gr.1865504 Blackman, R. L., & Eastop, V. F. (2020). *Aphids on the world's plants: An online identification and information guide.* John Wiley & Sons Ltd. http://www.aphidsonworldsplants.info/. Blanco, E., Parra, G., & Guigo, R. (2002). *Using geneid to Identify Genes.* John Wiley & Sons, Inc. Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120. doi:10.1093/bioinformatics/btu170 Carolan, J. C., D Caragea, Reardon, K. T., Mutti, N. S., & Edwards, O. R. (2011). Predicted effector molecules in the salivary secretome of the pea aphid (*Acyrthosiphon pisum*): a dual transcriptomic/proteomic approach. *Journal of Proteome Research*, 10(4), 1505-18. https://doi.org/ 10.1021/pr100881q Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology Evolution, 17*(4), 540-552.*https://doi.org/10.1093/oxfordjournals.molbev.a026334* Chan, P. P., & Lowe, T. M. (2019). tRNAscan-SE: Searching for tRNA genes in genomic sequences. *Methods in Molecular Biology, 1962*:1-14. In book: Gene Prediction.*https://doi.org/10.1007/978-1-4939-9173-0_1* Chen, C., Chen, H., Y Zhang, Thomas, H. R., & Xia, R. (2020). Tbtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant*, 13(8).*https://doi.org/10.1016/j.molp.2020.06.009* Chen, W., Shakir, S., Bigham, M., Fei, Z., & Jander, G. (2019). Genome sequence of the corn leaf aphid (*Rhopalosiphum maidis* Fitch).*GigaScience*, 8(4). https://doi.org/ 10.1093/gigascience/giz033 Chen, X. M., Yang, Z. X., Chen, H., Qi, Q., Liu, J., & Wang, C., Shao, S. X., Lu, Q., Li, Y., Wu, H. X., King-Jones, K., Chen, M. S. (2020). A complex nutrient exchange between a gall-forming aphid and its plant host. *Frontiers in Plant Science, 11*, 811. https://doi.org/10.3389/fpls.2020.00811 Clore, A. (2014). gBlocks gene fragments for gene construction and more. *Journal of Immunological Methods, 188*(1), 165-167. https://doi.org/10.1016/0022-1759(95)00229-4 Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., Shamim, M. S., Machol, I., P., Lander, E. S., Aiden, A. P., & Aiden, E. L. (2017). *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds.*Science, 356*, 92-95.*https://doi.org/10.1126/science.aal3327* Dudchenko, O., Shamim, M. S., Batra, S. S., Durand, N. C. & Aiden, E. L. (2018). The Juicebox Assembly Tools module facilitates de novo assembly

of mammalian genomes with chromosome-length scaffolds for under $1000. https://doi.org/ 10.1101/254797 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. *PNAS, 117*(17), 9451-9457.*https://doi.org/10.1073/pnas.1921046117*Ghosh, S., & Chan, C. K. (2016). Analysis of rna-seq data using tophat and cufflinks. *Methods in Molecular Biology*, 1374, 339-61. https://doi.org/ 10.1007/978-1-4939-3167-5_18 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 biology, 9*(1), R7. https://doi.org/10.1186/gb-2008-9-1-r7 Hahn, M. W., Demuth, J. P., & Han, S. G. (2007). Accelerated rate of gene gain and loss in primates. *Genetics, 177*(3). https://doi.org/10.1534/genetics.107.080077 Hirano, T., Kimura, S., Sakamoto, T., Okamoto, A., Nakayama, T., Matsuura, T., Ikeda, Y., Takeda, S., Suzuki, Y., OhshimaI., & Sato, M. H. Reprogramming of the developmental program of *Rhus javanica*during initial stage of gall induction by *Schlechtendalia chinensis. Frontiers in Plant Science, 2020, 11*, 471.*https://doi.org/10.3389/fpls.2020.00471*Hu, J., Fang, J. P., Su, Z. Y., & Liu, S. L. (2019). NextPolish: a fast and efficient genome polishing tool for long-read assembly.*Bioinformatics, (7)*, 7. https://doi.org/10.1093/bioinformatics/btz891 Huang, S. F., Kang, M. J., & Xu, A. L. (2017). HaploMerger2: rebuilding both haploid sub-assemblies from high-heterozygosity diploid genome assembly. *Bioinformatics 16*, 2577.*https://doi.org/10.1093/bioinformatics/btx220*International Aphid Genomics Consortium. (2010). Genome sequence of the pea aphid *Acyrthosiphon pisum. Plos Biology, 8*(2), 1-25. https://doi.org/10.1371/journal.pbio.1000313 Julca, I., Marcet-Houben, M., Cruz, F., Vargas-Chavez, C., Johnston, J. S., Gomez-Garrido, J., Frias, L., Corvelo, A., Loska, D., Camara, F., Gut, M., Alioto, T., Latorre, A., & Gabaldon, T. (2020). Phylogenomics identifies an ancestral burst of gene duplications predating the diversification of aphidomorpha. *Molecular Biology and Evolution,37*(3), 730-756. https://doi.org/10.1093/molbe v/msz261 Johnson, K. P., Dietrich, C. H., Friedrich, F., Beutel, R. G., Wipfler, B., & Peters, R. S., et al. (2018). Phylogenomics and the evolution of hemipteroid insects. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(50). https://doi.org/10.1073/pnas.1815820115 Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research, 30*(14), 3059-3066. https://doi.org/10.1093/nar/gkf436 Katoh, K., & Standley, D. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability.*Molecular Biology and Evolution, 30*(4), 772-780. https://doi.org/10.1093/molbev/mst010 Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E. P., Rivas, E., Eddy, S. R., Bateman, A., Finn, R. D., & Petrov, A. I. (2018). Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Research 46* (Database issue), D335-D342. http://dx.doi.org/10.1093/nar/gkx1038 Karin, L., Peter, H., Rodland, E. A., Staerfeldt, H. H., Rognes, T., & Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research, 35*(9), 3100-3108.*https://doi.org/10.1093/nar/gkm160*Korgaonkar, A., Han, C., Lemire, A. L., Siwanowicz, I., & Stern, D. L. (2021). A novel family of secreted insect proteins linked to plant gall development. *Current Biology* (D1). Kim, D., Landmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods,12*(4), 357-360. https://doi.org/ 10.1038/nmeth.3317 Kurosu, U., & Aoki, S. (1992). Gall cleaning by the aphid Hormaphis betulae. Journal of Ethology,*9*, 51-55. https://doi.org/10.1007/BF02350191.Li, F., Zhao, X., Li, M., He, K., Huang, C., Zhou, Y., Li, Z., & Walters, J. R. (2019). Insect genomes: progress and challenges. *Insect Molecular Biology, 28*(6), 739-758. https://doi.org/10.1111/imb.12599 Li, L., Stoeckert, C. J., & Roos, D. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research, 13*(9), 2178-2189. https://doi.org/10.1101/gr.1224503 Liu, P., Yang, Z. X., Chen, X. M., Foottit, R. G. (2014). The Effect of the gall-forming aphid *Schlechtendalia chinensis* (Hemiptera: Aphididae) on leaf wing ontogenesis in *Rhus chinensis*(Sapindales: Anacardiaceae). *Annals of the Entomological Society of America, 107*(1), 242-250. http://www.bioone.org/doi/full/10.1603/AN13118 Li, Y., Park, H., Smith, T. E., & Moran, N. A. (2019). Gene family evolution in the pea aphid based on chromosome-level genome assembly.*Molecular Biology and Evolution, 36*(10), 2143-2156.*https://doi.org/10.1093/molbev/msz138*Li, Y., Zhang, B., & Moran, N. A. (2020). The aphid x chromosome is a dangerous place for functionally im-

portant genes: diverse evolution of hemipteran genomes based on chromosome-level assemblies. *Molecular Biology and Evolution, 37*(8), 2357-2368. https://doi.org/ 10.1093/molbev/msaa095 Mathers, T. C. (2020). Improved genome assembly and annotation of the soybean aphid (*Aphis glycines* Matsumura). *G3: Genes, Genomes, Genetics, 10*(3), g3.400954.2019. https://doi.org/10.1534/ g3.119.400954 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., Gabaldon, T., Jordan, A., Julca, I., Kettles, G. J., Kowitwanich, K., Lavenier, 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. https://doi.org/10.1186/s1305 9-016-1145-3 Mathers, T. C., Mugford, S. T., Hogenhout, S. A. T., & Tripathi, L. (2020). Genome sequence of the banana aphid, *Pentalonia nigronervosa* Coquerel (Hemiptera: Aphididae) and its symbionts. *G3: Genes, Genomes, Genetics, 10*(12), 4315-4321. https://doi.org/10.1534/g3.120.401358 Mathers, T. C., Wouters, R. H. M., Mugford, S. T., Swarbreck, D., Van Oosterhout, C., & Hogenhout, S. A. (2020). Chromosome-scale genome assemblies of aphids reveal extensively rearranged autosomes and long-term conservation of the X chromosome. *Molecular Biology and Evolution, 38(3):856-875.https://doi.org/10.1093/molbev/msaa246*Marcais, Guillaume, Kingsford, & Carl. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, 27, 764-770. https://doi.org/10.1093/bioinformatics/btr011 Moran, N. A. (1989). A 48-million-year-old aphid-host plant association and complex life cycle: biogeographic evidence. *Science, 245*(4914), 173-175. https://doi.org/10.1126/science.245.4914.173 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*(1), 1-16. https:// doi.org/10.1186/s1286 4-015-1525-1 Quan, Q. M., Hu, X., Pan, B. H., Zeng, B. S., Wu, N. N., Fang, G. Q., Cao, Y. H., Chen, X. Y., Li, X., Huang, Y. P., & Zhan, S. (2019). Draft genome of the cotton aphid *Aphis gossypii. Insect Biochemistry and Molecular Biology, 105*, 25-32. https://doi.org/10.1016/j.ibmb.2018.12.007 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. (2014). A 3D map of the human genome at Kilobase resolution reveals principles of chromatin looping. *Cell, 158*, 1-6. https://doi.org/10.1016/j.cell.2014.11.021 Ranallo-Benavidez, T. R., Jaron, K. S., & Schatz, M. C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications, 11*(1), 1432. https://doi.org/10.1038/s41467-020-14998-3 Ruan, J., & Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. *Nature Methods, 17*(Supp l 6), 1-4.*https://doi.org/10.1101/530972*Roach, M. J., Schmidt, S. A., & Borneman, A. R. (2018). Purge haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics, 19*(1). https://doi.org/ 10.1186/s12859-018-2485-7 Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics, 30*(9), 1312-1313. https://doi.org/10.1093/bioinformatics/btu033 Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006). AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Research, 34* (Web Server issue), W435-439. https://doi.org/10.1093/nar/gkl200 Thorpe, P., Escudero-Martinez, C. M., Cock, P. J. A., Eves-van den Akker, S., & Bos, J. I. B. (2018). Shared transcriptional control and disparate gain and loss of aphid parasitism genes. *Genome Biology and Evolution, 10*(10), 2716-2733. https://doi.org/10.1093/gbe/evy183 Takeda, S., Yoza, M., Amano, T., Ohshima, I., Hirano, T., Sato, M. H., Sakamoto, T., & Seisuke Kimura, S. (2019) Comparative transcriptome analysis of galls from four different host plants suggests the molecular mechanism of gall development. *PLoS One, 14*(10), e0223686.*https://doi.org/10.1371/journal.pone.0223686*Tarailo-Graovac, M., & Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Current protocols in bioinformatics*, Chapter 4, Unit 4.10.*https://doi.org/10.1002/0471250953.bi0410s25*Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q. D., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One, 9*(11), e112963.*https://doi.org/10.1371/journal.pone.0112963*Waterhouse, R. M., Seppey, M., Simao, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., . . . Zdobnov, E. M. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics. Molecular Biology and Evolution, 35(3), 543-548. https://doi.org/10.1093/molbev/msx319 Wang, Z., Ge, J., Chen, H., Cheng, X., Yang, Y., Li, J., Whitworth, R. J., & Chen M. C. S. (2018). An insect nucleo-

side diphosphate kinase (NDK) functions as an effector protein in wheat - Hessian fly interactions. *Insect Biochemistry and Molecular Biology, 100*, 30-38. https://doi.org/10.1016/j.ibmb.2018.06.003 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. (2016). Whole genome sequence of the soybean aphid, *Aphis glycines.Insect Biochemistry and Molecular Biology*, *123*, 102917. https://doi.org/10.1016/j.ibmb.2017.01.005 Wool, D. Galling aphids: specialization, biological complexity, and variation. (2004). *Annual Review of Entomology, 49*(1), 175. https://doi.org/10.1146/annurev.ento.49.061802.123236 Yang, Z. H. (2007). PAML 4: phylogenetic analysis by maximum likelihood.*Molecular Biology and Evolution*, 24(8), 1586-1591. https://doi.org/10.1093/molbev/msm088 Yang, Z. X., Ma, L., Francis, F., Yang, Y., Chen, H., Wu, H. X., & Chen, X. M. (2018). Proteins identified from saliva and salivary glands of the Chinese gall aphid*Schlectendalia chinensis*. *Proteomics, 18*, 1700378. https://doi.org/10.1002/pmic.201700378 Zhang, C. X., Tang, X. D., & Cheng, J. A. (2008). The utilization and industrialization of insect resources in China. *Entomological research, 38*, S38-S47. https://doi.org/10.1111/j.1748-5967.2008.00173.x Zhang, G. X., Qiao, G. X., Zhong, T. S., & Zhang, W. Y. (1999). *Fauna Sinica, Insecta* Vol. 14 Homoptera, Mindaridae and Pemphigidae. Science Press, Beijing.

Zhao, C. Y., Escalante, L.N., Chen, H., Benatti, T. R., Qu, J. X., Chellapilla, S., Waterhouse, R. M., Wheeler, D., Andersson, M. N., Bao, R., Batterton, M., Behura, S. K., Blankenburg, K. P., Caragea, D., Carolan, J. C., Coyle, M., El-Bouhssini M., Francisco L., ... Richards S. (2015) A massive expansion of effector genes underlies gall-formation in the wheat pest *Mayetiola destructor* .*Current Biology 25* (5): 613-620. https://doi.org/10.1016/j.cub.2014.12.057

Hosted file

`Table.docx` available at https://authorea.com/users/439900/articles/540692-chromosome-level-genome-assembly-of-the-horned-gall-aphid-schlechtendalia-chinensis-hemiptera-aphididae-erisomatinae