Chromosome-level assembly of the melon thrips genome yields insights into evolution of a sap-sucking lifestyle and pesticide resistance

Shao-Kun Guo, Li-Jun Cao, Wei Song, Pan Shi, Yong-Fu Gao, Ya-Jun Gong, Jin-Cui Chen, Ary Anthony Hoffmann, Shu-Jun Wei

1. Institute of Plant and Environmental Protection, Beijing Academy of Agriculture and Forestry Sciences, 9 Shuguanghuayuan Middle Road, Haidian District, Beijing 100097, China
2. School of BioSciences, Bio21 Institute, University of Melbourne, Parkville, Victoria, Australia

Shao-Kun Guo, mscgsk@163.com (ORCID: 0000-0002-8941-9750)
Li-Jun Cao, gmatjhpl@163.com (ORCID: 0000-0002-4595-0136)
Wei Song, songw0513@163.com
Pan Shi, shipan0808@163.com

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1755-0998.13189

This article is protected by copyright. All rights reserved
Short title: The melon thrips genome

Abstract

Thrips are tiny insects from the order Thysanoptera (Hexapoda: Condylognatha), including many sap-sucking pests that are causing increasing damage to crops worldwide. In contrast to their closest relatives of Hemiptera (Hexapoda: Condylognatha), including numerous sap-sucking species, there are few genomic resources available for thrips. In this study, we assembled the first thrips genome at the chromosomal level from the melon thrips, *Thrips palmi*, a notorious pest in agriculture, using PacBio long-read and Illumina short-read sequences. The assembled genome was 270.43 Mb in size, with 4,120 contigs and a contig N50 of 426 kb. All contigs were assembled into 16 linkage groups assisted by the Hi-C technique. In total, 16,333 protein-coding genes were predicted, of which 88.13% were functionally annotated. Among sap-sucking insects, polyphagous species (e.g., *T. palmi* and *Bemisia tabaci*) usually possess more detoxification genes than oligophagous species (e.g., *Diaphorina citri*). The polyphagous thrips genomes characterized so far have relatively more detoxification genes in the GST and CCE families than polyphagous aphids, but they have fewer UGTs. HSP genes, especially from the Hsp70s group, have expanded in thrips.
compared to other hemipterans. These differences point to different genetic mechanisms associated with detoxification and stress responses in these two groups of sap-sucking insects. The expansion of these gene families may contribute to the rapid development of pesticide resistance in thrips, as supported by a transcriptome comparison of resistant and sensitive populations of T. palmi. The high-quality genome developed here provides an invaluable resource for understanding the ecology, genetics, and evolution of thrips as well as their relatives more generally.

**Keywords:** Thysanoptera, Thrips palmi, pesticide resistance, de novo assembly, genome annotation

**Introduction**

Many insects feed on plant sap by using their sucking and piercing mouthparts. A sap-sucking lifestyle evolved mainly in the superorder Condylognatha, including thrips from the order Thysanoptera, and psyllids, whiteflies, aphids, mealybugs and true bugs from the order Hemiptera. Sap-sucking insects include a large number of notorious agricultural pests that have developed pesticide resistance, such as the western flower thrips Frankliniella occidentalis, whitefly Bemisia tabaci, green peach aphid Myzus persicae and brown planthopper Nilaparvata lugens (De Barro, Liu, Boykin, & Dinsdale, 2011; Reitz et al., 2020).

Understanding the genomic basis of a sap-sucking lifestyle may help in developing targets to reduce damage from outbreaks of these pests as well as provide information on convergent and divergent evolution in these diverse groups.

Among the sap-sucking insects, thrips species are causing increasing damage and economic losses to agricultural and horticultural crops worldwide (Morse & Hoddle, 2006; Reitz et al., 2020). Due to their small body size and cryptic habits, thrips are easily transferred across regions, leading to them becoming global invaders outside their native ranges (Morse & Hoddle, 2006). Although many methods have been applied to counter outbreaks of pest thrips (Mouden, Sarmiento, Klinkhamer, & Leiss, 2017), field control remains heavily reliant on pesticides (Mouden et al., 2017), leading to pesticide resistance problems (Y. L. Gao, Lei, & Reitz, 2012). Despite the increasing economic impact of thrips,
there is a lack of understanding about their genetics, ecology, and evolution (Mound, 2005).

For example, it was only established in the 1990s that thrips feed by sucking the sap from individual plant cells, whereas prior to this time it was assumed that thrips fed by rasping the surface of plants (Heming, 1993; Mound, 2005).

One of the challenges in developing a further understanding of thrips is that there is a lack of genomic resources for these organisms. Only one scaffold-level genome from *F. occidentalis* is available for thrips (Thomas et al., 2018). In contrast, for sap-sucking insects from the Hemiptera, there are useful genomic resources available, with published genomes for the whitefly *B. tabaci* (Chen et al., 2016), psyllid *Diaphorina citri* (Saha et al., 2017), aphids *Aphis gossypii* (Quan et al., 2019), *Aphis glycines* (Wenger et al., 2017), *Diuraphis noxia* (Nicholson et al., 2015), *M. persicae* (Ramsey et al., 2007), and *Acyrthosiphon pisum* (International Aphid Genomics, 2010), scale insect *Ericerus pela* (Cryan & Urban, 2012), and planthopper *N. lugens* (Xue et al., 2014). These genomic resources in Hemiptera provide an opportunity to compare the evolution of the sap-sucking lifestyle in thysanopteran and hemipteran insects. A well-assembled reference genome is also essential in understanding other aspects of the ecology, evolution and control of thrips, and can be used to build on transcriptomic studies that have examined pesticide resistance and virus transmission in thrips (Berger et al., 2016; Gamage, Rotenberg, Schneweis, Tsai, & Dietzgen, 2018; Schneweis, Whitfield, & Rotenberg, 2017; Y. Wan et al., 2018).

The melon thrips, *Thrips palmi* Karny, is one of the most important pest thrips damaging a variety of vegetables, such as eggplant and melons (Fig. 1) (Cannon, Matthews, & Collins, 2007). It is native to Southeast Asia and has become established in South and East Asia, South America, the Caribbean, Florida, Australia, and West Africa during the second half of the twentieth century (Cannon et al., 2007; Kawai, 2001), becoming one of the most important pests of vegetables (Y. F. Gao et al., 2019; Kawai, 2001; Przybylska, Fiedler, Kucharczyk, & Obrepska-Steplowska, 2015). In Japan and China, some populations of *T. palmi* have rapidly developed a high level of resistance to spinetoram (Bao et al., 2014; Y. F. Gao et al., 2019). This species, therefore, provides an ideal model to understand the genetic basis of sap-sucking lifestyle and pesticide resistance in thrips.

In the present study, we report the first genome of thrips determined from *T. palmi*. The genome was de novo assembled based on sequences obtained from the PacBio and Illumina platforms and assembled at the chromosome level assisted by the Hi-C technique,
joining the growing list of insect genomes that have been sequenced and assembled to a high level. It also provides a reference for the study of small insect genomes. We compared genome features between thrips and several sap-sucking hemipteran insects (whitefly, psyllid, aphid, and planthopper) to explore the evolution and genomic signatures of a sap-sucking lifestyle. The genomic resource developed here for *T. palmi* provides an invaluable resource for understanding the genetics, ecology and evolution of insects, as well as providing information for the eventual development of novel management options for thrips pests.

**Materials and methods**

**Samples and DNA extraction**

For genome sequencing, about 100 individuals of *T. palmi* were collected from cultivated cucumber at Shouguang, Shandong province, and reared for about 20 generations in the laboratory. Adults were reared with cucumber seedlings at 25 ± 1 °C under a relative humidity of 70% and a photoperiod of 14 h:10 h L:D. During rearing, adults laid eggs in cucumber leaves and the mature nymphs were transferred to bean pods for pupation.

Genomic DNA for the Illumina paired-end DNA library construction was extracted from one female adult with DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany), following the manufacturer’s instructions. Genomic DNA used for the SMRTbell library preparation was extracted from about 2,000 adults with Blood & Cell Culture DNA Midi Kit (Qiagen, Hilden, Germany), following the manufacturer’s protocol. All DNA extracts were verified with NanoDrop (NanoDrop products, Wilmington, DE, USA) and a Qubit 3.0 Fluorometer (Life Technologies Corporation, Eugene, OR, USA) using the Qubit™ dsDNA HS Assay Kit (PN# Q32851) (Life Technologies Corporation, Eugene, OR, USA) to quantify purity and concentration.

For pesticide resistance analysis, we collected one field population (BJ) of *T. palmi* from a cucumber crop in the Fangshan District, Beijing (Y. F. Gao et al., 2019) and another population (SD) from an eggplant crop in Jitai town, Shouguang city, Shandong province, with about 3,000 individuals collected from each population. Based on previously collected data, the BJ population was expected to be more sensitive to spinetoram than the SD
population (Gao et al., 2019). The field-collected generations of the two populations are referred to as BJF0 and SDF0.

In total 640 of these field-collected adults per population were used for bioassays according to a published method (Y. F. Gao et al., 2019) to test the susceptibility of thrips to 6% spinetoram suspension concentrate (SC) (Dow AgroSciences Company, USA), a spinosyn class pesticide widely used in thrips control. Another 400 adults per population were preserved in liquid nitrogen for transcriptome library construction. For each transcriptome library, 100 adults were used; four biological replicates were set up for each population. The remaining individuals from the two field populations were used to set up cultures that were then reared in the laboratory for five generations (BJF5 and SDF5) as described above, without contact with any pesticide. These were then used for further bioassays and transcriptome analyses carried out as described for the field populations, to examine changes in susceptibility to spinetoram and transcription patterns following multiple generations of laboratory culture.

**Library construction and sequencing**

For long-read sequencing, SMRTbell libraries were constructed with Sequel® Sequencing Kit 3.0 (Pacific Biosciences, Menlo Park, CA, USA). Long DNA fragments of the approximately 20 kb library were sequenced on a PacBio Sequel sequencer (Pacific Biosciences, Menlo Park, CA, USA). Three SMRT cells were processed. After filtering, we obtained 17.67 Gb of short clean reads from the Illumina platform (coverage: 65.34X) and 33.49 Gb subreads (mean subread length: 6.44 kb, subread N50 length: 8.77 kb, coverage: 123.84X) from the PacBio platform for contig-level genome assembly (Table S1).

For short-read sequencing, a paired-end library with short insert sizes of about 500 bp was constructed using VAHTSTM Universal DNA Library Prep Kit for Illumina® V2 (Vazyme, Nanning, China). After passing quality inspection, the paired-end library was sequenced on an Illumina NovaSeq platform with the standard protocol offered by Illumina (San Diego, CA, USA).

To assist the chromosome-level assembly, the Hi-C (High-throughput chromosome conformation capture) technique was applied to capture genome-wide chromatin interactions (Belaghzal, Dekker, & Gibcus, 2017). After nearly 1,000 pupae were ground in 2%
formaldehyde to allow cross-linking of cellular protein, 50 μg DNA was extracted and the
chromosome integrity and cross-linked protein residue were assessed. Chromatin digestion
was performed with the restriction enzyme MboI, then Hi-C samples were extracted by
biotin labeling, flat end ligation and DNA purification. After passing the DNA quality
detection test, we performed the standard Hi-C library construction process. The Hi-C library
was sequenced using the Illumina NovaSeq platform with paired-end 150-bp reads.

For transcriptome sequencing, total RNA was isolated from spinetoram-resistant and
susceptible populations of T. palmi with TRizol Reagent (Invitrogen, Carlsbad, CA, USA) and
quantified with a NanoDrop ND-2000 spectrophotometer (NanoDrop products, Wilmington,
DE, USA). A cDNA library was constructed using a VAHTSTM mRNA-seq V3 Library Prep Kit for
Illumina (Vazyme, NR611, Nanjing, China). After validation by quantitative real-time PCR with
a library quantification kit/Illumina GA Universal (KAPA, Wilmington, MA, USA), libraries
were sequenced on an Illumina NovaSeq instrument (Illumina, San Diego, CA, USA) at the
BerryGenomics company (Beijing, China).

Genome assembly and evaluation
The quality of Illumina raw reads from the above two libraries was checked by FastQC
(Andrews, 2010) and the low-quality reads were filtered by Trimmomatic v0.38 (Bolger,
Lohse, & Usadel, 2014). The long reads were used for genome de novo assembly. Raw reads
were corrected using wtdbg v2.2 (Ruan & Li, 2020) to generate a draft assembly. A
consensus assembly was obtained through running wtpoa-cns implemented in wtdbg2
based on corrected reads and the draft assembly. The assembly was further polished by
Pilon v1.22 (Walker et al., 2014) based on Illumina short reads four times to obtain the
contig-level assembly. Clean reads sequenced from the Hi-C library were aligned to the
contig-level genome with an end-to-end algorithm implemented in bowtie2 according to
HiC-Pro strategy (Langmead & Salzberg, 2012; Servant et al., 2015). Juicer v1.5 and 3D de
ovo assembly (3D-DNA) pipelines were used to assemble the scaffolds into a
chromosome-level genome (Dudchenko et al., 2017; Durand et al., 2016).

The completeness of the genome was evaluated through estimating the genome size,
mapping the Illumina reads to a reference genome using Burrow-Wheeler Aligner (BWA)
v0.7.17 (Li & Durbin, 2009), calculation of coverage and BUSCO analysis of single-copy
orthologs (Simao, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015). Size,
heterozygosity, and duplication of the genome were estimated by the K-mer method.

This article is protected by copyright. All rights reserved
K-mers were counted by jellyfish v2.2.9 (Marçais & Kingsford, 2011) with 17-base oligonucleotide based on Illumina short reads. Parameters were determined by GenomeScope v1.0 (Vurture et al., 2017). Benchmarking Universal Single-Copy Orthologs (BUSCO) v3.0.2 (Simao et al., 2015) was used to evaluate the completeness of the assembly based on the insecta_odb9 database (1,658 genes). We compared the genome assembly features among 10 insect species (Table 1).

**Repetitive elements and noncoding RNA annotation**

Repetitive elements in scaffolds longer than 1,000 bp were detected by RepeatMasker v4.0.7 (Tarailo-Graovac & Chen, 2009) against the Insecta repeats within RepBase Update (http://www.girinst.org). The assembled genome was analyzed for potential DNA transposon sequences using the program RepeatModeler (http://www.repeatmasker.org/RepeatModeler.html, RRID: SCR_015027). Most noncoding RNAs (ncRNA) were annotated by aligning the genomic sequence against RFAM (http://rfam.xfam.org/) with BLASTN. Three types of noncoding RNAs (ncRNA) - transfer RNA (tRNA), ribosome RNA (rRNA), and small nuclear RNA - were annotated. Among them, tRNAs and rRNAs were predicted by tRNAscan-SE and RNAmmer (Lagesen et al., 2007; Lowe & Eddy, 1997).

**Protein-coding gene annotation**

Protein-coding genes were annotated under three lines of evidence including ab initio, RNA-seq-based, and homolog-based methods. For the ab initio method, we used the software packages Augustus v3.2.3 (Stanke & Waack, 2003) and SNAP v2013-02-16 (Korf, 2004) employed with default parameters. The homologous genes from *Drosophila melanogaster* and *F. occidentalis* were employed to predict gene structures. All the known genes of *T. palmi* in the database were used to train Augustus and SNAP three times. For the RNA-seq-based method, short reads from transcriptome sequencing were aligned to the genome with TopHat v2.1.1 (Trapnell, Pachter, & Salzberg, 2009) and the gene structure was built by PASA v2.0.2 (M. A. Campbell, Haas, Hamilton, Mount, & Buell, 2006) with default settings. For the homolog-based approach, Gene Model Mapper (GeMoMa) v1.4.2 (Keilwagen et al., 2016) was used with the protein sequences of *D. melanogaster*, *F. occidentalis*, *Acyrthosiphon pisum* and *Pediculus humanus* as references. Results from three
methods were integrated with EVidenceModeler (EVM) v1.1.1 (Haas et al., 2008). All analysis was conducted in Maker v2.31.10 genome annotation pipeline (Cantarel et al., 2008). The gene set was annotated by aligning protein sequences to NR (non-redundant sequence databases) (Deng et al., 2006), UniRef (Suzek et al., 2014), UniprotKB/Swiss-prot (Bairoch & Boeckmann, 1991), UniProtKB/TrEMBL (Boeckmann et al., 2003) and COG (Cluster of Orthologous Groups of proteins)/KOG (eukaryotic orthologous groups of proteins) (Tatusov et al., 2001) using BLAST (Altschul, Gish, Miller, Myers, & Lipman, 1990) with a threshold of 1e-5. The software eggnog-mapper v1.0.3 was applied to annotate gene functions in the GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) databases (Kanehisa & Goto, 2000).

Identification of orthology and inference of phylogenetic relationships
OrthoFinder (Emms & Kelly, 2015) was used to find orthologues and orthogroups. Phylogenetic relationships within Condylognatha, including thrips (Thysanoptera), psyllid, whitefly, aphid and planthopper (Hemiptera) were reconstructed based on single-copy orthologs of protein-coding genes (Table 2). Two dipteran species were used as outgroups. MAFFT v7.450 (Katoh & Standley, 2013) was employed to align amino acid sequences of 1:1:1 orthologous gene with the G-INS-I algorithm. The phylogenetic tree was inferred using an approximately maximum-likelihood method implemented in FastTree v2.1.10 (Price, Dehal, & Arkin, 2009) under default settings. FastTree uses the JTT (Jones-Taylor-Thorton) or WAG (Whelan Goldman) models of amino acid evolution. The most likely category for each site was set using a Bayesian approach with a gamma prior. This method is a standard workflow in OrthoFinder, which rapidly infers the species tree from a concatenated multiple sequence alignment (MSA) of single-copy genes.

Gene family annotation
To facilitate both the curation of inaccurate annotations and the identification of previously undetected gene family copies directly from DNA sequences, we manually annotated seven gene families, including detoxification genes of cytochrome P450 monoxygenase (P450s), glutathione S-transferase (GSTs), carboxyl/cholinesterases (CCEs), UDP-glycosyltransferases (UGTs) and ATP-binding cassette (ABC) transporter, heat shock proteins (HSPs), and nicotinic acetylcholine receptors (nAChRs). Hidden Markov models (HMMs) and orthologs from related species of F. occidentalis and model species of D. melanogaster provided evidence for gene identification, run with HMMER v3.3 (Finn, Clements, & Eddy, 2011) and BLAST.
v2.2.31 (Altschul et al., 1990). The HMMs were downloaded from Pfam 32.0 (September 2018, 17,929 entries) (El-Gebali et al., 2018). We used the bioinformatic pipeline BITACORA (full mode) to conduct HMMER and BLAST analyses (Vizueta, Sánchez-Gracia, & Rozas, 2019). Hits were filtered with a default cut-off E-value of 10e-5. The annotated genes were further filtered manually based on gene length and the presence of conserved domains. We removed genes with a length shorter than 80% of the average gene length. Orthologs were aligned with the G-INS-I algorithm implemented in MAFFT v7.450 (Katoh & Standley, 2013). A neighbor-joining tree was constructed for each gene family using MEGA7 (Kumar, Stecher, & Tamura, 2016) with 500 bootstrap replicates.

Transcriptome analysis of resistant and susceptible strains of T. palmi

Illumina short reads sequenced from transcriptome libraries were filtered by Trimmomatic v0.38 (Bolger et al., 2014) and then mapped to the assembled genome of T. palmi using STAR v2.6.0c with default parameters (Dobin et al., 2013). To detect differentially expressed genes (DEGs), we applied the empirical Bayes hierarchical model EBSeq (Leng et al., 2013). In this analysis, we adopted the well-established Benjamini-Hochberg method to calibrate p values from the original assumption test (Ferreira & Zwinderman, 2006). After calibration, the p-value was determined using the false discovery rate (FDR) approach to decrease false positives caused by independent statistical hypothesis testing on expression changes in a large number of genes. We used an FDR < 0.001 and a $|\log(\text{fold-change (FC)})| \geq 2$ as the criteria for a significant difference in expression. Hierarchical clustering analysis of DEGs was performed to cluster genes that exhibited the same or similar expression levels. DEGs were mapped to GO terms and KEGG pathways, and an enrichment analysis was performed to identify any over-representation of GO terms and KEGG pathways.

Results

Summary of sequencing and assembly of T. palmi genome

Based on the Illumina reads, the genome size of T. palmi is estimated to be 202-223 Mb through k-mer analysis (k = 17, 21, 27, 31). The k-mer distributions show double peaks, indicating that this genome has a high rate of duplication and heterozygosity. The estimated heterozygosity ranges from 1.01% to 1.32% and duplication ranges from 1.37 to 1.96% (Fig. 2b, Fig. S1).
At the contig level, we assembled the *T. palmi* genome into 270.43 Mb sequences, including 4,120 contigs, with a contig N50 length of 426.28 Kb (Table 1). When we mapped the Illumina short reads to the assembled genome, 97.21% of reads could be mapped, covering 97.31% of the assembled genome. The BUSCO analysis showed that 97.17% (single-copied gene: 95.78%, duplicated gene: 1.39%) of 1,658 single-copy genes in the *insecta_odb9* database were identified as complete, 0.60% of genes were fragmented, and 2.23% of genes were missing in the assembled genome (Table S2).

For the chromosome-level assembly, we obtained 86.43 Gb of clean reads after filtering from 101.29 Gb raw reads (150 bp paired-end) from the Hi-C library (coverage: 319.60X). There were 33.62% normal paired reads while the others were chimeric paired (0.1%) or chimeric ambiguous (66.28%) among the 50 million paired reads. All read pairs were properly mapped to the genome, and 19.29% of the read pairs showed Hi-C contacts (Table S3). Based on contig interaction frequency calculated from the pairs aligned to the contigs, the 4,120 contigs were clustered into 16 linkage groups (Fig. 2a). The longest contig group was 22.82 Mb while the shortest one was 9.98 Mb, with an N50 of 14.67 Mb. BUSCO analysis showed that 97.10% (single-copied gene: 96.32%, duplicated gene: 0.78%) of 1,658 genes were identified as complete, 0.60% of genes were fragmented, while 2.29% of genes were missing in the assembled genome (Table S2).

**Genome annotation**

We identified 16,333 protein-coding genes and 16,386 annotated proteins from the genome of *T. palmi* using *de novo*, homology- and transcriptome sequencing-based methods. The number of genes in the *T. palmi* genome is comparable to other insect species (Table 1). BUSCO analysis showed that 95.2% (single-copied gene: 94.1%, duplicated gene: 1.1%) of the evaluated single-copy genes were identified as complete, 1.6% of the genes were fragmented, while 3.2% of the genes were missing in the annotated genes. Functional annotation found that 14,357 (87.90%), 14,359 (88.13%), and 13,139 (80.44%) genes had significant hits with proteins catalogued in NR, SwissProt, and TrEMBL, respectively. There were 7,247 (44.37%) and 6,773 (41.47%) genes annotated to GO terms and KEGG KOs, respectively (Fig. S2). We predicted 309 rRNAs, 1,171 tRNAs, 67 small nuclear RNAs, and 114 micro RNAs in the *T. palmi* genome based on Rfam databases (Table S4).

In total, 15.34 Mb (6.45%) of the genome was identified to be repeat DNA, and the *T. palmi* genome comprised approximately 54.06% GC base pairs. Overall, 15,558 transposable...
elements (TEs) including 4,823 retroelements (446 short interspersed nuclear elements (SINEs), 2,105 long interspersed nuclear elements (LINEs) and 2,272 long terminal repeats (LTR)) and 10,735 DNA transposons were identified. There were 58 satellites and 232,896 simple repeats identified as tandem repeats (TRs), accounting for 4.74% of the *T. palmi* genome (Table S5).

**Orthology and phylogenetic relationships among sap-sucking insects**

OrthoFinder assigned 220,436 genes (83.5% of total) to 15,694 orthogroups for the 12 species compared (Fig. 3). Fifty percent of all genes were in orthogroups with 17 or more genes (G50 was 17) and were contained in the largest 4,238 orthogroups (O50 was 4,238). There were 4,145 orthogroups with all species present and 102 of these consisted entirely of single-copy genes. We identified 13,806 genes specific to *T. palmi*, fewer than in the case of *F. occidentalis* (14,755) (Fig. 3).

For the phylogenetic analysis, 1,224 single-copy genes were used (Table S6). The result supports the sister relationship between Thysanoptera and Hemiptera and monophyly of Sternorrhyncha, including psyllids, aphids and whiteflies (Fig. 3), congruent with currently accepted topologies of these lineages (Cryan & Urban, 2012; Johnson et al., 2018).

**Evolution of detoxification genes in sap-sucking insects**

Detoxification genes are key genes that allow herbivorous insects to adapt to their host plant chemicals (Heckel, 2018). We identified 96 P450s, 25 GSTs, 39 CCEs, 17 UGTs, and 49 ABCs in the *T. palmi* genome (Table 2; Fig. 5, Figs. S3-S5). Previous research has shown that the number of detoxification genes is associated with phenotypes of host usage in insects (Rane et al., 2019). We compared the number of detoxification genes among evolutionarily related species from Condylognatha. There are more detoxification genes in polyphagous species than in oligophagous species except for the ABC genes that shows the opposite pattern (Fig. 4a, Table 2). We also compared the number of detoxification genes among polyphagous species of thrips, aphids and whitefly. The whitefly has the highest number of P450 and CCE genes followed by thrips; thrips have the highest number of GST genes and lowest number of UGT genes; aphids have the highest number of ABC genes but the lowest number of P450, GST and CCE genes (Fig. 4b, Table 2). Compared with the cockroach, termite and fruit fly, the sap-sucking insects have a low number of detoxification genes as noted previously (Gloss, Abbot, & Whiteman, 2019; Rane et al., 2019) (Table 2).
We further analyzed the number of genes in subfamilies of P450 and ABCs among polyphagous species. For P450s, the CYP3 and CYP4 subfamilies comprise a large number of genes as compared to the CYP2 and Mito families. In comparison to *T. palmi*, *F. occidentalis* and *B. tabaci* have higher numbers of P450 genes, particularly in the subfamilies CYP3 and CYP4. The CYP2 and Mito subfamilies in thrips show obvious expansion. Compared with *F. occidentalis* (26 genes), the CYP4 subfamily has expanded in *T. palmi* (49 genes) (Fig. 4c). For ABCs, subfamilies ABCG and ABCH have a higher number of genes than the other subfamilies. *Thrips palmi* and *F. occidentalis* have a relatively higher number of genes in the subfamily ABCC. *F. occidentalis* has more ABC genes than *T. palmi*, *B. tabaci*, *M. persicae* and *A. gossypii* (Fig. 4e).

**Evolution of HSP genes in sap-sucking insects**

HSPs function in adaptation to proteotoxic stresses (Bedulina et al., 2013; Colinet, Siaussat, Bozzolan, & Bowler, 2013). We identified 51 HSP genes in *T. palmi*, including 11 Hsp60s, 32 Hsp70s, two Hsp90s, and six sHSPs (small heat shock protein genes) (Fig. 4b, Fig. 5b). Compared to hemipteran insects, thrips have high numbers of HSPs comparable to the cockroach (Fig. 4b, Table 2). The expansion of HSP genes in thrips is mainly attributed to an expansion of Hsp70 genes relative to *B. tabaci*, *M. persicae* and *A. gossypii* (Fig. 4d). Hsp70s represent one of the most conserved protein subfamilies of HSPs. They can respond to stimuli and interact with other proteins, resulting in changes in stress response phenotypes (Bettencourt, Hogan, Nimali, & Drohan, 2008).

**Evolution of spinosyn-targeted nAChR genes in sap-sucking insects**

The spinosyns are a class of spinosad insecticides, and spinetoram is effective against various insect pests, especially lepidopterans, thysanopterans, and dipterans, but not hemipterans (Thomas C Sparks, Crouse, & Durst, 2001). The spinosyns targets on the nicotinic acetylcholine receptor (nAChR) (T. C. Sparks, Dripps, Watson, & Paroonagian, 2012). We analyzed the evolution of nAChRs in sap-sucking insects. In *T. palmi*, a total of 10 nAChR genes were identified (Fig. 6a). The two thrips species have a complete set of α1-6 and beta1 subunit genes. Compared to *F. occidentalis*, *T. palmi* has two more nAChR beta2 and beta3 genes (Fig. 6b). Mutation of the α6 subunit gene was reported to be associated with spinosyn resistance of several pests (Bao et al., 2014; Y. Wan et al., 2018); there is no
evidence of this subunit gene in the two aphids *M. persicae* and *A. gossypii* (Fig. 6b). In other sap-sucking hemipteran insects, two or three of the α8, α9 and α10 subunits are present (Fig. 6b). We also found that α5, α6 and α7 form a monophyletic lineage in the phylogenetic analysis (Fig. 6a).

**Pesticide resistance and the reference T. palmi genome**

Based on our assembled genome as a reference, we identified genes differentially expressed between populations with different levels of resistance to spinetoram. Bioassay results showed the LC₅₀ value of the BJF0 field population to spinetoram was 1.69 (1.326-2.139) mg/L (Y. F. Gao et al., 2019) while the more resistant SDF0 field population had an LC₅₀ of 759.34 (433.741-1,916.574) mg/L. After rearing both populations in the laboratory for five generations, the LC₅₀ of BJF5 was reduced to 0.37 (0.191-0.557) mg/L, and that of SDF5 was reduced to 33.12 (18.437-137.664) mg/L. While the resistance levels of the two populations were therefore significantly decreased by laboratory culture, there remained a significant difference between the populations. We first compared DEGs between the BJF0 and SDF0 field populations. A total of 1,892 DEGs were detected, including 461 up-regulated and 1,431 down-regulated genes in the relatively more resistant population (SDF0) (Figs. S6a, S6b, and S7, Table S7). The GO system classified genes involved in three categories: biological process, cellular component and molecular function. Among the up-regulated genes in the SDF0 population, 22 were mapped to P450s, Hsp20/α crystallin family and Hsp70 proteins. For the down-regulated genes, 11, 7, and 11 genes were mapped to functions related to P450s, CCEs and ABC transporters, respectively. Moreover, nine genes mapped to the neurotransmitter gated ion channel, and four of them were nAChR genes.

Next, we compared the two populations after rearing them for five generations in the laboratory (BJF5 and SDF5) which had reduced resistance levels. A total of 86 DEGs were detected, including 27 up-regulated and 59 down-regulated genes in the relatively more resistant population (SDF5). We searched for genes that might be related to pesticide resistance and found that one CCE and three P450 genes were identified as up-regulated in
SDF5 compared to BJF5 (Figs. S6c, S6d, and S7, Table S7). The number of DEGs was significantly decreased compared to the comparison between the two field populations. To compare changes in DEGs between the field and laboratory-reared populations, DEGs were identified between SDF0 and SDF5, and between BJF0 and BJF5. A total of 821 DEGs (153 up-regulated and 668 down-regulated genes in SDF0) were detected in the comparison between SDF0 and SDF5 (Figs. S6e, S6f and S7, Table S7), among which seven up-regulated DEGs were mapped to the Hsp20/α crystallin and Hsp70 families, while 20 down-regulated DEGs were mapped to nAChRs, CCEs and ABC transporters. A total of 258 DEGs (77 up-regulated and 181 down-regulated genes in BJF0) were detected in the comparison between BJF0 and BJF5, including one up-regulated CCE and one up-regulated P450, as well as three down-regulated CCEs and four down-regulated p450s (Figs. S6g, S6h and S7, Table S7).

Discussion

Chromosome-level assembly of a heterozygous genome for T. palmi

We report the first chromosome-level genome sequence for a thrips based on Illumina and PacBio sequencing platforms and Hi-C technology. The k-mer analysis showed that the T. palmi genome harbors a high degree of complexity with high heterozygosity (1.01% to 1.32%) compared to the beet armyworm, Spodoptera exigua (0.59%) (Zhang, Zhang, Yang, & Wu, 2019), the invading fall webworm (0.75% and 0.83%) (Wu et al., 2018) and the caddisfly Stenopsyche tiemnushanensis (1.05%-1.10%) (Luo, Tang, Frandsen, Stewart, & Zhou, 2018). Genome assembly has usually been challenged by high heterozygosity and duplication, especially for small invertebrates that need multiple individuals to be processed for DNA extraction to construct sequencing libraries (Chen et al., 2016; You et al., 2013), and this includes the genome of T. palmi. We, therefore, used long-read sequencing strategies involving PacBio and Hi-C to assemble the T. palmi genome, which has been proven previously to produce high completeness and continuity in genome assembly (F. Wan et al., 2019; Yin et al., 2018; L. Zhang et al., 2019). Controlled breeding may help to reduce genome heterozygosity for genome assembly; however, it is often unfeasible in small insects such as T. palmi, which are difficult to rear and need a large sample size for sequencing library
construction. Nevertheless, we demonstrate that current methods are appropriate for high quality de novo assembly of the genome of small organisms.

The assembled genome of *T. palmi* was larger than the estimated genome size, as is found in other cases (Quan et al., 2019), which might be caused by the assembly of highly heterozygous regions into different genomic regions or problems in accurate size estimation given heterozygosity. The proportion of the *T. palmi* genome involving duplicated single-copy genes evaluated in BUSCO was very low (0.8%), indicating that duplication was not a major issue in assembling the genome. Among species of Blattodea, Thysanoptera, and Hemiptera, this is the second genome assembled to the chromosomal level following *A. gossypii*. At the contig-level, the contig N50 of *T. palmi* is higher compared to other genomes of species of Thysanoptera and Hemiptera (Table 1). The completeness estimated using BUSCO is higher than for all species of Hemiptera (Table 1), providing confidence in the quality of the assembled genome.

**Varied genomic basis of detoxification and stress response in sap-sucking insects**

Detoxification is one of the major strategies that insects have evolved to counter toxins in their foods (Despres, David, & Gallet, 2007). Insects feeding on different food resources have evolved a variable number of detoxification genes (Gloss et al., 2019; Rane et al., 2019). In general, insects feeding on chemically complex tissues tend to have relatively more detoxification genes than those on relatively simple diets (Rane et al., 2019). Among species using the same food type, detoxification of plant chemicals can be based on a variety of pathways that determine host plant adaptation (Heckel, 2018). Thrips and hemipteran insects provide useful systems to understand these diverse responses because they represent sister groups of Condylognatha that include many sap-sucking species.

When we compared the polyphagous thrips with whitefly and aphids, we found that thrips and whitefly have relatively more detoxification genes than aphids. In sap-sucking hemipteran species, there is a close association between insects and endosymbionts (Baumann, 2005) which may play a role in dealing with plant toxins and this might lead to loss of detoxification genes that have evolved in aphids. Our comparisons of sap-sucking Condylognatha support the notion that polyphagous species usually have more detoxification genes than oligophagous species, although for polyphagous sap-sucking thrips and aphids a more extensive range of host plant may not necessarily translate into more detoxification genes. Based on the CABI database (https://www.cabi.org/), there are > 28
plants from 11 families used by *T. palmi*, compared to > 250 plants from > 65 families for *F. occidentalis*, > 600 plants from 89 families for *B. tabaci*, > 120 species from > 40 families for *M. persicae*, and >93 species from > 92 families for *A. gossypii*. These host plant numbers do not correlate with differences in numbers of detoxification genes (Table 2).

Phylogenetic analysis indicated that thrips (Thysanoptera) are the closest living relatives to Hemiptera, with divergence times before the Carboniferous period, over 365 million years ago (Johnson et al., 2018). Phylogenetic and fossil evidence suggest that the earliest hemipterans fed on detritus, pollen, fungi, or spores, and sap-sucking evolved convergently in thrips and hemipterans (Johnson et al., 2018). Our results suggest that a different genetic basis of detoxification may have evolved in thrips and aphids although more species need to be included in comparisons.

The HSP family represents stress proteins and molecular chaperones (King & MacRae, 2015) and appears to be expanded in thrips when compared to hemipteran insects. Aphids may be adapted to cold climates, with the species richness of this group being relatively greater in temperate regions (Heie, 2013). Aphids are generally sensitive to changes in temperature (Hullé, d’Acier, Bankhead-Dronnet, & Harrington, 2010) and may rely on endosymbionts for high-temperature resistance (Dunbar, Wilson, Ferguson, & Moran, 2007). There are few studies of temperature effects on thrips, although *F. occidentalis* can survive temperatures as high as 41 °C for 12 hours (J. C. Wang, Zhang, Li, Wang, & Zheng, 2014) and perhaps HSPs play a role in high-temperature tolerance in thrips species.

**Implications for pesticide resistance**

Some detoxification enzymes in insects are involved in both metabolizing plant toxins and insecticides (Despres et al., 2007). Given the high number of cytochrome P450s, GSTs and CCEs found in thrips, there is a high potential for this group to evolve pesticide resistance through detoxification. In thrips, enhanced activity of metabolic enzymes is usually involved in resistance to broad-spectrum pesticides, such as the organochlorine endosulfan and carbamate methiocarb (Jensen, 2000; Maymo, Cervera, Sarabia, Martinez-Pardo, & Garcia, 2002). Resistance in thrips to other pesticides has also been related to metabolic detoxifying enzyme activity (Bao, Kataoka, Fukada, & Sonoda, 2015; Bao et al., 2014; Bao & Sonoda, 2012; Maymo et al., 2002). Although both thrips sequenced so far show an abundance of detoxification genes, we found the CYP4 subfamily was expanded in *T. palmi* compared to *F. occidentalis*; this may contribute to the relatively high level of resistance of *T. palmi* to...
spinetoram when compared to *F. occidentalis* as well as *Frankliniella intonsa* (Y. F. Gao et al., 2019). Detoxification genes have also been important for resistance in aphids, with the green peach aphid *Myzus persicae* and the cotton aphid *A. gossypii* being particularly well-known agriculture pests that have evolved metabolic resistance to multiple chemicals (Chris Bass et al., 2014; Cao, Zhang, Gao, Liang, & Guo, 2008). We predicted more P450 and ABC genes in *F. occidentalis* than predicted in a different study (Rotenberg et al., 2020), although this paper was only available in a preprint form and used different analyses. We combined HMMER and BLAST analyses using the bioinformatic pipeline BITACORA (full mode) to ensure the accuracy of identification results. However, the relative abundance of the identified genes is similar between the two studies (Rotenberg et al., 2020).

The relatively high number of HSP genes in *T. palmi* may also have an impact on pesticide resistance. Several studies have shown that HSP family genes are highly expressed under exposure to pesticides (Lu et al., 2017; Nazir, Mukhopadhyay, Saxena, & Kar Chowdhuri, 2001; Yoshimi et al., 2009), such as the five HSP genes induced by exposure to avermectin in *F. occidentalis*, and the marked up-regulation of HSP70 transcripts in the brown planthopper *N. lugens* treated with imidacloprid (Si et al., 2019; H. H. Wang et al., 2014).

The spinosyn class of pesticides has provided alternatives to broad-spectrum pesticides for the control of thrips in the past few years (Cannon et al., 2007; Mouden et al., 2017; Reitz et al., 2020). However, resistance of several thrips to spinosyns has now been reported in many areas (Bao et al., 2014; Espinosa, Bielza, Contreras, & Lacasa, 2002; Fu et al., 2018; Y. F. Gao et al., 2019; Z. H. Wang et al., 2016), which has led to difficulties in thrips control and may be contributing to an acceleration of the displacement of local species by resistant thrips (Y. F. Gao et al., 2019; Zhao et al., 2017). Previous studies showed that spinosad targets the α subunit of nAChR (Cisneros et al., 2002; Connolly & Wafford, 2004; Jones & Sattelle, 2006). Mutation of nAChR α6 subunit transmembrane region (G275E) confers resistance to spinosad in *F. occidentalis*, *T. palmi*, and *Tuta absoluta* (Povolny) (Bao et al., 2014; Puinean, Lansdell, Collins, Bielza, & Millar, 2013; Silva et al., 2016). Truncated transcripts of nAChR α6 subunit caused by alternative splicing can also lead to pesticide resistance (Y. Wan et al., 2018). Alternative splicing is a novel mechanism for organisms to quickly respond to environmental stresses (Filichkin, Priest, Megraw, & Mockler, 2015). More subunit genes in thrips may compensate for the function of nAChR when the target...
542 gene is transcribed in an alternative pattern under stresses produced by pesticides, but this 543 remains to be tested.

544 Conclusions
545 We successfully assembled a genome for the pest thrips *T. palmi*, providing the first 546 chromosome-level genome for a species from the insect order Thysanoptera. This novel 547 genomic resource allowed us to compare genomic changes in the evolution of sap-sucking 548 insects from the Thysanoptera and Hemiptera. We found differences in the two groups for 549 the number of genes potentially involved in detoxification across multiple detoxification 550 gene families. We also noted differences in HSP gene numbers between the groups. The high 551 diversity of detoxification genes in thrips may contribute to the rapid evolution of pesticide 552 resistance in this species, and the diversity in HSPs may contribute to its broad geographic 553 distribution. Thrips have completely different types of mouthparts and physically interact 554 with plant cells in a different manner compared to hemipterans (Steenbergen et al., 2018). 555 The genomic resource provide here may help to understand this different interaction 556 between thrips and plants. Thrips are important vectors of viral disease (Whitfield, Ullman, 557 & German, 2005), and the genome also may provide insights into viral interactions, 558 especially with regard to horizontal gene transfer which plays an important role in 559 adaptation and evolution of insects. Based on the genes we have identified that may be 560 involved in insecticide resistance, the genome described here will also be useful in 561 understanding the evolution of resistance in thrips, such as by investigating the differential 562 expression of genes following sub-lethal insecticide exposure and through resequencing of 563 resistant strains. Overall, the *T. palmi* genome provides a useful resource for understanding 564 the genetic basis of traits that underlie the ecology of thrips and evolutionary divergence 565 and convergence of sap-sucking insects more generally. This genomic resource may also 566 ultimately be useful for the management of *T. palmi* such as through the identification of 567 novel targets for chemical control and resistance monitoring.

568 Data Availability Statement
570 Raw reads obtained for genome assembly have been deposited in the Sequence Read 571 Archive (SRA) repository (accession numbers: SRR11148454, SRR11591408 and

This article is protected by copyright. All rights reserved
Raw RNA-seq reads have been deposited in the SRA repository under NCBI BioProject PRJNA607377.

**Author contributions**

Shu-Jun Wei conceived and designed the study; Shao-Kun Guo, Pan Shi, Yong-Fu Gao, Ya-Jun Gong, Jin-Cui Chen conducted the field and bioassay works; Shao-Kun Guo, Pan Shi, Yong-Fu Gao conducted molecular works; Shao-Kun Guo, Li-Jun Cao, Wei Song, and Shu-Jun Wei analyzed the data; Shao-Kun Guo, Shu-Jun Wei, Ary Hoffmann discussed the results; Shao-Kun Guo, Shu-Jun Wei and Ary Hoffmann wrote the manuscript.

**Acknowledgments**

We thank Qiang Gao for his help on the assembly of Hi-C data, Xing-Chen Liu, Ze-Hua Wang, and Ming-Liang Li for their assistance of bioassay and transcriptome works. This research was supported by the Promotion and Innovation Program of BAAFS (KJCX20180415), the Natural Science Foundation of Beijing Municipality (6184037), Beijing Academy of Agriculture and Forestry Sciences (BAAFS) - The University of Melbourne (UOM) Joint Laboratory on Pest Control Research, and the Beijing Key Laboratory of Environmentally Friendly Pest Management on Northern Fruits (BZ0432).

**Reference**


This article is protected by copyright. All rights reserved


This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved

doi:10.1093/molbev/msh013


doi:10.1002/ps.4531


doi:10.1007/s002440010270

doi:10.1186/s12864-015-1525-1


doi:10.1093/molbev/msp077


This article is protected by copyright. All rights reserved


This article is protected by copyright. All rights reserved


This article is protected by copyright. All rights reserved


# Tables

## Table 1 Assembly features for genomes of *Thrips palmi* and other insect species

<table>
<thead>
<tr>
<th>Feature</th>
<th>Tpal</th>
<th>Focc</th>
<th>Bger</th>
<th>Znev</th>
<th>Nlug</th>
<th>Btab</th>
<th>Apis</th>
<th>Mper</th>
<th>Agos</th>
<th>Agly</th>
<th>Dnox</th>
<th>Epel</th>
<th>Dcit</th>
<th>Dmel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size (Mb)</td>
<td>270.43</td>
<td>262.25</td>
<td>1,942.92</td>
<td>462.56</td>
<td>586.53</td>
<td>541.14</td>
<td>331.22</td>
<td>280.65</td>
<td>289.11</td>
<td>376.77</td>
<td>660.00</td>
<td>463.20</td>
<td>137.07</td>
<td></td>
</tr>
<tr>
<td>No. Scaf./Chr.</td>
<td>16</td>
<td>18,479</td>
<td>24,818</td>
<td>31,663</td>
<td>46,559</td>
<td>19,751</td>
<td>5</td>
<td>4,021</td>
<td>4,718</td>
<td>941</td>
<td>5,637</td>
<td>1,979</td>
<td>161,988</td>
<td>7</td>
</tr>
<tr>
<td>Scaf. N50 (Mb)</td>
<td>14.67</td>
<td>0.418</td>
<td>1.007</td>
<td>0.716</td>
<td>0.340</td>
<td>3.083</td>
<td>126.41</td>
<td>0.426</td>
<td>0.428</td>
<td>2.391</td>
<td>0.379</td>
<td>0.375</td>
<td>0.105</td>
<td>24.116</td>
</tr>
<tr>
<td>No. contig</td>
<td>4,120</td>
<td>34,226</td>
<td>317,827</td>
<td>64,772</td>
<td>121,137</td>
<td>31,571</td>
<td>68,187</td>
<td>6,044</td>
<td>12,178</td>
<td>1,022</td>
<td>50,913</td>
<td>2,173</td>
<td>176,470</td>
<td>2,442</td>
</tr>
<tr>
<td>Contig N50 (Mb)</td>
<td>0.426</td>
<td>0.024</td>
<td>0.012</td>
<td>0.022</td>
<td>0.022</td>
<td>0.081</td>
<td>0.025</td>
<td>0.214</td>
<td>0.075</td>
<td>1.907</td>
<td>0.013</td>
<td>0.660</td>
<td>0.033</td>
<td>20.490</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>97.2</td>
<td>98.5</td>
<td>97.7</td>
<td>99.2</td>
<td>94.0</td>
<td>94.5</td>
<td>93.5</td>
<td>95.5</td>
<td>94.4</td>
<td>83.2</td>
<td>88.4</td>
<td>99.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. gene</td>
<td>16,333</td>
<td>17,546</td>
<td>28,774</td>
<td>15,876</td>
<td>27,571</td>
<td>15,664</td>
<td>20,601</td>
<td>16,352</td>
<td>14,694</td>
<td>19,182</td>
<td>19,097</td>
<td>12,022</td>
<td>22,786</td>
<td>17,468</td>
</tr>
</tbody>
</table>

Tpal: *Thrips palmi* (Thysanoptera: Thripidae); Focc: *Frankliniella occidentalis* (Thysanoptera: Thripidae) (Thomas et al., 2018); Bger: *Blattella germanica* (Blattodea: Ectobiidae) (Harrison et al., 2018); Znev: *Zootermopsis nevadensis* (Blattodea: Termopsidae) (Terrapon et al., 2014); Nlug: *Nilaparvata lugens* (Hemiptera: Delphacidae) (Xue et al., 2014); Btab: *Bemisia tabaci* (Hemiptera: Aleyrodidae) (Chen et al., 2016); Dcit: *Diaphorina citri* (Hemiptera: Liviidae) (Saha et al., 2017); Apis: *Acyrthosiphon pisum* (Hemiptera: Aphididae) (International Aphid Genomics, 2010); Dnox: *Diuraphis noxia* (Hemiptera: Aphididae) (Nicholson et al., 2015); Mper: *Myzus persicae* (Hemiptera: Aphididae) (Ramsey et al., 2007); Agos: *Aphis gossypii* (Hemiptera: Aphididae) (Quan et al., 2019); Agly: *Aphis glycines* (Hemiptera: Aphididae) (Wenger et al., 2017); Epel, *Ericerus pela* (Hemiptera: Coccidae) (Cryan & Urban, 2012). Completeness was estimated by BUSCO. Scaf., scaffold; Chr., chromosome.
Table 2: Statistics on detoxification, heat-shock protein (HSP) and nicotinic acetylcholine receptor (nAChR) genes across sap-sucking insects and other insects

<table>
<thead>
<tr>
<th>Common name</th>
<th>Species code</th>
<th>Feeding pattern</th>
<th>P450</th>
<th>GST</th>
<th>CCE</th>
<th>UGT</th>
<th>ABC</th>
<th>HSP</th>
<th>nAChR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thrips</td>
<td>Tpal</td>
<td>Polyphagous</td>
<td>96*</td>
<td>25*</td>
<td>39*</td>
<td>17*</td>
<td>49*</td>
<td>51*</td>
<td>10*</td>
</tr>
<tr>
<td>Focc</td>
<td></td>
<td>Polyphagous</td>
<td>66*</td>
<td>27*</td>
<td>28*</td>
<td>17*</td>
<td>125*</td>
<td>55*</td>
<td>8</td>
</tr>
<tr>
<td>Cockroach</td>
<td>Bger</td>
<td>Oligophagous</td>
<td>135*</td>
<td>33*</td>
<td>76*</td>
<td>50*</td>
<td>118*</td>
<td>48*</td>
<td>9</td>
</tr>
<tr>
<td>Termite</td>
<td>Znev</td>
<td>Polyphagous</td>
<td>76*</td>
<td>16*</td>
<td>35*</td>
<td>32*</td>
<td>60*</td>
<td>28*</td>
<td>10*</td>
</tr>
<tr>
<td>Planthopper</td>
<td>Nlug</td>
<td>Oligophagous</td>
<td>67</td>
<td>11</td>
<td>25*</td>
<td>18*</td>
<td>76*</td>
<td>27*</td>
<td>13</td>
</tr>
<tr>
<td>Whitefly</td>
<td>Btab</td>
<td>Polyphagous</td>
<td>130</td>
<td>22</td>
<td>51</td>
<td>81</td>
<td>50</td>
<td>33*</td>
<td>11*</td>
</tr>
<tr>
<td>Psyllid</td>
<td>Dcit</td>
<td>Oligophagous</td>
<td>60*</td>
<td>19*</td>
<td>20*</td>
<td>17*</td>
<td>53</td>
<td>22*</td>
<td>11*</td>
</tr>
<tr>
<td>Aphid</td>
<td>Apis</td>
<td>Oligophagous</td>
<td>83</td>
<td>20</td>
<td>29</td>
<td>58</td>
<td>117*</td>
<td>19*</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>55</td>
<td>10</td>
<td>20</td>
<td>43</td>
<td>66</td>
<td>24*</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Dnox</td>
<td>Oligophagous</td>
<td>115</td>
<td>21</td>
<td>22</td>
<td>60</td>
<td>59*</td>
<td>32*</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Mper</td>
<td>Polyphagous</td>
<td>62</td>
<td>7</td>
<td>20</td>
<td>56</td>
<td>64*</td>
<td>32*</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Agos</td>
<td>Polyphagous</td>
<td>64</td>
<td>11</td>
<td>23</td>
<td>49</td>
<td>87</td>
<td>20*</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Agly</td>
<td>Oligophagous</td>
<td>85</td>
<td>38</td>
<td>39</td>
<td>36</td>
<td>56</td>
<td>41</td>
<td>10</td>
</tr>
</tbody>
</table>

See Table 1 for the abbreviation of species names. * indicates that the number of genes was annotated in this study; [1] (Rotenberg et al., 2020), [2] (Xue et al., 2014), [3] (Chen et al., 2016), [4] (Z. Wang et al., 2019), [5] (Ramsey et al., 2010), [6] (Ahn, Vogel, & Heckel, 2012), [7]
Figure legends

Fig. 1 Image of adult *Thrips palmi* (a) and the damage symptoms on the fruit of eggplant (b) and the leaf of cucumber (c). Photos were taken by Shu-Jun Wei.

Fig. 2 GenomeScope analysis of genome size, heterozygosity and duplicate rate (a) and genome-wide all-by-all Hi-C interaction (b) of *Thrips palmi* genome. The Illumina short-read sequencing data was used to count k-mers in DNA by using the software Jellyfish with $K = 17$. The estimated genome size of *Thrips palmi* was 202 Mb, heterozygosity was 1.34%, and the duplicate rate was 1.96% (a). In total 16 linkage groups were identified based on Hi-C contact, indicated by blue boxes (b).

Fig. 3 Orthology and phylogenetic relationships among thrips and hemipteran insects. The maximum-likelihood phylogenetic tree was built using the single-copy orthologs (a). Relationships of ortholog genes were compared among 12 insect species (right). "1:1:1" indicates single-copy genes in all species; "N:N:N" indicates multi-copy genes in all species; "Thrips" indicates thrips-specific genes and presence in two thrips species; "Diptera" indicates dipteran-specific genes; "Hemiptera" indicates hemipteran-specific genes; "SD" indicates species-specific duplicated genes; "ND" indicates species-specific genes; "Other" indicates all other orthologous groups.

Fig. 4 Boxplots of the number of genes in four detoxification families (P450, GST, CCE, UGT and ABC), HSPs and nAChRs in Thysanoptera and Hemiptera. Bar graphs comparing gene numbers in subfamilies are also included. (a) Numbers of genes in five oligophagous and five polyphagous species (Table 2); (b) number of genes among two thrips, two aphids and one whitefly species (Table 2); number of P450 (c), HSP (d) and ABC genes (e) among two thrips, two aphids, and one whitefly species compared at the subfamily level. See Table 1 for abbreviations of species names.
Fig. 5 Phylogenetic relationships of *Thrips palmi* (TP) cytochrome P450 (P450) (a), heat shock protein (HSP) (b) and ATP-binding cassette (ABC) transporter gene families (c) in comparison with *Drosophila melanogaster* (DM) and *Frankliniella occidentalis* (FOCC).

Fig. 6 Phylogenetic tree (a) and gene number (b) of nicotinic acetylcholine receptors (nAChRs) in sap-sucking insects. DM/Dmel, *Drosophila melanogaster*; TP/Tpal, *Thrips palmi*; FO/Focc, *Frankliniella occidentalis*; NL/Nlug, *Nilaparvata lugens*; BT/Btab, *Bemisia tabaci*; AG/Agos, *Aphis gossypii*; Dcit, *Diaphorina citri*; Apis, *Acyrthosiphon pisum*; Dnox, *Diuraphis noxia*; Mper, *Myzus persicae*; Agly, *Aphis glycines*.

Supplemental informations

Table S1 Statistics for sequencing data for *Thrips palmi* genome assembly

Table S2 Completeness of *Thrips palmi* genome assembly and annotation evaluated by BUSCO based on insecta_odb9 database (1658 genes)

Table S3 Summary of Hi-C data for chromosome-level assembly of *Thrips palmi* genome

Table S4 Statistics for noncoding RNA genes in the genome of *Thrips palmi*

Table S5 Statistics for repeat elements in the genome of *Thrips palmi*

Table S6 Gene orthology comparison involving *Thrips palmi* and eight other insect species

Table S7 Differentially expressed genes in populations of *Thrips palmi* potentially related to resistance to spinetoram

Figure S1 GenomeScope analysis of genome size, heterozygosity and duplicate rate of *Thrips palmi* genome.

Figure S2 Venn diagram of functional annotation of protein-coding genes in the genome of *Thrips palmi* based on five databases.

Figure S3 Phylogenetic tree of *Thrips palmi* carboxylesterase (CCE) gene family in comparison with other insects. DM, *Drosophila melanogaster*; tp, *Thrips palmi*; FOCC, *Frankliniella occidentalis*.

This article is protected by copyright. All rights reserved
**Figure S4** Phylogenetic tree of *Thrips palmi* glutathione S-transferase (GST) gene family in comparison with other insects. DM, *Drosophila melanogaster*; tp, *Thrips palmi*; Ag, *Anopheles gambiae*.

**Figure S5** Phylogenetic tree of *Thrips palmi* UDP-glycosyltransferases (UGT) family in comparison with other insects. DM, *Drosophila melanogaster*; tp, *Thrips palmi*; FOCC, *Frankliniella occidentalis*.

**Figure S6** Dot plot of GO (a, c, e, g) and KEGG enrichment (b, d, f, h) of DEGs between SDF0 and BJF0 (a, b), SDF5 and BJF5 (c, d), SDF0 and SDF5 (e, f) and BJF0 and BJF5 (g, h).

**Figure S7** Venn diagram of up- (pink, light orange, green and light blue) and down-regulated (dark blue, purple, dark orange and gray) DEGs and resistant genes that were differentially expressed (SDF0 vs BJF0, SDF5 vs BJF5, SDF0 vs SDF5 and BJF0 vs BJF5) in populations of *Thrips palmi*. 
<table>
<thead>
<tr>
<th>Species</th>
<th>α1</th>
<th>α2</th>
<th>α3</th>
<th>α4</th>
<th>α5</th>
<th>α6</th>
<th>α7</th>
<th>α8</th>
<th>α9</th>
<th>α10</th>
<th>β1</th>
<th>β2</th>
<th>β3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tpal</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>10</td>
</tr>
<tr>
<td>Focc</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>8</td>
</tr>
<tr>
<td>Nlug</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>13</td>
</tr>
<tr>
<td>Btab</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>11</td>
</tr>
<tr>
<td>Dcit</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>11</td>
</tr>
<tr>
<td>Apis</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>11</td>
</tr>
<tr>
<td>Dnox</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>8</td>
</tr>
<tr>
<td>Mper</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>6</td>
</tr>
<tr>
<td>Agos</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>7</td>
</tr>
<tr>
<td>Agly</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>7</td>
</tr>
<tr>
<td>Dmel</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>10</td>
</tr>
</tbody>
</table>
Author/s:
Guo, S-K; Cao, L-J; Song, W; Shi, P; Gao, Y-F; Gong, Y-J; Chen, J-C; Hoffmann, AA; Wei, S-J

Title:
Chromosome-level assembly of the melon thrips genome yields insights into evolution of a sap-sucking lifestyle and pesticide resistance

Date:
2020-07

Citation:

Persistent Link:
http://hdl.handle.net/11343/275938