Received: 2 March 2020 | Revised: 29 May 2020 | Accepted: 1 June 2020

DOI: 10.1111/1755-0998.13206

RESOURCE ARTICLE

MOLECULAR ECOLOGY RESOURCES WILEY

A chromosome-level genome assembly of the parasitoid wasp *Pteromalus puparum*

Xinhai Ye[1,2] | Zhichao Yan[1] | Yi Yang[1] | Shan Xiao[1] | Longfei Chen[1] | Jiale Wang[1] | Fei Wang[1] | Shijiao Xiong[1] | Yang Mei[1] | Fang Wang[1] | Hongwei Yao[1] | Qisheng Song[3] | Fei Li[1] | Qi Fang[1] | John H. Werren[2] | Gongyin Ye[1]

[1]State Key Laboratory of Rice Biology, Ministry of Agricultural and Rural Affairs Key Laboratory of Molecular Biology of Crop Pathogens and Insect Pests & Key Laboratory of Biology of Crop Pathogens and Insects of Zhejiang Province, Institute of Insect Sciences, Zhejiang University, Hangzhou, China

[2]Department of Biology, University of Rochester, Rochester, NY, USA

[3]Division of Plant Sciences, College of Agriculture, Food and Natural Resources, University of Missouri, Columbia, MO, USA

Correspondence
Gongyin Ye and Qi Fang, State Key Laboratory of Rice Biology, Ministry of Agricultural and Rural Affairs Key Laboratory of Molecular Biology of Crop Pathogens and Insect Pests & Key Laboratory of Biology of Crop Pathogens and Insects of Zhejiang Province, Institute of Insect Sciences, Zhejiang University, Hangzhou, China.
Email: chu@zju.edu.cn (G.Y.); fangqi@zju.edu.cn (Q.F.)

John H. Werren, Department of Biology, University of Rochester, Rochester, NY, USA.
Email: jack.werren@rochester.edu

Funding information
National Natural Science Foundation of China, Grant/Award Number: 31701843; Major International (Regional) Joint Research Project of NSFC, Grant/Award Number: 31620103915; National Science Foundation, Grant/Award Number: DEB-125703 and NSF-IOS-1456233; National Key Research and Development Program of China, Grant/Award Number: 2017YFD0200400; Key Program of National Natural Science Foundation of China, Grant/Award Number: 31830074

Abstract

Parasitoid wasps represent a large proportion of hymenopteran species. They have complex evolutionary histories and are important biocontrol agents. To advance parasitoid research, a combination of Illumina short-read, PacBio long-read and Hi-C scaffolding technologies was used to develop a high-quality chromosome-level genome assembly for *Pteromalus puparum*, which is an important pupal endoparasitoid of caterpillar pests. The chromosome-level assembly has aided in studies of venom and detoxification genes. The assembled genome size is 338 Mb with a contig N50 of 38.7 kb and a scaffold N50 of 1.16 Mb. Hi-C analysis assembled scaffolds onto five chromosomes and raised the scaffold N50 to 65.8 Mb, with more than 96% of assembled bases located on chromosomes. Gene annotation was assisted by RNA sequencing for the two sexes and four different life stages. Analysis detected 98% of the BUSCO (Benchmarking Universal Single-Copy Orthologs) gene set, supporting a high-quality assembly and annotation. In total, 40.1% (135.6 Mb) of the assembly is composed of repetitive sequences, and 14,946 protein-coding genes were identified. Although venom genes play important roles in parasitoid biology, their spatial distribution on chromosomes was poorly understood. Mapping has revealed venom gene tandem arrays for serine proteases, pancreatic lipase-related proteins and kynurenine–oxoglutarate transaminases, which have amplified in the *P. puparum* lineage after divergence from its common ancestor with *Nasonia vitripennis*. In addition, there is a large expansion of P450 genes in *P. puparum*. These examples illustrate how chromosome-level genome assembly can provide a valuable resource for molecular, evolutionary and biocontrol studies of parasitoid wasps.

KEYWORDS
chromosome-level genome assembly, P450s expansion, parasitoid wasp, *Pteromalus puparum*, venom evolution

Xinhai Ye, Zhichao Yan and Yi Yang contributed equally to this work.

1 | INTRODUCTION

Parasitoid wasps are abundant and diverse insects, with more than 100,000 identified species, and possibly over 500,000 species yet to be identified (Asgari & Rivers, 2011; Heraty, 2009). Most parasitoid wasps share a common ancestor that lived in the Permian or Triassic (Peters et al., 2017). Female adult parasitoids lay their eggs on or into other arthropods, where the young develop and typically kill the host (Godfray, 1994; Pennacchio & Strand, 2006; Quicke, 1997). Hence, they are referred to as 'parasitoids' because the adult stage is free-living while the young develop on hosts. During parasitization, females inject effectors into the host, which can include venom proteins, small molecules (i.e., nonprotein compounds with low molecular weight, such as dopamine and bradykinins), polydnaviruses, virus-like particles and parasitoid cells (Asgari & Rivers, 2011; Heavner et al., 2017; Mrinalini & Werren, 2016; Pennacchio & Strand, 2006; Pichon et al., 2015; Strand, 2014; Strand & Burke, 2013). These effectors regulate host physiological processes, including immunity, development and metabolism, in order to protect and ensure offspring survival and successful development in the host (Asgari & Rivers, 2011; Heavner et al., 2017; Lin et al., 2019; Martinson, Wheeler, Wright, Mrinalini, & Werren, 2014; Mateo Leach et al., 2009; Moreau & Asgari, 2015; Mrinalini, Wright, Martinson, Wheeler, & Werren, 2015; Pennacchio & Strand, 2006; Tan et al., 2018; Wang, Ye, et al., 2018). Parasitoid wasp development eventually kills hosts; therefore, parasitoids are widely used as biological control agents against various insect pests (Beckage & Gelman, 2004; Quicke, 1997). Different parasitoid species show a wide range of hosts utilized, host developmental stages attacked (e.g., egg, larva, pupa, adult), egg deposition strategies (on or within the host) and levels of host specialization (Godfray, 1994; Pennacchio & Strand, 2006; Quicke, 1997). Because of their megadiverse and complex biology, effects on other insects, and pest control implications, parasitoid wasps are the subject of many biological studies (Beckage & Gelman, 2004; Moreau & Asgari, 2015; Strand, 2014; Strand & Burke, 2013; Werren et al., 2010).

Parasitoid lifestyle, behaviour and living habit are reflected at the genomic level (Tellier, Moreno-Gamez, & Stephan, 2014; Werren et al., 2010; Yin et al., 2018). Since the first parasitoid genomes were published in 2010, for *Nasonia vitripennis* and two congeners, *Nasonia giraulti* and *Nasonia longicornis* (Werren et al., 2010), at least 15 parasitoid wasp genome projects have been completed, and the number of new genome projects is rapidly increasing (Branstetter et al., 2018). Given their diverse biology, parasitoid wasps are now emerging as a model system for comparative genomics (Branstetter et al., 2018; Chen, Lang, et al., 2018). With the development of new sequencing technologies, a number of insect genome assemblies have now been improved to the chromosome level, which provides information that was not available in more fragmented genomes (Benetta et al., 2020; Fallon et al., 2018; Ma et al., 2020; Wallberg et al., 2019; Wan et al., 2019).

The chromosomal evolution of Hymenoptera is complex, with a variable haploid chromosome number (n) ranging from 1 to 60 (Gokhman, 2009). For parasitoid wasps, the haploid number range is 3–23 (Gokhman, 2009). However, in the Hymenoptera, only six genomes have chromosome-level assemblies. These are the honey bee *Apis mellifera* (using Hi-C, $n = 16$), buff-tailed bumblebee *Bombus terrestris* (using Roche 454, $n = 18$), ant *Ooceraea biroi* (using Hi-C, $n = 14$), parasitoid *N. vitripennis* (using recombination mapping and long-read sequencing, $n = 5$) and redheaded pine sawfly *Neodiprion lecontei* (using quantitative trait locus [QTL] mapping, $n = 7$; Benetta et al., 2020; Desjardins et al., 2013; Linnen, O'Quin, Shackleford, Sears, & Lindstedt, 2018; McKenzie & Kronauer, 2018; Sadd et al., 2015; Wallberg et al., 2019; Werren et al., 2010). Recently, a chromosome-level genome assembly of the parasitoid *Cotesia congregata* (using Hi-C, $n = 10$) has been reported in the preprint server bioRxiv (Gauthier et al., 2020). Additional chromosomal-level genome assemblies will improve comparative genome analyses of the Hymenoptera.

Pteromalus puparum is a pupal endoparasitoid wasp from the family Pteromalidae (superfamily Chalcidoidea; Figure 1). It has high reproductive output (Takagi, 19855) and utilizes numerous hosts, including more than 20 species from Papilionidae, Pieridae, Nymphalidae, Geometridae, Coleophoridae, Psychidae, Braconidae, Ichneumonidae, Pteromalidae and Vespidae (Peck, 1963). Pupae of the butterfly *Pieris rapae* (Lepidoptera: Pieridae) are its major hosts, which *P. puparum* parasitizes on an almost global scale in North America, Europe, Oceania, Africa and Asia. It contributes significantly to natural control of *Pieris rapae*, and is also a laboratory model for parasite–host interactions (Barron, Barlow, & Wratten, 2003; Blunck, 1957; El-Fakharany & Hendawy, 2014; Hu, 1983, 1984; Wang et al., 2019). Like most chalcidoid wasps, *P. puparum* does not have polydnaviruses or other virus-like particles as part of its venom repertoire. However, it does utilize a diverse set of venom proteins, which promote parasitoid offspring development (Cai, Ye, & Hu, 2004; Fang, Wang, Ye, Wang, & Ye, 2016; Fang et al., 2010, 2011). Over the last two decades, the composition and biological functions of the venom protein components of *P. puparum* have been studied (Wang et al., 2013; Yan et al., 2016), as well as venom interactions with the pupal host immunity (Yan et al., 2017).

Here, we report a high-quality chromosome-level genome assembly for *P. puparum* using a combination of Illumina short-read sequencing, PacBio long-read sequencing and Hi-C technology. The assembly has high completeness, providing an excellent genomic resource for subsequent research. In this study, besides the basic genome description and comparative genomics, we also investigated the location and distribution of venom genes on chromosomes, and describe a large expansion of P450 genes in *P. puparum*. Our findings provide a valuable genomic resource for molecular and evolutionary studies in parasitoid wasps (e.g., parasitoid venom evolution), as well as future potential biological control applications.

2 | MATERIALS AND METHODS

2.1 | Genome sequencing

Parasitoid wasps *Pteromalus puparum* and hosts *Pieris rapae* were reared under laboratory conditions as previously described (Cai

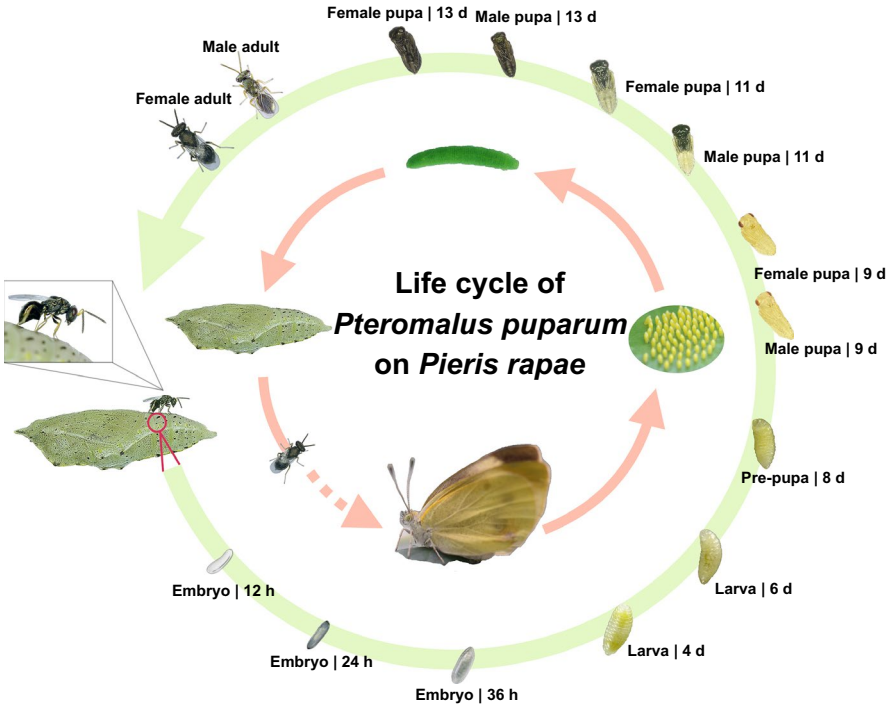


FIGURE 1 Life cycle of the parasitoid wasp *Pteromalus puparum* on its host *Pieris repae*. Female wasps lay eggs into the host chrysalis. The embryos hatch into larvae in 48 hr after parasitism, and the larval stage lasts 7 days. The pupal stage of *P. puparum* is around 7 days and then the adults emerge. The life span of *P. puparum* adults is ~20 days [Colour figure can be viewed at wileyonlinelibrary.com]

et al., 2004; Wang et al., 2017). The laboratory strain (Ppup-ZJU) used for DNA extraction was initially started with 50 wasps in 2001, and has been maintained since. We applied the Illumina HiSeq 2000 and PacBio platforms to sequence the genome of *P. puparum*. DNA was extracted from 50 yellow haploid male pupae of *P. puparum* as described by Werren et al. (2010). Sequencing libraries were generated using a Truseq Nano DNA HT Sample preparation Kit (Illumina). For paired-end libraries with 500-bp, 2- and 5-kb insert lengths, genomic DNA first sheared via the Covaris S2 system (Covaris, Inc.). For paired-end libraries with insert lengths of 10 and 15 kb, genomic DNA was sheared using the Covaris g-TUBE device (Covaris, Inc.). Then, DNA fragments were end-polished, A-tailed and ligated with the full-length adapter. After adapter ligation, DNA fragments were PCR-amplified. Amplified DNA fragments were purified with Agencourt AMPure XP beads and final library was analysed for size distribution using an Agilent 2100 Bioanalyzer and quantified using real-time PCR. In total, five libraries with average insert sizes of 500 bp, 2, 5, 10 and 15 kb were constructed, generating 83×, 38×, 33×, 32× and 37× coverage with the HiSeq 2000 platform, respectively. In addition, we generated 5.72 Gb (18× coverage) of PacBio data, with an average read length of 2.8 kb using PacBio P6-C4 chemistry by PacBio RS II. The sequencing depth was 241×, based on *P. puparum* genome size estimated by K-mer analysis.

2.2 | Estimation of genome size

Genome size was estimated by K-mer analysis (Marcais & Kingsford, 2011), using filtered reads from the library with short insert size of 500 bp and K-mer size of 17.

2.3 | Genome assembly

The *P. puparum* genome was assembled using SOAPDENOVO version 1.0 (Li et al., 2010), to first construct de Brujin graphs with automatically optimized K-mer sizes based on the short insert libraries. Then contigs were joined into scaffolds using paired-end or mate-pair information by SSPACE version 2.0 (Boetzer, Henkel, Jansen, Butler, & Pirovano, 2011) and unresolved gaps regions were filled by local reassembly using GAPCLOSER in SOAPDENOVO2 (Luo et al., 2012). In addition, we used PacBio long reads to close remaining gaps in scaffolds by PBJELLY in PBSUITE version 15.8.24 (English, Salerno, & Reid, 2014). Contaminating bacterial scaffolds were removed using the pipeline of Wheeler, Redding, and Werren (2013) and as further described in Benoit et al. (2016) and Ferguson et al. (2020).

2.4 | Hi-C library preparation

The Hi-C (van Berkum et al., 2010) library was prepared using the standard protocol modified for application to whole insects (Xiao et al., 2020). In brief, 50 freshly harvested parasitoid male yellow pupae were cut into small pieces and immersed in 2% formaldehyde for 10 min for crosslinking. Crosslinking was stopped by adding 2.5 m glycine to 0.2 m solution for 5 min. Fixed tissue was frozen in liquid nitrogen and ground to powder for the isolation of nuclei following the protocol descried by Xiao et al. (2020). The purified nuclei were digested with 100 units of *Hind*III and marked by incubating with biotin-14-dCTP. Biotin-14-dCTP from nonligated DNA ends was removed owing to the exonuclease activity of T4 DNA polymerase. The ligated DNA was sheared into fragments of 300–600 bp, and then was

blunt-end repaired and A-tailed, followed by purification through biotin–streptavidin-mediated pull down. Finally, the Hi-C libraries were quantified and sequenced using the Illumina Hiseq platform.

2.5 | Scaffolding with Hi-C

The chromosome-level genome was constructed with Hi-C data using the HIC-PRO and LACHESIS pipeline. In total, 370 million paired-end reads were generated from the libraries. Then, quality control of Hi-C raw data was performed using HIC-PRO version 2.8.0 (Servant et al., 2015). First, low-quality sequences (quality scores <20), adaptor sequences and sequences shorter than 30 bp were filtered out using FASTP version 0.12.6 (Chen, Zhou, Chen, & Gu, 2018), and then the clean paired-end reads were mapped to the draft assembled sequence using BOWTIE2 version 2.3.2 (Langmead & Salzberg, 2012). Read pairs are first independently aligned on the reference genome to avoid any constraint on the proximity between the two reads. The unmapped reads were mainly composed of the regions spanning the ligation junction. According to the Hi-C protocol, HIC-PRO version 2.8.0 was used to detect the ligation site of unmapped reads using an exact matching procedure and aligned its 5′ fractions back onto the genome. The results of two mapping steps were merged in a single alignment file. As a result, 107 million uniquely mapped pair-end reads were generated, of which 76.28% were determined to be valid interaction pairs following the HIC-PRO pipeline (Servant et al., 2015). Combined with the valid Hi-C data, we subsequently used the LACHESIS (https://github.com/shendurelab/LACHESIS) *de novo* assembly pipeline to produce chromosome-level scaffolds (Burton et al., 2013). The scaffolds were concatenated with 100 Ns to create five pseudochromosomes. These five pseudochromosomes correspond to the five chromosomes known for *P. puparum* (Dozortseva, 1936; Gokhman, 2009), although the genome assembled pseudochromosomes have not yet been mapped to chromosomes cytologically.

2.6 | Assessment of the genome completeness and quality

We used two different methods to check the completeness and quality of the assembly. First, all the RNA-seq reads from different tissues were mapped to the assembly genome using TOPHAT2 (Kim et al., 2013) with default parameters. Unigenes from different tissue transcriptomes were mapped to the assembly genome with BLAT version 3.2.22 (Kent, 2002) using default parameters. Second, BUSCO version 3.0 (Waterhouse et al., 2017) was used to search the 1,658 bench-marking universal single-copy orthologous genes in insecta_odb9.

2.7 | Repeat annotation

We detected repetitive sequences and transposable elements (TEs) in the genome using a combination of *de novo* and homology-based approaches. In brief, we first constructed a *de novo* repeat library using REPEATMODELER version 1.0.7 (http://www.repeatmasker.org/RepeatModeler.html) with default parameters. Then, REPEATMASKER version 4.0.5 (Tarailo-Graovac & Chen, 2009) was applied using both the repeat database Repbase (http://www.girinst.org/repbase) with default parameters. The repeat divergence rate was measured by the percentage of substitutions in the corresponding regions between annotated repeats and consensus sequences constructed by REPEATMASKER version 4.0.5. To identify simple sequence repeats (SSRs) in the *P. puparum* genome, we used MICROSATELLITE (MISA; http://misaweb.ipk-gatersleben.de/; Beier, Thiel, Munch, Scholz, & Mascher, 2017) with default parameters.

2.8 | Gene prediction and functional annotation

We used a combination of homology-based methods, *de novo* methods and RNA-seq-based methods to predict the genes in the *P. puparum* genome. For *de novo* prediction, AUGUSTUS version 3.1 (Stanke, Steinkamp, Waack, & Morgenstern, 2004), SNAP version 2006-07-28 (Korf, 2004) and GENEMARK-ET SUITE 4.21 (Lomsadze, Burns, & Borodovsky, 2014) were performed on the repeat masked genome with parameters trained on 1,000 randomly selected intact genes from the transcriptomes of *P. puparum* and *N. vitripennis*, respectively. For homology-based prediction, protein sequences from five sequenced insect species, including *Apis mellifera*, *Anopheles gambiae*, *Drosophila melanogaster*, *N. vitripennis* and *Tribolium castaneum*, were initially mapped onto the *P. puparum* genome using TBLASTN version 2.8.1 (E-value < 1e−5), and then the genome sequences were aligned to the matching protein using GENEWISE version 2.4.1 (Birney & Durbin, 2000) to define the gene models. Furthermore, the RNA-seq reads were aligned to the reference genome by TOPHAT2 to identify candidate exon regions, and the donor and acceptor sites. Then the alignments were assembled into transcripts using CUFFLINKS version 2.2.1 (Trapnell et al., 2012) under default parameters, and open reading frames were predicted to obtain reliable transcripts by using hidden Markov model-based training parameters. To finalize the gene set, all the predictions were combined by EVIDENCEMODELER (EVM) version 1.1.1 (Haas et al., 2008) to produce a consensus gene set. Gene Ontology (GO) analysis was carried out using the software BLAST2GO version 4.1 (Conesa et al., 2005). The *P. puparum* protein sequences were mapped to the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways using BLASTKOALA version 2.2 online service (Kanehisa, Sato, & Morishima, 2016). Motifs and domains were annotated using INTERPROSCAN version 5.13-52.0 (Jones et al., 2014) by searching against the publicly available database InterPro. We also searched the SwissProt and Trembl databases using BLASTP version 2.8.1 (*E*-value < 1e−5).

2.9 | RNA sequencing and analysis

In total, six RNA-seq libraries were constructed, including four developmental stages of *P. puparum*. Total RNA was isolated and libraries prepared from embryos (<10 hr after laying eggs, 400

individuals), larvae (combined 1–3-day-old larvae, 60 individuals), yellow male pupae (10 individuals), yellow female pupae (10 individuals), and newly emerged adult males (10 individuals) and adult females (10 individuals). The transcriptomes were sequenced using the Illumina HiSeq 2000 platform with paired-end libraries. Low-quality reads were filtered using TRIMMOMATIC version 0.38 (Bolger, Lohse, & Usadel, 2014). The RNA-seq clean reads were *de novo* assembled using TRINITY version 2.8.2 (Haas et al., 2013). In addition, the RNA-seq clean reads were mapped to the *P. puparum* genome using TOPHAT2. Then the alignments were assembled into transcripts using CUFFLINKS version 2.2.1 with default parameters (Trapnell et al., 2012). RSEM version 1.3.1 was used for estimating the gene expression level (Li & Dewey, 2011).

2.10 | Comparative genomics

Orthologous and paralogous gene groups were identified from the protein sequences of 14 insects, using ORTHOMCL version 2.0.9 (Li, Stoeckert, & Roos, 2003), and subsequently analysed using in-house Perl scripts. We compared PpOGS1 (official gene set 1 of *P. puparum*) and NvOGS2 (official gene set 2 of *N. vitripennis*) using BLASTP version 2.8.1 (*E*-value < 1e−10), then scanned NvOGS2 genes that had not been annotated as genes in PpOGS1 in the *P. puparum* genome using GENOMETHREADER version 1.7.1 (Gremme, Brendel, Sparks, & Kurtz, 2005) with the default parameters. Sequences from the following hymenopteran insects were employed: *P. puparum* (this study), *Athalia rosae* (RefSeq assembly accession: GCF_000344095.2), *Orussus abietinus* (RefSeq assembly accession: GCF_000612105.2), *Macrocentrus cingulum* (InsectBase, http://www.insect-genome.com), *Microplitis demolitor* (RefSeq assembly accession: GCF_000572035.2), *Diachasma alloeum* (RefSeq assembly accession: GCF_001412515.2), *Fopius arisanus* (RefSeq assembly accession: GCF_000806365.1), *Copidosoma floridanum* (RefSeq assembly accession: GCF_000648655.2), *Trichogramma pertiosum* (RefSeq assembly accession: GCF_000599845.2), *Ceratosolen solmsi* (RefSeq assembly accession: GCF_000503995.1), *N. vitripennis* (OGS2.0, Rago et al., 2016), *Atta cephalotes* (RefSeq assembly accession: GCF_000143395.1) and *Apis mellifera* (RefSeq assembly accession: GCF_003254395.2). The coleopteran species *Tribolium castaneum* (RefSeq assembly accession: GCF_000002335.3) was used as an outgroup.

2.11 | Chromosome synteny

Chromosome synteny between *P. puparum*, *N. vitripennis* and *Apis mellifera* was detected and plotted using MCSCAN (Python version) (https://github.com/tanghaibao/jcvi/wiki/MCscan-(Python-version)) with default parameters. The chromosome-level genome assembly of *Apis mellifera* (Amel_HAv3.1; Wallberg et al., 2019) was downloaded at https://www.ncbi.nlm.nih.gov/assembly/GCF_003254395.2. The chromosome-level genome assembly of *N. vitripennis* (Nvit_psr_1; Benetta et al., 2020) was downloaded at https://www.ncbi.nlm.nih.gov/assembly/GCA_009193385.1#/st.

2.12 | Phylogenetic analysis

We reconstructed a phylogeny for 14 insects using the single-copy genes from ORTHOMCL results. The protein sequences of each gene family were independently aligned by MAFFT version 7 (Katoh & Standley, 2013), filtered by TRIMAL version 1.2 (Capella-Gutierrez, Silla-Martinez, & Gabaldon, 2009) with the default parameters, and then concatenated into one super-sequence for phylogenomic analysis. The bootstrap phylogenetic tree was constructed by maximum likelihood (ML) using IQ-TREE package version 1.5.5 (Nguyen, Schmidt, von Haeseler, & Minh, 2015) with the best model (LG + I + G) estimated by MODELFINDER (Kalyaanamoorthy, Minh, Wong, von Haeseler, & Jermiin, 2017). The MCMCTREE program in the PAML package version 4.9e (Yang, 1997) was used to estimate divergence time based on protein sequences. Three calibration time points based on fossil records (https://www.paleobiodb.org/), stem Chalcidoidea: *Khutelchalcis gobiensis* (125.0–129.4 million years ago [Ma]), stem Braconidae: *Cretobraconus maculatus* (145.0–152.1 Ma), stem Hymenoptera: *Triassoxyela foveolata* (237.0–228.0 Ma), were used for divergence time calibration. The tree was visualized using FIGTREE version 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/). The resulting phylogeny is consistent with other published insect and hymenopteran analyses (Peters et al., 2017, 2018; Zhang et al., 2020).

2.13 | Gene family expansion and contraction

We used CAFÉ software version 4.2.1 (De Bie, Cristianini, Demuth, & Hahn, 2006) to analyse the gene family expansion and contraction. The gene family results from the TREEFAM version 9 pipeline (Li et al., 2006) and the estimated divergence time between species were used as inputs. For the specific expanded or contracted gene families, we conducted GO and KEGG pathway annotation by BLAST2GO version 4.1 and BLASTKOALA version 2.2 online service. Omicshare CLOUDTOOLS (http://www.omicshare.com/tools/?l=en-us) was used for enrichment analyses. GO results were summarized and visualized by REVIGO (http://revigo.irb.hr/; Supek, Bošnjak, Škunca, & Šmuc, 2011).

2.14 | Gene family analysis

For targeted gene family annotation, we first downloaded a set of reference protein sequences from NCBI GenBank and manually confirmed these sequences to obtain clean reference sequences for each gene family. TBLASTN version 2.8.1 was used to search candidate sequences in the *P. puparum* genome assembly (*E*-value < 1e−5). GENEWISE version 2.4.1 was used to define the gene structure. Then, we confirmed the candidate sequences using HMMER version 3.3

TABLE 1 Assembly statistics for five chalcidoid genomes and one braconid genome

Features	*Pteromalus puparum*	*Nasonia vitripennis* (Nvit_2.1)	*Nasonia vitripennis* (Nvit_psr_1)	*Ceratosolen solmsi*	*Copidosoma floridanum*	*Trichogramma pertiosum*	*Microplitis demolitor*
Taxonomy							
Superfamily	Chalcidoidea	Chalcidoidea	Chalcidoidea	Chalcidoidea	Chalcidoidea	Chalcidoidea	Ichneumonoidea
Family	Pteromalidae	Pteromalidae	Pteromalidae	Agaonidae	Encyrtidae	Trichogrammatidae	Braconidae
Sequencing info							
Sequencing technology	Pacbio + Illumina+Hi-C	Sanger	PacBio + Oxford Nanopore	Illumina	Illumina	Illumina	Illumina
Genome coverage	241× (Pacbio + Illumina) + 100x (Hi-C)	6.0x	142x	93.0x	139.0x	232.7x	26.0x
Assembly info							
Assembly level	Chromosomes ($n = 5$, using Hi-C)	Chromosomes ($n = 5$, using genetic markers)	Chromosomes ($n = 5$, using genetic markers)	Scaffolds	Scaffolds	Scaffolds	Scaffolds
Genome size (Mb)	338.10	295.78	297.31	277.94	553.96	195.09	241.19
Number of contigs	22,140 (22,172 after Hi-C)	25,484	444	9,928	31,515	1,475	27,508
Number of scaffolds	3,446 (2,738 after Hi-C)	6,169	—	2,457	4,840	925	1,794
Genome assembly quality							
Contig N50 (bp)	38,691 (38,626 after Hi-C)	18,840	7,180,486	74,395	40,744	78,655	14,116
Scaffold N50 (bp)	1,160,518 (65,768,243 after Hi-C)	708,988	—	9,558,897	1,210,516	3,706,225	1,139,389
BUSCO genes (%)	98	96.2	—	97.7	96	97.5	97
Genomic features							
GC content (%)	40.62	41.7	—	30.4	36	39.9	33.1
Repeat content (%)	40.11	20.63	—	3.17	36.85	30.3	36.15
Number of protein-coding genes	14,946	24,388	—	11,412	12,143	12,928	12,670

(Meng & Ji, 2013) against the Pfam database (Finn et al., 2016; *E*-value < 1e−5). For gene family tree reconstruction, we aligned protein sequences of each gene family using MAFFT version 7 and filtered by TRIMAL version 1.2 to get the conserved blocks. IQ-TREE version 1.5.5 was used to construct the phylogenetic tree with the best model estimated by MODELFINDER (1,000 ultrafast bootstrap approximation replicates). The tree was visualized using FIGTREE version 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/). The R package RIDEOGRAM version 0.1.1 (Hao et al., 2020) was used to map and visualize genes in chromosomes.

2.15 | Venom gene location on chromosome

To locate all 70 *P. puparum* venom genes identified by Yan et al. (2016) onto the chromosome, we first searched venom genes in PpOGS1 using BLASTP version 2.8.1 (*E*-value < 1e−10, identity > 90). For the venom genes that were misannotated, we manually annotated these genes in the *P. puparum* genome by TBLASTN version 2.8.1 (*E*-value < 1e−10) and GENEWISE version 2.4.1. The chromosomal gene location was transferred to an AGP file using an in-house python script. In addition, we also mapped 52 *N. vitripennis* venom genes (de Graaf et al., 2010) onto chromosomes (Nvit_2.1) using the same method. Venom gene synteny analysis was performed using MCSCAN (Python version) with default parameters.

3 | RESULTS AND DISCUSSION

3.1 | Chromosome-level genome assembly

The genome of the parasitoid wasp *Pteromalus puparum*, sequenced using both PacBio and Illumina HiSeq 2000 platforms, generated 5.72 Gb PacBio long reads and 71.51 Gb Illumina short reads, with 241 × genome coverage. Genome size was estimated to be 320 Mb based on 17 K-mer analysis (Tables S1–S3; Figure S1). After filtering the bacterial contaminated scaffolds (5.5 Mb), we obtained a reference *P. puparum* genome of 338 Mb with contig N50 of 38.69 kb and scaffold N50 of 1,160 kb. The GC content of the *P. puparum* genome was 40.62% (Tables 1 and S4). We next used Hi-C technology to improve the genome assembly to the chromosomal level. According to previous studies, the number of chromosomes (*n*) of *P. puparum* is five (Dozortseva, 1936; Gokhman, 2009). In total, Hi-C linking information supported 738 scaffolds being anchored, ordered and oriented to give five chromosomes, with more than 96.5% of assembled bases located on the chromosomes (Figures 2, S2; Table S5;). The length of the largest chromosome was 83.7 Mb, while the smallest one was 49.5 Mb (Table S5). The scaffold N50 of the chromosome-level genome assembly reached 65.8 Mb (Table S6).

This is a high-quality genome assembly, as >99% of the *de novo* assembled transcripts could be mapped to the reference genome (Table S7); we also assessed the completeness of the assembly using Benchmarking Universal Single-Copy Orthologs (BUSCO), 1,624 out of 1,658 (98%) conserved arthropod genes were found in the genome. This BUSCO score represented the highest among five sequenced wasps in the superfamily Chalcidoidea (Table 1). Another example of chromosome-level genome assembly of chalcidoid parasitoid wasps is *Nasonia vitripennis* (Nvit_2.1), which used recombination mapping by a long oligomer hybridization array (Desjardins et al., 2013). A new chromosome-level genome assembly of *N. vitripennis* (Nvit_psr_1) that was sequenced using PacBio and Nanopore technologies was recently released (Benetta et al., 2020). It uses a combination of long reads and mapping of the oligos from Desjardins et al. (2013) to place scaffolds onto the five linkage groups. The assembly of *P. puparum* provides an important addition to the set of reliable and high-completeness chromosome-level genome assemblies in parasitoid wasps, to be used for focused research on the biology of *P. puparum*, as well as for comparative genomics.

3.2 | Genome annotation

In total, 40.1% of the *P. puparum* genome consists of repeat sequences, which is a higher percentage than most for other sequenced insects in Hymenoptera (Tables 1 and S8). However, this may reflect the quality of the assembly, as repetitive DNA often fails to assemble into scaffolds using earlier methods. We also identified in total 207,894 SSRs in the *P. puparum* genome, with 45% of scaffolds (1,543 out of 3,446) containing SSR sequences (Table S9). The TEs in *P. puparum* are mostly DNA transposons (11.98% of genome) and long terminal repeat retrotransposons (LTRs, 9.89% of the genome) (Table S7). Four types of TE (SINE, LINE, LTR and DNA) of *P. puparum* are all more abundant than in *N. vitripennis* (Table S8; Figure S3a). Sequence divergence rate analyses showed that TE sequences of *P. puparum* form a peak at a low divergence rate of ~3%, indicating a recent expansion of TEs in *P. puparum* (Figure S3b). This TE divergence pattern can also be found in *N. vitripennis* and the braconid wasp *Microplitis demolitor*, but not in the sawfly *Athalia rosae* or honey bee *Apis mellifera* (Figure S3b).

We predicted 14,946 protein-coding genes (PpOGS1) in the genome of *P. puparum* based on three lines of evidence: *de novo* prediction, homology alignment and assembled RNA-seq transcripts. More than 91% of the predicted genes (13,571 genes) have homology in public databases of Swissprot, Trembl, KEGG, GO and InterPro or are supported by RNA-seq data from at least one stage of *P. puparum* development, including embryonic, larval, pupal and adult stages (Table S10). The protein-coding gene number of *P. puparum* is close to that of most parasitoid wasps, such as *Copidosoma floridanum* (family: Encyrtidae, 12,750 genes), *Trichogramma pretiosum* (family: Trichogrammatidae, 12,928 genes), *Ceratosolen solmsi* (family: Agaonidae, 11,412 genes) and *Microplitis demolitor* (family: Braconidae, 12,670 genes), but it has a large gap as compared to *N. vitripennis's* 24,388 predicted genes in NvOGS2 (Table 1). We reasoned that

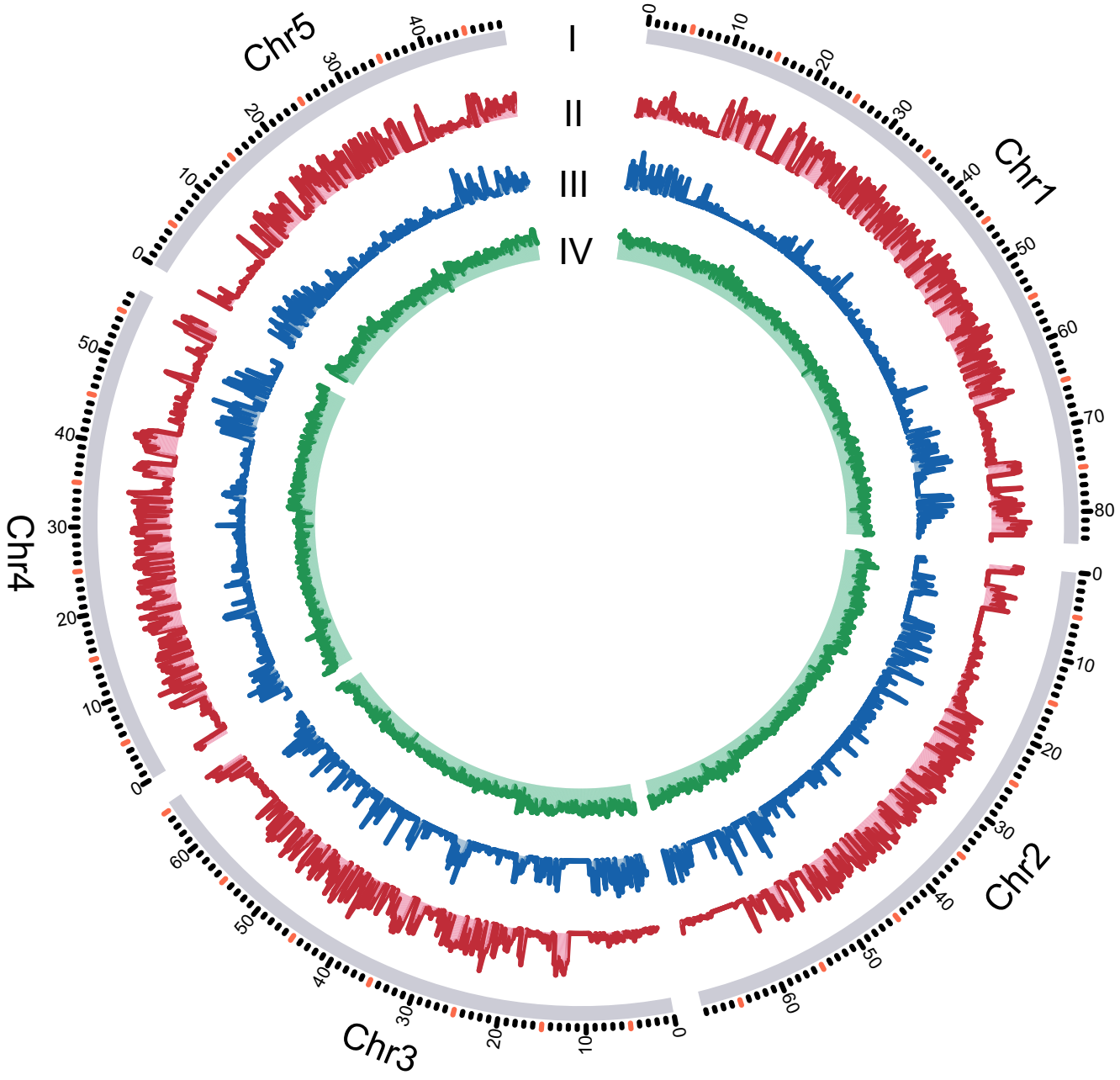


FIGURE 2 Genome landscape of the parasitoid wasp *Pteromalus puparum*. From outer to inner circles: I, five chromosomes at the Mb scale; II and III, repeat density (red) and gene density (blue) across the genome, respectively, drawn in 0.1-Mb nonoverlapping windows; IV, CG contents (green) across the genome, drawn in 0.1-Mb nonoverlapping windows [Colour figure can be viewed at wileyonlinelibrary.com]

the gap was mainly caused by different annotation methods. For better comparative analysis with *N. vitripennis*, we compared PpOGS1 and NvOGS2 by BLASTP version 2.8.1, then scanned 8,828 NvOGS2 genes that had not been annotated as genes in PpOGS1, in the *P. puparum* genome by GENOMETHREADER. The result shows that an additional 2,551 NvOGS2 genes have homology in the *P. puparum* genome (Table S11). These may be the genes 'missed' in the original annotation, and will need to be examined further in subsequent annotations as additional RNA sequencing data become available.

We compared the exon and intron lengths across five chalcid wasps, and found that exon length is significantly shorter in *P. puparum* (mean exon length is 241.9 bp) than *N. vitripennis* (259.3 bp; $p < 2.2e{-}16$,*** Mann–Whitney *U* test), while intron length is longer than *N. vitripennis* (mean intron length is 1,584.2 bp vs. 1,273.6 bp; $p < 2.2e{-}16$, Mann–Whitney *U* test). Mean exon length in *Copidosoma floridanum* is about two-fold higher longer than other four chalcid wasps (Table S12).

In addition, we also manually identified several gene families which are important in insect adaptation, including 117 cytochrome

P450s, 27 UDP-glucuronosyltransferases (UGTs), 26 glutathione *S*-transferases (GSTs), 64 ATP-binding cassette transporters (ABCs), 63 gustatory receptors (GRs), 51 ionotropic receptors (IRs), 232 olfactory receptors (ORs), 55 odorant binding proteins (OBPs), 12 sensory neuron membrane proteins (SNMPs) and 11 chemosensory proteins (CSPs; Table S13).

3.3 | Comparative genomics

We identified 17,802 gene family clusters using ORTHOMCL among *P. puparum* and the 13 other insect species used in our analysis (Table S14). The gene family clusters were divided into five categories, universal single-copy genes, multiple-copy genes, specific duplications, dispensable clusters (present in two to 13 species) and species-specific genes (unclustered genes). A total of 5,022 specific duplicated gene families and 2,042 species-specific genes were identified in the *P. puparum* genome (Table S14). The complexity of gene families was then compared between *P. puparum* and the other three hymenopteran insects: 6,447 gene families were shared by *P. puparum*, *N. vitripennis*, *Macrocentrus cingulum* and *Apis mellifera*; and 2,145 gene families were shared by *P. puparum* and *N. vitripennis* but not the other two hymenopterans. Only 68 gene families were shared between *P. puparum* and *Macrocentrus cingulum* but not with the other two; 110 gene families were shared by *P. puparum* and *Apis mellifera* but not with the other two (Figure S4).

3.4 | Phylogenetic reconstruction

The phylogenetic relationships between *P. puparum* and the 12 other representative hymenopteran insects used in this study were determined with a genome-wide set of 3,399 single-copy genes. The red flour beetle *Tribolium castaneum* was used as the outgroup (Figure 3a). As expected, the ML phylogenetic analysis showed that *P. puparum* has a closer relationship to *N. vitripennis* than the other three chalcidoids. The five chalcidoids (*P. puparum*, *N. vitripennis*, *Ceratosolen solmsi*, *Trichogramma pretiosum* and *Copidosoma floridanum*) cluster together, with four braconids (*F. arisanus*, *D. alloeum*, *Microplitis demolitor* and *Macrocentrus cingulum*) as a sister group. The parasitoid wood wasp *Orussus abietinus* was placed as a sister to the Apocrita clade, which includes nine wasps, the honey bee *Apis mellifera* and the ant *Atta cephalotes*. Estimated divergence times of *P. puparum* and other species (calculated using MCMCTREE) suggest that *P. puparum* diverged from the common ancestor of the other members in the subfamily Pteromalinae approximately 19 Ma (Figure 3a). The lineage to which *Copidosoma floridanum* (family Encyrtidae) belongs was estimated to have diverged from the other chalcidoid wasps ~113 Ma, which is earlier than the ~49 Ma estimated based on the phylogenetic analysis of Peters et al. (2018). This difference may reflect the more extensive taxon sampling and use of fossil data for outgroup time calibration in that study. If we recalibrate our estimates using the 49-Ma time point for the common ancestor of *P. puparum*–*Copidosoma floridanum*, then the time estimate for the

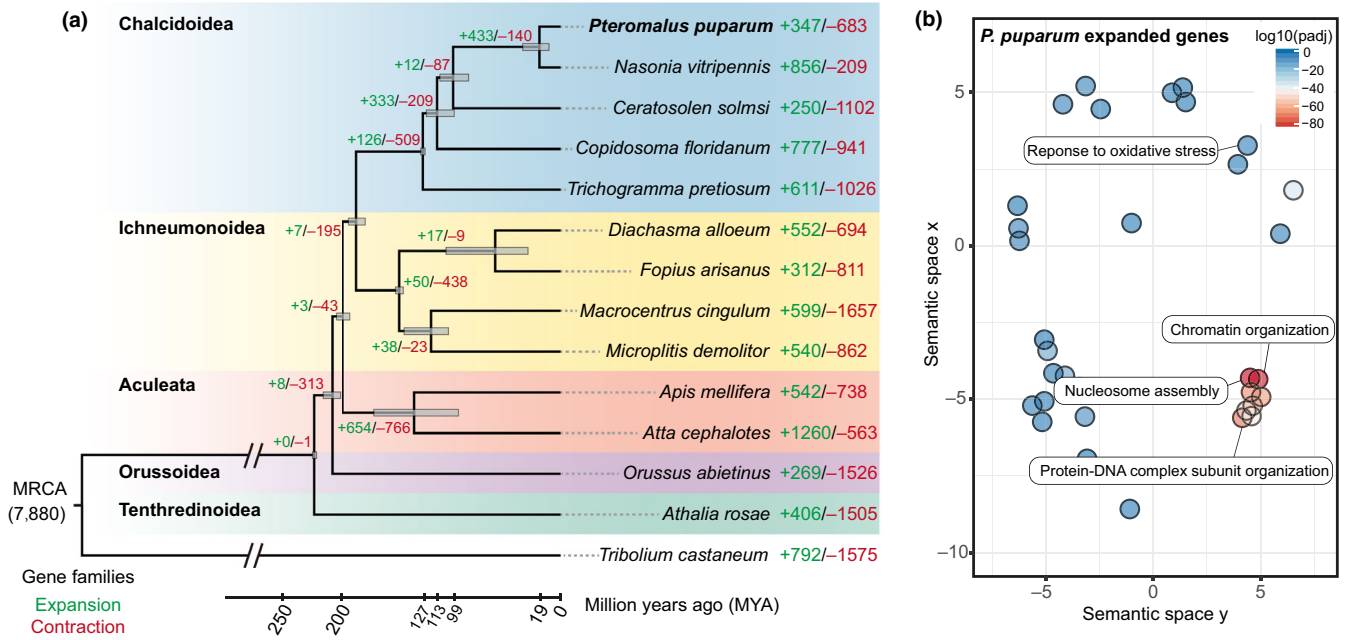


FIGURE 3 Comparative genomic analysis of the parasitoid wasp *Pteromalus puparum*. (a) A maximum-likelihood phylogenetic tree is shown for *P. puparum* with 12 hymenopterans and one beetle. The phylogenetic tree was based on 3,399 single-copy proteins. The red flour beetle *Tribolium castaneum* was used as the outgroup. The bootstrap value of all nodes is supported at 100/100. Divergence times are indicated by light grey bars at the internodes, and the bars indicates the 95% confidence interval of the divergence time. The expansion numbers of gene families (green) and contraction (red) are shown on the branches. (b) GO terms of *P. puparum* expanded genes are summarized and visualized as a scatter plot [Colour figure can be viewed at wileyonlinelibrary.com]

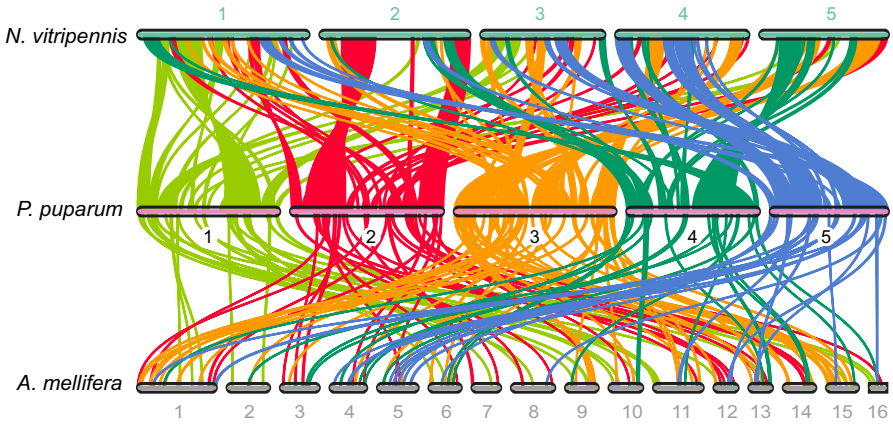


FIGURE 4 Chromosome synteny based on CDS pairwise alignment between *Pteromalus puparum*, *Nasonia vitripennis* and *Apis mellifera*. Coloured lines indicate shared syntenic blocks (containing at least 10 orthologous genes) [Colour figure can be viewed at wileyonlinelibrary.com]

common ancestor of *P. puparum*–*N.vitripennis* would be adjusted to ~8.2 Ma. The topology of this phylogeny is supported by previous phylogenetic studies for Chalcidoidea and Hymenoptera using transcriptome data (Peters et al., 2017, 2018; Zhang et al., 2020).

3.5 | Chromosome synteny

Synteny of the *P. puparum* assembly was compared to *N. vitripennis* and *Apis mellifera*. The chromosome synteny analysis was based on a CDS (coding region of a gene) pairwise synteny search using MCSCAN (Python version) pipeline. In our analysis, we defined that a syntenic block should include at least 10 orthologous genes. Because of the poor coverage of assembly Nvit_2.1 (more than 61 Mb, 20% unplaced length), in this analysis, we used a newly released genome assembly Nvit_psr_1, which was generated by PacBio and Nanopore sequencing technologies, as the reference genome assembly of *N. vitripennis* (Benetta et al., 2020). *P. puparum* showed a low level of synteny with the more distantly related hymenopteran *Apis mellifera* (Figure 4). However, when compared with *N. vitripennis*, we identified 118 syntenic blocks, and the gene numbers in these syntenic blocks were in the range 11–556 with an average of 97.8. This compares to an average of 19.2 genes per 136 synteny blocks for *Apis mellifera*. As expected, the synteny length is significantly greater between *P. puparum* and *N. vitripennis* than between *P. puparum* and *Apis mellifera* ($p < .0001$, Mann–Whitney *U* test). Comparison between *N. vitripennis* and *Apis mellifera* indicated 415 syntenic blocks with 10–101 genes and an average of 21.42 (Figure S5), similar to the comparison of *P. puparum* to *Apis mellifera*. The results suggest that many genome rearrangements have occurred since divergence of *P. puparum* and *N. vitripennis* (according to this study, about 19 Ma), which is reflected in the synteny map (Figure 4).

3.6 | Patterns of gene family expansion and contraction

We used TREEFAM and CAFÉ software to study gene family expansion and contraction during the evolution of *P. puparum* and related species (Figure 3a). By comparing 14 species, we found that the *P. puparum* genome displayed 347 expanded and 683 contracted gene families compared to the common ancestor of *P. puparum* and *N. vitripennis* (Figure 3a; Table S15). However, *N. vitripennis* contained 856 expanded and 209 contracted gene families. There were 301 specific expanded gene families and 440 specific contracted gene families for *P. puparum*, and 608 specific expanded gene families and 168 specific contracted gene families for *N. vitripennis*. Only five expanded gene families were identified in both species and no gene families were contracted in both (Table S15; Figure S6).

GO and KEGG enrichment analysis revealed that the *P. puparum* expanded gene families are enriched in nucleosome assembly (GO: 0006334, 113 genes, $p = 2.419e{-}85$, false discovery rate [FDR]-adjusted), chromatin organization (GO: 0006325, 125 genes, $p = 2.696e{-}85$), protein catabolic process (GO: 0030163, 25 genes, $p = .000367$), apoptotic process (GO: 0006915, 10 genes, $p = .000450$), and response to oxidative stress (GO: 0006979, 10 genes, $p = .00169$) (Tables S16 and S17; Figure 3b). Conversely, *P. puparum* showed significantly contracted gene families in sensory perception, chitin catabolic process and lipid metabolic process ($p = 2.616e{-}123$, $p = 6.277e{-}10$ and $p = 4.005e{-}17$, respectively; Tables S18 and S19). The *N. vitripennis* expanded gene families are enriched in histone methylation (GO: 0016571, 24 genes, $p = .000414$), protein alkylation (GO: 0008213, 26 genes, $p = .000414$), and response to chemicals (GO: 0042221, 143 genes, $p = .00148$) and salivary secretion (ko04970, eight genes, $p = .000172$) (Tables S20–S21). Our data show significant differences in genome features between *P. puparum* and *N. vitripennis*, and most of these are associated with epigenetic modification, metabolic pathways and chemoreception systems.

3.7 | Venom proteins in the *P. puparum* genome

Venom is one of the most important tools used by parasitoid wasps to affect successful parasitism. Parasitoid wasp venom contains numerous biologically active compounds, which can manipulate the metabolism and gene expression of the host to create a suitable environment for wasp larvae (Asgari & Rivers, 2011; Fang

FIGURE 5 Venom proteins in the *Pteromalus puparum* genome. (a) The genomic positions of 60 identified venom genes in the *P. puparum* genome are shown. Those genes with high expression in venom gland are shown in blue (FPKM > 10 in venom gland, FPKM < 0.5 in carcass). Three venom gene families involving tandem repeats are shown in rectangles with different colours. (b) Gene expression of venom genes is shown across development and for different tissues for each of the three tandem repeat venom gene families and for the flanking genes of each venom gene family (labelled in grey). (c) Gene synteny is indicated for three venom tandem array regions between *P. puparum* and *Nasonia vitripennis*

et al., 2016, 2010, 2011; de Graaf et al., 2010; Martinson et al., 2014; Moreau & Asgari, 2015; Mrinalini et al., 2015; Wang et al., 2013; Yan et al., 2017). Venom genes have been identified by two major criteria: (a) presence in the venom reservoir supported by proteomics, and (b) high expression in the venom gland (Martinson, Mrinalini, Kelkar, Chang, & Werren, 2017; Yan et al., 2016). The venom proteins of several parasitoid wasps have so far been reported (Colinet et al., 2013, 2014; Danneels, Rivers, & de Graaf, 2010; de Graaf et al., 2010; Martinson et al., 2017; Teng et al., 2017; Vincent et al., 2010; Yan et al., 2016). Here we utilize the venom repertoire identified in Yan et al. (2016). To adapt to different hosts, parasitoid wasp venom compositions turnover and evolve quickly, making parasitoid wasp venom an excellent system for studying gene function evolution (Cavigliasso et al., 2019; Colinet et al., 2013; Goecks et al., 2013; Martinson et al., 2017). In contrast to the classic duplication and neofunctionalization model of venom evolution described in some venomous animals, such as snakes (Wong & Belov, 2012), the mechanism of new venom gene evolution in parasitoids often involves single-copy genes acquiring a venom function, either by 'moonlighting' (i.e., when a protein evolves to perform more than one function) (Jeffery, 1999) or by co-option (i.e., shifting to a new function), through rapid *cis*-regulatory expression evolution (Martinson et al., 2017). Venom tandem arrays have been reported in several snakes (Schield et al., 2019; Vonk et al., 2013), but the distribution of venom genes in parasitoid wasps remains poorly understood. According to the co-option evolution model of parasitoid venom genes, the distribution of venom genes in the chromosome would probably be dispersive rather than clustered. With the high-quality chromosome-level genome assembly, we can explore the distribution of venom genes in the chromosome, which provides a nice resource for future studies on parasitoid venom evolution.

In this study, we located onto chromosomes 70 previously identified *P. puparum* venom genes (Yan et al., 2016). Of these, the location of 60 venom genes are shown in Figure 5a, and they are distributed across all five of the *P. puparum* chromosomes. The remaining 10 venom genes were in unanchored scaffolds (Tables S22 and S23). Most of these venom genes were much more highly expressed in the venom gland than in adult female tissues minus the venom gland (i.e., carcass), typically with FPKM (fragments per kilobase of transcript per million mapped reads) >5 in carcass, and a venom gland/carcass ratio >4 (Yan et al., 2016). Eight venom genes are highly expressed in the venom gland and very lowly expressed in the carcass (FPKM > 10 in venom gland, FPKM < 0.5 in carcass). These are probably highly specific venom genes (Figure 5a; Table S22).

Most of the venom genes (52) are scattered across the genome and do not occur in tandem arrays, which is consistent with the model of venoms evolving from single-copy genes. However, three venom gene families involving tandem repeats occur on three different chromosomes, indicating probable gene family expansion by tandem duplications. These include serine protease, pancreatic lipase-related protein and kynurenine–oxoglutarate transaminase (Figure 5a,b). Phylogenetic analyses confirm that each tandem repeat venom family was derived from a single ancestral gene by tandem duplication (Figure S7). A serine protease gene tandem array contains five genes in chromosome 2 (Chr2). Using our previous published RNA-seq data (Yan et al., 2016), we found that three of these genes have a similar expression pattern, with 201–1,228-fold higher expression in venom gland than carcass, and very low expression across development (Figure 5b). This is not due to cross-mapping of RNA-seq reads because they differ by an average 39.5% pairwise identity. The pattern suggests that serine protease genes expanded after evolution of venom function. Two of these were previously identified as venom genes, but another one (PPU12256) also shows a 627-fold higher expression in venom gland than carcass (Figure 5b). We confirmed the cellular export signal peptides of PPU12256. The other two members in the tandem array of serine protease genes are more highly expressed in male adult and larva (Figure 5b), but are not highly expressed in the venom gland.

We also found possible venom genes not detected in previous studies (Yan et al., 2016). For example, two tandem arrays contain five and three chymotrypsin genes, respectively, and are separated by one nonchymotrypsin gene (PPU12249; Figure 5b). Although none were previously identified as venom protein genes, three of five chymotrypsin genes in one tandem array show 15–448-fold higher expression in the venom gland than in the remaining carcass (Yan et al., 2016). The remaining two are more highly expressed in larva, and not highly expressed in the venom gland. In the tandem array which contains three chymotrypsin genes, one gene is highly expressed in the venom gland and the other two are more highly expressed in adult males and females (Figure 5b). This supports the view that the chymotrypsin genes highly expressed in venom gland might have potential function in venom (Figure 5b), although that function is currently unknown. The pattern is consistent with neofunctionalization to venom protein, followed by gene family expansion in each array. However, a more detailed phylogenetic and gene expression analysis in related species is needed to determine functional evolution in the two sets of paralogous genes. A third tandem array with new venom gene candidates contains three kynurenine–oxoglutarate transaminase genes that are located on Chr4 (Figure 5b). Two of these were previously identified as venom genes, but the third (PPU12438) also shows an expression pattern indicative of venom function, with 1,083-fold higher expression in the venom gland than in the remaining carcass (Figure 5b).

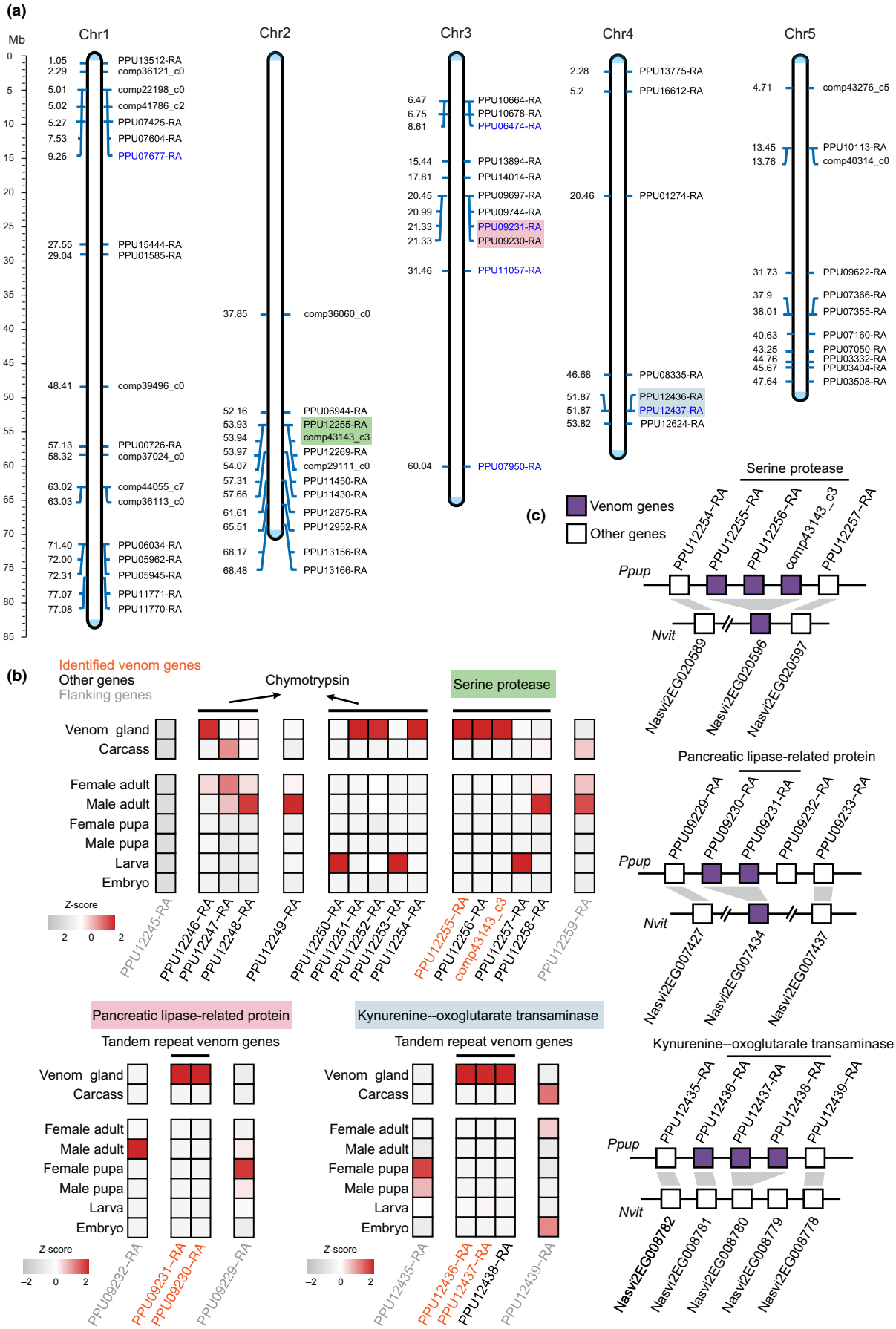

(a)
Mb
Chr1
Chr2
Chr3
Chr4
Chr5
1.05 PPU13512-RA
2.29 comp36121_c0
5.01 comp22198_c0
5.02 comp41786_c2
5.27 PPU07425-RA
7.53 PPU07604-RA
9.26 PPU07677-RA
27.55 PPU15444-RA
29.04 PPU01585-RA
48.41 comp39496_c0
57.13 PPU00726-RA
58.32 comp37024_c0
63.02 comp44055_c7
63.03 comp36113_c0
71.40 PPU06034-RA
72.00 PPU05962-RA
72.31 PPU05945-RA
77.07 PPU11771-RA
77.08 PPU11770-RA
37.85 comp36060_c0
52.16 PPU06944-RA
53.93 PPU12255-RA
53.94 comp43143_c3
53.97 PPU12269-RA
54.07 comp29111_c0
57.31 PPU11450-RA
57.66 PPU11430-RA
61.61 PPU12875-RA
65.51 PPU12952-RA
68.17 PPU13156-RA
68.48 PPU13166-RA
6.47 PPU10664-RA
6.75 PPU10678-RA
8.61 PPU06474-RA
15.44 PPU13894-RA
17.81 PPU14014-RA
20.45 PPU09697-RA
20.99 PPU09744-RA
21.33 PPU09231-RA
21.33 PPU09230-RA
31.46 PPU11057-RA
60.04 PPU07950-RA
2.28 PPU13775-RA
5.2 PPU16612-RA
20.46 PPU01274-RA
46.68 PPU08335-RA
51.87 PPU12436-RA
51.87 PPU12437-RA
53.82 PPU12624-RA
4.71 comp43276_c5
13.45 PPU10113-RA
13.76 comp40314_c0
31.73 PPU09622-RA
37.9 PPU07366-RA
38.01 PPU07355-RA
40.63 PPU07160-RA
43.25 PPU07050-RA
44.76 PPU03332-RA
45.67 PPU03404-RA
47.64 PPU03508-RA
(b)
Identified venom genes
Other genes
Flanking genes
Chymotrypsin
Serine protease
Venom gland
Carcass
Female adult
Male adult
Female pupa
Male pupa
Larva
Embryo
Z-score
−2 0 2
PPU12245-RA
PPU12246-RA
PPU12247-RA
PPU12248-RA
PPU12249-RA
PPU12250-RA
PPU12251-RA
PPU12252-RA
PPU12253-RA
PPU12254-RA
PPU12255-RA
PPU12256-RA
comp43143_c3
PPU12257-RA
PPU12258-RA
PPU12259-RA
(c)
Venom genes
Other genes
Serine protease
PPU12254-RA
PPU12255-RA
PPU12256-RA
comp43143_c3
PPU12257-RA
Ppup
Nvit
Nasvi2EG020589
Nasvi2EG020596
Nasvi2EG020597
Pancreatic lipase-related protein
PPU09229-RA
PPU09230-RA
PPU09231-RA
PPU09232-RA
PPU09233-RA
Nasvi2EG007427
Nasvi2EG007434
Nasvi2EG007437
Pancreatic lipase-related protein
Tandem repeat venom genes
PPU09232-RA
PPU09231-RA
PPU09230-RA
PPU09229-RA
Kynurenine--oxoglutarate transaminase
Tandem repeat venom genes
PPU12435-RA
PPU12436-RA
PPU12437-RA
PPU12438-RA
PPU12439-RA
Kynurenine--oxoglutarate transaminase
PPU12435-RA
PPU12436-RA
PPU12437-RA
PPU12438-RA
PPU12439-RA
Nasvi2EG008782
Nasvi2EG008781
Nasvi2EG008780
Nasvi2EG008779
Nasvi2EG008778

FIGURE 6 Cytochrome P450 genes in the *Pteromalus puparum* genome. (a) The maximum-likelihood phylogenetic tree of P450 proteins is shown for four chalcidoids (*P. puparum*, *Nasonia vitripennis*, *Copidosoma floridanum* and *Trichogramma pretiosum*) using the IQ-TREE package. Four main clades of P450 genes are indicated. Branch length scale indicates average residue substitutions per site. Genes in *CYP6* clusters are highlighted with stars. (b) The distribution of 92 predicted P450 genes in the *P. puparum* genome. The distribution analysis shows that there are 12 gene clusters which have three or more P450 genes. Gene density is visualized across the genome by a heat map [Colour figure can be viewed at wileyonlinelibrary.com]

PPU12438 has a cellular export signal peptide predicted by SIGNALP 5.0 software, suggesting it might be a new candidate venom protein. The pattern is consistent with gene duplications after neofunctionalization of the progenitor to venom function.

Based on the Nvit_2.1 assembly, 52 out of 79 identified venom genes (de Graaf et al., 2010) were located onto *N. vitripennis* chromosomes. In total, four venom gene tandem arrays were found on Chr1 and Chr3 of *N. vitripennis* (Figure S8). A large venom tandem array in Chr3 contained four venom protein genes (venom proteins Q, G, V and Z) and two cysteine-rich/pacifastin venom protein genes. We compared the venom distributions between *P. puparum* and *N. vitripennis* (two species in the family Pteromalidae divergent by ~19 million years, Figure 3a), and found that these two relatively closely related wasps did not share any ancestral venom tandem arrays.

We also examined an expanded venom repertoire reported for *N. vitripennis* which includes additional 21 genes (Martinson et al., 2017). In that expanded set, we found an orthologue to two of the three venom tandem arrays present in *P. puparum* (the serine protease and pancreatic lipase-related protein arrays). However, in both cases, there was only a single copy in *N. vitripennis*, indicating that the tandem arrays probably expanded in the lineage leading to *P. puparum*. However, although there are two kynurenine–oxoglutarate transaminase orthologues in *N. vitripennis*, they are not venom genes. Therefore, we concluded that these kynurenine–oxoglutarate transaminase genes both evolved into venom genes and then underwent an expansion to three venom paralogues in *P. puparum* (Figure 5c).

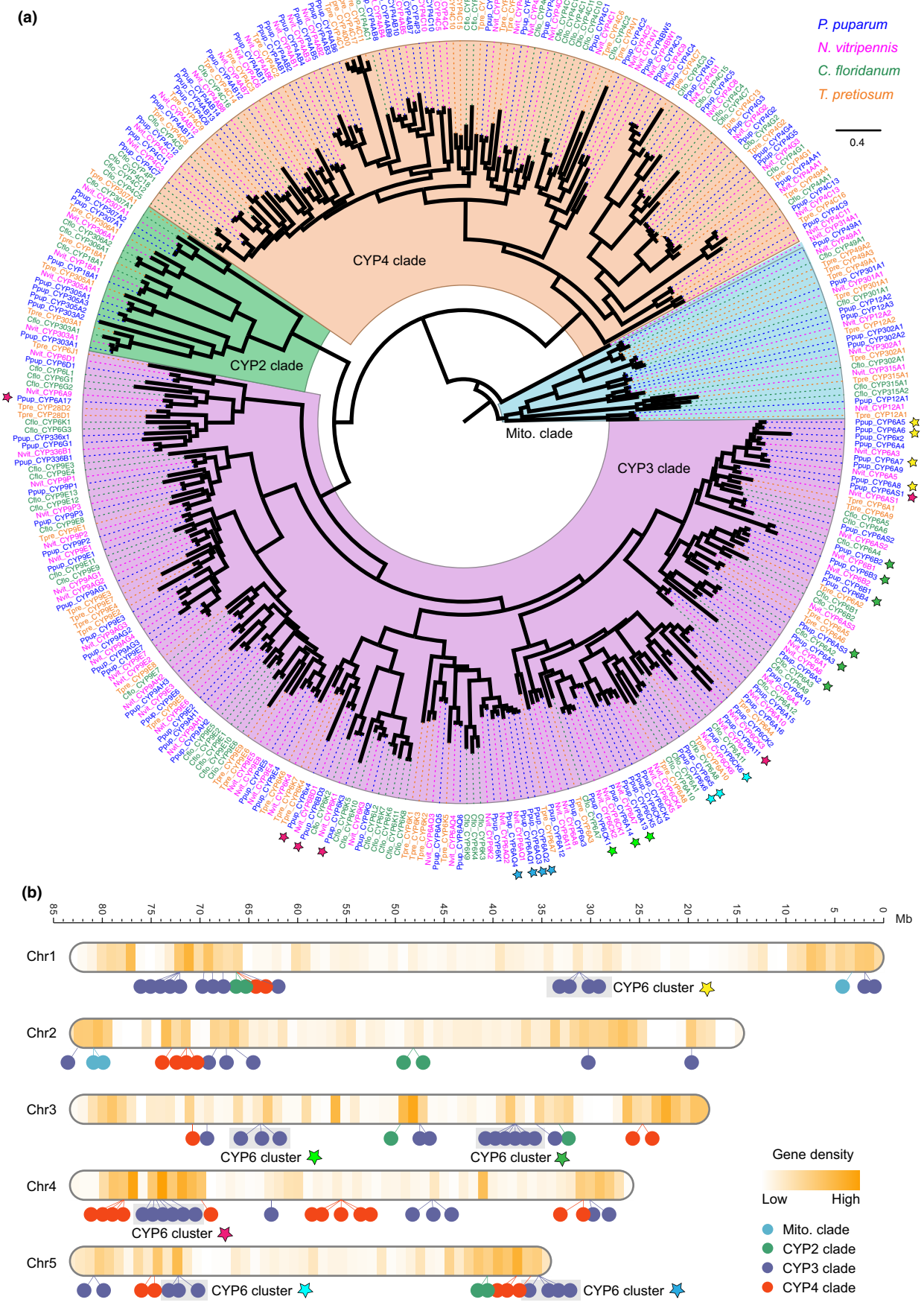
The result supports previous findings of rapid turnover and recruitment of venom genes from nonvenom orthologues in parasitoid wasps (Martinson et al., 2017). Tandem arrays are found in gene families involved in adaptation to the environment (Francino, 2005). The venom tandem arrays in *P. puparum* are different from those in *N. vitripennis*, suggesting that these venom tandem arrays might play important roles in *P. puparum's* adaptive evolution after diverging from *N. vitripennis*, such as to changes in host usage from Diptera to Lepidoptera. The primary hosts of *P. puparum* are the pupae of Lepidoptera, while the hosts of *N. vitripennis* are the pupae of Diptera. Given the development of high-quality parasitoid genome assemblies, future detailed analyses on the origins of venom tandem arrays and regulation of venom genes will soon be possible.

3.8 | Cytochrome P450 gene expansion

Cytochrome P450 is an ancient and large superfamily involved in the metabolism of endogenous and exogenous compounds (Scott & Wen, 2001). Because of its contribution to adaptation to exogenous compounds and pesticide resistance, this gene family has been well studied in insects (Cheng et al., 2017; Feyereisen, 1999; Wan et al., 2019; Wang et al., 2014; Wang, Shi, et al., 2018). In total, we annotated 117 P450 genes in the *P. puparum* genome, more than so far found in any other sequenced parasitoid wasp (Yin et al., 2018) (Table S13). The closely related species *N. vitripennis* has 93 P450 genes, the second largest number in parasitoid wasps. Phylogenetic analysis indicates that P450 clans 3 and 4 underwent large shared expansions in *P. puparum* and *N. vitripennis* compared to more basal wasps (*Copidosoma floridanum* and *Trichogramma pretiosum*) in Chalcidoidea, suggesting the expansion events occurred in the common ancestor of *P. puparum* and *N. vitripennis*. Additional independent expansions also occurred in the *P. puparum* branch (Figure 6a; Table S24). Mapping of 92 P450 genes to the *P. puparum* chromosomes shows 12 gene clusters which have three or more P450 genes (Figure 6b). The *CYP6* gene was greatly expanded and formed six clusters in *P. puparum* on Chr1, Chr3, Chr4 and Chr5 (Figure 6b). In several dipteran and lepidopteran insects, the expansion of *CYP6* is involved in resistance to pesticides and xenobiotic metabolism (Feyereisen, 1999, 2006; Müller et al., 2008; Wang, Shi, et al., 2018). The expansion of P450 genes (mainly in clans 3 and 4) in *P. puparum* might have evolved to overcome the different metabolites in the host body, such as plant-derived toxins and insecticides, or to adult exposure to these toxins. These expansions also might be related to its diverse host range, including insects from Papilionidae, Pieridae, Nymphalidae, Geometridae, Coleophoridae, Psychidae, Braconidae, Ichneumonidae, Pteromalidae and Vespidae (Peck, 1963).

4 | CONCLUSION

We have provided a high-quality chromosome-level genome assembly of *Pteromalus puparum* using a combined Illumina + PacBio + Hi-C assembly strategy in parasitoid wasps. Basic genomic analyses revealed genome features, phylogenetic position and gene family evolution of *P. puparum*. We mapped venom genes to the genome and identified three venom gene tandem arrays. We also found a large expansion of P450 genes in *P. puparum*. This genome assembly serves as a useful resource for further research into parasitoid biology, parasitoid–host interactions, comparative genomics, chromosome evolution and biological pest control.


(a)
P. puparum
N. vitripennis
C. floridanum
T. pretiosum
0.4
CYP4 clade
CYP2 clade
Mito. clade
CYP3 clade
(b)
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
Mb
Chr1
Chr2
Chr3
Chr4
Chr5
CYP6 cluster
CYP6 cluster
CYP6 cluster
CYP6 cluster
CYP6 cluster
Gene density
Low
High
Mito. clade
CYP2 clade
CYP3 clade
CYP4 clade

ACKNOWLEDGEMENTS

This work was supported by Major International (Regional) Joint Research Project of NSFC (Grant no. 31620103915 to G.Y.Y.), Key Program of National Natural Science Foundation of China (NSFC) (Grant no. 31830074 to G.Y.Y.), National Key Research and Development Program of China (Grant no. 2017YFD0200400 to H.W.Y.), and National Natural Science Foundation of China (NSFC) (Grant no. 31701843 to Z.C.Y.). X.H.Y. thanks 'Academic Star' Program for PhD Student of Zhejiang University for support. J.H.W. thanks the US NSF DEB-125703, NSF-IOS-1456233, and the Nathaniel & Helen Wisch Chair for support. We thank Dr Omar S. Akbari (University of California) and Dr Igor Antoshechkin (California Institute of Technology) for early access to the new assembly of Nasonia vitripennis. We thank the help from Nextomics Biosciences Institute for technical support in Illumina, PacBio and Hi-C. We thank Dr Floria M. K. Uy (University of Rochester) for her comments and suggestions on the article.

CONFLICT OF INTEREST

The authors declare no competing interests.

AUTHOR CONTRIBUTIONS

G.Y.Y. and J.H.W. supervised the project. Q.F., F.L., Q.S.S., J.H.W. and G.Y.Y. planned and coordinated the project. Z.C.Y., J.L.W., F.W., Q.F., F.W. and H.W.Y. prepared the samples for PacBio and Illumina sequencing. S.X. prepared the samples for Hi-C. X.H.Y. and Z.C.Y. performed the genome annotation. X.H.Y., Y.M. and Y.Y. conducted gene family annotation. X.H.Y. and Y.Y. performed chromosome, comparative genomics, phylogenetic, and gene family evolution analyses. J.H.W. and Z.C.Y. conducted the screen for bacterial scaffold contamination. X.H.Y., L.F.C. and Y.Y. performed venom gene analysis. X.H.Y., Y.Y., L.F.C. and S.J.X. prepared figures, tables and supporting information. X.H.Y. and Y.M. built the database for data release. X.H.Y., Z.C.Y., Y.Y., Q.F., Q.S.S., F.L., J.H.W. and G.Y.Y. wrote the manuscript. All authors read and approved the final manuscript.

DATA AVAILABILITY STATEMENT

All sequence data are available at the NCBI, Bioproject no. PRJNA542765. The genome assembly has been deposited at DDBJ/ENA/GenBank under accession no. VCDM00000000. The version described in this paper is version VCDM02000000. The raw genome sequencing data have been deposited in the NCBI Sequence Read Archive (SRA) database (SRR11444499–SRR11444504, SRR11580020). The *P. puparum* transcriptome data for genome annotation are available through NCBI SRA (SRR11306568–SRR11306573). In addition, genome assembly, RNA-seq data, and genome annotation information are available at database InsectBase (www.insect-genome.com/Pteromalus/). Custom scripts are available from https://github.com/yexinhai/Scripts4PteromalusGenome_Ye-et-al-2020-MER.

ORCID

Xinhai Ye https://orcid.org/0000-0002-0203-0663
Hongwei Yao https://orcid.org/0000-0002-6986-2181
Fei Li https://orcid.org/0000-0002-8410-5250
Gongyin Ye https://orcid.org/0000-0003-4937-8867

REFERENCES

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

Barron, M. C., Barlow, N. D., & Wratten, S. D. (2003). Non-target parasitism of the endemic New Zealand red admiral butterfly (*Bassaris gonerilla*) by the introduced biological control agent *Pteromalus puparum*. *Biological Control*, *27*, 329–335. https://doi.org/10.1016/S1049-9644(03)00028-8

Beckage, N. E., & Gelman, D. B. (2004). Wasp parasitoid disruption of host development: Implications for new biologically based strategies for insect control. *Annual Review of Entomology*, *49*, 299–330. https://doi.org/10.1146/annurev.ento.49.061802.123324

Beier, S., Thiel, T., Munch, T., Scholz, U., & Mascher, M. (2017). MISA-web: A web server for microsatellite prediction. *Bioinformatics*, *33*(16), 2583–2585. https://doi.org/10.1093/bioinformatics/btx198

Benetta, E. D., Antoshechkin, I., Yang, T., Nguyen, H. Q. M., Ferree, P. M., & Akbari, O. S. (2020). Genome elimination mediated by gene expression from a selfish chromosome. *Science Advances*, *6*(14), eaaz9808. https://doi.org/10.1126/sciadv.aaz9808

Benoit, J. B., Adelman, Z. N., Reinhardt, K., Dolan, A., Poelchau, M., Jennings, E. C., … Richards, S. (2016). Unique features of a global human ectoparasite identified through sequencing of the bed bug genome. *Nature Communications*, *7*, 10165. https://doi.org/10.1038/ncomms10165

Birney, E., & Durbin, R. (2000). Using GeneWise in the *Drosophila* annotation experiment. *Genome Research*, *10*(4), 547–548. https://doi.org/10.1101/gr.10.4.547

Blunck, H. (1957). *Pieris rapae* (L.), its parasites and predators in Canada and the United States. *Journal of Economic Entomology*, *50*, 835–836.

Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., & Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics*, *27*(4), 578–579. https://doi.org/10.1093/bioinformatics/btq683

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Branstetter, M. G., Childers, A. K., Cox-Foster, D., Hopper, K. R., Kapheim, K. M., Toth, A. L., & Worley, K. C. (2018). Genomes of the Hymenoptera. *Current Opinion Insect Science*, *25*, 65–75. https://doi.org/10.1016/j.cois.2017.11.008

Burton, J. N., Adey, A., Patwardhan, R. P., Qiu, R. L., Kitzman, J. O., & Shendure, J. (2013). Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nature Biotechnology*, *31*(12), 1119–1125. https://doi.org/10.1038/nbt.2727

Cai, J., Ye, G. Y., & Hu, C. (2004). Parasitism of *Pieris rapae* (Lepidoptera: Pieridae) by a pupal endoparasitoid, *Pteromalus puparum* (Hymenoptera: Pteromalidae): Effects of parasitization and venom on host hemocytes. *Journal of Insect Physiology*, *50*(4), 315–322. https://doi.org/10.1016/j.jinsphys.2004.01.007

Capella-Gutierrez, S., Silla-Martinez, J. M., & Gabaldon, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Cavigliasso, F., Mathé-Hubert, H., Kremmer, L., Rebuf, C., Gatti, J. L., Malausa, T., … Poirié, M. (2019). Rapid and differential evolution of

the venom composition of a parasitoid wasp depending on the host strain. *Toxins*, *11*(11), 629. https://doi.org/10.3390/toxins11110629

Chen, L., Lang, K., Bi, S., Luo, J., Liu, F., Ye, X., … Chen, X. (2018). WaspBase: A genomic resource for the interactions among parasitic wasps, insect hosts and plants. *Database (Oxford)*, *2018*, 1–9. https://doi.org/10.1093/database/bay081

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Cheng, T., Wu, J., Wu, Y., Chilukuri, R. V., Huang, L., Yamamoto, K., … Mita, K. (2017). Genomic adaptation to polyphagy and insecticides in a major East Asian noctuid pest. *Nature Ecology & Evolution*, *1*(11), 1747–1756. https://doi.org/10.1038/s41559-017-0314-4

Colinet, D., Anselme, C., Deleury, E., Mancini, D., Poulain, J., Azéma-Dossat, C., … Gatti, J.-L. (2014). Identification of the main venom protein components of *Aphidius ervi*, a parasitoid wasp of the aphid model *Acyrthosiphon pisum*. *BMC Genomics*, *15*, 342. https://doi.org/10.1186/1471-2164-15-342

Colinet, D., Deleury, E., Anselme, C., Cazes, D., Poulain, J., Azema-Dossat, C., … Poirié, M. (2013). Extensive inter- and intraspecific venom variation in closely related parasites targeting the same host: The case of *Leptopilina* parasitoids of *Drosophila*. *Insect Biochemistry and Molecular Biology*, *43*(7), 601–611. https://doi.org/10.1016/j.ibmb.2013.03.010

Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M., & Robles, M. (2005). BLAST2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, *21*(18), 3674–3676. https://doi.org/10.1093/bioinformatics/bti610

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

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

de Graaf, D. C., Aerts, M., Brunain, M., Desjardins, C. A., Jacobs, F. J., Werren, J. H., & Devreese, B. (2010). Insights into the venom composition of the ectoparasitoid wasp *Nasonia vitripennis* from bioinformatic and proteomic studies. *Insect Molecular Biology*, *19*, 11–26. https://doi.org/10.1111/j.1365-2583.2009.00914.x

Desjardins, C. A., Gadau, J., Lopez, J. A., Niehuis, O., Avery, A. R., Loehlin, D. W., … Werren, J. H. (2013). Fine-scale mapping of the *Nasonia* genome to chromosomes using a high-density genotyping microarray. *G3: Genes, Genomes, Genetics*, *3*(2), 205–215. https://doi.org/10.1534/g3.112.004739

Dozortseva, R. L. (1936). Chromosomal morphology of the parasitic wasp *Pteromalus puparum*. *Izvestiya Akademii Nauk SSSR Biology*, *6*, 1206–1221.

El-Fakharany, S. K. M., & Hendawy, A. S. (2014). Field studies on cabbage white butterfly, *Pieris rapae* (Linnaeus) and its associated parasitoid and predatory species in Egypt. *Egyptian Journal of Biological Pest Control*, *24*(2), 437–444.

English, A. C., Salerno, W. J., & Reid, J. G. (2014). PBHoney: Identifying genomic variants via long-read discordance and interrupted mapping. *BMC Bioinformatics*, *15*, 180. https://doi.org/10.1186/1471-2105-15-180

Fallon, T. R., Lower, S. E., Chang, C.-H., Bessho-Uehara, M., Martin, G. J., Bewick, A. J., … Weng, J.-K. (2018). Firefly genomes illuminate parallel origins of bioluminescence in beetles. *Elife*, *7*, e36495. https://doi.org/10.7554/eLife.36495

Fang, Q., Wang, B. B., Ye, X. H., Wang, F., & Ye, G. Y. (2016). Venom of parasitoid *Pteromalus puparum* impairs host humoral antimicrobial activity by decreasing host cecropin and lysozyme gene expression. *Toxins*, *8*(2), 52. https://doi.org/10.3390/toxins8020052

Fang, Q., Wang, F., Gatehouse, J. A., Gatehouse, A. M. R., Chen, X. X., Hu, C., & Ye, G. Y. (2011). Venom of parasitoid, *Pteromalus puparum*, suppresses host, *Pieris rapae*, immune promotion by decreasing host C-Type lectin gene expression. *PLoS One*, *6*(10), e26888. https://doi.org/10.1371/journal.pone.0026888

Fang, Q. I., Wang, L., Zhu, J., Li, Y., Song, Q., Stanley, D. W., … Ye, G. (2010). Expression of immune-response genes in lepidopteran host is suppressed by venom from an endoparasitoid, *Pteromalus puparum*. *BMC Genomics*, *11*, 484. https://doi.org/10.1186/1471-2164-11-484

Ferguson, K. B., Visser, S., Dalíková, M., Provazníková, I., Urbaneja, A., Pérez-Hedo, M., & Verhulst, E. C. (2020). Jekyll or Hyde? The genome (and more) of Nesidiocoris tenuis, azoophytophagous predatory bug that is both a biological control agent and a pest. *bioRxiv*. https://doi.org/10.1101/2020.02.27.967943

Feyereisen, R. (1999). Insect P450 enzymes. *Annual Review of Entomology*, *44*, 507–533. https://doi.org/10.1146/annurev.ento.44.1.507

Feyereisen, R. (2006). Evolution of insect P450. *Biochemical Society Transactions*, *34*(Pt 6), 1252–1255. https://doi.org/10.1042/BST0341252

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., … Bateman, A. (2016). The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Research*, *44*(D1), D279–285. https://doi.org/10.1093/nar/gkv1344

Francino, M. P. (2005). An adaptive radiation model for the origin of new gene functions. *Nature Genetics*, *37*(6), 573–577. https://doi.org/10.1038/ng1579

Gauthier, J., Boulain, H., van Vugt, J., Baudry, L., Persyn, E., Aury, J., & Drezen, J. (2020). Chromosomal resolution reveals symbiotic virus colonization of parasitic wasp genomes. *bioRxiv*. https://doi.org/10.1101/2020.03.19.994459

Godfray, H. C. J. (1994). *Parasitoids: Behavioral and evolutionary ecology*, Princeton, NJ: Princeton University Press.

Goecks, J., Mortimer, N. T., Mobley, J. A., Bowersock, G. J., Taylor, J., & Schlenke, T. A. (2013). Integrative approach reveals composition of endoparasitoid wasp venoms. *PLoS One*, *8*(5), e64125. https://doi.org/10.1371/journal.pone.0064125

Gokhman, V. E. (2009). *Karyotypes of parasitic hymenoptera*, Dordrecht: Springer. https://doi.org/10.1007/978-1-4020-9807-9

Gremme, G., Brendel, V., Sparks, M. E., & Kurtz, S. (2005). Engineering a software tool for gene structure prediction in higher organisms. *Information and Software Technology*, *47*(15), 965–978. https://doi.org/10.1016/j.infsof.2005.09.005

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., … Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, *8*(8), 1494–1512. https://doi.org/10.1038/nprot.2013.084

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., … 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

Hao, Z. D., Lv, D. K., Ge, Y., Shi, J. S., Yu, G. C., & Chen, J. H. (2020). RIDEOGRAM: Drawing SVG graphics to visualize and map genome-wide data in idiograms. *PeerJ Computer Science*, *6*:e251. http://doi.org/10.7717/peerj-cs.251

Heavner, M. E., Ramroop, J., Gueguen, G., Ramrattan, G., Dolios, G., Scarpati, M., … Govind, S. (2017). Novel organelles with elements of bacterial and eukaryotic secretion systems weaponize parasites of *Drosophila*. *Current Biology*, *27*(18), 2869–2877.e6. https://doi.org/10.1016/j.cub.2017.08.019

Heraty, J. (2009). Parasitoid diversity and insect pest management. In R. G. Foottit & P. H. Adler (Eds.), *Insect biodiversity: Science and society* (pp. 445–462). Hoboken, NJ: Wiley-Blackwell.

Hu, C. (1983). A survey of the parasites of the small white butterfly, *Pieris rapae* (L.) in China. *Acta Entomlogica Sinica*, *26*, 287–294.

Hu, C. (1984). Life history and occurrence of *Pteromalus puparum* L. in Hangzhou. *Acta Entomological Sinica*, *27*, 302–307.

Jeffery, C. J. (1999). Moonlighting proteins. *Trends in Biochemical Sciences*, *24*(1), 8–11. https://doi.org/10.1016/s0968-0004(98)01335-8

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., … Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *Journal of Molecular Biology*, *428*(4), 726–731. https://doi.org/10.1016/j.jmb.2015.11.006

Katoh, K., & Standley, D. M. (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

Kent, W. J. (2002). BLAT–the BLAST-like alignment tool. *Genome Research*, *12*(4), 656–664. https://doi.org/10.1101/gr.229202

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, *14*(4), R36. https://doi.org/10.1186/gb-2013-14-4-r36

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics*, *5*, 59. https://doi.org/10.1186/1471-2105-5-59

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Li, B., & Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*, 16. https://doi.org/10.1186/1471-2105-12-323

Li, H., Coghlan, A., Ruan, J., Coin, L. J., Heriche, J. K., Osmotherly, L., … Durbin, R. (2006). TREEFAM: A curated database of phylogenetic trees of animal gene families. *Nucleic Acids Research*, *34*, D572–D580. https://doi.org/10.1093/nar/gkj118

Li, L., Stoeckert, C. J. Jr., & Roos, D. S. (2003). OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Research*, *13*(9), 2178–2189. https://doi.org/10.1101/gr.1224503

Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., … Wang, J. (2010). De novo assembly of human genomes with massively parallel short read sequencing. *Genome Research*, *20*(2), 265–272. https://doi.org/10.1101/gr.097261.109

Lin, Z., Wang, R.-J., Cheng, Y., Du, J., Volovych, O., Han, L.-B., … Zou, Z. (2019). Insights into the venom protein components of *Microplitis mediator*, an endoparasitoid wasp. *Insect Biochemistry and Molecular Biology*, *105*, 33–42. https://doi.org/10.1016/j.ibmb.2018.12.013

Linnen, C. R., O'Quin, C. T., Shackleford, T., Sears, C. R., & Lindstedt, C. (2018). Genetic basis of body color and spotting pattern in red-headed pine sawfly larvae (*Neodiprion lecontei*). *Genetics*, *209*(1), 291–305. https://doi.org/10.1534/genetics.118.300793

Lomsadze, A., Burns, P. D., & Borodovsky, M. (2014). Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Research*, *42*(15), e119. https://doi.org/10.1093/nar/gku557

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., … Wang, J. (2012). SOAPdenovo2: An empirically improved memory-efficient short-read *de novo* assembler. *Gigascience*, *1*(1), 18. https://doi.org/10.1186/2047-217X-1-18

Ma, W. H., Zhao, X. X., Yin, C. L., Fan, J., Du, X. Y., Chen, T. Y., … Lin, Y. J. (2020). A chromosome-level genome assembly reveals the genetic basis of cold tolerance in a notorious rice insect pest *Chilo suppressalis*. *Molecular Ecology Resources*, *20*(1), 268–282. https://doi.org/10.1111/1755-0998.13078

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

Martinson, E. O., Mrinalini, Kelkar, Y. D., Chang, C.-H., & Werren, J. H. (2017). The evolution of venom by co-option of single-copy genes. *Current Biology*, *27*(13), 2007–2013.e8. https://doi.org/10.1016/j.cub.2017.05.032

Martinson, E. O., Wheeler, D., Wright, J., Mrinalini, Siebert, A. L., & Werren, J. H. (2014). *Nasonia vitripennis* venom causes targeted gene expression changes in its fly host. *Molecular Ecology*, *23*(23), 5918–5930. https://doi.org/10.1111/mec.12967

Mateo Leach, I., Hesseling, A., Huibers, W. H., Witsenboer, H., Beukeboom, L. W., & van de Zande, L. (2009). Transcriptome and proteome analysis of ovaries of arrhenotokous and thelytokous *Venturia canescens*. *Insect Molecular Biology*, *18*(4), 477–482. https://doi.org/10.1111/j.1365-2583.2009.00890.x

McKenzie, S. K., & Kronauer, D. J. C. (2018). The genomic architecture and molecular evolution of ant odorant receptors. *Genome Research*, *28*(11), 1757–1765. https://doi.org/10.1101/gr.237123.118

Meng, X., & Ji, Y. (2013). Modern computational techniques for the HMMER sequence analysis. *ISRN Bioinform*, *2013*, 252183. https://doi.org/10.1155/2013/252183

Moreau, S. J. M., & Asgari, S. (2015). Venom proteins from parasitoid wasps and their biological functions. *Toxins*, *7*(7), 2385–2412. https://doi.org/10.3390/toxins7072385

Mrinalini, Siebert, A. L., Wright, J., Martinson, E., Wheeler, D., & Werren, J. H. (2015). Parasitoid venom induces metabolic cascades in fly hosts. *Metabolomics*, *11*(2), 350–366. https://doi.org/10.1007/s11306-014-0697-z

Mrinalini, Werren, J.H. (2016). Parasitoid Wasps and Their Venoms. In P. Gopalakrishnakone & A. Malhotra (Eds.), *Evolution of Venomous Animals and Their Toxins* (pp. 1–26). Dordrecht, Netherlands: Springer.

Müller, P., Warr, E., Stevenson, B. J., Pignatelli, P. M., Morgan, J. C., Steven, A., … Donnelly, M. J. (2008). Field-caught permethrin-resistant *Anopheles gambiae* overexpress CYP6P3, a P450 that metabolises pyrethroids. *PLoS Genetics*, *4*(11), e1000286. https://doi.org/10.1371/journal.pgen.1000286

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Peck, O. (1963). A catalogue of the Nearctic Chalcidoidea (Insecta: Hymenoptera). *The Canadian Entomologists*, *Supplement*, *30*, 718–721. https://doi.org/10.4039/entm9530fv

Pennacchio, F., & Strand, M. R. (2006). Evolution of developmental strategies in parasitic hymenoptera. *Annual Review of Entomology*, *51*, 233–258. https://doi.org/10.1146/annurev.ento.51.110104.151029

Peters, R. S., Krogmann, L., Mayer, C., Donath, A., Gunkel, S., Meusemann, K., … Niehuis, O. (2017). Evolutionary history of the Hymenoptera. *Current Biology*, *27*(7), 1013–1018. https://doi.org/10.1016/j.cub.2017.01.027

Peters, R. S., Niehuis, O., Gunkel, S., Bläser, M., Mayer, C., Podsiadlowski, L., … Krogmann, L. (2018). Transcriptome sequence-based phylogeny of chalcidoid wasps (Hymenoptera: Chalcidoidea) reveals a history of rapid radiations, convergence, and evolutionary success. *Molecular Phylogenetics and Evolution*, *120*, 286–296. https://doi.org/10.1016/j.ympev.2017.12.005

Pichon, A., Bézier, A., Urbach, S., Aury, J.-M., Jouan, V., Ravallec, M., … Volkoff, A.-N. (2015). Recurrent DNA virus domestication leading to different parasite virulence strategies. *Science Advance*, *1*(10), e1501150. https://doi.org/10.1126/sciadv.1501150

Quicke, D. L. J. (1997). *Parasitic wasps*. London, UK: Chapman & Hall.

Rago, A., Gilbert, D. G., Choi, J.-H., Sackton, T. B., Wang, X. U., Kelkar, Y. D., … Colbourne, J. K. (2016). OGS2: Genome re-annotation of the jewel wasp *Nasonia vitripennis*. *BMC Genomics*, *17*, 678. https://doi.org/10.1186/s12864-016-2886-9

Sadd, B. M., Barribeau, S. M., Bloch, G., de Graaf, D. C., Dearden, P., Elsik, C. G., … Worley, K. C. (2015). The genomes of two key bumblebee species with primitive eusocial organization. *Genome Biology*, *16*, 76. https://doi.org/10.1186/s13059-015-0623-3

Schield, D. R., Card, D. C., Hales, N. R., Perry, B. W., Pasquesi, G. M., Blackmon, H., … Castoe, T. A. (2019). The origins and evolution of chromosomes, dosage compensation, and mechanisms underlying venom regulation in snakes. *Genome Research*, *29*(4), 590–601. https://doi.org/10.1101/gr.240952.118

Scott, J. G., & Wen, Z. (2001). Cytochromes P450 of insects: The tip of the iceberg. *Pest Management Science*, *57*(10), 958–967. https://doi.org/10.1002/ps.354

Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C.-J., Vert, J.-P., … Barillot, E. (2015). HiC-Pro: An optimized and flexible pipeline for Hi-C data processing. *Genome Biology*, *16*, 259. https://doi.org/10.1186/s13059-015-0831-x

Stanke, M., Steinkamp, R., Waack, S., & Morgenstern, B. (2004). AUGUSTUS: A web server for gene finding in eukaryotes. *Nucleic Acids Research*, *32*, W309–W312. https://doi.org/10.1093/nar/gkh379

Strand, M. R. (2014). Teratocytes and their functions in parasitoids. *Current Opinion in Insect Science*, *6*, 68–73. https://doi.org/10.1016/1049-9644(91)90110-L

Strand, M. R., & Burke, G. R. (2013). Polydnavirus-wasp associations: Evolution, genome organization, and function. *Current Opinion in Virology*, *3*(5), 587–594. https://doi.org/10.1016/j.coviro.2013.06.004

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*, *6*(7), e21800. https://doi.org/10.1371/journal.pone.0021800

Takagi, M. (1985). The reproductive strategy of the gregarioius parasitoid, *Pteromalus puparum* (Hymenoptera: Pteromalidae). 1. Optimal number of eggs in a single host. *Oecologia*, *68*, 1–6.

Tan, C. W., Peiffer, M., Hoover, K., Rosa, C., Acevedo, F. E., & Felton, G. W. (2018). Symbiotic polydnavirus of a parasite manipulates caterpillar and plant immunity. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(20), 5199–5204. https://doi.org/10.1073/pnas.1717934115

Tarailo-Graovac, M., & Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols Bioinformatics*, *5*(1), 4.10.1-4.10.14. https://doi.org/10.1002/0471250953.bi0410s25

Tellier, A., Moreno-Gamez, S., & Stephan, W. (2014). Speed of adaptation and genomic footprints of host-parasite coevolution under arms race and trench warfare dynamics. *Evolution*, *68*(8), 2211–2224. https://doi.org/10.1111/evo.12427

Teng, Z.-W., Xiong, S.-J., Xu, G., Gan, S.-Y., Chen, X., Stanley, D., … Fang, Q. I. (2017). Protein discovery: Combined transcriptomic and proteomic analyses of venom from the endoparasitoid *Cotesia chilonis* (Hymenoptera: Braconidae). *Toxins*, *9*, 135. https://doi.org/10.3390/toxins9040135

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., … Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, *7*(3), 562–578. https://doi.org/10.1038/nprot.2012.016

van Berkum, N. L., Lieberman-Aiden, E., Williams, L., Imakaev, M., Gnirke, A., Mirny, L. A., … Lander, E. S. (2010). Hi-C: A method to study the three-dimensional architecture of genomes. *Journal of Visualized Experiments*, *39*, e1869. https://doi.org/10.3791/1869

Vincent, B., Kaeslin, M., Roth, T., Heller, M., Poulain, J., Cousserans, F., … Moreau, S. J. M. (2010). The venom composition of the parasitic wasp *Chelonus inanitus* resolved by combined expressed sequence tags analysis and proteomic approach. *BMC Genomics*, *11*, 693. https://doi.org/10.1186/1471-2164-11-693

Vonk, F. J., Casewell, N. R., Henkel, C. V., Heimberg, A. M., Jansen, H. J., McCleary, R. J. R., … Richardson, M. K. (2013). The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(51), 20651–20656. https://doi.org/10.1073/pnas.1314702110

Wallberg, A., Bunikis, I., Pettersson, O. V., Mosbech, M.-B., Childers, A. K., Evans, J. D., … Webster, M. T. (2019). A hybrid *de novo* genome assembly of the honeybee, *Apis mellifera*, with chromosome-length scaffolds. *BMC Genomics*, *20*(1), 275. https://doi.org/10.1186/s12864-019-5642-0

Wan, F., Yin, C., Tang, R., Chen, M., Wu, Q., Huang, C., … Li, F. (2019). A chromosome-level genome assembly of *Cydia pomonella* provides insights into chemical ecology and insecticide resistance. *Nature Communications*, *10*(1), 4237. https://doi.org/10.1038/s41467-019-12175-9

Wang, B. J., Shahzad, M. F., Zhang, Z., Sun, H. N., Han, P., Li, F., & Han, Z. J. (2014). Genome-wide analysis reveals the expansion of Cytochrome P450 genes associated with xenobiotic metabolism in rice striped stem borer, Chilo suppressalis. *Biochemical and Biophysical Research Communications*, *443*(2), 756–760. https://doi.org/10.1016/j.bbrc.2013.12.045

Wang, B. B., Yang, Y., Liu, M. M., Yang, L., Stanley, D. W., Fang, Q., & Ye, G. Y. (2019). A digestive tract expressing 𝜶-amylase influences the adult lifespan of *Pteromalus puparum* revealed through RNAi and rescue analyses. *Pest Management Science*, *75*(12), 3346–3355. https://doi.org/10.1002/ps.5462

Wang, F., Fang, Q. I., Wang, B., Yan, Z., Hong, J., Bao, Y., … Ye, G. (2017). A novel negative-stranded RNA virus mediates sex ratio in its parasitoid host. *PLoS Path*, *13*(3), e1006201. https://doi.org/10.1371/journal.ppat.1006201

Wang, H., Shi, Y. U., Wang, L. U., Liu, S., Wu, S., Yang, Y., … Wu, Y. (2018). CYP6AE gene cluster knockout in Helicoverpa armigera reveals role in detoxification of phytochemicals and insecticides. *Nature Communications*, *9*(1), 4820. https://doi.org/10.1038/s41467-018-07226-6

Wang, L., Fang, Q., Qian, C., Wang, F., Yu, X. Q., & Ye, G. Y. (2013). Inhibition of host cell encapsulation through inhibiting immune gene expression by the parasitic wasp venom calreticulin. *Insect Biochemistry and Molecular Biology*, *43*(10), 936–946. https://doi.org/10.1016/j.ibmb.2013.07.010

Wang, Z.-Z., Ye, X.-Q., Shi, M., Li, F., Wang, Z.-H., Zhou, Y.-N., … Chen, X.-X. (2018). Parasitic insect-derived miRNAs modulate host development. *Nature Communications*, *9*(1), 2205. https://doi.org/10.1038/s41467-018-04504-1

Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., … Zdobnov, E. M. (2017). 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

Werren, J. H., Richards, S., Desjardins, C. A., Niehuis, O., Gadau, J., Colbourne, J. K., … Gibbs, R. A. (2010). Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. *Science*, *327*(5963), 343–348. https://doi.org/10.1126/science.1178028

Wheeler, D., Redding, A. J., & Werren, J. H. (2013). Characterization of an ancient lepidopteran lateral gene transfer. *PLoS One*, *8*(3), e59262. https://doi.org/10.1371/journal.pone.0059262

Wong, E. S., & Belov, K. (2012). Venom evolution through gene duplications. *Gene*, *496*, 1–7. https://doi.org/10.1016/j.gene.2012.01.009

Xiao, H. M., Ye, X. H., Xu, H. X., Mei, Y., Yang, Y., Chen, X., … Li, F. (2020). The genetic adaptations of fall armyworm Spodoptera frugiperda

facilitated its rapid global dispersal and invasion. *Molecular Ecology Resources*, *00*, 1–19. https://doi.org/10.1111/1755-0998.13182

Yan, Z. C., Fang, Q., Liu, Y., Xiao, S., Yang, L., Wang, F., … Ye, G. Y. (2017). A venom serpin splicing isoform of the endoparasitoid wasp Pteromalus puparum suppresses host prophenoloxidase cascade by forming complexes with host hemolymph proteinases. *The Journal of Biological Chemistry*, *292*(3), 1038–1051. https://doi.org/10.1074/jbc.M116.739565

Yan, Z., Fang, Q. I., Wang, L., Liu, J., Zhu, Y. U., Wang, F., … Ye, G. (2016). Insights into the venom composition and evolution of an endoparasitoid wasp by combining proteomic and transcriptomic analyses. *Scientific Reports*, *6*, 19604. https://doi.org/10.1038/srep19604

Yang, Z. H. (1997). PAML: A program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences*, *13*(5), 555–556. https://doi.org/10.1093/bioinformatics/13.5.555

Yin, C., Li, M., Hu, J., Lang, K., Chen, Q., Liu, J., … Chen, X. (2018). The genomic features of parasitism, polyembryony and immune evasion in the endoparasitic wasp *Macrocentrus cingulum*. *BMC Genomics*, *19*(1), 420. https://doi.org/10.1186/s12864-018-4783-x

Zhang, J. X., Lindsey, A. R. I., Peters, R. S., Heraty, J. M., Hopper, K. R., Werren, J. H., … Yoder, M. J. (2020). Conflicted signal in transcriptomic markers leads to a poorly resolved backbone phylogeny of chalcidoid wasps (Hymenoptera: Chalcidoidea). *Systematic Entomology*. https://doi.org/10.1111/syen.12427

SUPPORTING INFORMATION

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

How to cite this article: Ye X, Yan Z, Yang Y, et al. A chromosome-level genome assembly of the parasitoid wasp *Pteromalus puparum*. *Mol Ecol Resour*. 2020;20:1384–1402. https://doi.org/10.1111/1755-0998.13206