Received: 30 April 2021 | Revised: 4 September 2021 | Accepted: 9 September 2021
DOI: 10.1111/1755-0998.13503

RESOURCE ARTICLE

MOLECULAR ECOLOGY RESOURCES WILEY

The first chromosome-level genome assembly of a green lacewing *Chrysopa pallens* and its implication for biological control

Yuyu Wang[1] | Ruyue Zhang[1] | Mengqing Wang[2] | Lisheng Zhang[2] | Cheng-Min Shi[1] | Jing Li[1] | Fan Fan[1] | Shuo Geng[1] | Xingyue Liu[3] | Ding Yang[3]

[1]College of Plant Protection, Hebei Agricultural University, Baoding, China
[2]Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, China
[3]Department of Entomology, China Agricultural University, Beijing, China

Correspondence
Yuyu Wang, College of Plant Protection, Hebei Agricultural University, Baoding, China.
Email: wangyy_amy@126.com

and
Xingyue Liu and Ding Yang, Department of Entomology, China Agricultural University, Beijing, China.
Emails: xingyue_liu@yahoo.com (X.L.); dyangcau@126.com (D.Y.)

Funding information
National Natural Science Foundation of China, Grant/Award Number: 31900341, 32170451 and 32130012; Hebei Natural Science Foundation for Excellent Young Scholars, Grant/Award Number: C2019204149; Hebei Youth Talent Support Programme, Grant/Award Number: BJ2019039; Hebei Importing Overseas Students Support Programme, Grant/Award Number: C20190341; Starting Scientific Research Foundation for the Introduced Talents of Hebei Agricultural University, Grant/Award Number: YJ201817; Hebei Key Research and Development Program, Grant/Award Number: 19226508D; Earmarked Fund for China Agriculture Research System, Grant/Award Number: Cars-27; National Key Research and Minister of Science and Technology of China, Grant/Award Number: 2019YFD0300104

Abstract

Many lacewing species (Insecta: Neuroptera) are important predators of pests with great potential in biological control. So far, there is no chromosome-level published genome available for Neuroptera. Here we report a high-quality chromosome-level reference genome for a green lacewing species *Chrysopa pallens* (Neuroptera: Chrysopidae), which is one of the most important insect natural enemies used in pest biocontrol. The genome was sequenced using a combination of PacBio and Hi-C technologies and assembled into seven chromosomes with a total size of 517.21 Mb, occupying 96.07% of the genome sequence. A total of 12,840 protein-coding genes were identified and approximately 206.21 Mb of repeated sequences were annotated. Phylogenetic analyses indicated that *C. pallens* diverged from its common ancestor with *Tribolium castaneum* (Coleoptera) approximately 300 million years ago. The gene families involved in digestion, detoxification, chemoreception, carbohydrate metabolism, immunity, nerves and development were significantly expanded, revealing the potential genomic basis for the polyphagia of *C. pallens* and its role as an excellent biocontrol agent. This high-quality genome of *C. pallens* will provide an important genomic resource for future population genetics, evolutionary and phylogenetic investigations of Chrysopidae as well as comparative genomic studies of Neuropterida and other insects.

KEYWORDS
chromosome-level genome, *Chrysopa pallens*, Hi-C, PacBio sequencing

Yuyu Wang, Ruyue Zhang and Mengqing Wang contributed equally to this work.

1 | INTRODUCTION

With the development of new sequencing technologies, the genomic information of more and more species has been obtained and analysed. Deep sequencing data can provide extensive information about genomes and gene expression profiles. Since the completion of the Human Genome Project, obtaining high-quality reference genome sequence maps has become the basis for functional gene studies of various species (Venter et al., 2001). Especially, high-quality genomes of many insects have helped us to understand the molecular mechanisms of their ecological and evolutionary characteristics. For example, the analyses of the digestion and detoxification genes based on the genome of *Hypothenemus hampei* revealed the mechanism of food seeking and pathogenic bacteria resistance, which can help develop insecticides, biological control agents and traps (Vega et al., 2015). The relationships between the expansion of the gene families of *Spodoptera litura* (gustatory receptor, cytochrome P450, carboxy-lesterase as well as glutathione-S-transferase) and the resistance to various insecticides were analysed based on the high-quality genome data, providing new insights into the evolutionary mechanism, host plant selection and ecological adaptability of Lepidoptera (Cheng et al., 2017). The high-quality draft genome of *Propylea japonica* provides invaluable resource for understanding the molecular mechanisms of stress resistance in Coleoptera (Zhang et al., 2020). Two gene families associated with environmental adaptation (odorant receptor and cytochrome P450) were analysed, and a putative biosynthesis pathway of the defence alkaloid harmonine was successfully constructed based on the chromosome-level genome assembly of *Harmonia axyridis* (Chen et al., 2021).

The fauna of extant Neuropterida comprises nearly 6600 species of 20 families from three orders (Neuroptera, Megaloptera and Raphidioptera) (Yang et al., 2018), which is the key group connecting Holometabola and Hemimetabola. Many species of this group are important predators of pests with great potential in biological pest management such as green lacewings, brown lacewings, dusty lacewings and so on. *Chrysopa pallens* (Rambur, 1838) (Neuropterida: Neuroptera: Chrysopidae) is one of the most important predatory species widely used in agricultural and forestry ecosystems (Figure 1). It is the dominant species widely distributed in China except for the Tibet Plateau (Yang et al., 2018) and also distributed in Japan, Korea and Europe (Yang et al., 2005). Both larvae and adults of this species are predacious, feeding on aphids, coccids, thrips, planthoppers, whiteflies and some lepidopteran insects, as well as mites in agroforestry ecosystems (Bai et al., 2005; Winterton & de Freitas, 2006). In addition, laboratory experiments showed that *C. pallens* demonstrated an effective control capacity on the eggs and larvae of *S. frugiperda* (Xu et al., 2019). The high pest controlling potentials of *C. pallens* were underlain by its wide diet, high predation, strong adaptability and high egg production (about 2000 eggs per female) (Miller et al., 2004; Tauber et al., 2000). However, the molecular mechanisms of why *C. pallens* serve as such an excellent biocontrol agent remain unclear.

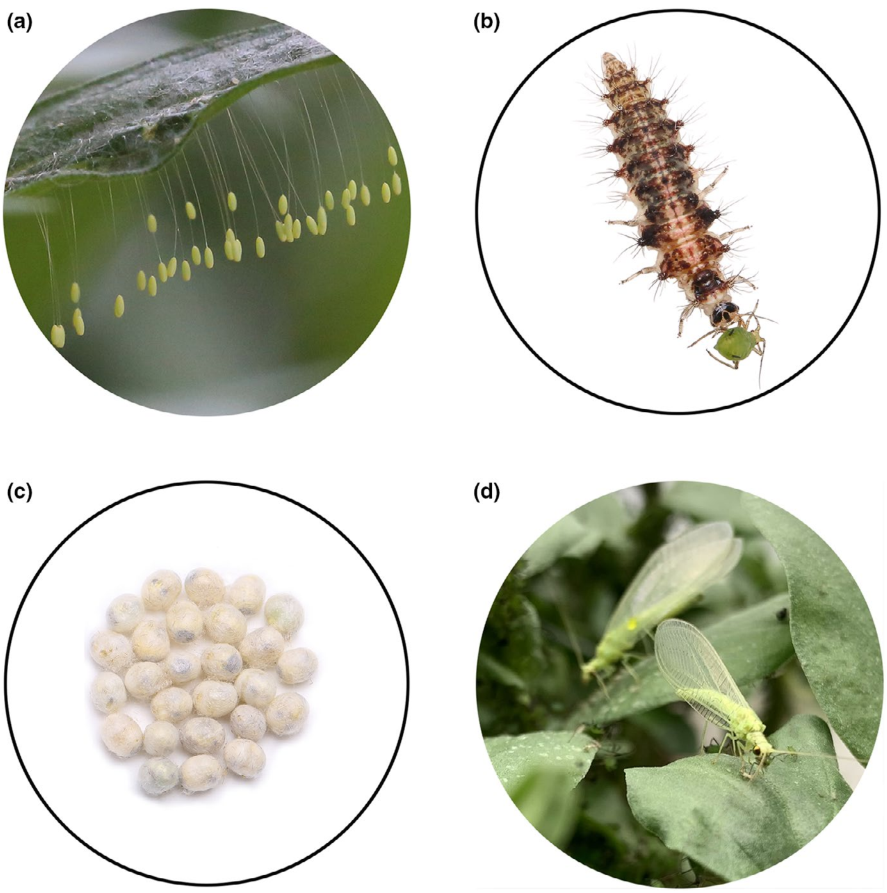

FIGURE 1 Ecological photos of *Chrysopa pallens*. (a) Egg, (b) larva, (c) pupa, (d) adult

Scholars from all over the world had carried out researches on the olfactory tropism, artificial feeding (Ali et al., 2018), mechanisms of resistance to chemical pesticides (Liu & Zeng, 2014) and pest-control abilities of *C. pallens* (Xu et al., 2019), laying foundations for further exploring their potential application in biological control. *C. pallens* had evolved a sophisticated olfactory system to accurately sense a large amount of chemical information in order to seek prey, locate host plants, find mates and spot dangers (Li et al., 2019). Previous studies have shown that using chemicals to guide *C. pallens* predation on pests could effectively improve their abilities to search for prey (Cui et al., 2015; Koczor et al., 2010). It is necessary to study the potential mechanism of their olfactory system for effective biological control (Hesler, 2016). A normalized transcriptome of *C. pallens* was sequenced by Li et al. (2013), and a large number of candidate chemosensory genes were identified. A full-length cDNA library from the antenna of *C. pallens* was constructed by Wang et al. (2014), and the functions of OBPs (CpalOBPs), the clone of olfaction-related genes of antennae as well as the mechanism of information recognition of *C. pallens* were studied. All the results mentioned above provided a basis for exploring the mechanisms of chemoreception in *C. pallens*. However, the olfactory mechanism of *C. pallens* at the genetic level is still not clarified. The absence of genomic information about *C. pallens* has limited further study into the molecular mechanisms of its ecological and evolutionary characteristics.

To date, there are more than 600 insect genomes sequenced and available at GenBank (accession date, 15 April 2021). However, no chromosome-level genome has been published for any species of Neuroptera so far except one of *Chrysoperla carnea* (GCA_905475395.1) uploaded to NCBI during the preparation of this manuscript, which remain unpublished. Therefore, we generated a high-quality genome assembly of a green lacewing species *C. pallens*, which would provide invaluable information for further study of *C. pallens* and promote large-scale breeding and commercial utilization of *C. pallens*. This study will also facilitate the development of modern pest control strategies. Besides, it will also serve as a reference for further comparative genomic studies of Neuropterida and even Insecta.

2 | MATERIALS AND METHODS

2.1 | Sampling and genome sequencing

The specimens used in the present study were reared at Langfang station of the Chinese Academy of Agricultural Sciences and inbred under laboratory conditions. All samples were selected from newly-emerged unfed adults to avoid contamination. There were 50 female and 10 male adults used for genome sequencing. The whole bodies of samples were used for the construction of each library. High-quality genomic DNA was extracted with the Qiagen DNeasy Blood & Tissue kit. DNA was quantified by 0.75% agarose gel electrophoresis, Nanodrop spectrophotometry (Thermo Fisher) and Qubit 3.0 fluorometry (Invitrogen). RNA was extracted with the TRIzol Reagent kit.

A Single Molecule Real-Time DNA library was prepared for sequencing using SMRTbell Template Prep Kit 2.0 with an insert size of 40 kb. The genome of *C. pallens* was sequenced with PacBio Sequel II System using 25 male adults. Illumina paired-end library (with 350 bp insert size) was prepared with Truseq DNA PCR-free kit using five male adults. Illumina paired-end reads were used for error checking and as references to Hi-C. The RNA library was constructed with Illumina TruSeq RNA v2 Kit according to the manufacturer's instructions using 10 female and 10 male adults. Hi-C library was constructed using 10 male adults. The construction of a Hi-C library included cross-linking of formaldehyde, restriction enzyme digestion, ends repair of fragments, DNA cyclization, DNA purification and other steps with MboI as the restriction enzyme.

2.2 | Genome size estimation and assembly

Quality control of the second-generation Illumina data was carried out by BBTools v38.67 (qtrim=rl trimq =20 minlen =15 ecco =t maxns =5 trimpolya =10 trimpolyg =10 trimpolyc =10) (Khan et al., 2012). Genome survey was conducted based on K-mer distribution analyses using GenomeScope v2.0 (Ranallo-Benavidez et al., 2020) to predict the genome size, heterozygosity and proportion of repeat content to select appropriate assembly tools and adjust corresponding parameter (-k 21 -p 2 -m 10,000) (Kirkness et al., 2003). The k-mer frequency was evaluated using BBTools v38.67 (Khan et al., 2012) and the length was set to 21 k-mer.

For genomic contig assembly, the long PacBio reads were corrected by NextDenovo v2.3.0 (seed_cutoff =20,000) (https://github.com/Nextomics/NextDenovo). The corrected PacBio sequences were assembled using Raven v1.2.2 (--weaken -p 0) (Vaser & Šikić, 2021). After the preliminary assembly of the genome, NextPolish v1.3.0 (Hu et al., 2019) was used to perform one round of long-read polishing and two rounds of short-read polishing to get the corrected genome sequence, which further improves the assembly accuracy. In this process, Minimap2 v2.17 (Li, 2018) with default parameters were used to align the second-generation/third-generation data to the third-generation assembled genome. SAMTools v1.9 (samtools view) (Li et al., 2009) was used for format conversion (SAM to BAM).

3D-DNA v180922 (Dudchenko et al., 2017; Nene et al., 2007) process was used to conduct chromosome anchoring based on the Hi-C sequences. The adapter sequences of raw reads were trimmed and low-quality PE reads were removed using Juicer v1.6.2 (Durand et al., 2016) to obtain clean data. Then, the clean reads were mapped to the draft genome using 3D-DNA v180922 (Dudchenko et al., 2017; Nene et al., 2007). Contigs in the same chromosome were separated by 100 Ns. A heat map was constructed based on the interaction signals revealed by the effective mapping read pairs. The sequencing depth of each locus and the average depth of chromosomes were calculated using SAMTools (Li et al., 2009). Illumina paired-end reads of male adults were sequenced separately and then were aligned to the *C. pallens* assembled genome to identify the sex chromosomes (X and Y) in *C. pallens*. The average sequencing depth and length were based to identify chromosome X and Y.

Potential contamination sequences were searched and removed from the nucleotide sequence database (NT) (Stoesser et al., 2001) and UniVec (https://www.ncbi.nlm.nih.gov/tools/vecscreen/univec/) using BLAST+ (i.e., Blastn) v2.9.1 (Camacho et al., 2009) and uploaded to GenBank for reconfirmation. The completeness of the final draft genome was assessed using BUSCO v3.0.1 (Simão et al., 2015) with insecta_odb10 reference, predicting the integrity of the genome with the percentage of single-copy orthologues founded in the insect orthologues library.

2.3 | Genome annotation

The repeat sequences were identified by both homology-based and de novo prediction methods. Firstly, the LTR search (-LTRStruct) was processed under RepeatModeler v2.0.1 (https://github.com/Dfam-consortium/RepeatModeler) and the de novo repeat library was built based on the specific structure of repeat sequence and the principle of de novo repeat library. A custom library was then formed by combining Dfam_3.3 (http://www.dfam.org/) (Storer et al., 2021) and RepBase-20181026 databases (https://www.girinst.org/repbase/) under the default parameters (Jurka et al., 2005). Finally, the repeat sequences were searched by RepeatMasker v4.0.9 with the default commands (http://repeatmasker.org/cgi-bin/WEBRepeatMasker) according to the custom library.

The protein coding genes were annotated by integrating the evidences of ab initio, transcriptome-based prediction and homology-based annotations. Firstly, the protein coding gene structures were predicted using MAKER v3.01.03 with the default commands (Cantarel et al., 2008). Different pieces of evidences were weighted using EVidenceModeler (EVM) (Haas et al., 2008). Augustus v3.3.4 (http://augustus.gobics.de/) (Stanke et al., 2006) and GeneMark-ES/ET/EP 4.59_lic (Lomsadze et al., 2014) were automatically trained by BRAKER (Tomas et al., 2021) to improve the prediction accuracy combined with transcriptome and protein homology. Then, the transcriptome information was used to align to the genome by HISAT2 v2.2.0 (Kim et al., 2015) to generate BAM files. The arthropod protein sequences were extracted from the OrthoDB10 v1 data (Kriventseva et al., 2018). Transcriptome assembly with reference genome was performed using StringTie v2.1.4 (Kovaka et al., 2019). For the homology-based approach, proteins of related species were downloaded from GenBank, such as *Drosophila melanogaster* (Diptera), *Tribolium castaneum* (Coleoptera), *Apis mellifera* (Hymenoptera), *Bombyx mori* (Lepidoptera) and *Rhopalosiphum maidis* (Hemiptera) (Table 1). Finally, the preparation files obtained from the above three approaches were imported into MAKER v3.01.03 (Cantarel et al., 2008) for integrated annotations. EVM weights for the three types of evidences were set as: transcriptome 8, protein homologue 2, and ab initio 1.

Two strategies were used for the annotation of gene functions. In the first strategy, the gene functions were predicted by aligning with existing databases, that is, UniprotKB (Swissprot +Trembl) (https://www.uniprot.org/) (Morgat et al., 2019) and the nonredundant protein sequence database (NR) using the sensitive mode of Diamond v0.9.24 (--more -sensitive -e 1e-5) (Buchfink et al., 2015). In the second strategy, the gene functions were predicted by aligning with a comprehensive database, that is, gene ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome. Five databases including protein families (Pfam) (El-Gebali et al., 2019), Smart (http://smart.embl-heidelberg.de/) (Letunic et al., 2021), Gene3D v21.0 (http://gene3d.biochem.ucl.ac.uk/) (Lewis et al., 2018), Superfamily (Wilson et al., 2009) and CDD (Marchler-Bauer et al., 2017) were searched by InterProScan 5.41–78.0 (Quevillon et al., 2005). At the same time, Eggnog v5.0 (http://eggnog.embl.de) database was searched with EggNOG-mapper v2.0.1 (Huerta-Cepas et al., 2017). The results of these two strategies were combined to get the final prediction of gene functions.

Non-coding RNAs including transfer RNAs (tRNAs), microRNAs (miRNAs), ribosome RNAs (rRNAs) and small nuclear RNAs (snRNAs)

TABLE 1 Information of species used in this study

Order	Species	NCBI accession	References
Ephemeropterp	*Cloeon dipterum*	GCA_902829235.1	Almudi et al. (2020)
Ispptera	*Coptotermes formosanus*	GCA_013340265.1	Itakura et al. (2020)
Hemiptera	*Rhopalosiphum maidis*	GCA_003676215.3	Chen et al. (2019)
Thysanoptera	*Thrips palmi*	GCA_012932325.1	Guo et al. (2020)
Coleptera	*Tribolium castaneum*	GCA_000002335.3	Richards et al. (2008)
Diptera	*Drosophila melanogaster*	GCA_000001215.4	Adams et al. (2000)
Hymenoptera	*Apis mellifera*	GCA_003254395.2	Solignac et al. (2007)
Trichoptera	*Stenopsyche tienmushanensis*	GCA_008973525.1	Luo et al. (2018)
Lepidoptera	*Bombyx mori*	GCA_014905235.2	International Silkworm Genome Consortium (2008)
Siphonaptera	*Ctenocephalides felis*	GCA_003426905.1	Driscoll et al. (2020)
Neuroptera	*Chrysopa pallens*	this study	this study

were also identified. The rRNAs, snRNAs and miRNAs were detected from the RFAM database (release 13.0) (Kalvari et al., 2018) using INFERNAL v1.1.3 (Nawrocki & Eddy, 2013). The tRNAs were predicted using TRNASCAN-SE v2.0 with "EukHighConfidenceFilter" (Lowe & Eddy, 1997). The rRNAs and subunits were predicted using RNAMMER v1.2 (Lagesen et al., 2007).

2.4 | Orthology, synteny and phylogenetic reconstruction

Orthologues from 11 insect species were clustered using the ORTHOFINDER v2.5.2 (Emms & Kelly, 2015) pipelines with default parameters (Table 1). According to the results of gene family clustering obtained from ORTHOFINDER v2.5.2 (Emms & Kelly, 2015), single-copy genes were used for phylogenetic reconstruction. Firstly, MAFFT v7.394 (Katoh & Standley, 2013) was used to align the homologous region of the sequences with the L-INS-I strategy. The unreliable homologous alignments were filtered using BMGE v1.12 (-m BLOSUM90 -h 0.4) (Criscuolo & Gribaldo, 2010). Alignment sequences were then combined into a supermatrix using FASCONCAT-G v1.04 (Kück & Longo, 2014). Phylogenetic reconstruction was performed with IQ-TREE v2.0-rc1 (-m MFP --mset LG --msub nuclear --rclusterf 10) (Nguyen et al., 2015) based on the supermatrix and support values for nodes were assessed by Ultrafast Boostrap (Minh et al., 2013) and SH-ALRT algorithm (-B 1000 --alrt 1000) (Guindon et al., 2010). Calibration points applied to estimate the divergence time of *C. pallens* were set based on the fossil records in the PBDB database (https://paleobiodb.org/navigator/) as well as from Misof et al. (2014) and Nel et al. (2013) (Table 2).

Tribolium castaneum was used as the reference to obtain relevant information to study the variation of interspecific chromosomes due to the lack of published high quality genomes of Neuropterida. There are only one chromosome-level genome of *Chrysoperla carnea* (GCA_905475395.1) uploaded to NCBI during the preparation of this manuscript, which remain unpublished and another scaffold-level genome of *Neoneuromus ignobilis* (GCA_014529405.1). The protein sequences of *C. pallens* and *T. castaneum* were compared with "blastp-like" using MMSEQ2 v11-E1A1C with default parameters (-s 7.5 --alignment-mode 3 --num-iterations 4 -e 1e-5 --max-accept 5) (Steinegger & Söding, 2017). In addition, the all.blast result file and the integrated annotation file (all.gff file) were used as input files to obtain the results of collinearity analyses by MCScanX (Wang et al., 2012). Finally, the diagram was drawn by TBTOOLS v1.0692 (Chen et al., 2020).

2.5 | Gene family expansion and contraction analyses

Expansions and contractions of orthologous gene families were determined using COMPUTATIONAL ANALYSIS OF FAMILY EVOLUTION (CAFÉ) v4.2.1 (Han et al., 2013) with default parameters ($p = .01$) according to the gene families and phylogenetic tree. The random genetic birth and death rate was simulated to predict the evolution of gene families in different species at various branches of evolution to assess the expansion and contraction of gene families at each node on the phylogenetic tree. The significantly expanded gene families were classified through R package CLUSTERPROFILER v3.14.3 (Yu, Wang, et al., 2012) according to the gene annotation database. Then, the redundancy was removed to get final results of the gene enrichment.

The selection of rapidly expanded families were analysed using PAML 4 (Yang, 2007) based on the CDS with site models to further understand the evolution of the expansion of gene families. The phylogenetic tree used was reconstructed with IQ-TREE v2.0-rc1 (Nguyen et al., 2015) with ambiguity data filtered before the formal calculation. The specific model used were M0 (one rate), M1a (neutral) – M2a (selection), M7 (beta) – M8 (beta & w) (i.e., "NSsites=0 1 2 7 8"). The results of the M1a–M2a and M7–M8 pairs of models were compared using the likelihood ratio test (LRT) ($p = .05$). The gene family was identified as being positively selected only if M2a and M8 confirmed that the family had positive selection sites at the same time.

3 | RESULTS AND DISCUSSION

3.1 | Genome sequencing and assembly

About 78.59 Gb clean reads were generated by Illumina paired-ended short-read sequencing and 113.72 Gb filtered subreads were obtained by PacBio long-read sequencing (Table 3). The average length and N50 of PacBio long reads were 32.86 kb and 18.31 kb, respectively. Hi-C fragment libraries sequencing resulted in about 75.48 Gb (Table 3).

The genome of *C. pallens* was predicted to be about 526.01–529.11 Mb indicated by the distribution frequency of 21-mers (Table 4). The kmer analysis shows that the genome of *C. pallens* has high heterozygosity ranged from (1.48%–1.54%) (Table 4). The genome size was reduced from 555.40 to 539.70 Mb after the removal of heterozygous sequences (Table S1). These characteristics implied that the *C. pallens* genome harboured a high degree of complexity. The mitochondrial genome of *C. pallens* was assembled and removed by aligning with the genome. An assembled genome of 538.35 Mb

TABLE 2 Information of calibration times used in this study

Placement	Minimum age (Ma)	Maximum age (Ma)
Diptera	242	252
Pterygota	–	443
Paraneoptera	315	323
Holometabola	315	383
Coleoptera +Neuroptera	307	323
Trichoptera +Lepidoptera+ Diptera +Siphonaptera	311	323

TABLE 3 Sequencing data used for the *Chrysopa pallens* genome assembly

The type of libraries	Number of reads	Raw data (bp)	Read length (bp)	N50 contig length (bp)
Survey	523,921,116	78,588,167,400	150	–
RNA	112,484,672	16,872,700,800	150	–
PacBio	6,210,590	113,715,446,179	18,309.93	32,855
Hi-C	251,600,926	75,480,277,800	150	–

TABLE 4 Statistical results from the genome survey

Property	Minimum	Maximum
Homozygous (aa)	98.46%	98.52%
Heterozygous (ab)	1.48%	1.54%
Genome haploid length (bp)	526,010,433	529,108,099
Genome repeat length (bp)	87,197,893	87,711,400
Genome unique length (bp)	438,812,540	441,396,699
Model fit	79.74%	92.05%
Read error rate	0.59%	0.59%

TABLE 5 Genome assembly and annotation statistics of *Chrysopa pallens*

Statistic	Value
Genome assembly	
Assembly size (Mb)	538.35
Number of scaffolds/contigs	552/1639
Longest scaffold/contig (Mb)	141.00/12.21
N50 scaffold/contig length (Mb)	89.78/1.32
GC (%)	27.08
Gaps (%)	0.02
BUSCO completeness (%)	98.10
Gene annotation	
Protein-coding genes	12,840
Mean protein length (aa)	540.79
Mean gene length (bp)	7555.93
Exons/introns per gene	6.03/4.75
Exon (%)	5.89
Mean exon length	409.18
Intron (%)	12.13
Mean intron length	1070.23
BUSCO completeness (%)	97.50

slightly larger than predicted was finally obtained including 1639 contigs with contig N50 of 1.32 Mb long and the longest contig of 12.21 Mb. The average GC content was 27.08% (Table 5). The assembly was further improved with Hi-C data, resulting in 552 scaffolds with a scaffold N50 of 89.78 Mb long and the longest scaffold of 141.00 Mb (Table 5). The results of each assembly step are shown in Table S1. Our genome assembly is highly complete with 98.10% BUSCO completeness, whereas only 0.30% and 1.60% BUSCO genes are fragmented and missing, respectively. This assembly includes seven pseudochromosomes which is consistent with the chromosome number identified by karyotype analyses (Figure 2a). In total, the length of pseudochromosomes was 517.21 Mb and contained 96.07% of the genome sequence. The pseudochromosomes ranged from 10.14 Mb to 141.00 Mb in length. The average sequencing depth of X chromosome (73.16x) was significantly lower than other chromosomes (deeper than 100x) and the length of Y chromosome was extremely short (10.14 Mb) (Table 6). The GC content of Y chromosome was the lowest (Figure 2b). The final genome sequences were submitted to GenBank under accession number JAEMTY000000000.

3.2 | Gene annotation

A total of 206.21 Mb of repeat sequences were identified constituting 38.31% of the genome of *C. pallens*. The three most abundant classes of repeat sequences included unclassified (24.65%), DNA elements (3.82%) and simple repeats (2.13%), respectively (Table S2). The high proportion of unclassified elements may be due to the lack of studies on the repeat sequences of Neuropterida.

In total, 12,840 protein-coding genes were predicted in the genome of *C. pallens* (Table 5). The average lengths of gene, CDS, exon, and intron regions were 7555.93 bp, 1625.29 bp, 409.18 bp, and 1070.23 bp, respectively. The average exon number per gene was 6.03. The completeness of protein-coding gene predictions was further evaluated by BUSCO (Simão et al., 2015) (insecta_odb10) to be 97.50%, including 93.90% single-copy and 3.60% duplicated BUSCO genes (Table 5). 11,583 annotated genes (90.21%) matched one or more records in the UniProtKB database, while 10,879 genes (84.73%) encode proteins with at least one known domain in the InterPro database (Quevillon et al., 2005). Additional functional annotations identified 9503 GO terms, 7569 KO terms, 2662 enzyme codes, 8503 KEGG and 9580 reactome pathways, and 10,759 COG categories.

There were 2361 noncoding RNA sequences annotated totally, including 1271 tRNAs, 607 ribosomal RNAs (rRNAs), 182 small nucleolar RNAs (snRNAs), 52 microRNAs (miRNAs), 15 small RNA (sRNAs) and others (Table S3). SnRNAs included 142 spliceosomal RNAs, four minor spliceosomal RNAs, one (small Cajal body-specific RNA) scaRNA (SCARNA8) and others (Table S3). The numbers of rRNAs and tRNAs are significantly higher than that of other insects, which might indicate vigorous protein synthesis of *C. pallens*.

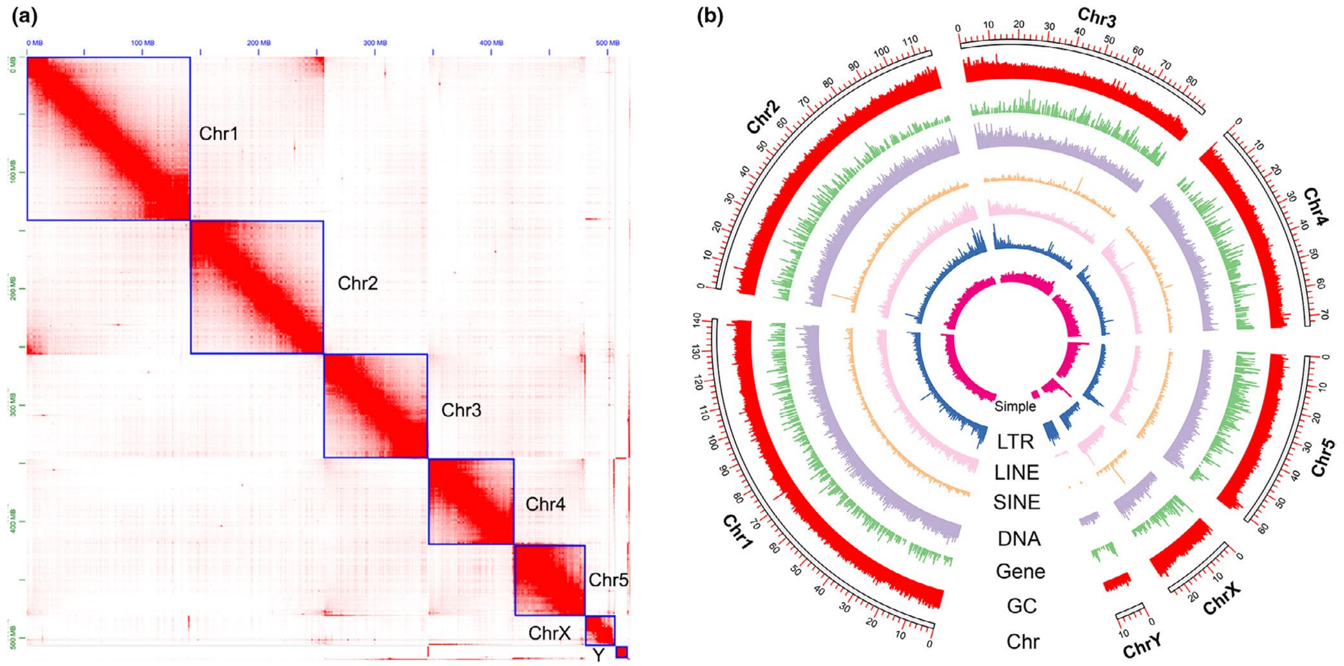

FIGURE 2 Heatmap of genome-wide Hi-C data and overview of the genomic landscape of *Chrysopa pallens*. (a) The heatmap shows all-by-all interactions among seven chromosomes of *C. pallens*. Resolution: 500 kb. There were strong intrachromosomal interactions (blocks on the diagonal line), while interchromosomal interactions were weaker. The frequencies of Hi-C interaction links are represented by the colour, which ranges from white (low) to red (high). (b) Blocks on the outmost circle represent all seven chromosomes of *C. pallens*. Peak plots from outer to inner circles represent the length of each chromosome, the GC content of each chromosome, protein coding genes, the density of repeat sequences (SINE, short interspersed elements; LINE, long interspersed elements; LTR, long terminal repeat elements; simple repeats), respectively

TABLE 6 Chromosome length and average sequencing depth

Chromosome	Length (bp)	Average sequencing depth (x)
Chr1	140,998,244	105.33
Chr2	115,703,577	106.76
Chr3	89,781,665	107.80
Chr4	73,683,371	108.56
Chr5	61,463,935	107.43
ChrX	25,444,600	73.16
ChrY	10,136,392	113.63

3.3 | Orthology, synteny and phylogenetic relationships

Comparative genomic analyses were performed between *C. pallens* and 10 other insects representing major insect orders (i.e., Ephemeroptera, Isoptera, Hemiptera, Thysanoptera, Hymenoptera, Siphonaptera, Diptera, Trichoptera, Lepidoptera, Coleoptera and Neuroptera). In total, 138,491 genes were clustered into 13,857 gene families using OrthoFinder (Table S4). Among them, 967 were single-copy and 3053 were multicopy in all 11 insect genomes (Figure 3a, Table S4). In the genome of *C. pallens*, 11,444 of the 12,840 genes were assigned to 8357 families that are also present in other insects, whereas the remaining 936 genes (223 gene families) were specific for this species (Table S5).

Among the 967 single-copy genes, 879 genes (321,918 amino acid sites) passed the symmetry test in IQ-Tree (Nguyen et al., 2015) and were used for phylogenetic reconstruction. All nodes in the estimated ML tree were fully supported (100/100 for UFB/SH-aLRT). The phylogenetic tree supported Neuropterida being the sister group to Coleoptera which was consistent with former studies based on different molecular markers, such as *18S rRNA* (Kjer, 2004), Vitellogenin sequences (Liu et al., 2015), transcriptome (Misof et al., 2014; Wang et al., 2019) and mitochondrial genomes (Song et al., 2016) (Figure 3a). Neuropterida and Coleoptera separated from each other about 300 million years ago (Figure 3a).

Comparison of the chromosome-level genome assemblies of *C. pallens* and *T. cataneum* revealed poor collinearity in general, which may be due to the relatively long evolutionary distance between the two species. Only X chromosome had an obvious large syntenic block (Figure 3b). There is one chromosome-level genome of *Ch. carnea* (GCA_905475395.1) uploaded to NCBI during the preparation of this manuscript, which remain unpublished. Further analyses will be performed in the future.

3.4 | Gene family expansion and contraction

The analysis of gene family evolution showed that 912 gene families were expanded and 3704 gene families were contracted in the genome of *C. pallens* compared to that of *T. cataneum* (Coleoptera)

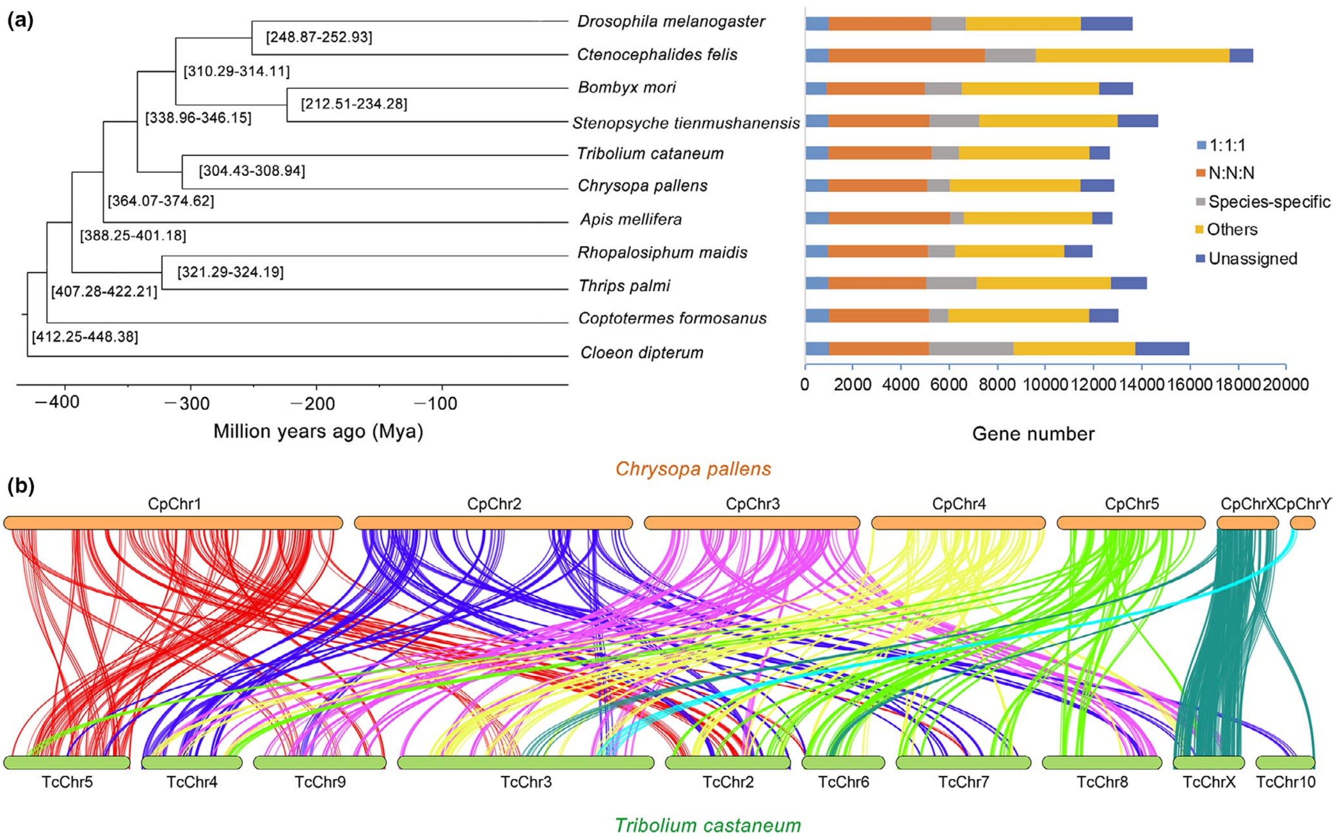

FIGURE 3 Phylogenetic tree, gene orthology, and synteny blocks. The phylogenetic tree was constructed based on 879 single-copy gene families with 11 insects using maximum likelihood methods by IQ-TREE (shown in Table 1). The length of branch indicated the divergence time. Bootstrap values (UFB/SH-aLRT) were 100 at all nodes. Bars are subdivided to represent different types of orthologues with different colours. 1:1:1 (single copy orthologous genes in common gene families); N:N:N (mutiple copy orthologous genes in common gene families); Specific (genes from unique gene family from each species); Other (genes that do not belong to any above-mentioned orthologue categories); Unassigned (orthologues which cannot be assigned into any orthogroups) (b) Synteny blocks between *Chrysopa pallens* and *Tribolium cataneum*

(Figure 4). Among them, 49 orthologous groups were expanded significantly ($p < .05$), including genes encoding C2H2-type zinc finger proteins, trypsin, zinc finger Y-chromosomal protein, acyltransferase, facilitated trehalose transporter, histone-lysine N-methyltransferase PRDM9-like protein, longitudinals lacking protein, ionotropic receptor, cytochrome P450, leucine rich repeat and so on (Figure S1). On the other hand, only one orthologous group was contracted significantly ($p < .05$).

GO analyses showed that the expanded gene families were enriched for functions related to digestion, immunity and nerves (Figure 5a), which might explain why *C. pallens* could serve as an excellent biocontrol agent. For instance, the successful capture of prey is attributed to the ability of *C. pallens* to effectively perceive information in the environment, in which the olfactory system on antennae plays a key role (Li et al., 2019). Li et al. (2015) cloned five OBP genes and analysed the tissue expression profiles of CpalOBPs in five strains of *C. pallens* using quantitative RT-PCR and demonstrated that *CpalOBP10* and the *OBP7* of aphids have close evolutionary relationships and can both mediate the perception of green leaf volatiles and (E)-β-farnesene later (Li et al., 2017). However, the underlying molecular mechanisms were not clarified. Further researches on olfactory related genes in *C. pallens* based on the high-quality genome will help to clarify the olfactory recognition mechanism of insects at the genetic level and to control the behavior of natural enemies.

The reproduction of *C. pallens* was proven to be closely related to the weight of females and the activities of trypsin-like as well as chymotrypsine-like enzymes (Liu, Mao, & Zeng, et al., 2015). Han et al. (2017) analysed the transcriptome of female adults of *C. pallens* under starvation as well as feeding conditions and found that different expressed genes (DEGs) were mainly involved in ribosome biogenesis, protein processing in endoplasmic reticulum, biosynthesis of amino acids, and carbon metabolism. They also annotated four vitellogenins, three insulin-like peptides and two insulin receptors genes. However, these earlier studies of *C. pallens* genes mainly relied on the transcriptome data which carry incomplete information. The chromosome-level genome of *C. pallens* will provide more information, which will contribute to the further study of the molecular mechanisms of its vitellogenin protein, nutrient metabolism and reproduction.

Interestingly, the gene families involved in cold acclimation and wax synthesis were also expanded in the genome of *C. pallens*, which

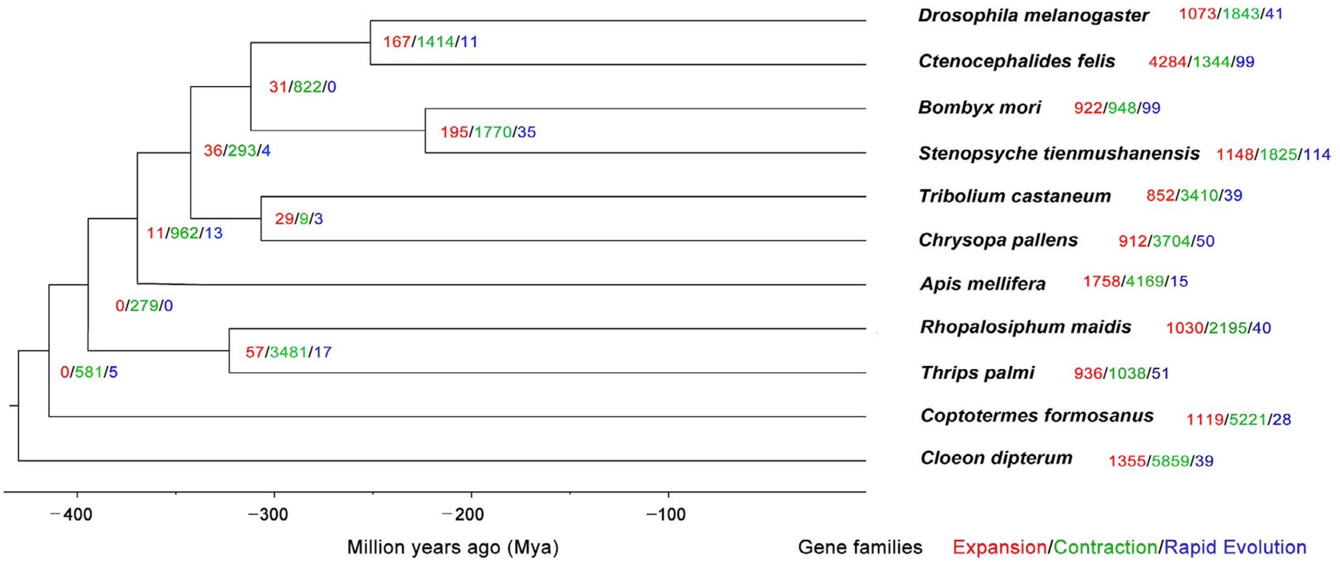

FIGURE 4 Gene family evolutions among genomes of *Chrysopa pallens* and other species. Red indicates gene family expansions, green indicates gene family contractions, and blue indicates gene family rapid evolutions. The length of branch indicated the divergence time

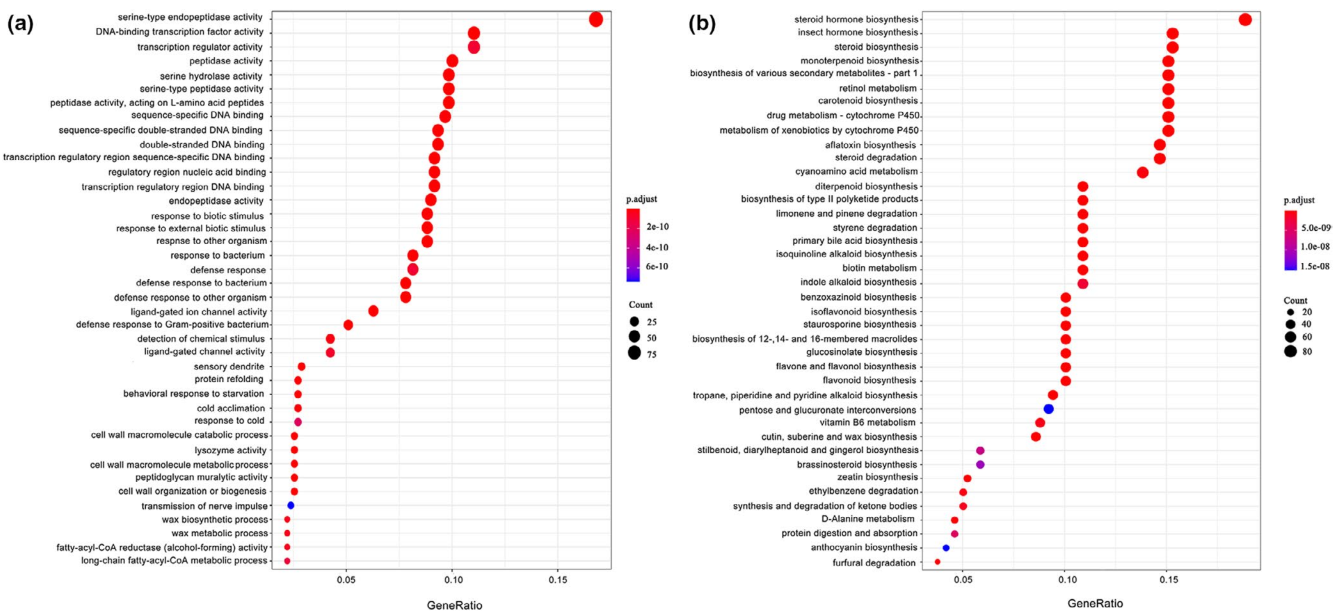

FIGURE 5 Enrichment analyses of expanded gene families of *Chrysopa pallens*. (a) GO enrichment map of gene families expanded rapidly. The vertical axis represents the path name and the horizontal axis represents gene ratio. The size of the dot indicates the number of differentially expressed genes in the pathway and the colour of the point corresponds to different *q*-value ranges. (b) KEGG enrichment map of gene families expanded rapidly. The vertical axis represents the path name and the horizontal axis represents gene ratio. The size of the dot indicates the number of differentially expressed genes in the pathway and the colour of the point corresponds to different *q*-value ranges

may be related to its overwintering, spinning and cocooning as well as the formation of egg stalks. KEGG pathways enriched in expanded gene families mainly included insect pheromone synthesis, detoxification metabolism, digestion and other related pathways as well as wax synthesis (cutin, suberine and wax biosynthesis) (Figure 5b). In the long evolutionary process of adaptation, insects have gradually developed adaptations to low temperatures. Many insects are able to build up their tolerance before cold winter temperatures set in to improve the survival of their populations (Bale, 1987; Danks, 2006; Lee, 1989). The cold hardiness of the prepupae of *C. pallens* show obvious seasonal characteristics. The prewinter and midwinter prepupae had significantly higher cold tolerance than those in summer and post-wintering (Yu, Shi, et al., 2012). *C. pallens* overwintered by adopting a freeze avoidance strategy, and the diapausing prepupae increased their cold tolerance by reducing the supercooling point. As the short-day diapause type, the prepupae of *C. pallens*, such as *Ostrinia furnacalis* (Gong et al., 1984) and *Chrysoperla sinica* (Guo et al., 2006), can improve the cold resistance through their own part-time diapause. The gene families responding to cold acclimation were found to be expanded in *C. pallens* in this study (Figure 5a). However, the mechanism of how *C. pallens* goes through the winter still needs more specialized researches in the future.

Chrysopa pallens is a complete metamorphosis insect as its developmental stages consist of egg, larva, prepupa, pupa and adult (Yang et al., 2005). Interestingly, each egg of Chrysopidae is attached to the top of a filamentous petiole while the eggs of other neuropteridan insects are not. The mature larvae often cocoon on the wall of the back of the leaf. There are filaments attached to the attachment outside the cocoon. The expansion of gene families involved in wax synthesis in the genome of *C. pallens* may be related to these biological habits mentioned above.

To further understand the evolution of expanded gene families, the 49 rapidly expanded gene families were analysed for their selective pressures. Three families related to digestion were found to be positively selected, namely trypsin (OG0001146), chymotrypsin (OG0001336) and zinc finger MYM-type protein 1 (OG0000501), which may be one of the important driving forces for the wide feeding behaviours of *C. pallens*.

4 | CONCLUSIONS

In this study, a chromosome-level genome assembly was reported for *C. pallens*, an important predatory biological control agent. The key functional genes expanded in *C. pallens* were identified, including those involved in digestion, detoxification and chemoreception, implying why green lacewings can serve as an excellent biocontrol agent. The chromosome-level genome of *C. pallens* also provides insights into the genomics of Neuropterida and would serve as an invaluable reference for further comparative genomic studies of Neuropterida and even Insecta. In addition, it will support more accurate and reliable studies about the habits of *C. pallens* at the genomic level which may help to improve its control efficiency. The high quality genome will also provide a solid foundation to promote the development and utilization of natural enemies in biological pest control in the future.

ACKNOWLEDGEMENTS

We thank Professor Feng Zhang from Nanjing Agricultural University for the help in data analyses, Professor Xiaofan Zhou from South China Agricultural University, Dr Yan Yan and Dr Scott Williams for comments on this manuscript. This work was supported by the National Natural Science Foundation of China (31900341, 32170451, 32130012), the Hebei Natural Science Foundation for Excellent Young Scholars (C2019204149), the Hebei Youth Talent Support Programme (BJ2019039), the Hebei Importing Overseas Students Support Programme (C20190341), the Starting Scientific Research Foundation for the Introduced Talents of Hebei Agricultural University (YJ201817), the Hebei Key Research and Development Program (19226508D), the Earmarked Fund for China Agriculture Research System (Cars-27) and the National Key Research and Minister of Science and Technology of China (2019YFD0300104).

CONFLICT OF INTEREST

The authors have declared that no competing interest exits.

AUTHOR CONTRIBUTIONS

Yuyu Wang, Xingyue Liu and Ding Yang designed the research. Ruyue Zhang, Mengqing Wang, Jing Li and Lisheng Zhang performed research. Yuyu Wang, Cheng-Min Shi, Mengqing Wang and Ruyue Zhang analysed data., Ruyue Zhang, Fan Fan, Jing Li, Shuo Geng, Xingyue Liu and Ding Yang wrote the manuscript. All authors reviewed the manuscript.

DATA AVAILABILITY STATEMENT

All raw data have been deposited into the GenBank Sequence Read Archive (SRA) database with the BioProject accession PRJNA685652 and the BioSample accession SAMN17088963. The whole-genome sequencing and transcriptome sequencing data are also available under accession numbers SRR14085663, SRR14085664, SRR14085665, and SRR14085666. In addition, the genome assembly and annotation information are available under the accession number of JAEMTY000000000.

ORCID

Yuyu Wang https://orcid.org/0000-0002-8645-303X

REFERENCES

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

Ali, I., Zhang, S., Muhammad, M. S., Iqbal, M., & Cui, J.-J.- J.-J. (2018). Bt proteins have no detrimental effects on larvae of the green lacewing, *Chrysopa pallens* (Rambur) (Neuroptera: Chrysopidae). *Neotropical Entomology*, *47*(3), 336–343. https://doi.org/10.1007/s13744-017-0526-y

Almudi, I., Vizueta, J., Wyatt, C. D. R., de Mendoza, A., Marlétaz, F., Firbas, P. N., Feuda, R., Masiero, G., Medina, P., Alcaina-Caro, A., Cruz, F., Gómez-Garrido, J., Gut, M., Alioto, T. S., Vargas-Chavez, C., Davie, K., Misof, B., González, J., Aerts, S., … Casares, F. (2020). Genomic adaptations to aquatic and aerial life in mayflies and the origin of insect wings. *Nature Communications*, *11*(1), 2631. https://doi.org/10.1038/s41467-020-16284-8

Bai, Y. Y., Jiang, M. X., & Cheng, J. A. (2005). Effects of transgenic *cry1Ab* rice pollen on the oviposition and adult longevity of *Chrysoperla sinica* Tjeder. *Acta Phytophylacica Sinica*, *32*(3), 225–230. https://doi.org/10.13802/j.cnki.zwbhxb.2005.03.001

Bale, J. S. (1987). Insect cold hardiness: Freezing and supercooling-an ecophysiological perspective. *Journal of Insect Physiology*, *33*(12), 899–908. https://doi.org/10.1016/0022-1910(87)90001-1

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59–60. https://doi.org/10.1038/nmeth.3176

Camacho, C., Coulouris, G., Avagyan, V., Ning, M., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*(1), 421. https://doi.org/10.1186/1471-2105-10-421

Cantarel, B. L., Korf, I., Robb, S. M., Parra, G., Ross, E., Moore, B., Holt, C., Sanchez Alvarado, A., & Yandell, M. (2008). MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research*, *18*(1), 188–196. https://doi.org/10.1101/gr.6743907

Chen, C., Chen, H., Zhang, Y., Thomas, H., Frank, M. H., He, Y., & Xia, R. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant*, *13*(8), 1194–1202. https://doi.org/10.1016/j.molp.2020.06.009

Chen, M., Mei, Y., Chen, X., Chen, X., Xiao, D., He, K., Li, Q., Mengmeng, W., Wang, S., Zhang, F., & Li, F. (2021). A chromosome-level assembly of the harlequin ladybird *Harmonia axyridis* as a genomic resource to study beetle and invasion biology. *Molecular Ecology Resources*, *21*(4), 1318–1332. https://doi.org/10.1111/1755-0998.13342

Chen, W., Shakir, S., Bigham, M., Richter, A., Fei, Z., & Jander, G. (2019). Genome sequence of the corn leaf aphid (*Rhopalosiphum maidis* Fitch). *Gigascience*, *8*(4), giz033. https://doi.org/10.1093/gigascience/giz033

Cheng, T., Wu, J., Wu, Y., Chilukuri, R. V., Huang, L., Yamamoto, K., Feng, L. I., Li, W., Chen, Z., Guo, H., Liu, J., Li, S., Wang, X., Peng, L. I., Liu, D., Guo, Y., Fu, B., Li, Z., Liu, C., … Mita, K. (2017). Genomic adaptation to polyphagy and insecticides in a major East Asian noctuid pest. *Nature Ecology and Evolution*, *1*(11), 1747–1756. https://doi.org/10.1038/s41559-017-0314-4

Criscuolo, A., & Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy): A new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evolutionary Biology*, *10*(1), 210. https://doi.org/10.1186/1471-2148-10-210

Cui, L., Zhang, X. T., Zhou, N. N., Ye, H. X., Yu, J. X., Zhu, Y., & Han, B. Y. (2015). Behavioral responses of *Chrysopa septempunctata* to synomones of tea plants and sex pheromones of aphids: Effectiveness on tea aphid control. *Acta Ecologica Sinica*, *35*(5), 1537–1546.

Danks, H. (2006). Insect adaptations to cold and changing environments. *The Canadian Entomologist*, *138*(1), 1–23. https://doi.org/10.4039/n05-802

Driscoll, T. P., Verhoeve, V. I., Gillespie, J. J., Johnston, J. S., Guillotte, M. L., Rennoll-Bankert, K. E., Rahman, M. S., Hagen, D., Elsik, C. G., Macaluso, K. R., & Azad, A. F. (2020). A chromosome-level assembly of the cat flea genome uncovers rampant gene duplication and genome size plasticity. *BMC Biology*, *18*(1), 70. https://doi.org/10.1186/s12915-020-00802-7

Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., Shamim, M. S., Machol, I., Lander, E. S., Aiden, A. P., & Aiden, A. P. (2017). De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*, *356*(6333), 92–95. https://doi.org/10.1126/science.aal3327

Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S., Huntley, M. H., Lander, E. S., & Aiden, E. L. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Systems*, *3*(1), 95–98. https://doi.org/10.1016/j.cels.2016.07.002

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Sonnhammer, E. L. L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S. C. E., & Finn, R. D. (2019). The Pfam protein families database in 2019. *Nucleic Acids Research*, *47*(D1), D427–D432. https://doi.org/10.1093/nar/gky995

Emms, D. M., & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, *16*(1), 157. https://doi.org/10.1186/s13059-015-0721-2

Gong, H. F., Chen, P., Wang, R., Lian, M. L., Xia, Z. H., & Yan, Y. (1984). The influences of photoperiod and temperature on the diapause of the Asian corn borer *Ostrinia furnacallis* (Guenee). *Acta Entomologica Sinica*, *27*(3), 280–285.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307–321. https://doi.org/10.1093/sysbio/syq010

Guo, H. B., Xu, Y. Y., Ju, Z., & Li, M. G. (2006). Seasonal changes of cold hardiness of the green lacewing, *Chrysoperla sinica* (Tjeder) (Neuroptera: Chrysopidae). *Acta Ecologica Sinica*, *26*(10), 3238–3244.

Guo, S.-K., Cao, L.-J., Song, W., Shi, P., Gao, Y.-F., Gong, Y.-J., Chen, J.-C., Hoffmann, A. A., & Wei, S.-J. (2020). Chromosome-level assembly of the melon thrips genome yields insights into evolution of a sap-sucking lifestyle and pesticide resistance. *Molecular Ecology Resources*, *20*(4), 1110–1125. https://doi.org/10.1111/1755-0998.13189

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., White, O., Buell, C. R., & Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biology*, *9*(1), R7. https://doi.org/10.1186/gb-2008-9-1-r7

Han, B., Zhang, S., Zeng, F., & Mao, J. (2017). Nutritional and reproductive signaling revealed by comparative gene expression analysis in *Chrysopa pallens* (Rambur) at different nutritional statuses. *PLoS One*, *12*(7), e0180373. https://doi.org/10.1371/journal.pone.0180373

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

Hesler, L. S. (2016). Volatile semiochemicals increase trap catch of green lacewings (Neuroptera: Chrysopidae) and flower flies (Diptera: Syrphidae) in corn andsoybean plots. *Journal of Insect Science*, *16*(1), 77. https://doi.org/10.1093/jisesa/iew057

Hu, J., Fan, F., Sun, Z., & Liu, S. (2019). NextPolish: A fast and efficient genome polishing tool for long-read assembly. *Bioinformatics*, *36*(7), 2253–2255. https://doi.org/10.1093/bioinformatics/btz891

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., & Bork, P. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Molecular Biology and Evolution*, *34*(8), 2115–2122. https://doi.org/10.1093/molbev/msx148

International Silkworm Genome Consortium. (2008). The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. *Insect Biochemistry and Molecular Biology*, *38*(12), 1036–1045. https://doi.org/10.1016/j.ibmb.2008.11.004

Itakura, S., Yoshikawa, Y., Togami, Y., & Umezawa, K. (2020). Draft genome sequence of the termite, *Coptotermes formosanus*: Genetic insights into the pyruvate dehydrogenase complex of the termite. *Journal of Asia-Pacific Entomology*, *23*(3), 666–674. https://doi.org/10.1016/j.aspen.2020.05.004

Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., & Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic Genome Research*, *110*(1–4), 462–467. https://doi.org/10.1159/000084979

Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E. P., Rivas, E., Eddy, S. R., Bateman, A., Finn, R. D., & Petrov, A. I. (2018). Rfam 13.0: Shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Research*, *46*(D1), D335–D342. https://doi.org/10.1093/nar/gkx1038

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

Khan, M. A., Bhatia, P., & Sadiq, M. (2012). BBTool: A tool to generate the test cases. *International Journal of Recent Technology & Engineering*, *1*(2), 192–197.

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357–360. https://doi.org/10.1038/nmeth.3317

Kirkness, E. F., Bafna, V., Halpern, A. L., Remington, K., & Rusch, D. B. (2003). The dog genome: Survey sequencing and comparative analysis. *Science*, *301*(5641), 1898–1903. https://doi.org/10.1126/science.1086432

Kjer, K. M. (2004). Aligned 18S and insect phylogeny. *Systematic Biology*, *53*(3), 506–514. https://doi.org/10.1080/10635150490445922

Koczor, S., Szentkirályi, F., Birkett, M. A., Pickett, J. A., Voigt, E., & Tóth, M. (2010). Attraction of *Chrysoperla carnea* complex and *Chrysopa* spp. lacewings (Neuroptera: Chrysopidae) to aphid sex pheromone components and a synthetic blend of floral compounds in Hungary. *Pest Management Science*, *66*(12), 1374–1379. https://doi.org/10.1002/ps.2030

Kovaka, S., Zimin, A. V., Pertea, G. M., Razaghi, R., Salzberg, S. L., & Pertea, M. (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biology*, *20*, 278. https://doi.org/10.1186/s13059-019-1910-1

Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A., & Zdobnov, E. M. (2018). OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Research*, *47*(D1), D807–D811. https://doi.org/10.1093/nar/gky1053

Kück, P., & Longo, G. C. (2014). FASconCAT-G: Extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Frontiers in Zoology*, *11*(1), 81. https://doi.org/10.1186/s12983-014-0081-x

Lagesen, K., Hallin, P., Rødland, E. A., Staerfeldt, H. H., Rognes, T., & Ussery, D. W. (2007). RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, *35*(9), 3100–3108. https://doi.org/10.1093/nar/gkm160

Lee, R. E. (1989). Insect cold hardiness: To freeze or not to freeze. *BioScience*, *39*(5), 308–313. https://doi.org/10.2307/1311113

Letunic, I., Khedkar, S., & Bork, P. (2021). SMART: Recent updates, new developments and status in 2020. *Nucleic Acids Research*, *49*(D1), 458–460. https://doi.org/10.1093/nar/gkaa937

Lewis, T. E., Sillitoe, I., Dawson, N., Lam, S. D., Clarke, T., Lee, D., Orengo, C., & Lees, J. (2018). Gene3D: Extensive prediction of globular domains in proteins. *Nucleic Acids Research*, *46*(D1), 435–439. https://doi.org/10.1093/nar/gkx1069

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, S., Wang, J., Guo, X., Tian, R., Wang, S., & Zhang, F. (2019). Research progress and prospects of *Chrysopa pallens* (Rambur) (Neuroptera: Chrysopidae). *Journal of Environmental Entomology*, *41*(2), 241–252. https://doi.org/10.3969/j.issn.1674-0858.2019.02.2

Li, Z. Q., Zhang, S., Cai, X. M., Luo, J. Y., Dong, S. L., Cui, J. J., & Chen, Z. M. (2017). Three odorant binding proteins may regulate the behavioural response of *Chrysopa pallens* to plant volatiles and the aphid alarm pheromone (E)-β-farnesene. *Insect Molecular Biology*, *26*(3), 255–265. https://doi.org/10.1111/imb.12295

Li, Z.-Q., Zhang, S., Luo, J.-Y., Wang, S.-B., Dong, S.-L., & Cui, J.-J. (2015). Odorant-binding proteins display high affinities for behavioral attractants and repellents in the natural predator *Chrysopa pallens*. *Comparative Biochemistry and Physiology Part A: Molecular and Integrative Physiology*, *185*, 51–57. https://doi.org/10.1016/j.cbpa.2015.03.011

Li, Z.-Q., Zhang, S., Ma, Y., Luo, J.-Y., Wang, C.-Y., Lv, L.-M., Dong, S.-L., & Cui, J.-J. (2013). First transcriptome and digital gene expression analysis in Neuroptera with an emphasis on chemoreception genes in *Chrysopa pallens* (Rambur). *PLoS One*, *8*(6), e67151. https://doi.org/10.1371/journal.pone.0067151

Liu, C., Mao, J., & Zeng, F. (2015). *Chrysopa septempunctata* (Neuroptera: Chrysopidae) vitellogenin functions through effects on egg production and hatching. *Journal of Economic Entomology*, *108*, 2779–2788. https://doi.org/10.1093/jee/tov229

Liu, C. Y., & Zeng, F. R. (2014). Cloning of *Chrysopa septempunctata* P450 cDNA fragment and the expression induced by imidacloprid. *Chinese Journal of Biological Control*, *30*(03), 427–433. https://doi.org/10.16409/j.cnki.2095-039x.2014.03.006

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

Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, *25*(5), 955–964. https://doi.org/10.1093/nar/25.5.955

Luo, S., Tang, M., Frandsen, P. B., Stewart, R. J., & Zhou, X. (2018). The genome of an underwater architect, the caddisfly *Stenopsyche tienmushanensis* Hwang (Insecta. *Trichoptera). Gigascience*, *7*(12), giy143. https://doi.org/10.1093/gigascience/giy143

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C., … Bryant, S. H. (2017). CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research*, *45*(D1), D200–D203. https://doi.org/10.1093/nar/gkw1129

Miller, G. L., Oswald, J. D., & Miller, D. R. (2004). Lacewings and scale insects: A review of predator/prey associations between the Neuropterida and Coccoidea (Insecta: Neuroptera, Rphidioptera, Hemiptera). *Annals of the Entomological Society of America*, *97*(6), 1103–1254. https://doi.org/10.1603/0013-8746(2004)097[1103:LASIAR]2.0.CO;2

Minh, B. Q., Nguyen, M. A. T., & von Haeseler, A. (2013). Ultrafast approximation for phylogenetic Bootstrap. *Molecular Biology and Evolution*, *30*(5), 1188–1195. https://doi.org/10.1093/molbev/mst024

Misof, B., Liu, S., Meusemann, K., Peters, R. S., Donath, A., Mayer, C., Frandsen, P. B., Ware, J., Flouri, T., Beutel, R. G., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J., Aberer, A. J., Aspock, U., Aspock, H., Bartel, D., … Zhou, X. (2014). Phylogenomics resolves the timing and pattern of insect evolution. *Science*, *346*(6210), 763–767. https://doi.org/10.1126/science.1257570

Morgat, A., Lombardot, T., Coudert, E., Axelsen, K., Neto, T. B., Gehant, S., Bansal, P., Bolleman, J., Gasteiger, E., de Castro, E., Baratin, D., Pozzato, M., Xenarios, I., Poux, S., Redaschi, N., & Bridge, A. (2019). Enzyme annotation in UniProtKB using Rhea. *Bioinformatics*, *36*(6), 1896–1901. https://doi.org/10.1093/bioinformatics/btz817

Nawrocki, E. P., & Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, *29*(22), 2933–2935. https://doi.org/10.1093/bioinformatics/btt509

Nel, A., Roques, P., Nel, P., Prokin, A. A., Bourgoin, T., Prokop, J., Szwedo, J., Azar, D., Desutter-Grandcolas, L., Wappler, T., Garrouste, R., Coty, D., Huang, D., Engel, M. S., & Kirejtshuk, A. G. (2013). The earliest known holometabolous insects. *Nature*, *503*(7475), 257–261. https://doi.org/10.1038/nature12629

Nene, V., Wortman, J. R., Lawson, D., Haas, B., Kodira, C., Tu, Z., Loftus, B., Xi, Z., Megy, K., Grabherr, M., Ren, Q., Zdobnov, E. M., Lobo, N. F., Campbell, K. S., Brown, S. E., Bonaldo, M. F., Zhu, J., Sinkins, S. P., Hogenkamp, D. G., … Severson, D. W. (2007). Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science*, *316*(5832), 1718–1723. https://doi.org/10.1126/science.1138878

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 Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., & Lopez, R. (2005). InterProScan: Protein domains identifier.

Nucleic Acids Research, *33*(Web Server), W116–W120. https://doi.org/10.1093/nar/gki442

Ranallo-Benavidez, T. R., Jaron, K. S., & Schatz, M. C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications*, *11*(1), 1432. https://doi.org/10.1038/s41467-020-14998-3

Richards, S., Gibbs, R. A., Weinstock, G. M., Brown, S. J., Denell, R., Beeman, R. W., Gibbs, R., Beeman, R. W., Brown, S. J., Bucher, G., Friedrich, M., Grimmelikhuijzen, C. J. P., Klingler, M., Lorenzen, M., Richards, S., Roth, S., Schröder, R., Tautz, D., Zdobnov, E. M., … Bucher, G. (2008). The genome of the model beetle and pest *Tribolium castaneum*. *Nature*, *452*(7190), 949–955. https://doi.org/10.1038/nature06784

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Solignac, M., Zhang, L., Mougel, F., Li, B., Vautrin, D., Monnerot, M., Cornuet, J.-M., Worley, K. C., Weinstock, G. M., & Gibbs, R. A. (2007). The genome of *Apis mellifera*: Dialog between linkage mapping and sequence assembly. *Genome Biology*, *8*(3), 403. https://doi.org/10.1186/gb-2007-8-3-403

Song, F., Li, H. U., Jiang, P., Zhou, X., Liu, J., Sun, C., Vogler, A. P., & Cai, W. (2016). Capturing the phylogeny of holometabola with mitochondrial genome data and bayesian site-heterogeneous mixture models. *Genome Biology and Evolution*, *8*(5), 1411–1426. https://doi.org/10.1093/gbe/evw086

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006). AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Research*, *34*(Web Server), W435–W439. https://doi.org/10.1093/nar/gkl200

Steinegger, M., & Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, *35*(11), 1026–1028. https://doi.org/10.1038/nbt.3988

Stoesser, G., Baker, W., van den Broek, A., Camon, E., Garcia-Pastor, M., Kanz, C., & Tuli, M. A. (2001). The EMBL nucleotide sequence database. *Nucleic Acids Research*, *29*(1), 17–21. https://doi.org/10.1093/nar/29.1.17

Storer, J., Hubley, R., Rosen, J., Wheeler, J. R., & Smit, A. R. (2021). The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mobile DNA*, *12*(1), 2. https://doi.org/10.1186/s13100-020-00230-y

Tauber, M. J., Tauber, C. A., Daane, K. M., & Hagen, K. S. (2000). Commercialization of predators: Recent lessons from green lacewings (Neuroptera: Chrysopidae: *Chrysoperla*). *American Entomologist*, *46*(1), 26–39. https://doi.org/10.1093/ae/46.1.26

Tomas, B., Katharina, J. H., Alexandre, L., Mario, S., & Mark, B. (2021). BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics*, *3*(1), lqaa108. https://doi.org/10.1093/nargab/lqaa108

Vaser, R., & Šikić, M. (2021). Raven: A de novo genome assembler for long reads. *Nature Computational Science*, 2020.2008.2007.242461. https://doi.org/10.1101/2020.08.07.242461

Vega, F. E., Brown, S. M., Chen, H., Shen, E., Nair, M. B., Ceja-Navarro, J. A., Brodie, E. L., Infante, F., Dowd, P. F., & Pain, A. (2015). Draft genome of the most devastating insect pest of coffee worldwide: The coffee berry borer. *Hypothenemus Hampei*. *Scientific Reports*, *5*(12525), 1–17. https://doi.org/10.1038/srep12525

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., … Zhu, X. (2001). The sequence of the human genome. *Science*, *291*(5507), 1304–1351. https://doi.org/10.1126/science.1058040

Wang, J., Zhang, L. S., Zhang, H. P., Liu, C. X., Wang, M. Q., & Chen, H. Y. (2014). Construction of cDNA library of *Chrysopa pallens* (Rambur) antenna. *Journal of Environmental Entomology*, *36*(6), 898–904. https://doi.org/10.3969/j.issn.1674-0858.2014.06.6

Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., Lee, T.-H., Jin, H., Marler, B., Guo, H., Kissinger, J. C., & Paterson, A. H. (2012). MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, *40*(7), e49. https://doi.org/10.1093/nar/gkr1293

Wang, Y., Zhou, X., Wang, L., Liu, X., Yang, D., & Rokas, A. (2019). Gene selection and evolutionary modeling affect phylogenomic inference of Neuropterida based on transcriptome data. *International Journal of Molecular Sciences*, *20*(5), 1072. https://doi.org/10.3390/ijms20051072

Wilson, D., Pethica, R., Zhou, Y., Talbot, C., Vogel, C., Madera, M., Chothia, C., & Gough, J. (2009). SUPERFAMILY-sophisticated comparative genomics, data mining, visualization and phylogeny. *Nucleic Acids Research*, *37*(D1), 380–386. https://doi.org/10.1093/nar/gkn762

Winterton, S., & De Freitas, S. (2006). Molecular phylogeny of the green lacewings (Neuroptera: Chrysopidae). *Australian Journal of Entomology*, *45*(3), 235–243. https://doi.org/10.1111/j.1440-6055.2006.00537.x

Xu, Q. X., Wang, S., Tian, R. B., Wang, S., Zhang, F., Ma, J., & Di, N. (2019). Study on the predation potential of *Chrysopa pallens* on *Spodoptera frugiperda*. *Journal of Environmental Entomology*, *41*(4), 754–759. https://doi.org/10.3969/j.issn.1674-0858.2019.04.10

Yang, D., Liu, X. Y., & Yang, X. K. (2018). *Species Catalogue of China.Vol.2. Animals, Insecta (II), Neuropterida*. Science Press.

Yang, X. K., Yang, J. K., & Li, W. Z. (2005). *Chinese animal records: Chrysopidae*. Science Press.

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

Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS A Journal of Integrative Biology*, *16*(5), 284–287. https://doi.org/10.1089/omi.20110118

Yu, L., Shi, A., Zheng, F., Lu, H., Zhang, F., & Xu, Y. (2012). Seasonal changes in the cold hardiness of *Chrysopa pallens* (Rambur) prepupa. *Scientia Agricultura Sinica*, *45*(9), 1723–1730. https://doi.org/10.3864/j.issn.0578-1752.2012.09.007

Zhang, L., Li, S., Luo, J., Du, P., Wu, L., Li, Y., Zhu, X., Wang, L., Zhang, S. & Cui, J. (2020). Chromosome-level genome assembly of the predator *Propylea japonica* to understand its tolerance to insecticides and high temperatures. *Molecular Ecology Resources*, *20*(1), 292–307. https://doi.org/10.1111/1755-0998.13100

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Wang, Y., Zhang, R., Wang, M., Zhang, L., Shi, C.-M., Li, J., Fan, F., Geng, S., Liu, X., & Yang, D. (2022). The first chromosome-level genome assembly of a green lacewing *Chrysopa pallens* and its implication for biological control. *Molecular Ecology Resources*, 22, 755–767. https://doi.org/10.1111/1755-0998.13503