Received: 17 December 2019 | Revised: 24 January 2021 | Accepted: 25 January 2021

DOI: 10.1111/1755-0998.13342

MOLECULAR ECOLOGY RESOURCES WILEY

RESOURCE ARTICLE

A chromosome-level assembly of the harlequin ladybird *Harmonia axyridis* as a genomic resource to study beetle and invasion biology

Mengyao Chen[1] | Yang Mei[1] | Xu Chen[2] | Xi Chen[1] | Da Xiao[2] | Kang He[1] | Qiang Li[1] | Mengmeng Wu[2] | Su Wang[2] | Fan Zhang[2] | Fei Li[1]

[1]Ministry of Agriculture Key Lab of Molecular Biology of Crop Pathogens and Insects, Institute of Insect Science, Zhejiang University, Hangzhou, China

[2]Institute of Plant and Environment Protection, Beijing Academy of Agricultural and Forestry Sciences, Beijing, China

Correspondence
Fei Li, Ministry of Agriculture Key Lab of Molecular Biology of Crop Pathogens and Insects, Institute of Insect Science, Zhejiang University, Hangzhou, China.
Email: lifei18@zju.edu.cn

Funding information
This work was funded by grants from the National Key Research and Development Program (2017YFC1200602, 2017YFD0201000, 2018YFD0200402 and 2016YFC1200602), Zhejiang Science Foundation of China (LZ18C060001), and Youth Scientific Research Funds of Beijing Academy of Agricultural and Forestry Sciences (QNJJ201725).

Abstract

The harlequin ladybird, *Harmonia axyridis* (Pallas), is a well-known model organism for genetic studies and is also a well-studied natural enemy used for pest control. It became an invasive species after being introduced to North America and Europe as a pest control agent. Though two genome assemblies for this insect have been previously reported, a high-quality genome assembly at the chromosome level is still not available. Here, we obtained a new chromosome-level genome assembly of *H. axyridis* by combining various sequencing technologies, namely Illumina short reads, PacBio long reads, 10X Genomics and Hi-C. The chromosome-level genome assembly is 423 Mb with a scaffold N50 of 45.92 Mb. Using Hi-C data 1,897 scaffolds were anchored to eight chromosomes. A total of 730,068 repeat sequences were identified, making up 51.2% of the assembled genome. After masking these repeat sequences, we annotated 22,810 protein-encoding genes. The X chromosome and Y-linked scaffolds were also identified by resequencing male and female genomes and calculating the male to female coverage ratios. Two gene families associated with environmental adaptation, odorant receptor and cytochrome P450, were analysed and showed no obvious expansion in *H. axyridis*. We successfully constructed a putative biosynthesis pathway of harmonine, a defence compound in the haemolymph of *H. axyridis*, which is a key factor for *H. axyridis* strong immunity. The chromosome-level genome assembly of *H. axyridis* is a helpful resource for studies of beetle biology and invasive biology.

KEYWORDS
cytochrome P450, genome assembly, *Harmonia axyridis*, harmonine biosynthesis, odorant receptor, sex chromosome

1 | INTRODUCTION

The harlequin ladybird, *Harmonia axyridis* Pallas (Coleoptera: Coccinellidae), is a well-known model organism for genetic studies because of its colour pattern polymorphisms, which are known to be regulated by the *pannier* gene in the *h* locus (Ando et al., 2018; Gautier et al., 2018). This insect has also been widely used as a biological control agent (Koch, 2003) and it preys on many pests, such as scale insects, citrus pests, lepidopteran larvae, mites and numerous aphids (Koch, 2003; Roy & Wajnberg, 2008). Because *H. axyridis* had been successfully used as a biological control agent in various systems, it was intentionally introduced to North America and Europe in the last century (Ferran et al., 1996; Koch, 2003). However, it is highly competitive against native ladybirds and

threatens local ladybird populations (Ferran et al., 1996; Koch, 2003; Roy & Wajnberg, 2008). Thus, *H. axyridis* has become a notorious invasive insect species in North America and Europe in the past two decades, but the invasive mechanism is still largely unclear.

It has been reported that the invasiveness of some insects is caused by the expansion of some gene families such as the odorant receptor (OR) and cytochrome P450 (P450) families (Wan et al., 2019). ORs, which perceive volatile chemicals, are amongst the largest gene families in insect genomes and have contributed to insect adaptation to terrestrial conditions during evolution (Eyun et al., 2017; Mitchell et al., 2020). Originating from gustatory receptors, the OR family appeared after the emergence of the Hexapoda, and then greatly expanded in the pterygote insects (Eyun et al., 2017). ORs exhibit high diversity in Coleoptera, with the number of ORs varying from hundreds to 22 in coleopteran insects (Mitchell et al., 2020; Zhang et al. 2016). The P450 gene family is one of the largest and oldest gene superfamilies and is found in almost all kingdoms of life (Feyereisen, 1999, 2012). The functional diversity of P450 genes has shaped the success of insects, especially in detoxifying xenobiotics including insecticides (Feyereisen, 2012; Scott, 1999).

Harmonine ((17*R*,9*Z*)-1,17-diaminooctadec-9-ene), a defence alkaloid compound in the haemolymph, also contributes to the high adaptability and successful invasiveness of *H. axyridis* (Braconnier et al., 1985; Haulotte et al., 2012; Rohrich et al., 2012). Harmonine exhibits high antimicrobial activity against a broad range of pathogens, some of which are harmful to humans such as *Mycobacterium tuberculosis* and *Plasmodium falciparum* (Rohrich et al., 2012; Vilcinskas et al., 2013). Harmonine was found to exhibit antiproliferative and cytotoxic activity against both human and lepidopteran cell lines, with immunity research potential (Rohrich et al., 2012). Harmonine is also associated with the more aggressive intraguild and interspecific behaviours of *H. axyridis* compared with other ladybird beetles (Kajita et al., 2010; Sloggett et al., 2011). Moreover, harmonine allows *H. axyridis* to survive obligate parasitic microsporidia, which can cause high mortality when used to treat another Coccinellidae ladybird beetle, *Coccinella septempunctata* (Vilcinskas et al., 2013). This suggests that harmonine has an important role in *H. axyridis* immunity (Rohrich et al., 2012).

Though two genome assemblies have been previously reported (Ando et al., 2018; Gautier et al., 2018), a chromosome-level assembly is still not available at present. Coleoptera insects have great variation in the number of chromosomes, such as the number of chromosomes in the suborder Polyphaga, which ranges from four to 70 (Blackmon & Demuth, 2015; Smith, 1960). Compared with autosomes, the sex chromosomes have been less well studied. Sex-linked information in Coleoptera has been reported for several beetles (Fallon et al., 2018; Tribolium Genome Sequencing Consortium, 2008; Van Belleghem et al., 2018). However, sex chromosomes have been only well studied in the red flour beetle *Tribolium castaneum* (XY system) and big dipper firefly *Photinus pyralis* (XO system) (Fallon et al., 2018; Tribolium Genome Sequencing Consortium, 2008). *H. axyridis* has an XY system (Du et al., 2010; Tang et al., 2012), but its sex chromosome sequences have not yet been identified.

Here, we used a hybrid sequencing approach combining Illumina HiSeq X and PacBio RS II reads, 10X Genomics long reads and Hi-C scaffolding to generate a chromosome-level genome with a scaffold N50 of 45.92 Mb. Using Hi-C scaffolding we assigned ~90% of the bases to eight chromosomes. Using this high-quality genome assembly, we successfully identified the X chromosome, constructed the harmonine biosynthesis pathway, and analysed the OR and P450 gene families. A chromosome-level genome assembly will be helpful for studying aspects of ladybird biology such as sex determination, harmonine biosynthesis and genetic determination of colour pattern.

2 | MATERIALS AND METHODS

2.1 | Insects for sequencing

Harmonia axyridis were provided by the Beijing Academy of Agriculture and Forestry Sciences (BAAFS) (Chen et al., 2019). The beetles were maintained on *Aphis craccivora* Koch (Hemiptera: Aphididae) in aluminium frame cages (50 cm × 50 cm × 50 cm) covered with 100-mesh plastic gauze and maintained at 25 ± 1°C, 55 ± 5% relative humidity, and a 16-hr:8-hr light–dark photoperiod. Genomic DNA was extracted from 25 two-spot (i.e., morph *conspicua*) male adults, and 26 *H. axyridis* RNA samples were downloaded from NCBI (National Center for Biotechnology Information) (Table S1).

2.2 | Genome size estimation by flow cytometry

Nuclear suspension samples were prepared according to a standard procedure (Doležel et al., 2007; Galbraith et al., 1983; He et al., 2016). Briefly, 10–20 male or female adults were anaesthetized with carbon dioxide. Their head parts were then dissected and minced in ice-cold Galbraith's buffer (pH 7.0). The homogenate was centrifuged at 5,000 r.p.m. at 25°C for 5 min, suspended in 400 μl phosphate buffer (pH 7.4) and incubated with RNaseA (TaKaRa Bio Inc., Kusatsu, Shiga, Japan) at a final concentration of 20 μg ml^{-1} at 25°C for 10 min. The nuclei were stained with 50 μg ml^{-1} propidium iodide stock solution at 4°C for 10 min in the dark. A FACSCalibur platform (BD Biosciences) was used to analyse the samples. Fluorescence intensity was detected by laser at 488 nm excitation and 320 V. The FL2-A channel in linear amplification mode was applied with FACScomp version 4.0 (BD Biosciences), and the signals of 5,000 nuclei were counted. FlowJo version 7.6.1 (FlowJo, LLC) was used to analyse the nuclear peaks (2C-value). The genome size was calculated as follows: sample genome size = reference genome size × sample FL2-A mean/reference FL2-A mean. A nuclear suspension prepared from 20 heads of wild-type *Drosophila melanogaster* (Canton-S strain) was the external reference and nuclei were analysed with the same parameter settings as those used for *H. axyridis*. All experiments were repeated in triplicate.

2.3 | K-mer analysis estimation of genome size

A pair-end library with an insert size of 350 bp was sequenced on an Illumina HiSeq X Ten platform. Quality control of raw data was done using FASTQC to ensure that pair-end reads had more than 90% of bases with quality ≥Q20. A total of 25.86 Gb of clean reads were used to construct a k-mer distribution map (k = 17) using SOAPDENOVO (https://github.com/aquaskyline/SOAPdenovo2). The distribution of 17 k-mers obeyed a Poisson distribution, and this distribution was used to estimate the genome size based on the Lander–Waterman algorithm (Lander & Waterman, 1988). The genome size was estimated by calculating K-mer number/K-mer depth.

2.4 | Genome sequencing and *de novo* assembly

Libraries (20-kb inserts) were sequenced using a single molecule real-time (SMRT) PacBio system. Libraries were prepared using a standard PacBio library preparation protocol (Pacific Biosciences) followed by damage and end repair, blunt-end adaptor ligation and SMRTbell library quality assessment (McCarthy, 2010). One SMRT cell was run for genome sequencing on the PacBio RS II sequencing platform and yielded a total 54.8 Gb of clean data after removing low-quality reads. This content corresponded to ~121-fold genome coverage. PacBio subreads (mean subread length: 9.99 kb, subread N50 length: 13.41 kb) were initially cleaned and processed by FALCON version 1.8.7 (https://github.com/PacificBiosciences/falcon) to generate consensus sequences. Linked reads generated by the 10X GemCode Platform were aligned by FRAGSCAFF software (https://github.com/adeylab/fragScaff) to the consensus sequence of the PacBio assembly output to obtain extended scaffolds. Using the 10X Genomics Chromium System, we obtained 75.48 Gb of sequencing data and achieved a 168-fold coverage depth. The raw contigs were polished by mapping Illumina reads back to the improved contigs, and post-processing error corrected with PILON version 1.20 (https://github.com/broadinstitute/pilon) (Chin et al., 2016; Walker et al., 2014).

2.5 | Hi-C

High quality Hi-C data were obtained for genome assembly. Newly hatched first instar larvae were used to prepare Hi-C libraries using a standard procedure (Shi et al., 2019). Hi-C libraries were constructed and sequenced on the Illumina HiSeq X Ten platform with 2 × 150-bp reads. The clean Hi-C data were aligned by end-to-end modelling to the primary assembly with BOWTIE2 version 2.2.3. Only read pairs with both reads aligned to contigs were retained for downstream analysis. The HiC-Pro version 2.7.8 (https://github. com/nservant/HiC-Pro) pipeline was used to detect the unmapped read ligation sites and align the 5′ ends of the reads back to the genome (Servant et al., 2015). LACHESIS (https://github.com/shendurelab/LACHESIS) was then used to cluster, order and orient the contigs to pseudo-chromosomes (chromosome-scale scaffolds) (Burton et al., 2013). The pseudo-chromosomes were then cut into equal 100-kb bins. A heatmap was constructed based on the interaction signals between bins revealed by valid mapped read pairs (Lieberman-Aiden et al., 2009). The chromosome matrix was visualized by a heatmap in the form of diagonal patches of strong linkage (Figure S1).

2.6 | Assessment of genome assembly quality

BUSCO version 3.0 (Benchmarking Universal Single-Copy Orthologs) was used to assess the quality of the *H. axyridis* assembly; 1,658 universal single-copy orthologous genes from insecta_db 9 were used as queries in searches against the assembly using the default parameters (Simao et al., 2015).

2.7 | Gene prediction and annotation

Repeat sequences were masked and novel repeat sequences were predicted with REPEATMODELER version 1.0.7 (Tempel, 2012). Transposable elements (TEs) in the assemblies were predicted by performing a homology-based search against RepBase using REPEATMASKER version 4.0.5 (Tempel, 2012). The *H. axyridis* genome was annotated by integrating evidence from homologue comparison, transcriptome sequencing and *de novo* protein structure predictions (Liu et al., 2014). Homologous protein sequences were downloaded from the NCBI invertebrate RefSeq database. *H. axyridis* transcripts were obtained using the HISAT2 version 2.1.0 and STRINGTIE version 1.3.4 pipelines using the default parameters (D. Kim et al., 2015; Pertea et al., 2015). All selected transcripts were used as queries in BLAST searches against the Swiss-Prot protein database (E-value =1e-5) and Pfam to identify protein domains and then filtered with TRANSDECODER version 2.0.1 (Gremme et al., 2005). AUGUSTUS version 2.5.5 and SNAP 2013–02–16 were used to generate the *de novo* annotation, and EXONERATE version 2.2.0 and GENOMETHREADER version 1.7.1 were used to align the homologous proteins obtained from NCBI RefSeq (https://www.ncbi.nlm.nih.gov/refseq/) (Gremme et al., 2005; Korf, 2004; Slater & Birney, 2005; Stanke et al., 2006). Finally, we integrated these three types of evidence with different weights (*de novo* annotation was "1," homology-based annotation was "5" and RNA-Seq-based annotation was "10") using EVIDENCEMODELER (EVM) to obtain the official gene set (OGS) (Haas et al., 2008).

2.8 | Phylogenetic analysis

A total of 18 insects were selected to construct a phylogenetic tree: *D. melanogaster* (GCF_000001215.4; from NCBI), *Aedes aegypti*

(GCF_002204515.2; from NCBI), *Bombyx mori* (GCF_000151625.1; from NCBI), *Spodoptera litura* (GCF_002706865.1; from NCBI), *Aethina tumida* (GCF_001937115.1; from NCBI), *Agrilus planipennis* (GCF_000699045.2; from NCBI), *Anoplophora glabripennis* (GCF_000390285.2; from NCBI), *Aquatica lateralis* (AQULA_OGS1.0; downloaded from https://github.com/photocyte/AQULA_OGS/releases), *Diabrotica virgifera* (GCF_003013835.1; from NCBI), *Dendroctonus ponderosae* (GCF_000355655.1; from NCBI), *Onthophagus taurus* (GCF_000648695.1; from NCBI), *Oryctes borbonicus* (GCA_001443705.1; from NCBI), *Photinus pyralis* (PPYR_OGS1.1; downloaded from https://github.com/photocyte/PPYR_OGS/releases), *Tribolium castaneum* (GCF_000002335.3; from NCBI), *Apis mellifera* (GCF_003254395.2; from NCBI), *Nasonia vitripennis* (GCF_000002325.3; from NCBI) and *Rhodnius prolixus* (downloaded from VectorBase, https://www.vectorbase.org/organisms/rhodnius-prolixus). Protein sequences for these species were constructed with the ORTHOMCL version 2.0.9 pipeline using the default parameters (L. Li et al., 2003).

MAFFT version 7 and TRIMAL version 1.2 were used for multiple sequence alignments of all single-copy genes (Capella-Gutiérrez et al., 2009; Katoh & Standley, 2013). PROTEST 3.4.2 was used to select the best-fitting model (Darriba et al., 2011). The maximum likelihood tree was reconstructed using RAXML version 1.5.5 with the VT +GAMMA (4) model and 1,000 rapid bootstrap replicates (Stamatakis, 2014). The sequences were used as queries in searches against the OrthoMCL database, and a gene family was defined as a set of genes descended from a single gene in the last common ancestor of the species under consideration. In total, 836 single-copy genes obtained from OrthoMCL were clustered to reconstruct the phylogenetic tree. To estimate the divergence time of *H. axyridis*, one calibration point, namely stem Lampyridae (*Protoluciola albertalleni*) at 93.5–99.6 million years ago, was applied based on the fossil records in the Paleobiology Database (https://www.paleobiodb.org/#/). The divergence time was estimated with RelTime ML in MEGA version 10.0.5 based on the topology of the phylogenetic tree (Kumar et al., 2018). The phylogenetic tree was visualized with ITOL version 4 (https://itol.embl.de/).

The protein sequences of the 18 insect species were aligned to the TreeFam database, and a TreeFam ID was defined for each protein (H. Li et al., 2006). CAFÉ version 4.2.1 was used to perform gene family expansion and contraction analyses (De Bie et al., 2006). p <.05 was the accepted level indicating significantly altered gene families.

2.9 | Genome synteny analysis

Whole-genome synteny among *H. axyridis*, *T. castaneum* and *P.pyralis* was detected with SATSUMA version 3.1.0, which is a package in SPINES (https://www.broadinstitute.org/genome-sequencing-and-analysis/spines) with default parameters (Grabherr et al., 2010). Synteny blocks were plotted across chromosomes using CIRCOS version 0.69–9 (Krzywinski et al., 2009).

2.10 | Identifying sex chromosomes

To detect the X and Y chromosomes in *H. axyridis*, genomic DNA from 10 male and 10 female adults (mixed morph types) was sequenced on an Illumina HiSeq X Ten platform with ≥50× coverage. Clean resequencing data were aligned to the reference genome with BOWTIE2 version 2.3.5 using the default parameters. Sequencing coverage differences were compared between male and female samples to localize the sex-linked genomic regions. The corresponding male to female (M:F) read counts ratio was used to identify sex-linked scaffolds. The autosomes are expected to have equal coverage, while the X chromosomes should have approximately two-fold greater coverage in females and Y chromosomes should have approximately two-fold greater coverage in males. Chromosomes were defined as an autosome if the $\log_2$(M:F) value was approximately 0, an X chromosome if the value was less than or equal to −1, and a Y chromosome if the value was greater than or equal to 1. A custom script (unpublished) was used to identify sex-linked scaffolds based on the M:F ratio.

2.11 | Gene family analysis

For analysis of the OR and P450 gene families, we collected *T.castaneum* P450 and all known Coleopteran OR protein sequences from previously published papers (Mitchell et al., 2020; Zhu et al., 2013). Pseudogenes were removed manually and the remaining protein sequences were used as queries in TBLASTN (BLAST version 2.9.0+) searches (E-value =1e-5) against the *H. axyridis* genome to find candidate genes. GENEWISE version 2.4.1 was used to detect the gene structure (Birney et al., 2004). Then we used HMMER version 3.1b2 to perform searches against the Pfam database to confirm the identity of the candidate genes (Finn et al., 2014; Potter et al., 2018). For P450 candidate genes we used the p450 (PF00067) HMM profile, and for OR candidate genes we used the 7tm_6 (PF02949) or 7tm_4 (PF13853) HMM profile. The sequences containing the HMM profiles were regarded as confirmed genes. To identify more candidate OR genes, we carried out HMM pattern scans using HMMER version 3.1b2 at genome level without homology searching function. The identified candidate genes were merged with those identified by homology searching at the protein level.

For phylogenetic analysis of gene families, the protein sequences of each gene family were aligned and trimmed by MAFFT version 7 and TRIMAL version 1.2 with default options. IQ-TREE version 1.6.10 was used to construct the trees, which were formatted in Adobe Illustrator (Adobe Systems) (Nguyen et al., 2015).

2.12 | Predicting the putative harmonine biosynthesis pathway

Harmonine biosynthesis requires stearic acid (C18), oleic acid (C18) and their precursor, palmitic acid (C16), all of which participate in fatty

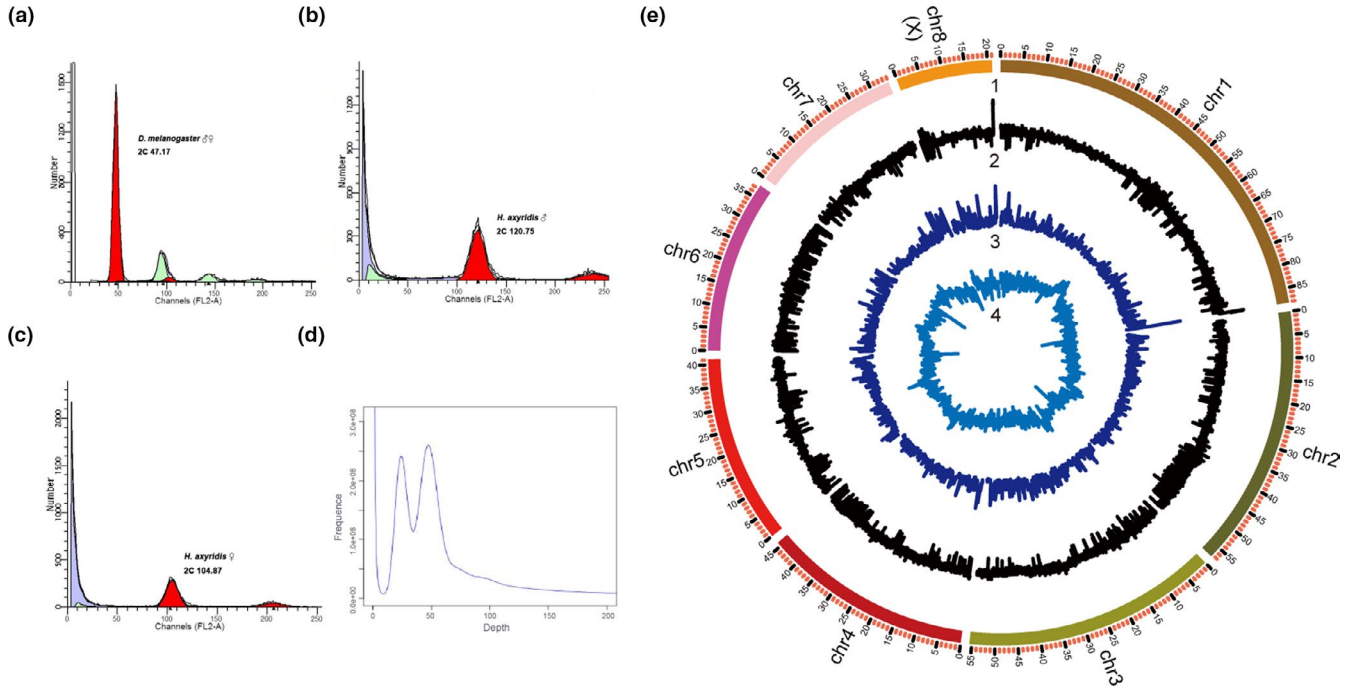


FIGURE 1 Estimation of *Harmonia axyridis* genome size and the genomic landscape of eight chromosomes. Flow cytometry and k-mer results are shown in (a) and (b). The information of eight chromosomes is shown in (e). (a) *Drosophila melanogaster* was used as a reference genome standard. (b) *H. axyridis* female (391.38 ± 2.75 Mb, 2C peak channel is 104.87). (c) *H. axyridis* male (450.44 ± 4.38 Mb, 2C peak channel is 120.75). (d) Distribution of 17-mer frequency in *H. axyridis* sequencing reads at different sequencing depths. *x*-axis: sequencing depth; *y*-axis: the 17-mer frequency (percentage). (e) Track 1 (from the outer ring): chromosome ideograms (Mb scale). Track 2: distribution of repeat density (sliding window size =200 kb; overlap =0 kb). Track 3: distribution of gene density (sliding window size =200 kb; overlap =0 kb). Track 4: distribution of genomic GC content (%; sliding window size =200 kb; overlap =0 kb)

acid biosynthesis. We used the KEGG PATHWAY and ORTHOLOGY tools in Kyoto Encyclopedia of Genes and Genomes (KEGG) (https://www.genome.jp/kegg/) to detect all participated components and KEGG orthology. All available KEGG pathway maps related to fatty acid biosynthesis (map01040, map01212, map01040, map00073, map00062 and map04212) were investigated including the C16 and C18 pathways and reactions involving KEGG orthologues (KO) of the enzymes. ORTHOMCL version 2.0.9 with default parameters was used to detect orthologous groups of harmonine biosynthesis KOs in the sequenced *H. axyridis* genome. BLASTP (E-value <1e-5) was used to determine the locations of the KOs. All steps related to harmonine biosynthesis were linked and a KEGG pathway was proposed for harmonine biosynthesis. RSEM version 1.2.31 with default parameters was used for harmonine biosynthesis gene expression level (FPKM) calculation (Li & Dewey, 2011).

2.13 | PCR validation of candidate genes of the harmonine biosynthesis pathway

H. axyridis samples were prepared from embryos (1-, 2-, 3-day-old eggs), larvae (first, second, third and fourth instar larvae), pupae (1-, 2-, 3-, 4-, 5-day-old) and adults (1-, 5-, 10- and 20-day-old). Total RNA was extracted from the mixed sample of *H. axyridis* using TRIzol reagent (Invitrogen) according to the manufacturer's instructions. RNA concentration was measured using a NanodropOne spectrophotometer (Thermo Fisher Scientific) at 260 nm. Then, 1 μg of total RNA was used for cDNA synthesis by using a First Strand cDNA Synthesis Kit (Takara) with oligo (dT)18 as the primer in a 20-μl reaction system. Specific primers for the genes on the biosynthesis pathway of harmonine were designed using PRIMER 5.0 (Table S10). "I-5 2×High-Fidelity Master Mix" was used for amplifying genes. The touchdown PCR conditions were as follow: 98℃ for 5 min, six cycles at 98℃ for 10 s, 61℃ for 30 s and 72℃ for 30 s, 35 cycles at 98℃ for 10 s, 55℃ for 20 s and 72℃ for 20 s and finally extended at 72℃ for 5 min. The PCR products were detected by 1% agarose gel electrophoresis.

3 | RESULTS AND DISCUSSION

3.1 | Chromosome-level genome assembly of *H. axyridis*

We first estimated the genome size of *Harmonia axyridis* using flow cytometry, which showed that the genome size of male adults is 450.44 ± 4.38 Mb (Figure 1). We also estimated different genomic features by performing K-mer analysis (K = 17). We found that the

laboratory-maintained *H. axyridis* population has a high level of heterozygosity (1.93%). K-mer analysis also showed that repeat sequences account for 49.75% of the genome, which is estimated to be 446.34 Mb in size (Figure 1; Table S2).

Next, we used PacBio long reads and 10X Genomics linked reads to generate a genome assembly of *H. axyridis*, which we named the ZJU-BJ version. We obtained 155.39 Gb of clean data with ~346-fold coverage (Tables S3–S7). A draft genome assembly of 440 Mb was obtained, yielding 2,813 contigs with a contig N50 of 1.25 Mb and 2,171 scaffolds with a scaffold N50 of 2.01 Mb. BUSCO analysis indicated that 98.5% of the BUSCO gene orthologues were captured (97.5% complete and 1% fragmented, Table 1), suggesting that this draft genome assembly has a high level of completeness.

Hi-C scaffolding was used to anchor, order and orient 1,897 scaffolds to seven autosomes and one X chromosome because the karyotype of *H. axyridis* is 2*n* = 16 (14A + XY) (Gotoh et al., 2015; J. Li, 1940; Smith, 1953) (Figure 1; Table S9). This yielded a chromosome-level genome assembly with a scaffold N50 of 45.92 Mb (Table 1). The Hi-C linking information indicated that more than 99% of the assembled bases were anchored to the chromosomes. We noticed that the BUSCO completeness estimate decreased from 98.5% to 93.1% after Hi-C scaffolding, a trend that is sometimes observed when genome contiguity increases and may be influenced by unanchored scaffolds, assembly errors or structural features of the assembly.

To evaluate our chromosome-level assembly, we compared with published genomes produced by Ando *et al.* and Gautier *et al.* The result indicated three genome assemblies have good alignments, suggesting the reliability of chromosome-level assembly. Compared with the previously published *H. axyridis* genome assemblies, we hence produced a chromosome-level genome for *H. axyridis* (Table 1). In addition, the chromosome-level genome has a higher N50 than that of chromosome-level genomes of two other coleoptera insects, *Tribolium castaneum* and *Photinus pyralis* (Table 2) (Fallon et al., 2018; Tribolium Genome Sequencing Consortium, 2008).

3.2 | Genome annotation

The genome assembly before Hi-C scaffolding was used for genome annotation. A total of 730,068 repeat sequences occupying 51.15% of the assembled genome were identified. Among these repeats, DNA classes were the most abundant repeat element and 18.65% could not be classified into any known repeat elements (Table 3). The *H. axyridis* genome had a much higher level of repetitive sequence than the genomes of the other two sequenced coleopterans *T.castaneum* and *P.pyralis* (25.85% and 42.6%, respectively) (Table 2) (Fallon et al., 2018; Kim et al., 2009; Wasserman & Ehrman, 1986). We used the EVM pipeline to annotate the genome, yielding 22,810 protein-encoding genes, which is larger than the number in *T.castaneum* (16,590 genes) and *P.pyralis* (15,773 genes) (Table 2).

TABLE 1 Comparison of three *Harmonia axyridis* version assemblies

Features	ZJU-BJ	ZJU-BJ chromosome-level	HaxR[a]	Assembly version 1[Δ]
Sequencing technologies	PacBio reads (121X) 10X Genomics (168X) Illumina reads (58X)	Hi-C	MinION long reads (65X) Illumina PE reads (100X)	Illumina HiSeq2500 sequencers
Genome size (Mb)	440	423	429	423
Numbers of contigs	2,813	2,094	1,071	n/a
Numbers of scaffolds	2,171	1,897	n/a	18,515
Contig N50 (kb)	1,254	881	1,434	63.5
Scaffold N50 (Mb)	2.71	45.92	n/a	1.616
Number of assembled chromosomes	n/a	7A+X	n/a	n/a
BUSCO genes (%)	98.5 [C: 97.5% (S: 92.3%, D: 5.2%), M:1.5%]	93.1 [C: 92.1% (S: 89.4%, D:2.7%), M: 6.9%]	97.3 [C: 95.6% (S: 91.9%, D: 3.7%), M: 2.0%]	98.9 [C: 97.9% (S: 94.9%, D: 3.0%), M: 1.1%] *
G + C (%)	34.87	35.05	34.84[b]	34.84[b]
Number of genes	22,810	18,765	n/a	n/a
Repeat (%)	51.15	51.38	53.22[b]	51.56[b]
N (%)	1.0	1.0	0.0[b]	6.3[b]

Abbreviations: A, autosome; n/a: Not available; X, X chromosome.

[a]Two previously published versions were HaxR and Assembly version 1. The data of HaxR were from Gautier et al. 2018, and the data of Assembly version 1 were from Ando et al. 2018.

[b]These data were calculated by us as it was not provided in Gautier et al. 2018 and Ando et al. 2018.

TABLE 2 Summary of chromosome-level assemblies for three coleoptera insects

Features	*Harmonia axyridis*	*Tribolium castaneum*	*Photinus pyralis*
Genome size (Mb)	423	166	448
Karyotype	2*n* = 16 with XY	2*n* = 20 with XY	2*n* = 20 with XO
Numbers of contigs	2,094	7,059	7,533
Numbers of scaffolds	1,897	2,149	2,160
Number of assembled chromosomes	7A + X	10A + X	10A + X
Chromosome-level contig N50 (kb)	881	73	192.5
Chromosome-level scaffold N50 (Mb)	45.92	4.46	49.17
BUSCO genes (%)	93.1	95.7[a]	97.2
G + C (%)	35.39	35.19	36.41[a]
Number of genes	18,765	16,590	15,773
Repeat (%)	51.38	25.85	42.6

Abbreviations: A, autosome; X, X chromosome; Y, Y chromosome.

[a]These data was calculated by us as they were not provided in Tribolium Genome Sequencing Consortium, 2008 and Fallon et al., 2018.

TABLE 3 Statistics of repeat elements of *Harmonia axyridis*

Repeat types	Number of elements	Length occupied (bp)	Percentage of sequence
Interspersed repeats			
SINE	2,984	342,897	0.08
LINE	79,149	31,519,012	7.16
LTR	20,393	11,765,441	2.67
DNA elements	288,822	99,185,101	22.54
Unclassified	337,401	82,109,997	18.65
Small RNA	973	973	157,915
satellites	346	346	65,484
Total base masked	730,068	225,145,847	51.15

3.3 | Phylogenetic analysis of *H. axyridis* and other 17 insects

We constructed a phylogenetic tree using protein-coding gene sets of *H. axyridis* and 17 other species from five orders, namely Coleoptera, Diptera, Hemiptera, Hymenoptera and Lepidoptera (Figure 2; Table S8). A total of 836 single-copy genes derived from ORTHOMCL gene family analysis were used to infer the phylogeny. *H. axyridis* clustered with five other coleopterans, and diverged from other coleopterans 180.89 million years ago (Mya). This was consistent with the previously estimated time of divergence of the Coccinelloidea superfamily from the other clades in Cucujiformia (Mckenna et al., 2015). A recent phylogenetic reconstruction showed that Coccinelloidea appeared at the Triassic/Jurassic transition (171 Mya, McKenna et al., 2015; 202 Mya, Hunt et al., 2007) or even earlier (252 Mya, Toussaint et al., 2017) (Hunt et al., 2007; Mckenna et al., 2015; Toussaint et al., 2017).

In total, 22,521 gene clusters were identified across the 18 species. A total of 1,714 gene clusters were found only in *H. axyridis*, accounting for 7.9% of all identified gene clusters (Figure S2). This was higher than the percentages of gene clusters unique to *D. virgifera*, *A. tumida*, *D. ponderosae* and *T.castaneum* (6.1%, 1.9%, 1.2% and 1.8%, respectively). In addition, *H. axyridis* has 1,975 expanded and 5,518 contracted gene families when compared with the estimated sizes in the common ancestor of Cucujiformia beetles.

3.4 | Sex chromosomes in *H. axyridis*

To identify the X and Y chromosomes of *H. axyridis*, we performed genome resequencing of 10 male and 10 female adults. Because females should have twice the X chromosome DNA content of males and no Y chromosome DNA content, it is possible to identify the X and Y chromosomes by calculating the coverage ratios of the DNA-sequencing reads obtained from sex-specific genome resequencing. The ~0.5 M:F ratio was used to localize the sex-linked scaffolds, and chromosome 8 was found to the X chromosome in *H. axyridis* (Figure S1). Because the Y chromosomes of insects have a high repeat content, we failed to identify the Y chromosome, but several Y-linked short scaffolds were found (Chang & Larracuente, 2019). The total length of all Y-linked scaffolds was about 4.4 Mb. This result was consistent with a previous report that the Y chromosome in *H. axyridis* is punctiform (Blackmon & Demuth, 2015).

3.5 | Chromosome fusion events in *H. axyridis*

Synteny analyses were performed with the chromosome-level genomes of *H. axyridis*, *T.castaneum*, and *P.pyralis* (Kim et al., 2009; Wasserman & Ehrman, 1986) (Figure 3). *T.castaneum* has 18

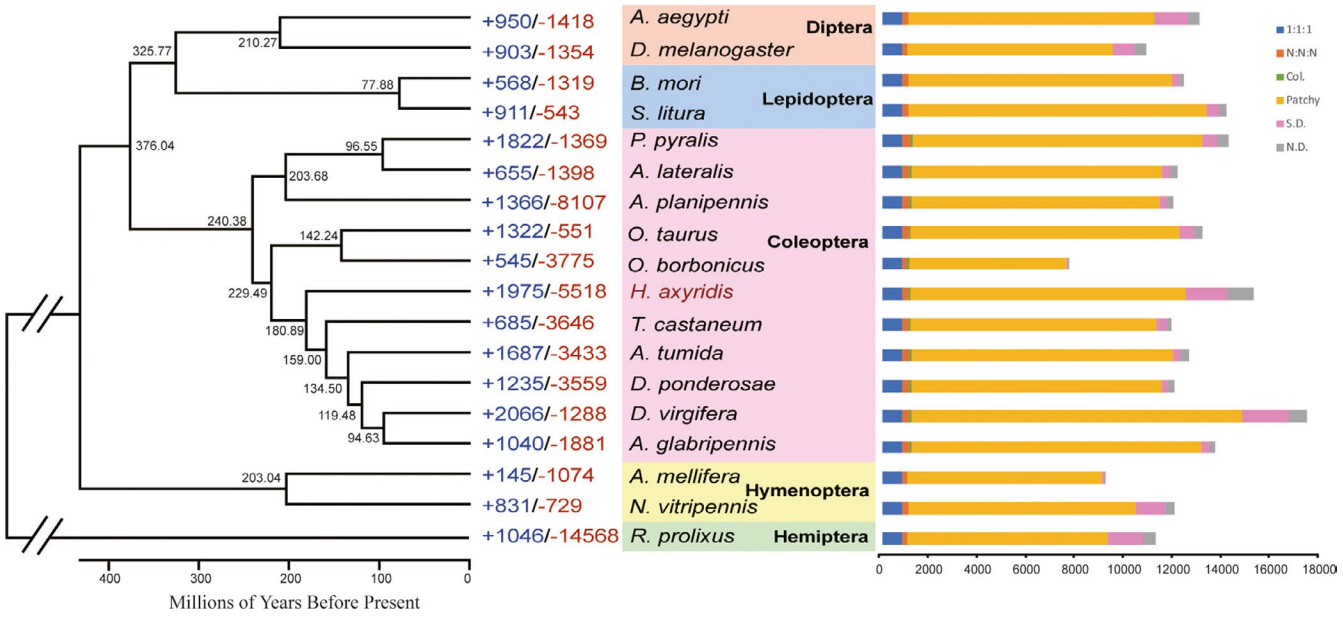


FIGURE 2 Phylogenetic tree and orthology between genomes of *Harmonia axyridis* and 17 other insects. Phylogeny inferred from 836 strict single-copy genes with 417,361 reliable sites using RAXML maximum likelihood methods employing the VT +GAMMA (4) model and 1,000 bootstrap replicates. Divergence times calculated by RelTime ML in MEGA are labelled at internodes. The 95% confidence intervals of the divergence times are listed in Table S8. "1:1:1": universal single-copy gene families across all examined species; duplication in a single genome and absence in <2 species. "N: N: N": other universal genes. "Col.": common gene families unique to Coleoptera. "SD": species-specific duplicated genes. "ND": species-specific genes. "Patchy": all other orthologous groups. Blue numbers: number of expanded gene families; red numbers: number of contracted gene families

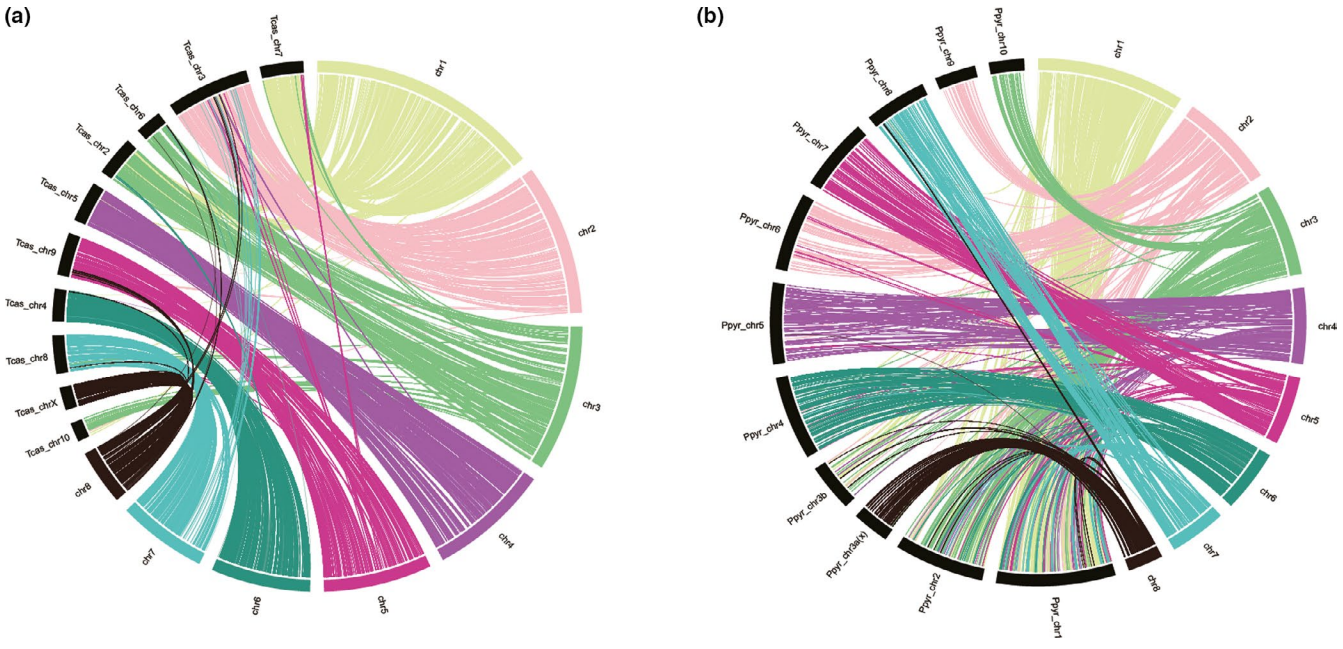
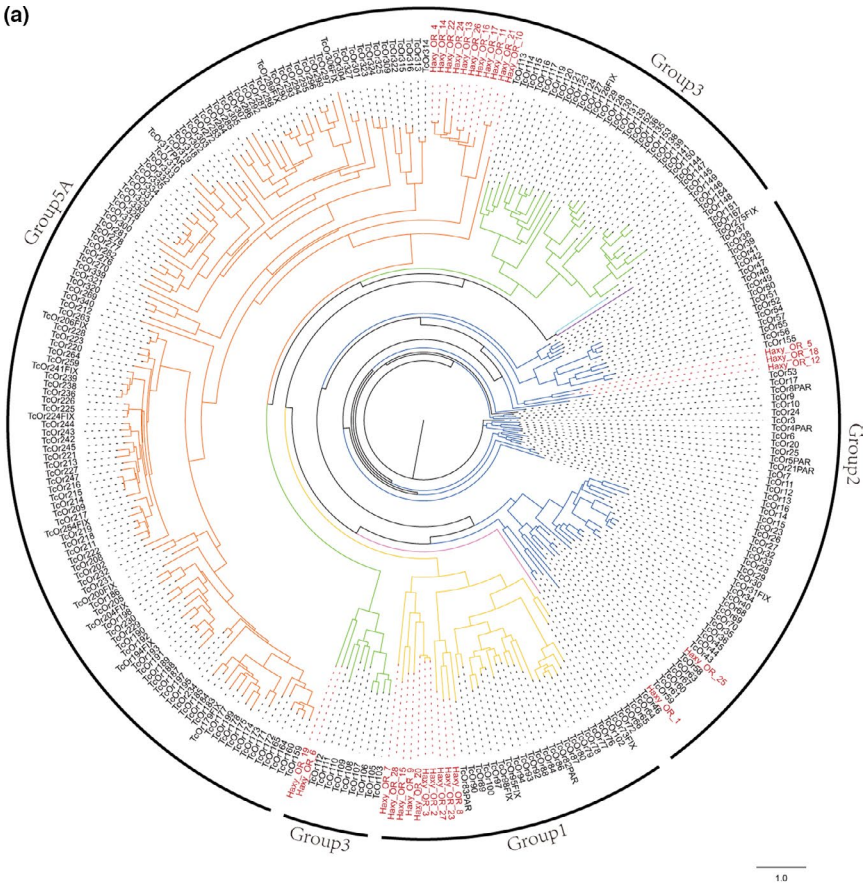


FIGURE 3 Comparative analysis of synteny between *Harmonia axyridis*, *Tribolium castaneum* and *Photinus pyralis*. (a) Whole-genome synteny between *H. axyridis* and *T.castaneum*. (b) Whole-genome synteny between *H. axyridis* and *P. pyralis*

autosomes, one X chromosome and one Y chromosome, which has not yet been identified (Stuart & Mocelin, 1995). *P.pyralis* males have 18 autosomes and one X chromosome (Fallon et al., 2018; Tribolium Genome Sequencing Consortium, 2008). Synteny analysis showed that Chr8 of *H. axyridis* aligned to ChrX of *T.castaneum* and Chr3a (X) of *P.pyralis*, confirming that Chr8 is the sex chromosome in *H. axyridis*.

Comparisons with the *T.castaneum* genome revealed that *H. axyridis* has undergone two fusion events. Chr1 of *H. axyridis* arose from the fusion of sequences found on Chr7 and some segments of Chr2



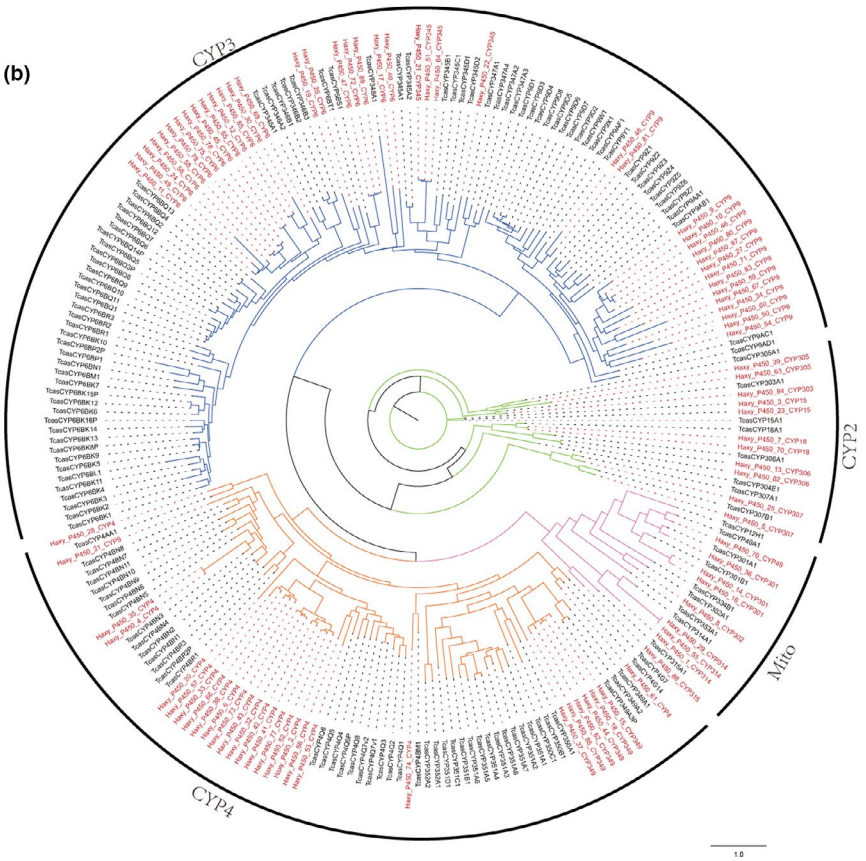


FIGURE 4 Odorant receptors and cytochrome P450 s in *Harmonia axyridis*. *H. axyridis* genes are labelled in red and *Tribolium castaneum* genes are labelled in black. (a) Maximum-likelihood phylogenetic analysis of ORs in *H. axyridis* and *T.castaneum*. Yellow branch: Group 1; blue branches: Group 2; green branches: Group 3; light blue branches: Group 4; orange branches: Group 5A; purple branches: Group 5B; pink branches: Group 6. (b) Maximum-likelihood phylogenetic analysis of P450 s in *H. axyridis* and *T.castaneum*. Green branches: CYP2 clan; pink branches: mitochondrial P450 clade; blue branches: CYP3 clan; orange branches: CYP4

in *T.castaneum*. Chr3 of *H. axyridis* was derived from the fusion of sequences found on Chr2, Chr 6 and Chr10 of *T.castaneum* (Figure 3a). From comparisons with *P.pyrali*, three fusion events were found in *H. axyridis*. Chr1 of *H. axyridis* is derived from the fusion of sequences found in Chr1 and Chr3b of *P.pyralis*. Chr2 of *H. axyridis* arose from the fusion of sequences found on Chr6, Chr9 and some segments of Chr1 in *P.pyralis*. Chr3 of *H. axyridis* is fused to sequences found on Chr2, Chr8 and Chr10 of *P.pyralis* (Figure 3b). These results indicated that the lower autosome number in *H. axyridis* (14) compared with those in *T.castaneum* and *P.pyralis* (18) is mainly due to the fusion events observed in Chr1 and Chr3 in *H. axyridis*.

3.6 | No obvious expansion of OR and P450 gene families in *H. axyridis*

We sought to determine whether the expansions of OR and P450 potentially contributed to the high adaptability and successful invasiveness of *H. axyridis*, as suggested for some insects by Wan et al. (2019). To this end, we annotated these two gene families by homology searching and with confirmatory HMM pattern scans.

With homology searching followed by HMM pattern, a total of 28 OR genes belonging to four subfamilies were identified (Figure 4a). These *H. axyridis* ORs were mainly clustered in Group 1 and Group 5A. Group 5A ORs are evident only in cucujiform beetles (Mitchell et al., 2020). Most published Coleopteran genome assemblies (suborder Polyphaga) encode more than 50 ORs (Mitchell et al., 2020), although two leaf beetles in the superfamily Chrysomelidae, *Pyrrhalta maculicollis* and *Pyrrhalta aenescens*, have only 26 and 22 OR genes, respectively (Mitchell et al., 2020; Zhang et al. 2016). *H. axyridis* therefore appears to have fewer OR genes compared to most other Coleopterans. To seek more candidate OR genes, we also carried out an HMM pattern scan at genome level without homology searching. After removing the redundancy, 75 OR genes belonging to five subfamilies were obtained (Figure S3). The first strategy predicted fewer OR genes but with high reliability, while the second method identified more putative OR genes but with lower reliability. Therefore, it remains possible that *H. axyridis* has relatively few OR genes and the hypothesis of OR contraction in these beetles is worthy of further investigations.

We also annotated 89 P450 genes, among which 42 genes were assigned to the CYP3 clade, 28 genes to the CYP4 clade, 10 genes to the CYP2 clade and nine genes to the mitochondrial P450 clade (Figure 4b). The mitochondrial P450 clade and CYP2 clan showed a high number of 1:1 orthologues between *H. axyridis* and *T.castaneum*, suggesting the functional conservation of these P450 s (Zhu et al., 2013). Compared with *T.castaneum*, no obvious expansion of the CYP3 or CYP4 clans was observed in *H. axyridis*. However, *H. axyridis* can still develop resistance to various insecticides quickly, and this might be due to the other mechanisms such as elevated esterase, target-site mutations, etc. (Hemingway, 2000; Koch, 2003; Wilson & Ashok, 1998). Our findings on OR and P450 genes in *H. axyridis* are not consistent with the results in Wan et al. (2019), suggesting that expansions of these gene families do not underlie the high adaptability and successful invasiveness of *H. axyridis*.

3.7 | The putative harmonine biosynthesis pathway

Harmonine is a defence alkaloid compound that can protect *H. axyridis* from a broad range of pathogens because it has a high antimicrobial activity (Braconnier et al., 1985; Haulotte et al., 2012; Rohrich et al., 2012). A previous study on broad-spectrum antimicrobial activity of harmonine suggests that it has contributed to invasive success of *H. axyridis* (Rohrich et al., 2012). Although *in vitro* chemical syntheses of ladybird alkaloids have been reported (Attygalle et al., 1994; Braconnier et al., 1985; Enders & Bartzen, 1991; Haulotte et al., 2012; Röhrich et al., 2011), the *in vivo* biosynthetic pathway remains elusive.

To construct a putative harmonine biosynthesis pathway in *H. axyridis*, six related KEGG pathways were analysed: biosynthesis of unsaturated fatty acids (map01040), fatty acid metabolism (map01212), biosynthesis of unsaturated fatty acids (map01040), cutin, suberin and wax biosynthesis (map00073), fatty acid elongation (map00062), and C16, C18 and all related enzymes involved in the longevity regulating pathway (map04212). Based on these analyses, we proposed a putative pathway for harmonine biosynthesis, which integrates most of the synthetic steps of harmonine (Figure 5). All KO enzymes was searched by ORTHOMCL and BLASTP in the protein coding genes of *H. axyridis* (E-value <1e-5) (Table 3). The transcript sequences of these genes were validated by PCR (Figure 6a). Analysis of the expression levels using two larvae, one pupa and 23 adult transcriptome data sets (Table S1) showed that these genes tend to be highly expressed during the larvae, pupa and most adult stages (Figure 6b), which is consistent with PCR results. Though further validation is required, this putative pathway integrates most steps of harmonine biosynthesis, which is a useful resource for elucidating the biosynthesis mechanism of harmonine.

4 | CONCLUSIONS

We provide a chromosome-level genome of *Harmonia axyridis* based on the Illumina and PacBio sequencing platforms and 10X Genomics and Hi-C technologies. The X chromosome and Y-linked scaffolds were identified in *H. axyridis*. Genome synteny analysis showed that frequent chromosome fusion events happened in the *H. axyridis* lineage. Surprisingly, there was no apparent expansion in the OR and P450 gene families. A putative *in vivo* harmonine biosynthetic pathway was constructed, and most of genes in this pathway were highly expressed in almost all tested samples, suggesting that harmonine might contribute to the high adaptability and successful invasiveness of *H. axyridis* in North America and Europe.

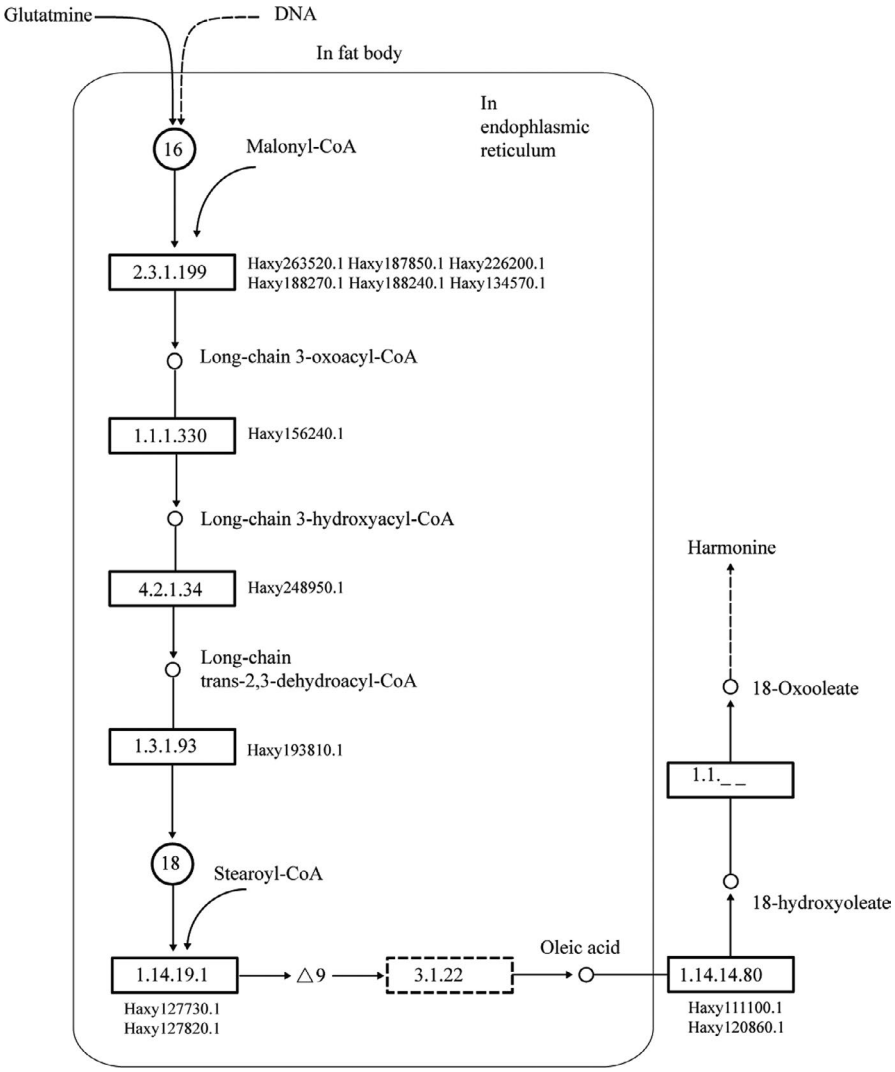


FIGURE 5 Schematic of the harmonine biosynthetic pathway in *Harmonia axyridis*. Dashed lines: unclear/unreported processes and biological substrates. 16 and 18: C16 (palmitic acid) and C18 (stearic and oleic acids), respectively. △: desaturase. Corresponding *H. axyridis* transcripts are listed next to each enzyme (Table 4)

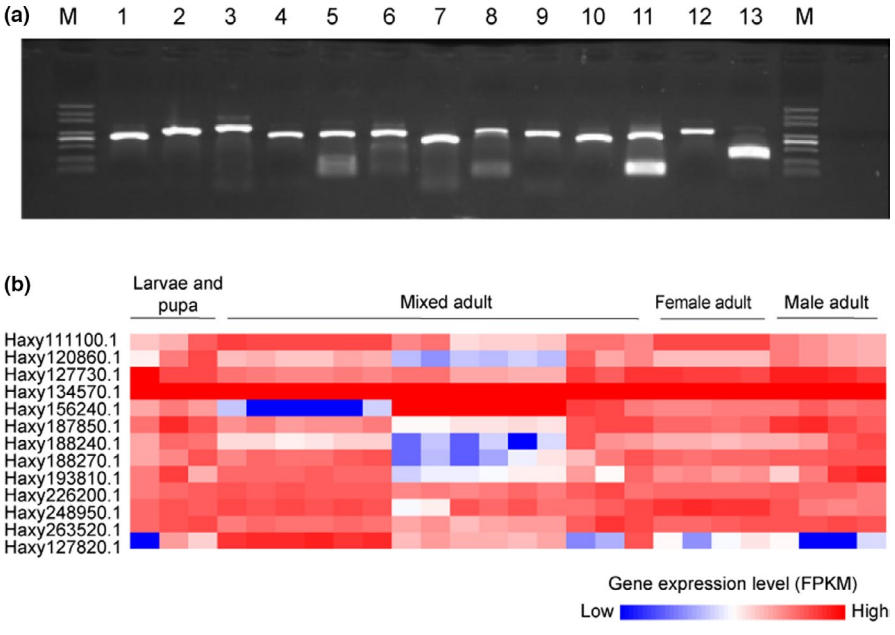


FIGURE 6 PCR validation and expression profiles of *Harmonia axyridis* harmonine biosynthesis candidate genes. (a) PCR validation of harmonine biosynthesis genes in *H. axyridis*. M: marker (5000 bp); 1–13: Haxy193810.1, Haxy127730.1, Haxy127820.1, Haxy263520.1, Haxy134570.1, Haxy187850.1, Haxy226200.1, Haxy188270.1, Haxy188240.1, Haxy156240.1, Haxy248950.1, Haxy111100.1 and Haxy120860.1. (b) The genes are highly expressed in most examined samples. Gene functions are listed in Table 4. Transcriptome information is shown in Table S1

TABLE 4 Summary of predicted *Harmonia axyridis* genes engaged in harmonine biosynthetic pathway

Enzymes	KEGG orthologues	*H.axyrids* transcripts	Locations in *H.axyrids* genome
EC1.3.1.93 (very-long-chain enoyl-CoA reductase)	K10258	Haxy193810.1	fragScaff_scaffold_198
EC1.14.19.1 (stearoyl-CoA desaturase)	K00507	Haxy127730.1 Haxy127820.1	fragScaff_scaffold_160
EC3.1.2.2 (long-chain fatty-acyl-CoA hydrolase)	K01068 K00659 K10804 K16339 K22554 K17360	n/a	n/a
EC2.3.1.199 (elongation of very long chain fatty acids protein 6)	K10203 K10205 K10244 K10245 K10246 K10247 K10248 K10249 K10250 K15397	Haxy263520.1 Haxy187850.1 Haxy226200.1 Haxy188270.1 Haxy188240.1 Haxy134570.1	original_scaffold_302 fragScaff_scaffold_210 original_scaffold_851 original_scaffold_1665
EC1.1.1.330 (17beta-estradiol 17-dehydrogenase/very-long-chain 3-oxoacyl-CoA reductase)	K10251	Haxy156240.1	fragScaff_scaffold_202
EC4.2.1.134 (very-long-chain (3*R*)−3-hydroxyacyl-CoA dehydratase)	K10703	Haxy248950.1	fragScaff_scaffold_109
EC1.14.14.80 (long-chain fatty acid omega-monooxygenase)	K07425 K15401 K20495 K22887	Haxy111100.1 Haxy120860.1	fragScaff_scaffold_237 fragScaff_scaffold_81

Abbreviation: n/a, Not applicable.

ACKNOWLEDGEMENTS

We thank Dr James R. Walters from the University of Kansas for providing the script for sex chromosome identification and useful suggestions. The authors thank Dr Xiao Huamei, Ye Xinhai, Xu Le, Zhao Xianxin and Li Meizhen for their helpful suggestions and Annoroad Gene Technology Co. Ltd for their support with Hi-C technology.

CONFLICT OF INTEREST

The authors declare no competing interests.

AUTHOR CONTRIBUTIONS

F. L. designed the experiments. S. W. and F. Z. participated in discussions and provided the beetle samples. M. C., D. X. and X. C. collected the samples and prepared the samples for sequencing. Y. M., M. C. and X. C. analysed the data. M. C., K. H., X. C., Q. L. and M. W. performed the experiments. M. C. and F. L. wrote the manuscript. F. L. revised the manuscript. All authors read and approved the final manuscript.

DATA AVAILABILITY STATEMENT

The genome assembly has been deposited in NCBI under BioProject accession no. PRJNA579455 and BioSample accession no. SAMN13111588. The Whole Genome Shotgun project data have been deposited in DDBJ/ENA/GenBank under accession no. WMPA00000000. The version described in this paper is No. WMPA01000000. The chromosome-level assembly and scripts presented in this paper and updated data can be accessed at InsectBase (http://insect-genome.com/Haxy/).

ORCID

Fei Li https://orcid.org/0000-0002-8410-5250

REFERENCES

Ando, T., Matsuda, T., Goto, K., Hara, K., Ito, A., Hirata, J., Yatomi, J., Kajitani, R., Okuno, M., Yamaguchi, K., Kobayashi, M., Takano, T., Minakuchi, Y., Seki, M., Suzuki, Y., Yano, K., Itoh, T., Shigenobu, S., Toyoda, A., & Niimi, T. (2018). Repeated inversions within a pannier intron drive diversification of intraspecific colour patterns of ladybird beetles. *Nature Communications*, *9*(1), 3843.

Attygalle, A. B., Blankespoor, C. L., Eisner, T., & Meinwald, J. (1994). Biosynthesis of a defensive insect alkaloid: Epilachnene from oleic acid and serine. *Proceedings of the National Academy of Sciences*, *91*(26), 12790–12793.

Birney, E., Clamp, M., & Durbin, R. (2004). GeneWise and genomewise. *Genome Research*, *14*(5), 988–995.

Blackmon, H., & Demuth, J. P. (2015). Coleoptera karyotype database. *The Coleopterists' Bulletin*, *69*, 174–175.

Braconnier, M. F., Braekman, J. C., & Daloze, D. (1985). Synthesis of the racemic form of (Z)-1, 17-diaminooctadec-9-Ene, an aliphatic diamine from coccinellidae, determination of the absolute configuration of the (+)-naturally-occurring antipode. *Bulletin Des Sociétés Chimiques Belges*, *94*(8), 605–613.

Burton, J. N., Adey, A., Patwardhan, R. P., Qiu, R., Kitzman, J. O., & Shendure, J. (2013). Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nature Biotechnology*, *31*(12), 1119.

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973.

Chang, C. H., & Larracuente, A. M. (2019). Heterochromatin-enriched assemblies reveal the sequence and organization of the Drosophila melanogaster Y chromosome. *Genetics*, *211*(1), 333–348.

Chen, X. U., Xiao, D. A., Du, X., Guo, X., Zhang, F., Desneux, N., Zang, L., & Wang, S. U. (2019). The role of the dopamine melanin pathway in the ontogeny of Elytral Melanization in *Harmonia axyridis*. *Frontiers in Physiology*, *10*, 1066.

Chin, C.-S., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Clum, A., Dunn, C., O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., Cramer, G. R., Delledonne, M., Luo, C., Ecker, J. R., Cantu, D., Rank, D. R., & Schatz, M. C. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. *Nature Methods*, *13*(12), 1050.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*, *27*(8), 1164–1165.

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.

Doležel, J., Greilhuber, J., & Suda, J. (2007). Estimation of nuclear DNA content in plants using flow cytometry. *Nature Protocols*, *2*(9), 2233.

Du, W., Chen, I., Ruan, C., & Sun, G. (2010). Investigations of the color variations of *Harmonia axyridis* (Pallas) in Changchun. *Chinese Agricultural Science Bulletin*, *4*, 57.

Enders, D., & Bartzen, D. (1991). Enantioselective total synthesis of harmonine, a defence alkaloid of ladybugs (Coleoptera: Coccinellidae). *Liebigs Annalen Der Chemie*, *1991*(6), 569–574.

Eyun, S.-I., Soh, H. Y., Posavi, M., Munro, J. B., Hughes, D. S. T., Murali, S. C., Qu, J., Dugan, S., Lee, S. L., Chao, H., Dinh, H., Han, Y. I., Doddapaneni, H. V., Worley, K. C., Muzny, D. M., Park, E.-O., Silva, J. C., Gibbs, R. A., Richards, S., & Lee, C. E. (2017). Evolutionary history of chemosensory-related gene families across the arthropoda. *Molecular Biology and Evolution*, *34*(8), 1838–1862. https://doi.org/10.1093/molbev/msx147.

Fallon, T. R., Lower, S. E., Chang, C.-H., Bessho-Uehara, M., Martin, G. J., Bewick, A. J., Behringer, M., Debat, H. J., Wong, I., Day, J. C., Suvorov, A., Silva, C. J., Stanger-Hall, K. F., Hall, D. W., Schmitz, R. J., Nelson, D. R., Lewis, S. M., Shigenobu, S., Bybee, S. M., … Weng, J.-K. (2018). Firefly genomes illuminate parallel origins of bioluminescence in beetles. *elife*, *7*, e36495. https://doi.org/10.7554/eLife.36495

Ferran, A., Niknam, H., Kabiri, F., Picart, J. L., De Herce, C., Brun, J., & Lapchin, L. (1996). The use of *Harmonia axyridis* larvae (Coleoptera: Coccinellidae) against *Macrosiphum rosae* (Hemiptera: Sternorrhyncha: Aphididae) on rose bushes. *European Journal of Entomology*, *93*, 59–68.

Feyereisen, R. (1999). Insect P450 enzymes. *Annual Review of Entomology*, *44*(1), 507–533.

Feyereisen, R. (2012). Insect CYP genes and P450 enzymes. In L. I. Gilbert (Ed.), *Insect molecular biology and biochemistry* (pp. 236–316). Elsevier.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L. L., Tate, J., & Punta, M. (2014). Pfam: the protein families database. *Nucleic Acids Research*, *42*(D1), D222–D230.

Galbraith, D. W., Harkins, K. R., Maddox, J. M., Ayres, N. M., Sharma, D. P., & Firoozabady, E. (1983). Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science*, *220*(4601), 1049–1051.

Gautier, M., Yamaguchi, J., Foucaud, J., Loiseau, A., Ausset, A., Facon, B., Gschloessl, B., Lagnel, J., Loire, E., Parrinello, H., Severac, D., Lopez-Roques, C., Donnadieu, C., Manno, M., Berges, H., Gharbi, K., Lawson-Handley, L., Zang, L.-S., Vogel, H., … Prud'homme, B. (2018). The genomic basis of color pattern polymorphism in the harlequin ladybird. *Current Biology*, *28*(20), 3296–3302.e7.

Gotoh, H., Nishikawa, H., Sahara, K., Yaginuma, T., & Niimi, T. (2015). A new molecular technique for determining the sex of *Harmonia axyridis*. *Journal of Insect Biotechnology and Sericology*, *84*(1), 1_009-001_015.

Grabherr, M. G., Russell, P., Meyer, M., Mauceli, E., Alföldi, J., Di Palma, F., & Lindblad-Toh, K. (2010). Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics*, *26*(9), 1145–1151.

Gremme, G., Brendel, V., Sparks, M. E., & Kurtz, S. (2005). Engineering a software tool for gene structure prediction in higher organisms. *Information Software Technology*, *47*(15), 965–978.

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.

Haulotte, E., Laurent, P., & Braekman, J. C. (2012). Biosynthesis of defensive coccinellidae alkaloids: incorporation of fatty acids in adaline, coccinelline, and harmonine. *European Journal of Organic Chemistry*, *2012*(10), 1907–1912.

He, K., Lin, K., Wang, G., & Li, F. (2016). Genome sizes of nine insect species determined by flow cytometry and k-mer analysis. *Frontiers in Physiology*, *7*, 569.

Hemingway, J. (2000). The molecular basis of two contrasting metabolic mechanisms of insecticide resistance. *Insect Biochemistry and Molecular Biology*, *30*(11), 1009–1015.

Hunt, T., Bergsten, J., Levkanicova, Z., Papadopoulou, A., John, O. S., Wild, R., & Caterino, M. S. (2007). A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. *Science*, *318*(5858), 1913–1916.

Kajita, Y., Obrycki, J., Sloggett, J., & Haynes, K. (2010). Intraspecific alkaloid variation in ladybird eggs and its effects on con-and heterospecific intraguild predators. *Oecologia*, *163*(2), 313–322.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology Evolution*, *30*(4), 772–780.

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

Kim, H. S., Murphy, T., Xia, J., Caragea, D., Park, Y., Beeman, R. W., & Brown, S. J. (2009). BeetleBase in 2010: Revisions to provide

comprehensive genomic information for *Tribolium castaneum*. *Nucleic Acids Research*, *38*(suppl_1), D437–D442.

Koch, R. (2003). The multicolored Asian lady beetle, Harmonia axyridis: a review of its biology, uses in biological control, and non-target impacts. *Journal of Insect Science*, *3*(1), 1–16.

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

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., & Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research*, *19*(9), 1639–1645.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096.

Lander, E. S., & Waterman, M. S. (1988). Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics*, *2*(3), 231–239.

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*(1), 323.

Li, H., Coghlan, A., Ruan, J., Coin, L. J., Heriche, J. K., Osmotherly, L., & Bolund, L. (2006). TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Research*, *34*(suppl_1), D572–D580.

Li, J. (1940). The chromosomes of the ladybird beetle, Hermonia axyridis Pallas. Peking Nat. His. Bull, 15.

Li, L., Stoeckert, C. J., & Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research*, *13*(9), 2178–2189.

Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., & Dorschner, M. O. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, *326*(5950), 289–293.

Liu, J., Xiao, H., Huang, S., & Li, F. (2014). OMIGA: Optimized maker-based insect genome annotation. *Molecular Genetics Genomics*, *289*(4), 567–573.

McCarthy, A. (2010). Third generation DNA sequencing: pacific biosciences' single molecule real time technology. *Chemistry Biology*, *17*(7), 675–676.

Mckenna, D. D., Wild, A. L., Kanda, K., Bellamy, C. L., Beutel, R. G., Caterino, M. S., & Jameson, M. L. (2015). The beetle tree of life reveals that Coleoptera survived end-Permian mass extinction to diversify during the Cretaceous terrestrial revolution. *Systematic Entomology*, *40*(4), 835–880.

Mitchell, R. F., Schneider, T. M., Schwartz, A. M., Andersson, M. N., & McKenna, D. D. (2020). The diversity and evolution of odorant receptors in beetles (Coleoptera). *Insect Molecular Biology*, *29*(1), 77–91.

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.

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, *33*(3), 290–295. https://doi.org/10.1038/nbt.3122.

Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., & Finn, R. D. (2018). HMMER web server: 2018 update. *Nucleic Acids Research*, *46*(W1), W200–W204.

Röhrich, C. R., Ngwa, C. J., Wiesner, J., Schmidtberg, H., Degenkolb, T., Kollewe, C., Fischer, R., Pradel, G., & Vilcinskas, A. (2012). Harmonine, a defence compound from the harlequin ladybird, inhibits mycobacterial growth and demonstrates multi-stage antimalarial activity. *Biology Letters*, *8*(2), 308–311. https://doi.org/10.1098/rsbl.2011.0760.

Roy, H., & Wajnberg, E. (2008). From biological control to invasion: the ladybird *Harmonia axyridis* as a model species. *BioControl*, *53*(1), 1–4.

Scott, J. G. (1999). Cytochromes P450 and insecticide resistance. *Insect Biochemistry and Molecular Biology*, *29*(9), 757–777.

Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C.-J., Vert, J.-P., Heard, E., Dekker, J., & Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biology*, *16*(1), 259.

Shi, J., Ma, X., Zhang, J., Zhou, Y., Liu, M., Huang, L., Sun, S., Zhang, X., Gao, X., Zhan, W., Li, P., Wang, L., Lu, P., Zhao, H., Song, W., & Lai, J. (2019). Chromosome conformation capture resolved near complete genome assembly of broomcorn millet. *Nature Communications*, *10*(1), 464.

Simao, 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.

Slater, G. S., & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, *6*, 31. https://doi.org/10.1186/1471-2105-6-31.

Sloggett, J. J., Magro, A., Verheggen, F. J., Hemptinne, J. L., Hutchison, W. D., & Riddick, E. W. (2011). The chemical ecology of *Harmonia axyridis*. *BioControl*, *56*(4), 643–661.

Smith, S. G. (1953). Chromosome numbers of Coleoptera. *Heredity*, *7*(1), 31.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313.

Stanke, M., Schöffmann, O., Morgenstern, B., & Waack, S. (2006). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics*, *7*(1), 62.

Stuart, J. J., & Mocelin, G. (1995). Cytogenetics of chromosome rearrangements in *Tribolium castaneum*. *Genome*, *38*(4), 673–680. https://doi.org/10.1139/g95-085.

Tang, B., Zhu, J., Guo, H. S., Fang, D., Shen, Q. D., Zheng, X. X., & Wang, S. (2012). Studies of the diversity of multiple elytral color morphs of *Harmonia axyridis* Pallas (Coleoptera: Coccinellidae). *Journal of Hangzhou Normal University (Natural. Science, Edition)*, *11*(2), 132–136.

Tempel, S. (2012). Using and understanding RepeatMasker. In Y. Bigot (ed.), *Mobile genetic elements* (pp. 29–51). Springer.

Toussaint, E. F. A., Seidel, M., Arriaga-varela, E., Hájek, J., Král, D., Sekerka, L., Short, A. E. Z., & Fikáček, M. (2017). The peril of dating beetles. *Systematic Entomology*, *42*(1), 1–10.

Tribolium Genome Sequencing Consortium (2008). The genome of the model beetle and pest *Tribolium castaneum*. *Nature*, *452*(7190), 949.

Van Belleghem, S. M., Vangestel, C., De Wolf, K., De Corte, Z., Möst, M., Rastas, P., De Meester, L., & Hendrickx, F. (2018). Evolution at two time frames: Polymorphisms from an ancient singular divergence event fuel contemporary parallel evolution. *PLoS Genetics*, *14*(11), e1007796.

Vilcinskas, A., Stoecker, K., Schmidtberg, H., Röhrich, C. R., & Vogel, H. (2013). Invasive harlequin ladybird carries biological weapons against native competitors. *Science*, *340*(6134), 862–863.

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, *9*(11), e112963.

Wan, F., Yin, C., Tang, R., Chen, M., Wu, Q., Huang, C., Qian, W., Rota-Stabelli, O., Yang, N., Wang, S., Wang, G., Zhang, G., Guo, J., Gu,

L. A., Chen, L., Xing, L., Xi, Y. U., Liu, F., Lin, K., … Li, F. (2019). A chromosome-level genome assembly of Cydia pomonella provides insights into chemical ecology and insecticide resistance. *Nature Communications*, *10*(1), 1–14.

Wasserman, M., & Ehrman, L. (1986). Firefly chromosomes, II. (Lampyridae: Coleoptera). *The Florida Entomologist*, *69*(4), 755–757.

Wilson, T. G., & Ashok, M. (1998). Insecticide resistance resulting from an absence of target-site gene product. *Proceedings of the National Academy of Sciences*, *95*(24), 14040–14044.

Zhang, B., Zhang, W., Nie, R. E., Li, W. Z., Segraves, K. A., Yang, X. K., & Xue, H. J. (2016). Comparative transcriptome analysis of chemosensory genes in two sister leaf beetles provides insights into chemosensory speciation. *Insect Biochemistry and Molecular Biology*, *79*, 108–118.

Zhu, F., Moural, T. W., Shah, K., & Palli, S. R. (2013). Integrated analysis of cytochrome P450 gene superfamily in the red flour beetle, *Tribolium castaneum*. *BMC Genomics*, *14*(1), 174.

SUPPORTING INFORMATION

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

How to cite this article: Chen M, Mei Y, Chen X, et al. A chromosome-level assembly of the harlequin ladybird *Harmonia axyridis* as a genomic resource to study beetle and invasion biology. *Mol Ecol Resour.* 2021;21:1318–1332. https://doi.org/10.1111/1755-0998.13342