DNA Research, 2023, **30**, 1–11
https://doi.org/10.1093/dnares/dsad001
Advance access publication 16 January 2023
Resource Article: Genomes Explored

Resource Article: Genomes Explored

A chromosome-level genome assembly of the *Henosepilachna vigintioctomaculata* provides insights into the evolution of ladybird beetles

Wenbo Zhu[1,†], Shengqi Chi[2,†], Yanchun Wang[3,†], Haorong Li[4,†], Zhongkai Wang[4], Songdong Gu[2], Ting Sun[1], Hui Xiang[5], Ping You[1,*], and Yandong Ren[1,*]

[1]College of Life Sciences, Shaanxi Normal University, Xi'an 710062, China,
[2]Key Laboratory of Integrated Crop Pest Management of Shandong Province, College of Plant Health and Medicine, Qingdao Agricultural University, Qingdao 266109, China,
[3]College of Science and Information, Qingdao Agricultural University, Qingdao 266109, China,
[4]School of Ecology and Environment, Northwestern Polytechnical University, Xi'an 710072, China, and
[5]Guangdong Provincial Key Laboratory of Insect Developmental Biology and Applied Technology, Institute of Insect Science and Technology, School of Life Sciences, South China Normal University, Guangzhou 510631, China

*To whom correspondence should be addressed. Tel. +86-029-85310266; Email: renyandong90@126.com (Y.R.); Tel. +86-029-85310266; Email: youping@snnu.edu.cn (P.Y.)
†These authors contributed equally to this work.

Abstract

The ladybird beetle *Henosepilachna vigintioctomaculata* is an economically significant oligophagous pest that induces damage to many Solanaceae crops. An increasing number of studies have examined the population and phenotype diversity of ladybird beetles. However, few comparative genome analyses of ladybird beetle species have been conducted. Here, we obtained a high-quality chromosome-level genome assembly of *H. vigintioctomaculata* using various sequencing technologies, and the chromosome-level genome assembly was ~581.63 Mb, with 11 chromosomes successfully assembled. The phylogenetic analysis showed that *H. vigintioctomaculata* is a more ancient lineage than the other three sequenced ladybird beetles, *Harmonia axyridis*, *Propylea japonica*, and *Coccinella septempunctata*. We also compared positively selected genes (PSGs), transposable elements (TEs) ratios and insertion times, and key gene families associated with environmental adaptation among these ladybird beetles. The pattern of TEs evolution of *H. vigintioctomaculata* differs from the other three ladybird beetles. The PSGs were associated with ladybird beetles development. However, the key gene families associated with environmental adaptation in ladybird beetles varied. Overall, the high-quality draft genome sequence of *H. vigintioctomaculata* provides a useful resource for studies of beetle biology, especially for the invasive biology of ladybird beetles.

Key words: Henosepilachna vigintioctomaculata, sex chromosome, transposable elements, positively selected genes

1. Introduction

The ladybird beetles (Insecta: Coleoptera: Coccinellidae) are distributed across the globe and comprise more than 6,000 species, belonging to 360 genera.[1] Their wide distribution and abundant diversity of likely stem from their widely environmental adaptability.[1] Most ladybird beetles are well-known predators of crop pests,[2,3] and some are harmful pests.[4,5] The 28-spotted ladybird beetle, *Henosepilachna vigintioctomaculata*, is an oligophagous pest that can induce damage to Solanaceae crops.[6] It feeds exclusively on mesophyll (i.e. potato, tomato, and eggplant leaves), which leading to the remaining veins and epidermis of the leaves to take on a russet browning appearance, retarding crop growth, decreasing fruit production, and eventually reducing farmer's income.[4] The genomes of several ladybird beetles, *Coccinella septempunctata*,[7] *Propylea japonica*,[8] *Harmonia axyridis*,[9–12] *Adalia bipunctata* (unpublished, https://www.ncbi.nlm.nih.gov/assembly/GCA_910592335.1), *Halyzia sedecimguttata* (unpublished, https://www.ncbi.nlm.nih.gov/assembly/GCA_937662695.1), and *Cryptolaemus montrouzieri*[13] have been sequenced to date. Previous studies supplied five high-quality chromosome-level genomes. However, the comparative genomic analyses are still lacking. The sex chromosomes of ladybird beetles are still poorly studied, despite that the sex-determining system has been elucidated and genome sequences have been sequenced in *H. axyridis*[9,14,15] and *C. septempunctata*.[7]

Due to the critical roles of ladybird beetles in the agricultural environment, many studies focussed on tolerance to insecticides and high temperature. Ladybird beetles showed great variation in resistance to insecticides and high temperature. For instance, many insecticides could entail both acute and sublethal toxic effects for *C. septempunctata*, such

Received 15 July 2022; Revised 9 January 2023; Accepted 13 January 2023

as thiamethoxam,[16] tolfenpyrad,[17] imidacloprid,[18] anthranilic diamide insecticides,[19] pyriproxifen, deltamethrin + heptenophos,[20] and lambda-cyhalothrin.[20] Its ability to resist insecticides is quite weak and its ability to resist other insecticides has not been reported. *Harmonia axyridis* also has weak ability to resist insecticides. Many insecticides could affect its physiological or behavioural on the orientation, predation, longevity, reproduction, such as imidacloprid,[21,22] chlorantraniliprole,[23] Beta-cypermethrin,[24] chlorpyrifos,[25] and thiamethoxam.[26] However, *H. axyridis* has a strong ability to resist emamectin benzoate and two fungicides (tebuconazole and myclobutanil).[26] Previous study showed that even temporary high temperature can affect its survival, body size, and colouration.[27] As to *H. vigintioctomaculata*, only a few studies showed that *Bacillus thuringiensis* Cry7Ab3 toxin has insecticidal activity against its larvae,[28] and its tolerance to other insecticides and high temperatures is not tested. Contrastly, *P. japonica* has a strong ability to resist high temperatures and insecticides, due to expansion of cytochrome P450 and *Hsp70* genes.[8,29] *Propylea japonica* have developed resistance to high dosage of insecticides, such as imidacloprid, methamidophos, fenvalerate, avermectin, paraoxon, malaoxon, mathamidophos, Beta-cypermethrin, and chlorpyrifos.[30–32] Previous studies have reported that the cytochrome P450 and *Hsp* (heat shock protein) genes are involved in the detoxification of pesticides and heat stress tolerance and might be related to the invasiveness of some insects, especially in some pests.[8,9,33,34] The P450 gene family is a diverse class of enzymes found in virtually all insect tissues, which is involved in the growth, development, and reproduction of insects.[35,36] The functional diversity of P450 genes is crucial to the adaptation of insect herbivores to their host plants and the detoxification of xenobiotics,[35,36] which facilitates their ability to invade new environments. *Hsps* are important molecular chaperones that regulate stress tolerance and lifespan.[37,38] Previous study showed that *Hsps* are also involved in thermal adaptation and resistance to some proteotoxic stresses.[39] One of the most widely known *Hsp* genes is *Hsp70*, which encodes one of the most well-studied stress proteins in insects.[37,40] Although the variable resistance of different ladybird beetles to insecticides and high temperature remain unclear, P450 and *Hsp* genes may play critical roles in the resistance to insecticides and high temperature.

Here, we used a hybrid sequencing approach combining Illumina short reads, Nanopore long reads, and Hi-C scaffolding to generate a chromosome-level genome of *H. vigintioctomaculata*. Using Hi-C scaffolding, we assigned 98.16% of the bases to 11 chromosomes and the N50 is ~56.17 Mb. We used all sequenced ladybird beetles and other related species in this study to better understand biological features of ladybird beetles. In the comparative genome analysis, we investigated the transposable elements (TEs) insertion pattern, sex chromosome evolution, phylogenetic relationships, positively selected genes (PSGs), and genes associated with environmental adaption in ladybird beetles.

2. Materials and methods

2.1. Sampling and genome sequencing

Four male 28-spotted ladybird beetle (*H. vigintioctomaculata*) adults were collected from cultivated potato fields in Yantai City, Shandong Province, China. To avoid contamination of the genome samples from other organisms, such as microbes and food sources, *H. vigintioctomaculata* samples were rinsed with distilled water for 2 min, and midguts were removed. To obtain sufficient high-quality DNA for the Nanopore platform (Oxford Nanopore, Oxford, UK) and Illumina platform (San Diego, CA, USA), genomic DNA was extracted from four ladybirds using a Qiagen Blood & Cell Culture DNA Mini Kit. RNA from muscle tissue was extracted using Trizol (Invitrogen) according to the manufacturer's instructions. To obtain an overview of the transcriptome, polyadenylated RNA was selected by oligo (dT) purification, reverse-transcribed to cDNA, and sequenced using the Illumina sequencing platform.

Raw reads were run through strict quality control before the further assembly and other analyses. For Nanopore long reads, any reads less than 1 kb or with mean quality values less than 7 were removed. For Illumina short reads, any reads with more than 10% unknown reads or 50% low-quality bases along with their paired-end reads were removed. All the adaptor sequences and duplicated reads produced by PCR step were removed.

2.2. K-mer analysis estimation of genome size

The k-mer distribution analysis was performed to estimate the genome size of *H. vigintioctomaculata*. All the filtered Illumina short reads were used to estimate genome size. 17-mer was used for k-mer analysis, and the 17-mer depth frequency distribution was calculated using GCE (v2).[41] Genome size was estimated using the following formula: Genome size = TKN17-mer / PKFD17-mer, where TKN17-mer is the total 17-mer number and PKFD17-mer is the peak 17-mer frequency depth. The estimated genome size was used in subsequent genome assembly analyses.

2.3. Genome assembly and correction

To ensure the high quality of the assembled genome and reduce the error ratio from sequencing, we used a hierarchical approach by combining Nanopore long reads, Illumina short-insert reads, and Hi-C reads. First, all Nanopore long reads were used as input for NextDenovo (v2.4; https://github.com/Nextomics/NextDenovo) with default parameters. Second, Illumina short-insert reads were used to polish the genome using NextPolish (v1.4.0) with default parameters. Third, after the error-correction steps, the Hi-C data were used to construct a chromosome-level genome assembly. All the Hi-C sequencing data were mapped to the polished genome to cluster, order, and orient the Contigs to pseudo-chromosomes. 3D *de novo* assembly (3D-DNA) software (v180419)[42] with default parameters was used for this step. The pseudo-chromosomes were cut into 100-kb bins, and the heatmap was constructed based on the interaction signals between bins.[43]

2.4. Assessment of genome assembly quality

To evaluate the completeness and accuracy of the assembled genome, all the filtered short reads generated were mapped to the assembled genome using BWA software (v0.7.12)[44] to detect the genome integrity. The assembled transcripts were mapped to the assembled genome using BLAT software (V35).[45] We evaluated the quality of the genome using Benchmarking Universal Single-Copy Orthologs using BUSCO software (V5.2.2)[46] with the core gene set of the eukaryote and metazoan databases. Genome synteny between *H. vigintioctomaculata* and the two ladybird beetles was

investigated using LAST software (version 802)[47] and plotted using CIRCOS (v0.69-9).[48]

2.5. Repetitive elements annotation

Tandem repeats in the genome were analysed using Tandem Repeat Finder (v4.09)[49] with default parameters. To identify the TEs in the assembled genome, we conducted annotations at both the protein and DNA levels. At the protein level, RepeatProteinMask (RM-BLASTX) was used to search TEs in the protein database. At the DNA level, RepeatModeler (v-1.0.11) (http://www.repeatmasker.org/RepeatModeler/) was used to construct the *de novo* repeat library, and RepeatMasker (v-4.1.0)[50] was used to produce a homologue-based repeat library with default parameters. The insertion time of each TE sequence was estimated using a Kimura distance-based analysis[51] with the parseRM pipeline (https://github.com/4ureliek/Parsing-RepeatMasker-Outputs).

2.6. Gene prediction and function annotation of protein-coding genes

Repeat sequences were masked, and the repeat-masked genome was used for gene set annotation with three methods: *ab initio* annotation, RNA-seq-based annotation, and homologue-based annotation.

In the *ab initio* method, the software packages Augustus (v2.5.5) and SNAP (v2006-07-28) were employed with default settings. Genes with incomplete open reading frames (ORFs) or a protein-coding length less than 300 bp were filtered out. In the RNA-seq-based method, published gene sets from *Aedes aegypti* (GCF_002204515.2; from NCBI), *Spodoptera frugiperda* (GCF_011064685.1; from NCBI), *P. japonica* (GCA_013421045.1; from NCBI), *Anoplophora glabripennis* (GCF_000390285.2; from NCBI), *Drosophila melanogaster* (GCF_000001215.4; from NCBI), *Bombyx mori* (GCF_014905235.1; from

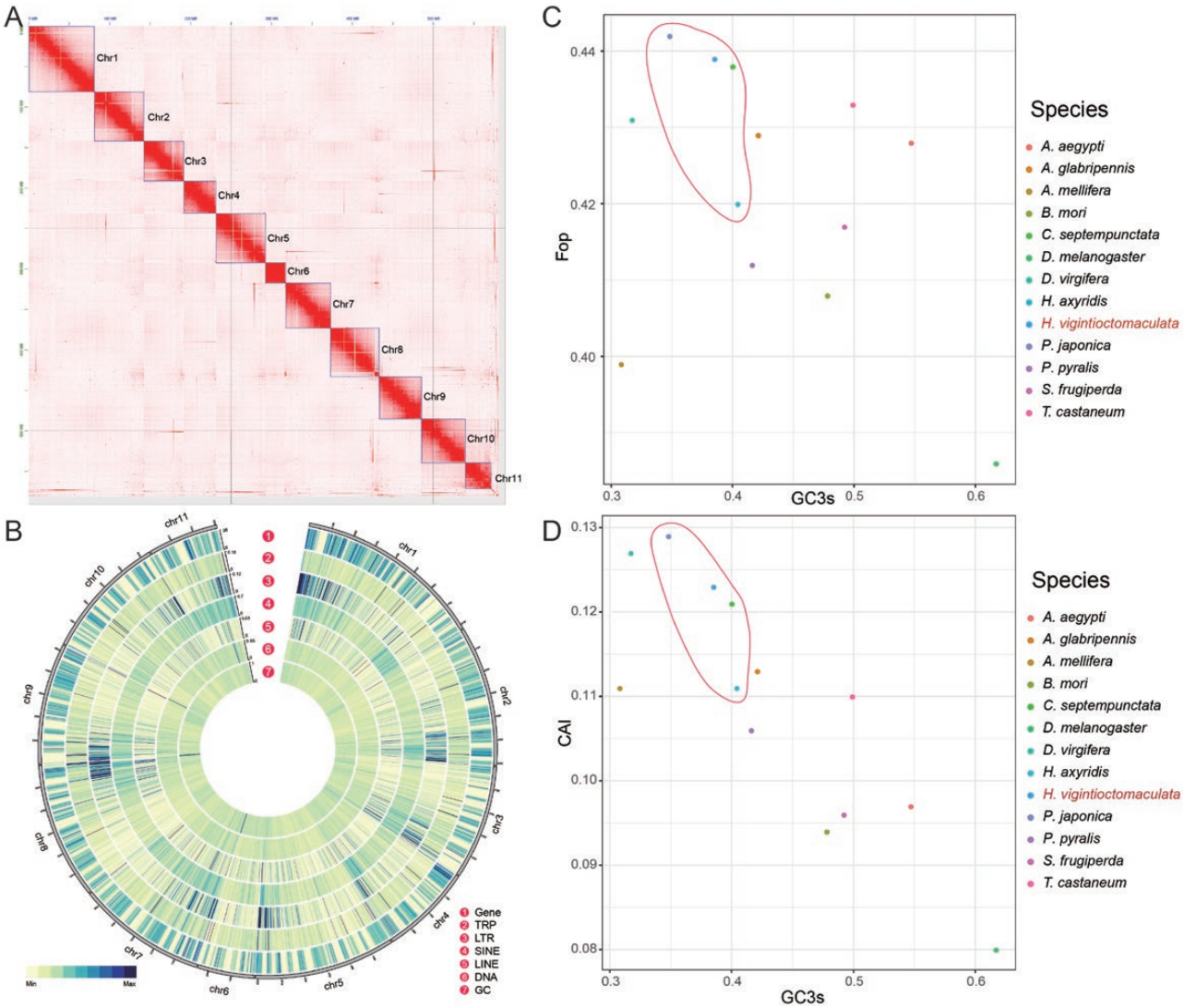

Figure 1. Genome assembly and comparative analysis of *H. vigintioctomaculata*. (A) Heatmap of chromosome interactions in *H. vigintioctomaculata*. (B) Circos plot of distribution of genomic elements in *H. vigintioctomaculata*. From the outer to the inner ring are the distributions of genes density, tandem repeats (TRP), long tandem repeats (LTR), short interspersed nuclear elements (SINE), long interspersed nuclear elements (LINE), DNA elements, and GC content. (C) Codon usage bias. Values of the frequency of optimal codon (FOP) on the frequency of guanine + cytosine at the synonymous third position of codons (GC3s) were determined using the nucleotide sequences of all predicted genes concatenated for individual species. (D) Codon usage bias. Values of the codon adaption index (CAI) on the frequency of guanine + cytosine at the synonymous third position of codons (GC3s) were determined using the nucleotide sequences of all predicted genes concatenated for individual species.

NCBI), *Photinus pyralis* (GCF_008802855.1; from NCBI), *Diabrotica virgifera* (GCF_003013835.1; from NCBI), *Apis mellifera* (GCF_003254395.2; from NCBI), *Tribolium castaneum* (GCF_000002335.3; from NCBI), and *H. axyridis* (GCA_011033045.1; from NCBI) were downloaded from NCBI or their own databases and used for homology-based annotation. The longest transcript of each protein-coding gene was aligned to the *H. vigintioctomaculata* genome using BLAST (tblastn, v2.6.0) with an *e*-value of 1×10^{-5}, and gene structures were predicted using GeneWise (v2.2.0).[52] In the RNA-seq-based gene approach, the *de novo* assembled transcripts were aligned to the *H. vigintioctomaculata* genome using BLAT (v35),[45] and PASA (v2.1.0)[53] was used to link the spliced alignments. Finally, the gene prediction results were integrated into a final gene set using EVidenceModeler (v1.1.1) software.[54]

For gene function annotation, the predicted protein-coding genes were searched against the following public databases: Gene Ontology (http://geneontology.org/), the Integrated Resource of Protein Domains and Functional Sites (InterPro: https://www.ebi.ac.uk/interpro/), Kyoto Encyclopedia of Genes and Genomes (KEGG: https://www.kegg.jp/), Clusters of Orthologous Groups of proteins (COG: https://www.ncbi.nlm.nih.gov/COG/), Swiss-Prot (www.uniprot.org), TrEMBL (www.uniprot.org), and NCBI non-redundant proteins database (NR: https://ftp.ncbi.nlm.nih.gov/blast/db).

2.7. Species phylogeny analysis

A total of 13 insects were used to construct a phylogenetic tree. *Harmonia axyridis*, *P. japonica*, *T. castaneum*, *D. virgifera*, *A. glabripennis*, *S. frugiperda*, *A. aegypti*, *B. mori*, *P. pyralis*, *D. melanogaster*, *A. mellifera*, *C. septempunctata*, and *H. vigintioctomaculata* were all used for the identification of orthologous genes with the OrthoMCL pipeline (v2.0.9)[55] using the default parameters. The single-copy genes of all these 13 species obtained from OrthoMCL were clustered for phylogenetic tree construction and divergence time estimation.

2.8. Phylogenetic analysis and divergence time estimation

A total of 418 single-copy genes were used to reconstruct the phylogenetic tree. The single-copy genes were extracted and aligned using MUSCLE software (v3.8.31), and maximum likelihood-based phylogenetic analysis was conducted using RAxML (v8.2.10)[56] with default parameters. To estimate the accuracy of the divergence time, the fossil records of Coccinellidae species were downloaded from previous study[57] and the fossil records of other species were downloaded from TIMETREE website (http://www.timetree.org) for calibration. The divergence time was estimated using the MCMCtree program in the PAML package (v4.8)[58] based on the topology of the above phylogenetic tree.

2.9. Analysis of the rate of molecular evolution

The well-aligned concatenated protein sequences of these single-copy orthologues were used to determine the rate of molecular evolution via two different methods: Tajima's relative rate test and two cluster analysis. *Apis mellifera* was specified as the outgroup species and tested the relative rate of evolution between *H. vigintioctomaculata* and other insects for both methods. Tajima's relative rate test was performed using MEGA software (V10).[59] A Chi-square test was used to identify species with faster rates of evolution compared with other species. Two cluster analysis was conducted in LINTRE software (V1) via the TPCV model.[60] A faster or slower rate of evolution of particular species was examined using *Z*-statistics.

2.10. Gene family expansion and contraction

Based on the gene families identified by OrthoMCL, we analysed the expansion and contraction of gene families in each

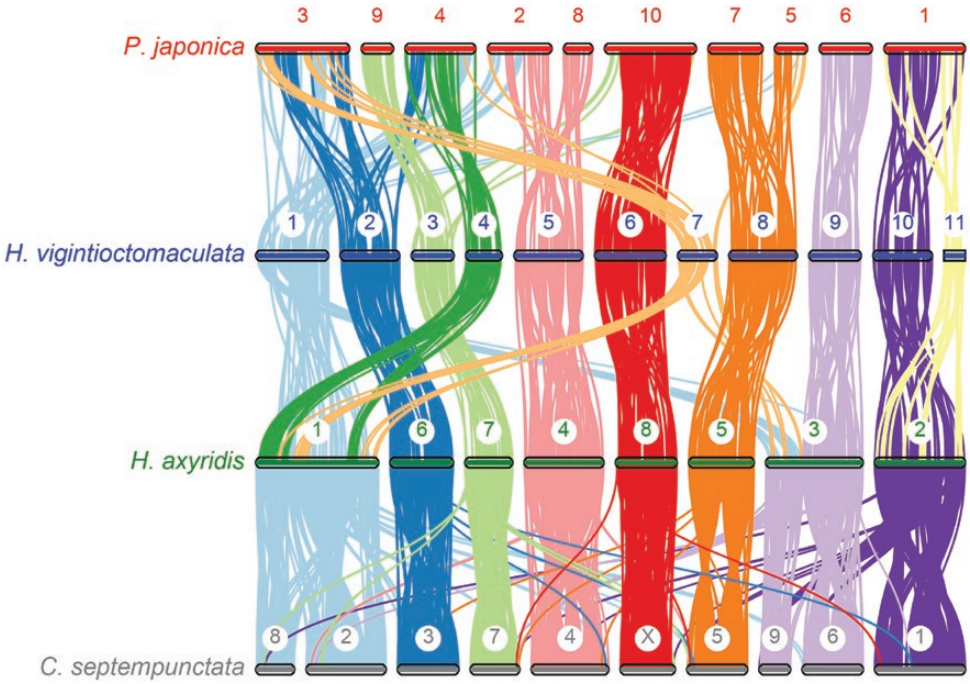

Figure 2. Comparative analysis of synteny among *H. vigintioctomaculata*, *P. japonica*, *H. Axyridis*, and *C. septempunctata*.

node of the phylogenetic tree using CAFÉ (v3.1).[61] Genes in expanded or contracted families in ladybird beetles were subject to GO/KEGG enrichment analysis. GO enrichment analysis was performed using the EnrichGO package in R (3.2.5), and KEGG analysis was performed using an R script.[62,63]

2.11. Positive selection analysis

All single-copy orthologues identified among the 13 species were extracted and aligned using MUSCLE software (v3.8.31) with default parameters. The branch model in the codeml tool of the PAML package (v4.8)[64] was used to identify genes that

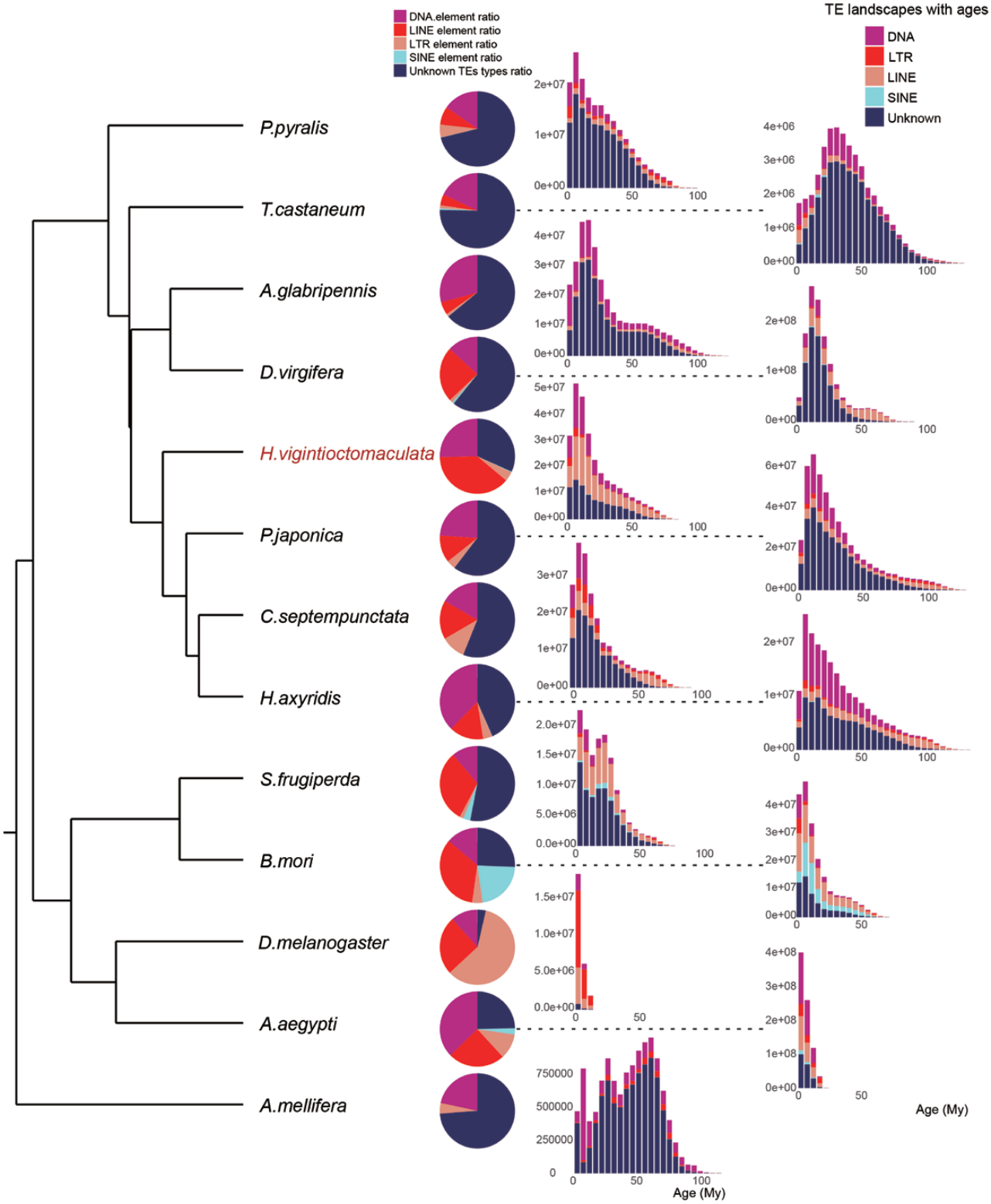

Figure 3. Comparison of the insertion history of TEs among species. The *x*-axis indicates the inferred insertion time (unit: million years ago) of TEs in the genome. The *y*-axis indicates the total length of TEs and the length of each TE in each species.

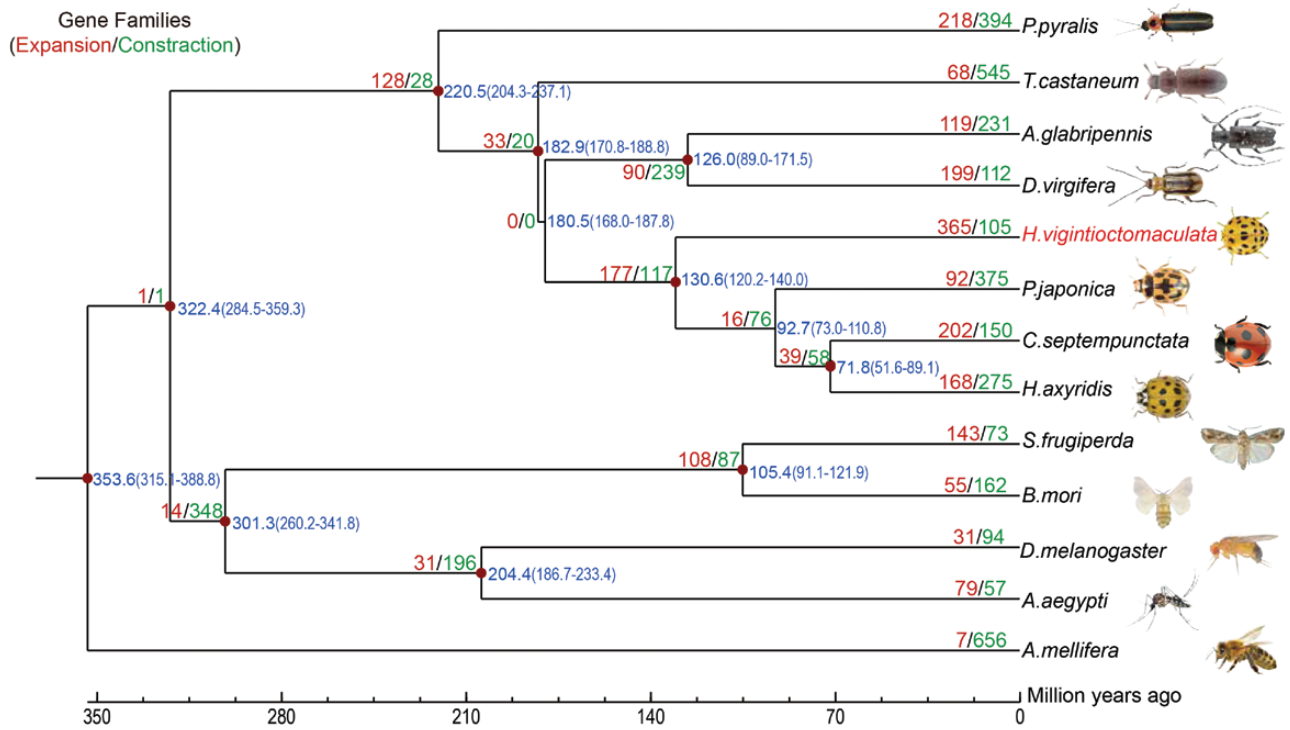

Figure 4. Phylogenetic relationships among these species. Red dots at nodes indicate that fossil records were used to calibrate divergence times. Blue numbers in each node indicate the divergence time and the 95% mean confidence interval (CI). Red and green numbers in each node/species indicate expanded/contracted gene families, respectively.

had experienced positive selection or rapid evolution. Briefly, the rate ratio (ω) of non-synonymous to synonymous nucleotide substitutions was estimated. The one-ratio model was used to detect the average ω across the species tree (ω0). For each gene, the two-ratio branch model was used to detect the ω of the appointed branch (the lineage of Coccinellidae) to test the (ω1) and ω of all other branches (ω_background). A likelihood ratio test was performed to compare the fit of the two-ratio model with the one-ratio model to determine whether the gene was positively selected in the appointed branch (ω1 > ω0; ω1 > ω_background; *P*-value <0.05).

3. Results and discussion

3.1. Chromosome-level genome assembly of *H. vigintioctomaculata*

We used multiple whole-genome sequencing strategies to assemble the genome of *H. vigintioctomaculata* (Supplementary Tables S1–S3). We first estimated the genome size of *H. vigintioctomaculata* using k-mer analysis (K = 17), which was estimated to be ~656.74 Mb (Supplementary Fig. S1). Next, we used Nanopore long reads and Illumina short reads to generate a Contig-level draft genome assembly of *H. vigintioctomaculata*. A 581.59-Mb draft genome assembly was obtained, yielding 493 Contigs with an N50 of ~3.51 Mb (Supplementary Table S3). The Hi-C reads were used to anchor, order, and orient 493 Contigs into 11 chromosomes (Fig. 1A, Supplementary Tables S3–S5). The Hi-C linking information indicated that more than 98.16% of the assembled bases were anchored to the chromosomes (Supplementary Table S5).

The completeness of the genome was evaluated by mapping the assembled transcripts and the Illumina short reads to the *H. vigintioctomaculata* genome and using BUSCO (v5.2.2). Approximately 94.91% of all assembled transcripts (total of 68,304) could be mapped to the assembled genome (Supplementary Tables S6–S8). A total of approximately 99.21% of reads and 93.98% of PE reads could be mapped to the assembled genome (Supplementary Table S9). The BUSCO analysis revealed that 98.1% of the 255 expected conserved core genes in the eukaryote database were identified as complete, and 97.7% of the 954 expected conserved core genes in the Metazoa database were identified as complete (Supplementary Table S10). The completeness of the *H. vigintioctomaculata* genome (98.1%) was comparable to that of genomes of other related insects (88.2–100.0%) (Supplementary Table S11). In Coccinellidae, the BUSCO analysis of *H. vigintioctomaculata* genome (98.1%) showed comparable gene-completeness to that of other high-quality Coccinellidae species (94.9–100.0%). Overall, the results of the above analyses indicated that the genome of *H. vigintioctomaculata* was of high quality and complete.

3.2. Genome annotation and synteny analyses

A total of 318,293,454 bp sequences occupying ~54.72% of the assembled genome were identified as repeat sequences (Supplementary Table S12). Among these repeats, LINE sequences (21.02%) were the most abundant repeat element, with 16.36% of these sequences not be classified into any known repeat elements (Supplementary Fig. S2; Supplementary Table S13). We used three methods to annotate protein-coding genes in the *H. vigintioctomaculata* genome: *ab initio* annotation, RNA-seq-based annotation, and homologue-based annotation. A total of 20,304 protein-coding genes were predicted in the *H. vigintioctomaculata* genome, which were comparable to those reported in other

species (Supplementary Fig. S3). Distributions of the genomic elements, such as gene density, tandem repeats, LINEs, SINEs, LTRs, and GC content, are shown in Fig. 1B. The codon usage of ladybird beetles slightly varied compared with other insects (Fig. 1C and D).

Previous studies showed that *H. axyridis* has an XY sex system, and the X chromosome has been identified.[9,14,15] Synteny analyses between *H. axyridis*, *T. castaneum*, and *P. pyralis* showed that ChrX of *H. axyridis* was aligned to ChrX of *T. castaneum* and Chr3a (X) of *P. pyralis*, respectively. Synteny analyses between *H. vigintioctomaculata*, *T. castaneum*, and *P. pyralis* showed that the candidate sex chromosome of *H. vigintioctomaculata* and *P. japonica* aligned to ChrX of *T. castaneum* and Chr3a (X) of *P. pyralis* (Supplementary Fig. S4), which was consistent with the above results. We further performed synteny analyses of all ladybird beetles and found that the ChrX in *H. axyridis* aligned to chromosome 6 in *H. vigintioctomaculata,* chromosome 10 in *P. japonica*, and X chromosome in *C. septempunctata* (Fig. 2). All the synteny analyses showed that the chromosome 6 in *H. vigintioctomaculata* might be X chromosome, BAC-FISH and genomic qPCR analysis are still needed.

3.3. TE evolution in ladybird beetles

We examined the types of TEs and the TEs insertion times of all 13 species and found that the TEs insertion time was similar in all ladybird beetles. TE insertion time was concentrated at approximately 2–4 million years ago. The TE insertion time of all ladybird beetles was much more recent than their divergence time, which suggests that TEs insertion events might have taken place after they diverged (Fig. 3). However, the LINE elements ratio of *H. vigintioctomaculata* (19.0640%) was much higher than that of the other three ladybird beetles (*H. axyridis*: 7.7098%; *P. japonica*: 6.9492%; *C. septempunctata*: 9.1957%) and other insects (from 0.003396% to 15.5970%) (Supplementary Table S14). LINE elements insertion pattern in *H. vigintioctomaculata* was quite different from that in other three ladybird beetles,

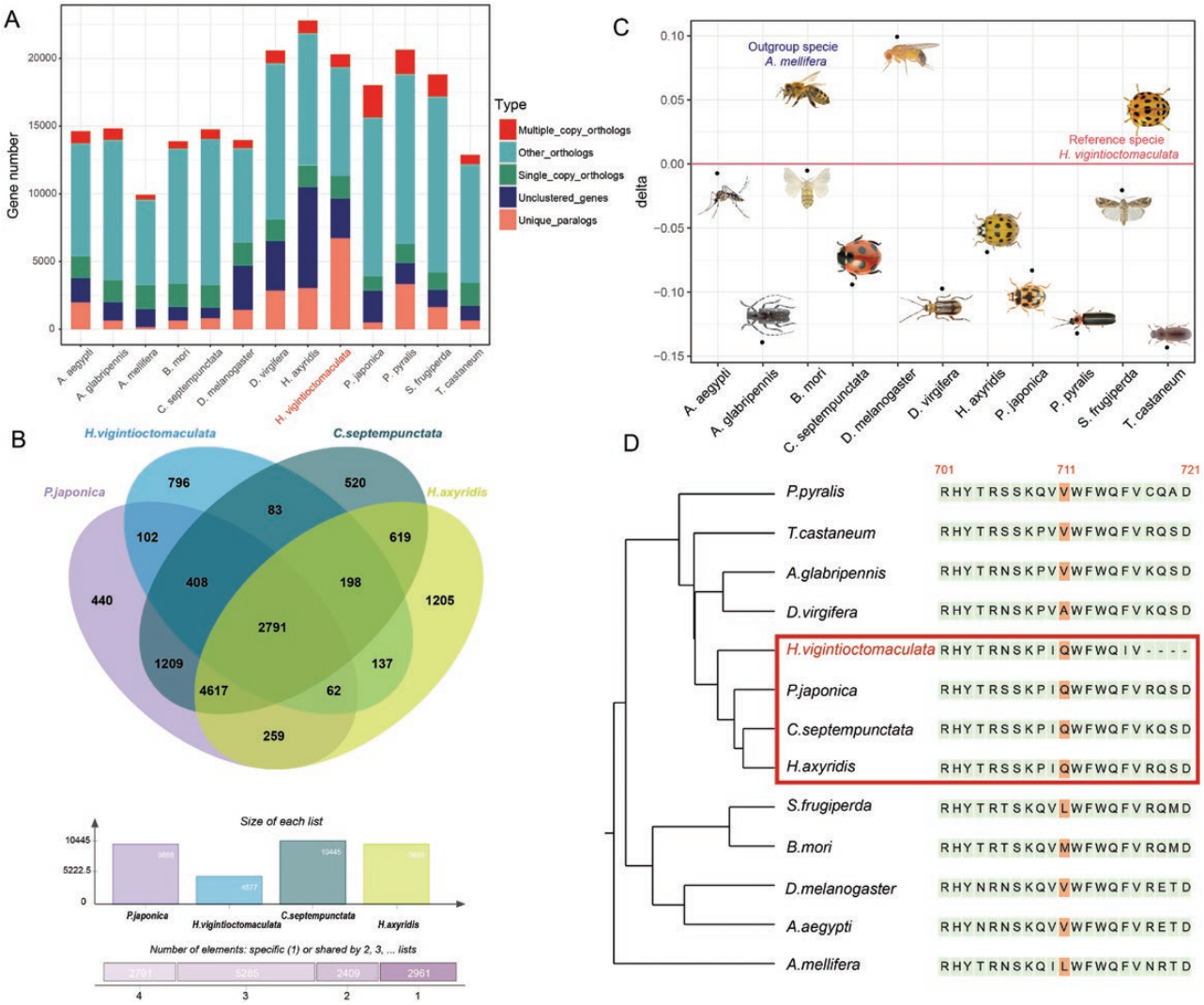

Figure 5. Comparative genomics of *H. vigintioctomaculata* and other related species. (A) Numbers of orthologous/paralogous genes in sampled species. (B) A Venn diagram showing overlap in orthologous genes in all ladybird beetles. (C) Relative evolutionary rates of these species. The analysis was performed using single-copy protein-coding genes with *H. vigintioctomaculata* as the reference species and *A. mellifera* as the outgroup species. The *y*-axis shows the relative evolutionary rates of the species, and the black dots show the specific relative evolutionary rates for each species. (D) The positively selected sites and three-dimensional structure of the protein encoded by *Su(dx)* in sampled species. Positively selected site of the *Su(dx)* gene was showed.

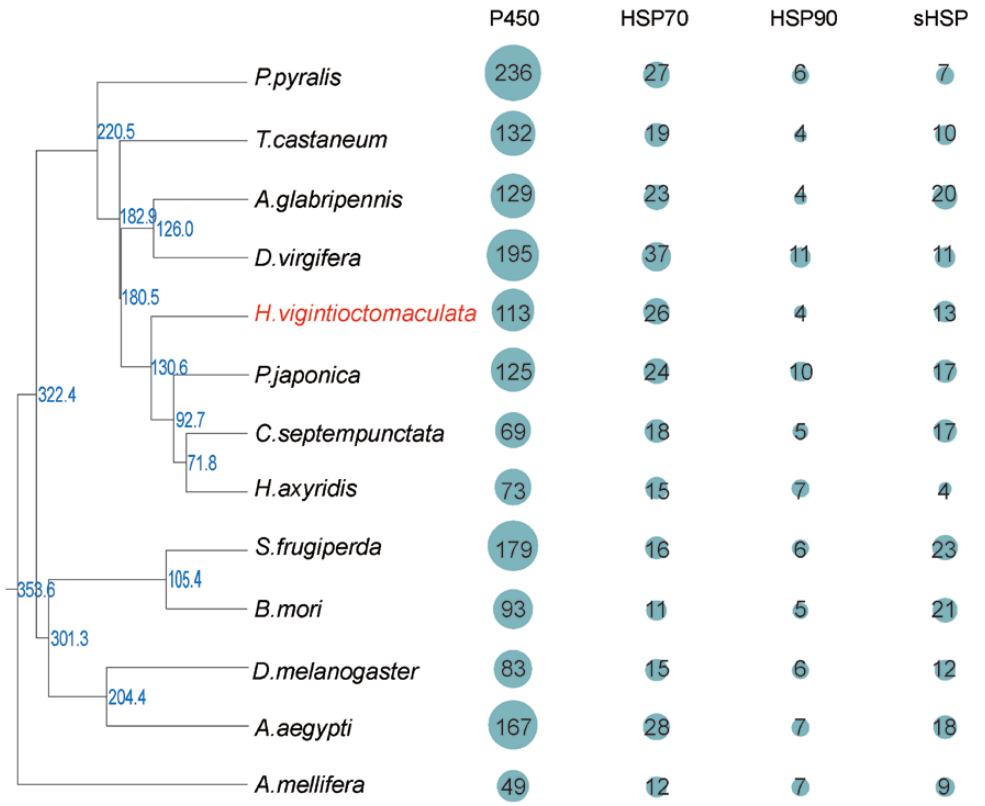

Figure 6. Number of cytochrome P450, HSP70, HSP90, and sHSP genes in these species. The divergence time of these species is indicated on the node of the phylogenetic tree.

i.e. *H. axyridis*, *P. japonica* and *C. septempunctata*, while these latter three showed similar patterns (Supplementary Fig. S5). These findings indicate that the TE elements and TEs insertion times of *H. vigintioctomaculata* differ from those in the other ladybird beetles.

3.4. Phylogenetic relationships of ladybird beetles

We constructed a phylogenetic tree using the protein-coding gene sets of *H. vigintioctomaculata* and 12 other species from 4 orders: Coleoptera, Diptera, Hymenoptera, and Lepidoptera (Fig. 4, Supplementary Figs S6 and S7). The phylogenetic relationship revealed that *H. vigintioctomaculata* clustered with other three ladybird beetles (*C. septempunctata*, *H. Axyridis*, and *P. japonica*) and diverged from them 130.6 million years ago (Mya). Our estimated divergence time is consistent with the previous study on Coccinellidae.[57] As shown in Fig. 5A, a total of 537 single-copy genes were identified in *H. vigintioctomaculata*. A total of 2,791 gene families were shared among all ladybird beetles. *Henosepilachna vigintioctomaculata* and *P. japonica* shared 3,363 gene families, *H. vigintioctomaculata* and *H. axyridis* shared 3,188 gene families, *H. vigintioctomaculata* and *C. septempunctata* shared 3,480 gene families (Fig. 5B). The GO/KEGG enrichment analyses of all expanded gene families in ladybird beetles revealed that these genes are enriched in various biosynthetic pathways and signalling pathways, including p53 signalling pathway (map04115, *P*-value = 7.61E–05), Wnt signalling pathway (map04310, *P*-value = 0.04434701), and Glycosaminoglycan biosynthesis (map00534, *P*-value = 0.000559532) (Supplementary Tables S15 and S16).

3.5. Rate of molecular evolution

Variation in the rate of molecular evolution among different species reflects differences in selection pressures of environments. Diptera (*A. aegypti* and *D. melanogaster*) had the fastest rate of evolution. *Henosepilachna vigintioctomaculata* had the fastest evolutionary rate among ladybird beetles. Chrysomelidae (*T. castaneum*), Lampyridae (*P. pyralis*), and Cerambycidae (*D. virgifera* and *A. glabripennis*) had slower rate of evolution than other groups (Fig. 5C; Supplementary Tables S17 and S18). The rate of evolution of ladybird beetles was intermediate among all species examined.

3.6. PSGs of ladybird beetles

We identified nine PSGs in the ladybird beetles lineage. One interesting PSG is *Su(dx)* (*P* = 0.00096935) (Fig. 5D, Supplementary Table S19). This gene can down-regulate the Notch/N signalling pathway and thus likely promotes Notch ubiquitination, endocytosis, and degradation.[65–67] It has also been shown to be involved in lateral inhibition, intercellular synchronization, neural development, wing development, leg joint formation, colour pattern formation, and so forth.[65–67] *Su(dx)* thus may play important roles in ladybird beetle development.

3.7. Cytochrome P450 monooxygenase and heat shock proteins in ladybird beetles

Different ladybird beetles showed variable resistance to insecticides and high temperature. Previous studies had proved cytochrome P450 monooxygenase (P450) and heat shock proteins (HSPs) was involved in the tolerance to heat stress in ladybird beetles.[8,9] To evaluate the insecticide resistance of all

ladybird beetles, P450s, which are key detoxification enzymes, were annotated in all species. The number of P450s in the *H. vigintioctomaculata* genome (113) was similar to that in the *P. japonica* genome (125) and much greater than that in the *H. axyridis* genome (73) and *C. septempunctata* (69) (Fig. 6), indicating that P450 genes have not undergone an expansion in all ladybird beetles. These results were consistent with the reports in previous study that P450 genes have not undergone an expansion in *H. axyridis*,[9] and an expansion in P450 genes has been documented in *P. japonica*.[8] The gene duplications events were observed in *H. vigintioctomaculata*, suggesting *H. vigintioctomaculata* may resist to higher dosage of insecticides. However, further insecticides resistance testing is need. *Hsps* are important molecular chaperones that are involved in thermal adaptation and resistance to some proteotoxic stresses.[39] We annotated Hsp genes in all ladybird beetle species. In *H. vigintioctomaculata*, a total of 43 Hsp genes were detected, including 26 *Hsp70*, 13 *sHsp* (small heat shock protein), and 4 *Hsp90* genes (Fig. 6). The number of Hsp superfamily genes in *H. vigintioctomaculata* was similar to *P. japonica* (24 *Hsp70*, 17 *sHsp*, and 10 *Hsp90* genes) and a little higher than other two ladybird beetles (*H. axyridis* and *C. septempunctata*). Besides, the number of *Hsp70* family was highest in HSPs. *Hsp70* proteins are involved in the response to a set of stress stimuli, which results in phenotypic changes under stress conditions.[68]

4. Conclusions

In this study, we assembled a chromosome-level genome of *H. vigintioctomaculata* using Illumina sequencing, Nanopore sequencing, and Hi-C technology. This high-quality genome provides an important genomic resource that will aid research on Coccinellidae species and their invasion biology.

The size of the assembled genome using NextDenovo assembly and Nextpolish correction was approximately 581.59 Mb, and the N50 was 3.51 Mb. A total of 493 scaffolds were anchored, ordered, and oriented to 11 chromosomes, and the final N50 was 56.17 Mb. This assembly was comparable to other chromosome-level ladybird beetle genomes and Coleoptera genomes. In addition, the X chromosome was identified in all three ladybird beetles examined. A total of 318,293,454 bp repeat sequences (54.72%) and 20,304 protein-coding genes were identified based on this assembly. We also found that the LINE elements ratio was much higher in *H. vigintioctomaculata* than in the other three ladybird beetles, and the LINE element insertion pattern in *H. vigintioctomaculata* was distinct. Some important gene families, e.g. *P450* and *Hsp*, have undergone expansions in some ladybird beetles. Using comparative analysis methods, we found that the PSG *Su(dx)* might be involved in the adaptive evolution of development of ladybird beetles.

Funding

This study was supported by National Natural Science Foundation of China (32000383, 31672002) and Modern Agriculture Industrial Technology System Funding of Shandong Province (SDAIT-16-06).

Authors' contributions

Y.R. and P.Y. conceived and designed the investigation. W.Z., S.C., and S.G. performed field and laboratory work. Z.W. and W.Z. assembled the genome. H.L. performed the Hi-C scaffold. Y.R., W.Z., Y.W., and S.C. analysed the data. S.C., P.Y., and Y.W. contributed materials and reagents. W.Z. and S.C. wrote the paper. Y.R., T.S., and H.X. revised the manuscript. All the authors read and approved the final manuscript.

Conflict of interest

The authors declare that there is no conflict of interest.

Data availability

All data were deposited at the NCBI in the sequence read archive (SRA) under accession number (BioProject Number: PRJNA828333). The assembled genome and gene annotation were deposited in the Genome Warehouse in National Genomics Data Center, under accession number GWHBJYP00000000.1 that is publicly accessible at https://ngdc.cncb.ac.cn/gwh. The assembled genome and gene annotation were also available at https://figshare.com/articles/dataset/Henosepilachna_vigintioctomaculata_genome_assembly_and_gene_annotation/21502014.

Supplementary data

Supplementary data are available at *DNARES* online.

References

1. Navarrete-Heredia, J.L. 2003, American Beetles, vol. 2: polyphaga: scarabaeoidea through curculionoidea, *Acta Zool. Mex. (N. S.)*, 277–81.
2. Finlay-Doney. 2013, Ecology and behaviour of the ladybird beetles (Coccinellidae), vol. 8: natural enemies pf ladybird beetles, *Austral. Ecol*, 375–443.
3. Obrycki, J.J. and Kring, T.J. 1998, Predaceous Coccinellidae in biological control, *Annu. Rev. Entomol.*, **43**, 295–321.
4. Miyashita, K. 1963, Outbreaks and population fluctuations of insects, with special reference to agricultural insect pests in Japan, *Bullet. Nat. Inst. Agric. Sci., Series C*, **15**, 99–170.
5. Li, C.S. 1993, Review of the Australian Epilachninae (Coleoptera: Coccinellidae), *Aust. J. Entomol.*, **32**, 209–24.
6. Hori, M., Nakamura, H., Fujii, Y., Suzuki, Y. and Matsuda, K. 2011, Chemicals affecting the feeding preference of the Solanaceae-feeding lady beetle *Henosepilachna vigintioctomaculata* (Coleoptera: Coccinellidae), *J. Appl. Entomol.*, **135**, 121–31.
7. Crowley, L.; University of Oxford and Wytham Woods Genome Acquisition Lab. 2021, The genome sequence of the seven-spotted ladybird, *Coccinella septempunctata* Linnaeus, 1758, *Wellcome Open Res.*, **6**, 319–28.
8. Zhang, L., Li, S., Luo, J., et al. 2020, Chromosome-level genome assembly of the predator *Propylea japonica* to understand its tolerance to insecticides and high temperatures, *Mol. Ecol. Resour.*, **20**, 292–307.
9. Chen, M., Mei, Y., Chen, X., et al. 2021, A chromosome-level assembly of the harlequin ladybird *Harmonia axyridis* as a genomic resource to study beetle and invasion biology, *Mol. Ecol. Resour.*, **21**, 1318–32.
10. Boyes, D., Crowley, L., Null, N., et al. 2021, The genome sequence of the harlequin ladybird, *Harmonia axyridis* (Pallas, 1773) [version 1; peer review: 1 approved with reservations], *Wellcome Open Res.*, **6**, 300–15.
11. Gautier, M., Yamaguchi, J., Foucaud, J., et al. 2018, The genomic basis of color pattern polymorphism in the harlequin ladybird, *Curr. Biol.*, **28**, 3296–302.e7.

12. Ando, T., Matsuda, T., Goto, K., et al. 2018, Repeated inversions within a pannier intron drive diversification of intraspecific colour patterns of ladybird beetles, *Nat. Commun.*, **9**, 3843–5.
13. Li, H.S., Tang, X.F., Huang, Y.H., et al. 2021, Horizontally acquired antibacterial genes associated with adaptive radiation of ladybird beetles, *BMC Biol.*, **19**, 7–22.
14. Du, W., Chen, L., Changchun, R. and Sun, G. 2010, Investigations of the color variations of *Harmonia axyridis* (Pallas) in Changchun, *Chin. Agric. Sci. Bull.*, **26**, 258–66.
15. Tang, B., Zhu, J., Guo, H.S., et al. 2012, Studies of the diversity of multiple elytral color morphs of *Harmonia axyridis* Pallas (Coleoptera: Coccinellidae), *J. Hangzhou Normal Univ. (Nat. Sci. Ed.)*, **11**, 132–6.
16. Jiang, J., Zhang, Z., Yu, X., Yu, C., Liu, F. and Mu, W. 2019, Sublethal and transgenerational effects of thiamethoxam on the demographic fitness and predation performance of the seven-spot ladybeetle *Coccinella septempunctata* L. (Coleoptera: Coccinellidae), *Chemosphere*, **216**, 168–78.
17. Chi, W., Mingyuan, H., Fengshou, D., et al. 2021, The influence of tolfenpyrad on fitness, development, and reproduction in parents and offspring of *Coccinella septempunctata*, *Ecotoxicol. Environ. Saf.*, **210**, 111875.
18. Xiao, D., Zhao, J., Guo, X., et al. 2016, Sublethal effects of imidacloprid on the predatory seven-spot ladybird beetle *Coccinella septempunctata*, *Ecotoxicology*, **25**, 1782–93.
19. Jiang, J., Wang, Y., Mu, W. and Zhang, Z. 2020, Sublethal effects of anthranilic diamide insecticides on the demographic fitness and consumption rates of the *Coccinella septempunctata* (Coleoptera: Coccinellidae) fed on *Aphis craccivora*, *Environ. Sci. Pollut. Res.*, **27**, 4178–89.
20. Bozsik, A. 2006, Susceptibility of adult *Coccinella septempunctata* (Coleoptera: Coccinellidae) to insecticides with different modes of action, *Pest Manag. Sci.*, **62**, 651–4.
21. Vincent, C., Ferran, A., Guige, L., Gambier, J. and Brun, J. 2000, Effects of imidacloprid on *Harmonia axyridis* (Coleoptera: Coccinellidae) larval biology and locomotory behavior, *Eur. J. Entomol.*, **97**, 501–6.
22. Twardowski, J.P., Hurej, M. and Twardowska, K. 2021, Effect of ingestion exposure of selected insecticides on *Coccinella septempunctata* and *Harmonia axyridis* (Coleoptera: Coccinellidae), *Insects*, **12**, 434–43.
23. Nawaz, M., Cai, W., Jing, Z., Zhou, X., Mabubu, J.I. and Hua, H. 2017, Toxicity and sublethal effects of chlorantraniliprole on the development and fecundity of a non-specific predator, the multicolored Asian lady beetle, *Harmonia axyridis* (Pallas), *Chemosphere*, **178**, 496–503.
24. Xiao, D., Tan, X., Wang, W., Zhang, F., Desneux, N. and Wang, S. 2017, Modification of flight and locomotion performances, respiratory metabolism, and transcriptome expression in the lady beetle *Harmonia axyridis* through sublethal pesticide exposure, *Front. Physiol.*, **8**, 33–44.
25. Rasheed, M.A., Khan, M.M., Hafeez, M., et al. 2020, Lethal and sublethal effects of chlorpyrifos on biological traits and feeding of the aphidophagous predator *Harmonia axyridis*, *Insects*, **11**, 491–505.
26. Cheng, S., Yu, C., Xue, M., et al. 2022, Toxicity and risk assessment of nine pesticides on nontarget natural predator *Harmonia axyridis* (Coleoptera: Coccinellidae), *Pest Manag. Sci.*, **78**, 5124–32.
27. Knapp, M. and Nedvěd, O. 2013, Gender and timing during ontogeny matter: effects of a temporary high temperature on survival, body size and colouration in *Harmonia axyridis*, *PLoS One*, **8**, e74984.
28. Song, P., Wang, Q., Nangong, Z., Su, J. and Ge, D. 2012, Identification of *Henosepilachna vigintioctomaculata* (Coleoptera: Coccinellidae) midgut putative receptor for *Bacillus thuringiensis* insecticidal Cry7Ab3 toxin, *J. Invertebr. Pathol.*, **109**, 318–22.
29. Zhang, S., Fu, W., Li, N., Zhang, F. and Liu, T.-X. 2015, Antioxidant responses of *Propylaea japonica* (Coleoptera: Coccinellidae) exposed to high temperature stress, *J. Insect Physiol.*, **73**, 47–52.
30. Wu, G., Miyata, T., Kang, C.Y. and Xie, L.H. 2007, Insecticide toxicity and synergism by enzyme inhibitors in 18 species of pest insect and natural enemies in crucifer vegetable crops, *Pest Manag. Sci.*, **63**, 500–10.
31. Wang, M.M., Xing, L.Y., Ni, Z.W. and Wu, G. 2018, Identification and characterization of ace1-type acetylcholinesterase in insecticide-resistant and -susceptible *Propylaea japonica* (Thunberg), *Bull. Entomol. Res.*, **108**, 253–62.
32. Tang, L.D., Qiu, B.L., Cuthbertson, A.G. and Ren, S.X. 2015, Status of insecticide resistance and selection for imidacloprid resistance in the ladybird beetle *Propylaea japonica* (Thunberg), *Pestic. Biochem. Physiol.*, **123**, 87–92.
33. Chen, X., Xiao, D., Du, X., et al. 2019, The role of the dopamine melanin pathway in the ontogeny of elytral melanization in *Harmonia axyridis*, *Front. Physiol.*, **10**, 1066–73.
34. Nauen, R., Zimmer, C.T. and Vontas, J. 2021, Heterologous expression of insect P450 enzymes that metabolize xenobiotics, *Curr. Opin. Insect Sci.*, **43**, 78–84.
35. Feyereisen, R. 1999, Insect P450 enzymes, *Annu. Rev. Entomol.*, **44**, 507–33.
36. Feyereisen, R. 2012, Insect CYP genes and P450 enzymes. In: L. I. Gilbert (Ed.), *Elsevier, Insect molecular Biology and Biochemistry*, 236–316.
37. Srensen, J.G., Kristensen, T.N. and Loeschcke, V. 2003, The evolutionary and ecological role of heat shock proteins, *Ecol. Lett.*, **6**, 1025–37.
38. Tower, J. 2009, Hsps and aging, *Trends Endocrinol. Metab.*, **20**, 216–22.
39. Colinet, H., Siaussat, D., Bozzolan, F. and Bowler, K. 2013, Rapid decline of cold tolerance at young age is associated with expression of stress genes in *Drosophila melanogaster*, *J. Exp. Biol.*, **216**, 253–9.
40. Sørensen, J.G. and Loeschcke, V. 2007, Studying stress responses in the post-genomic era: its ecological and evolutionary role, *J. Biosci.*, **32**, 447–56.
41. Liu, B., Shi, Y., Yuan, J., et al. 2013, Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects, *Quant. Biol*, 32–5.
42. Dudchenko, O., Batra, S.S., Omer, A.D., et al. 2017, De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds, *Science*, **356**, 92–5.
43. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., et al. 2009, Comprehensive mapping of long-range interactions reveals folding principles of the human genome, *Science*, **326**, 289–93.
44. Li, H. and Durbin, R. 2009, Fast and accurate short read alignment with Burrows-Wheeler transform, *Bioinformatics*, **25**, 1754–60.
45. Kent, W.J. 2002, BLAT—the BLAST-like alignment tool, *Genome Res.*, **12**, 656–64.
46. Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov, E.M. 2015, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, *Bioinformatics*, **31**, 3210–2.
47. Kielbasa, S.M., Wan, R., Sato, K., Horton, P. and Frith, M.C. 2011, Adaptive seeds tame genomic sequence comparison, *Genome Res.*, **21**, 487–93.
48. Krzywinski, M., Schein, J., Birol, I., et al. 2009, Circos: an information aesthetic for comparative genomics, *Genome Res.*, **19**, 1639–45.
49. Benson, G. 1999, Tandem repeats finder: a program to analyze DNA sequences, *Nucleic Acids Res.*, **27**, 573–80.
50. Bedell, J.A., Korf, I. and Gish, W. 2000, MaskerAid: a performance enhancement to RepeatMasker, *Bioinformatics*, **16**, 1040–1.
51. Chalopin, D., Naville, M., Plard, F., Galiana, D. and Volff, J.N. 2015, Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates, *Genome Biol. Evol.*, **7**, 567–80.
52. Birney, E., Clamp, M. and Durbin, R. 2004, GeneWise and Genomewise, *Genome Res.*, **14**, 988–95.
53. Haas, B.J., Delcher, A.L., Mount, S.M., et al. 2003, Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies, *Nucleic Acids Res.*, **31**, 5654–66.

54. Haas, B.J., Salzberg, S.L., Zhu, W., et al. 2008, Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments, *Genome Biol.*, **9**, R7.
55. Li, L., Stoeckert, C.J. and Roos, D.S. 2003, OrthoMCL: identification of ortholog groups for eukaryotic genomes, *Genome Res.*, **13**, 2178–89.
56. Stamatakis, A. 2014, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, *Bioinformatics*, **30**, 1312–3.
57. Tomaszewska, W., Escalona, H.E., Hartley, D., et al. 2021, Phylogeny of true ladybird beetles (Coccinellidae: Coccinellini) reveals pervasive convergent evolution and a rapid Cenozoic radiation, *Syst. Entomol.*, **46**, 611–31.
58. Yang, Z. 1997, PAML: a program package for phylogenetic analysis by maximum likelihood, *Comput. Appl. Biosci.*, **13**, 555–6.
59. Kumar, S., Tamura, K. and Nei, M. 1994, Mega—molecular evolutionary genetics analysis software for microcomputers, *Comput. Appl. Biosci.*, **10**, 189–91.
60. Takezaki, N., Rzhetsky, A. and Nei, M. 1995, Phylogenetic test of the molecular clock and linearized trees, *Mol. Biol. Evol.*, **12**, 823–33.
61. De Bie, T., Cristianini, N., Demuth, J.P. and Hahn, M.W. 2006, CAFE: a computational tool for the study of gene family evolution, *Bioinformatics*, **22**, 1269–71.
62. Beissbarth, T. and Speed, T.P. 2004, GOstat: find statistically overrepresented Gene Ontologies within a group of genes, *Bioinformatics*, **20**, 1464–5.
63. Huang, D.W., Sherman, B.T. and Lempicki, R.A. 2009, Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists, *Nucleic Acids Res.*, **37**, 1–13.
64. Yang, Z.H. 1997, PAML: a program package for phylogenetic analysis by maximum likelihood, *Comput. Appl. Biosci.*, **13**, 555–6.
65. Cornell, M., Evans, D.A., Mann, R., et al. 1999, The *Drosophila melanogaster* Suppressor of deltex gene, a regulator of the Notch receptor signaling pathway, is an E3 class ubiquitin ligase, *Genetics*, **152**, 567–76.
66. Wilkin, M.B., Carbery, A.M., Fostier, M., et al. 2004, Regulation of notch endosomal sorting and signaling by *Drosophila* Nedd4 family proteins, *Curr. Biol.*, **14**, 2237–44.
67. Mazaleyrat, S.L., Fostier, M., Wilkin, M.B., et al. 2003, Down-regulation of Notch target gene expression by Suppressor of deltex, *Dev. Biol.*, **255**, 363–72.
68. Bettencourt, B.R., Hogan, C.C., Nimali, M. and Drohan, B.W. 2008, Inducible and constitutive heat shock gene expression responds to modification of Hsp70 copy number in *Drosophila melanogaster* but does not compensate for loss of thermotolerance in Hsp70null flies, *BMC Biol.*, **6**, 5–19.