

Apolygus lucorum genome provides insights into omnivorousness and mesophyll feeding

Yang Liu¹ | Hangwei Liu² | Hengchao Wang² | Tianyu Huang¹ | Bo Liu² | Bin Yang¹ | Lijuan Yin² | Bin Li¹ | Yan Zhang² | Sai Zhang¹ | Fan Jiang² | Xiaxuan Zhang¹ | Yuwei Ren² | Bing Wang¹ | Sen Wang² | Yanhui Lu¹ | Kongming Wu¹ | Wei Fan² | Guirong Wang^{1,2}

¹State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, China

²Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China

Correspondence

Kongming Wu and Guirong Wang, State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, China.
Emails: kmwu@ippcaas.cn; wangguirong@caas.cn

Wei Fan, Guangdong Laboratory of Lingnan Modern Agriculture, Shenzhen; Genome Analysis Laboratory of the Ministry of Agriculture; Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China.
Email: fanwei@caas.cn

Funding information

National Natural Science Foundation of China, Grant/Award Number: 31621064 and 31672095; National Key R&D Program of China, Grant/Award Number: 2016YFC1200600 and 2017YFD0200400; Agricultural Science and Technology Innovation Program & The Elite Young Scientists Program of CAAS, Fundamental Research Funds for Central Non-profit Scientific Institution, Grant/Award Number: Y2017JC01; Agricultural Science and Technology Innovation Program Cooperation and Innovation Mission, Grant/Award Number: CAAS-XTX2016; Fund of Key Laboratory of Shenzhen, Grant/Award Number: ZDSYS20141118170111640

Abstract

Apolygus lucorum (Miridae) is an omnivorous pest that occurs worldwide and is notorious for the serious damage it causes to various crops and substantial economic losses. Although some studies have examined the biological characteristics of the mirid bug, no reference genome is available in Miridae, limiting in-depth studies of this pest. Here, we present a chromosome-scale reference genome of *A. lucorum*, the first sequenced Miridae species. The assembled genome size was 1.02 Gb with a contig N50 of 785 kb. With Hi-C scaffolding, 1,016 Mb contig sequences were clustered, ordered and assembled into 17 large scaffolds with scaffold N50 length 68 Mb, each corresponding to a natural chromosome. Numerous transposable elements occur in this genome and contribute to the large genome size. Expansions of genes associated with omnivorousness and mesophyll feeding such as those related to digestion, chemosensory perception, and detoxification were observed in *A. lucorum*, suggesting that gene expansion contributed to its strong environmental adaptability and severe harm to crops. We clarified that a salivary enzyme polygalacturonase is unique in mirid bugs and has significantly expanded in *A. lucorum*, which may contribute to leaf damage from this pest. The reference genome of *A. lucorum* not only facilitates biological studies of Hemiptera as well as an understanding of the damage mechanism of mesophyll feeding, but also provides a basis on which to develop efficient control technologies for mirid bugs.

KEYWORDS

Apolygus lucorum, chromosome-level genome, gene expansion, mesophyll feeding, omnivorousness

1 | INTRODUCTION

Apolygus lucorum, a global Miridae pest, occurs throughout Asia, Europe, Africa and America, and causes large economic losses each year (Jiang et al., 2015; Lu & Wu, 2008, 2011b). This notorious pest has attracted renewed research attention due to its resurgence, and it is gradually increasing from a secondary to a major pest of cotton due to the drastically decreased usage of insecticides as a result of commercial adoption of Bt cotton (Lu et al., 2010; Wu et al., 2002). The area affected by mirid bugs and the associated yield losses in cotton both increased 3- to 5- fold from 1991 to 2010 (Lu & Wu, 2011b). To make things worse, *A. lucorum* not only damages cotton, but has also becomes a primary pest of other crops such as cereals, vegetables, and fruits, leading to more severe economic losses (Jiang et al., 2015; Lu & Wu, 2011b; Lu et al., 2010; Pan et al., 2015). In order to control *A. lucorum*, farmers increased the use of insecticides, which reduced or even counterbalanced the benefits of Bt cotton (Lu et al., 2010; Zhao et al., 2011). However, efficient environmentally friendly measures to control this pest are still lacking.

Apolygus lucorum possesses a variety of biological characteristics associated with successful insect pests, including strong environmental adaptability, high population growth rate, and strong dispersal ability (Lu & Wu, 2008, 2011b; Lu et al., 2007, 2009, 2010). The wide host range of *A. lucorum* is likely to be critical for its success, which increases the range of damaged crops and the difficulty of prevention. *A. lucorum* is omnivorous and is specifically phytozoophagous, and mainly phytophagous, feeding on over 200 plant species that comprise the majority of crops (Lu et al., 2012). As a polyphagous pest, *A. lucorum* grows and develops normally on different hosts, and it often switches food plants in the field depending on its highly developed chemosensory system (Pan et al., 2015), which has been found to considerably increase the population growth and survival of many phytophagous pests (Kennedy & Storer, 2000). In addition to feeding on plants, *A. lucorum* can also prey on some small insects or insect eggs, which further increases its environmental adaptation (Li et al., 2016; Lu et al., 2008; Yuan et al., 2013).

Due to its special mesophyll feeding pattern, *A. lucorum* causes severe damage to crops. Both nymphs and adults of *A. lucorum* can suck the sap of terminal meristems, tender leaves, flowers, bolls, and fruits, resulting in serious reduction in yields and quality of numerous crops (Jiang et al., 2015; Wu & Guo, 2005). *A. lucorum* is a typical mesophyll feeder, using the cytoplasm and nucleus of plant tissue as its food source (Sharma et al., 2014; Zhang et al., 2013). In the process of feeding, *A. lucorum* uses its stylets to pierce plant tissue and ingest the liquefied plant material, which is preliminarily degraded by enzymes in saliva the insect injects into the plant tissue (Lu & Wu, 2008; Wheeler, 2001; Zhang et al., 2015). This 'lacerate and flush' feeding action results in stunting, abscission of squares and bolls, and fruit malformation of plants. The key factor eliciting symptoms of the feeding damage is the saliva containing various digestive enzymes, rather than the mechanical damage (Sharma et al., 2014; Wheeler, 2001; Zhang et al., 2013), which might be the

dominant reason for the economic importance of mirid pests (Fрати et al., 2006).

Hemiptera is the most species-rich hemimetabolous order, including many notorious agricultural pests and human disease vectors (Cassis & Schuh, 2012; Lu et al., 2010; Wheeler, 2001). Several genome sequences have been reported, including from aphids (Li et al., 2019; Mathers et al., 2017; Richards et al., 2010; Wenger et al., 2017), plant hoppers (Wang et al., 2017; Xue et al., 2014; Zhu et al., 2017), kissing bug (Mesquita et al., 2015), seed bugs (Panfilio et al., 2019), bed bug (Benoit et al., 2016; Rosenfeld et al., 2016) and psyllids (Sloan et al., 2014), paving the way to uncovering the molecular and genetic mechanisms of many biological problems. Miridae (Hemiptera: Heteroptera: Cimicomorpha), also called plant bugs, is one of two most species-rich families in Hemiptera (with family Cicadellidae) and can be found in all major biogeographic regions of the world (Cassis & Schuh, 2012; Jung & Lee, 2012). With a wide range of food preferences and behaviours, some mirid bugs are pests of crops, whereas others are predatory natural enemies (Cassis & Schuh, 2012; Wheeler, 2001). Although studies have examined the feeding characteristics of many mirid bugs (Lu et al., 2007; Lu & Wu, 2011a), no reference genome data is available in Miridae, limiting in-depth studies of these pests. Here, we generated the first reference genome for a mirid bug, *A. lucorum*, to explore the genetic processes and molecular mechanisms of its omnivorousness and mesophyll feeding.

2 | MATERIALS AND METHODS

2.1 | Insect rearing and genomic sequencing

Apolygus lucorum nymphs and adults were collected from cotton fields at the Langfang Experimental Station of Chinese Academy of Agricultural Sciences, Hebei Province, China. A laboratory colony was established and maintained at $28 \pm 1^\circ\text{C}$, relative humidity (RH) $60 \pm 5\%$, and 14:10 light: dark (L:D) and reared on green beans and corn. An inbred strain of *A. lucorum* was developed by successive single-pair sib mating for 12 generations from this laboratory colony. This inbred strain was used for all genomic sequencing experiments in this study.

For Illumina sequencing, a short paired-end DNA library with a 400 bp insert size from a female adult *A. lucorum* was constructed using standard Illumina protocols, and sequenced on the ILLUMINA HISEQ 2500 platform. For PacBio sequencing, two single-end libraries with ~20 kb insert size were constructed using PacBio SMRT from 100 female siblings of *A. lucorum*. PacBio long reads were sequenced using 18 SMRT Cells on the PACBIO SEQUEL SYSTEM (Pacific Biosciences).

2.2 | Assembly and polishing of contigs

A contig assembly was first carried out using CANU (version 1.6; Koren et al., 2017) with parameters (ovsMethod = sequential

genomeSize = 1g) based on PacBio long reads from pooled samples. This resulted in 18,403 contigs of total length ~1.97 Gb, contig N50 of 259 Kb.

Three rounds of contig polishing were performed. For the first round, contigs were polished using PacBio reads with the Arrow consensus caller in SMRT-LINK version 5.1.0. The original bam files generated from PacBio Sequel were aligned with contig assembly by PBALIGN (version 0.3.1). Then, using ARROW (version 2.2.2), we polished the assembly. For the second and third rounds, we used our in-house clean_adapter (-s 15 -a R1/2-adapter -r 75) and clean_lowqual (-e 0.001 -r 75) to filter out adaptors and low quality sequences in raw Illumina reads. After quality control, 140 Gb raw Illumina data from female *A. lucorum* generated 100 Gb clean data. The clean data were mapped to the contigs using BWA (version 0.7.12; Li & Durbin, 2009) and the assembly errors were corrected using PILON (version 1.22; Walker et al., 2014) with parameters (--fix bases --nonpf --threads 112 --minqual 20).

To filter haplotypic duplication in the contig assembly, we used PURGE_DUPS (version 1.2.3; Guan et al., 2020) with parameters (-2 -a 50) on the polished assembly. This resulted in a purged primary assembly of total length 1.03 Gb, contig N50 of 785 kb and a haplotig assembly of total length 936 Mb, contig N50 of 88 kb.

To assess the completeness of genome assembly, we run BUSCO (version 3.0.2) using the INSECTA database (OrthoDB version 9), which contains 1,658 conserved insecta genes, with parameters '-m genome -sp pea_aphid -l insecta_odb9'. For the gene set completeness assessment, parameters was changed to '-m proteins -l insecta_odb9'.

2.3 | Filtering contamination contigs

Because samples for PacBio sequencing were whole insects that may contain bacteria and parasites in the gut, and samples for Illumina sequencing were gut removed, we used clean Illumina data to filter possible contaminations in assembly. We used bwa (version 0.7.12) to align clean Illumina data with the assembly and if any contig had an Illumina coverage rate lower than 5%, it was removed. This resulted in a total of 80 contigs being removed from the genome. In order to identify further potential contaminations, we submitted the assembly to NCBI. After validation from contamination screening of NCBI, 22 additional contigs were reported as contaminated sequences from *Lactococcus lactis*, a common bacterium within the digestive tract of *A. lucorum* and were filtered.

2.4 | Scaffolding with LACHESIS

About 5 g of fresh tissues from living insects were macerated and crosslinked using paraformaldehyde to capture the interacting DNA segments. Chromatin was subsequently digested with MboI (NEB), and biotinylated nucleotides were used to fill in the resulting sticky ends. Following ligation, a protease was used to remove

the crosslinks. Finally, genomic DNA was extracted, sheared into 350 bp fragments using a focused ultrasonicator (Covaris), and fragments into which biotin had been incorporated were pulled down with streptavidin-coated magnetic beads. Purified DNA was then prepared and sequenced on an ILLUMINA HISEQ instrument according to the manufacturer's recommendations.

After quality control, clean Hi-C paired-end reads were first mapped to the contig assembly by BOWTIE2 (version 2.3.4.3; Langmead & Salzberg, 2012), and then HIC-PRO (version 2.11.0; Servant et al., 2015) used the alignment to detect valid alignments and filter multiple hits and singletons. Finally, LACHESIS (Burton et al., 2013) was used to cluster, order and orient the contigs.

2.5 | Transcriptome library preparation and sequencing

Total RNA (1 µg) was extracted by TRIzol reagent (Invitrogen) and was used to construct cDNA libraries. A total of 60 (20 samples × 3 replicates) individual unstranded cDNA libraries were prepared by ligating sequencing adaptors to cDNA fragments synthesized using random hexamer primers with NEBNext Ultra RNA Library Prep Kit for Illumina (NEB). The 20 samples included spawn, eight tissues (antenna, mouthpart, salivary gland, head, gut, leg, wing, body) from third instar nymphs, and 11 tissues (male antenna, female antenna, mouthpart, salivary gland, head, gut, leg, wing, male genital, female genital, body) from adults. Raw sequencing data were generated using an ILLUMINA HISEQ 4000 system with the paired-end 150 bp (PE150) strategy. The average length of the insert sequence was 350 bp.

RNA samples from the whole body of *A. lucorum* in six different developmental stages including first to fifth nymph, and adult were also prepared for full-length transcriptome sequencing using the PacBio Iso-Seq protocol. Total RNA was extracted separately from the six samples, equal amounts of which were then mixed. The synthetic full-length cDNAs were selected to prepare three 20 kb SMRTbell template libraries for sequencing on a PACBIO SEQUEL instrument.

2.6 | Genome annotation

Tandem repeats were identified by TANDEM REPEATS FINDER (version 4.07b; parameters: 2 7 7 80 10 50 2000 -f -m -d; Benson, 1999). Transposable elements (TEs) were identified using a combination of two methods: searching against the TE database (DFAM 3.0, RepBase 20170127) by REPEATMASKER (version 4.0.9; with parameter: -nolow -no_is -norna -engine ncbi; <http://repeatmasker.org>) and searching against the TE protein database by REPEAT-PROTEINMASK (with parameter: -engine ncbi -noLowSimple); and constructing a de novo repeat library by REPEATMODELER (version 1.0.8; parameters: -engine ncbi -database), followed by REPEAT-MASKER to find TE repeats. Because some repeat sequences in

the REPEATMODELER de novo library may come from simple repeats, low complexity regions, and duplicated protein-coding genes, we used two steps for filtering the non-TE repeat sequence to get a *A. lucorum* specific TE library: first, the 1,062 repeat sequence of 'unknown' type were aligned to NR database by blastx (v2.7.1+) using $1e-5$ as cutoff, and 14 of them were found to have homology with known non-TE protein-coding genes; second, the repeat sequence of 'unknown' type were subjected to sequence structure analysis, and 80 of them were found to have more than 20% regions covered by simple repeat or low-complexity sequences. In total, 94 'unknown' repeat sequences in the REPEATMODELER de novo library were filtered. After that, REPEATMASKER was used to find TEs based on the filtered de novo TE library.

The gene models in *A. lucorum* were predicted using AUGUSTUS (version 3.3.2; Stanke et al., 2006) on the TE soft-masked genome, integrating evidence from RNA sequencing alignments, Isoform sequencing alignments and protein homology searches. For RNA-Seq, 20 paired-end datasets from different tissues were aligned with the genome using STAR (version 2.7.1a; Dobin et al., 2013). After filtering by filterBam in AUGUSTUS, the sorted bam file was transferred to a hints file by bam2hints in AUGUSTUS. Moreover, we used Iso-Seq to assist in gene prediction. GMAP (version 2018-03-25) was used to align Iso-Seq sequences with the genome and blat2hints.pl in AUGUSTUS was used to generate a hints file. Additionally, for protein homology evidence, all Hemiptera proteins in NCBI RefSeq were download. We aligned the proteins with the genome by TBLASTN (version 2.7.1+) using $1e-5$ as cutoff and filtered those with less than 50% identity. We used exonerate (v2.2.0) to align the remaining proteins with the genome and used exonerate2hints.pl in AUGUSTUS to generate a hints file. Finally, we combined all hints files from RNA-Seq, Iso-Seq and protein homology, and used AUGUSTUS with the combined hints file to predict gene models, resulting in 23,106 gene models.

To get accurate gene sets, we filtered genes with less than 35 amino acids. We aligned protein sequences of gene models with the NR database in DIAMOND (version 0.8.28) blastp using $1e-5$ as a cutoff, and 16,187 gene models had homologous proteins in NR. We also aligned 20 RNA-Seq data sets with coding sequences of gene models in bwa (0.7.12) and 17,953 gene models had a coverage rate higher than 95%. We retained genes that either had homologous proteins in NR or had RNA support, resulting in 20,386 genes. After that, we detected 33 genes with two or more errors in start codons, stop codons or nontriplet length. We filtered those wrong genes and got the final official gene set (OGS) including 20,353 gene models.

For gene functional annotation, we aligned the protein sequences of genes with KEGG, EGGNOG, NR, SWISS-PROT databases by diamond, an alternative replacement of BLAST, using $1e-5$ as a cutoff and got the best hit. We also used INTERPROSCAN (version 5.38-76.0) to search INTERPRO databases to find motifs and domains. Taken together, 18,721 (91.98%) genes had homologous information in those databases, indicating that the OGS is reasonably accurate. Moreover, TRNASCAN-SE (versus 2.0) was used to find tRNAs with default parameters.

2.7 | Evolutionary analysis

Nine sequenced hemipteran insects including *Acyrtosiphon pisum*, *Myzus persicae*, *Bemisia tabaci*, *A. lucorum*, *Cimex lectularius*, *Rhodnius prolixus*, *Oncopeltus fasciatus*, *Laodelphax striatella*, *Nilaparvata lugens* and *Drosophila melanogaster* as an outgroup were used to infer gene orthology in OrthoFinder (Emms & Kelly, 2015) with default parameters. Phylogenetic tree and gene orthology results were displayed and annotated using Evolview (He et al., 2016). Expanded orthologous groups in *A. lucorum* were determined using a rank sum test compared to other eight insects in Hemiptera. Protein sequences of single copy genes from each species were aligned in MUSCLE (version 3.8.1551; Edgar, 2004), then concatenated into one super-sequence. PhyML was used to reconstruct the phylogenetic tree based on the concatenated super-sequence with the LG + I + G + F model (Guindon & Gascuel, 2003). Divergence times among species were calculated in MCMCTREE (PAML package, version 4.9; Yang, 2007). Calibration times were set according to a previous paper, minimum = 320 Ma and maximum = 390 Ma for *D. melanogaster* and *A. lucorum* (Misof et al., 2014). GO (Gene Ontology) annotation results were obtained from INTERPRO (Quevillon et al., 2005). GO enrichment analysis was performed using the OmicShare tools, a free online platform for data analysis (<https://www.omicshare.com/tools/>). The reciprocal BLAST best hit was used to calculate the synonymous mutation rate (Ks) by KAKS_CALCULATOR 2.0 (Wang et al., 2010) with default parameters. Duplicate_gene_classifier in MCscanX (Wang et al., 2012) was implemented to classify the origins of the duplicated genes into different types.

2.8 | Analysis of the digestive enzyme, chemosensory receptor, and detoxification enzyme genes

A set of described Hemiptera odorant receptors (ORs) and gustatory receptors (GRs) was used to search the *A. lucorum* gene sets by BLASTP with the cutoff *e*-value $1e-5$. Multiple PSI-BLASTP searches were initiated with divergent ORs and GRs to find any additional annotated proteins that might belong to these families, and up to four iterations were used. Finally, some ORs and GRs were corrected manually. Ionotropic receptors (IRs), digestive enzymes, and detoxification enzymes were annotated using DIAMOND (version 0.9.21.122) results compared to the NR database, UNIPROT database and KEGG database with *e*-value $1e-5$ and confirmed by InterProScan or eggNOG. To get a complete gene family set, reannotation of the gene family was performed. First, all digestive enzyme, chemosensory receptor, and detoxification enzyme genes got from former gene set was mapped to eight Hemiptera genomes by exonerate (version 2.2.0) with identity >35%, and exonerate2hints.pl was used to generate a hints file. Then, the region where these genes can map was used to predict gene models by AUGUSTUS with the hints file, and short gene models (less than 200 bp) were filtered. The predicted gene model that doesn't exist in former gene set was added to the gene family sets.

The protein sequences were aligned using MUSCLE (version 3.8.1551; Edgar, 2004). MEGA X (Knyaz et al., 2018) was used to reconstruct the phylogenetic tree for digestive enzymes, chemosensory receptors, and glutathione S-transferases (GSTs) with the neighbour-joining method. PhyML was used to reconstruct the phylogenetic tree for cytochrome P450 monooxygenases (P450s). To analyse gene expression, clean reads of each sample were mapped to *A. lucorum* gene sets using BOWTIE2 (version 2.2.5; Langmead & Salzberg, 2012), and then RSEM (version 1.2.12; Li & Dewey, 2011) was used to count the number of mapped reads and estimate the TPM (transcripts per million) value of each gene. The phylogenetic tree and heatmap were generated by ITOL (Letunic & Bork, 2016).

3 | RESULTS

3.1 | Chromosome-level genome assembly and recent expansion of DNA and LINE TEs

Due to the high heterozygosity of wild *Apolygus lucorum*, we used female adults from a 12-generation inbred strain for genomic sequencing. First, we produced 140 Gb paired-end 150 Illumina reads, and estimated the genome size of *A. lucorum* to be 1.03 Gb based on k-mer analysis (Figure S1). Then, we produced 103 Gb (~100× coverage) single molecule real-time (SMRT) sequences with an N50 read length of 14 kb on the PacBio Sequel platform (Table S1), and generated an assembly comprising 3,818 contigs with a total length of 1.02 Gb and contig N50 length of 785 kb (Tables 1, S2, S3). Finally, with Hi-C scaffolding, 1,016 Mb (99%) contig sequences were clustered, ordered and assembled into 17 large scaffolds with scaffold N50 length 68 Mb (Figure 1a, S2 and Table S4), the longest of which was 117 Mb and the shortest 16 Mb. To validate this assembly, we mapped paired-end clean Illumina reads onto it, resulting in a mapping rate of 99.39%. Furthermore, the assembled genome size was comparable to the estimated genome size. Based on ORTHODB version 9 of Insecta, 1,610 (97.1%) BUSCO genes were found in the

genome, including 1,564 (94.3%) complete and 46 (2.8%) fragmented BUSCO genes. Moreover, 75 (4.5%) 'duplicate' BUSCO genes were found in *A. lucorum*, comparable or lower than most compared species, suggesting that the current reference assembly contains few redundant fragments derived from highly heterozygous genomic regions (Figure 1b). Apart from the assembled nuclear genome, we also assembled the mitochondrial genome that has 99.7% identity with the published mitochondrion (Wang et al., 2014; Figure S3). In total, read mapping, genome size estimation, BUSCO assessment and assembly of the mitochondrion demonstrated that our genome assembly is reasonably accurate, and the majority of the genome has been successfully assembled.

A total of 20,353 annotated gene models were regarded as the official gene set (OGS), with average CDS length of 1.4 kb and average exon number of 6.6, which are comparable to other published hemipteran insects (Table S5; Figure S4). Furthermore, based on ORTHODB version 9 of Insecta, 1,617 (97.5%) BUSCO genes were found in OGS (Figure S5), including 1,579 (95.2%) complete and 38 (2.3%) fragmented BUSCO genes, which was consistent with the BUSCO assessment result for the genome assembly (Simao et al., 2015; Figure 1b). This indicated that our gene prediction pipeline has successfully obtained nearly all potential genes in the assembled genome (Table S6). In the OGS, 18,721 (91.98%) genes have homology evidence in at least one database among KEGG, NR, EGGNOG, INTERPRO and SWISS-PROT (Figure S6). Additionally, a total of 427 tRNAs for all 20 essential amino acids and suppressor are detected (Table S7).

The highly continuous genome enables us to study TEs comprehensively. In total, 656 Mb (64.10%) TEs were annotated in the *A. lucorum* genome, which is much more than the closest relatives *C. lectularius* (218 Mb) and *N. lugens* (489 Mb) with similar genome sizes (Figure 1c and Table S5, S8). Moreover, the genome size of *A. lucorum* is also the largest of compared species, indicating that TEs extensively contribute to the genome size. In *A. lucorum*, LTR (98 Mb), LINE (73 Mb) and DNA (88 Mb) elements are the major types of TEs, and LTR is considerably in excess of that from other compared insects (Figure 1c). Considering that most other genomes are poorly continuous, the comparison of TE content here may be biased. The top three families of LTR are Gypsy (59 Mb), Pao (16 Mb) and Copia (9 Mb; Figure S7). In addition, we found that DNA and LINE transposons had a recent activity burst at 3% and 4% divergence rates, respectively (Figure S8). This could be essential in promoting adaptive evolution for *A. lucorum*, as the recent explosion of active TEs can bring more genetic variations for adaptive selection and are often related to the gene enrichment of functions such as stress response and adaptations (Liu et al., 2018).

3.2 | Gene expansion and recent gene burst promote environmental adaptability

To gain insights into an evolutionary perspective for *A. lucorum*, a whole genome-based phylogenetic analysis was performed with nine hemipteran insects *A. pisum*, *M. persicae*, *B. tabaci*, *A. lucorum*,

TABLE 1 Major indicators of the *Apolygus lucorum* genome

	Number	Size (bp)
Assembly feature		
Assembled contigs	3,818	1,023,255,709
Contigs N50		785,215
Anchored scaffolds	17	1,015,881,696
Scaffolds N50		68,132,828
Estimated genome size (by k-mer analysis)		1,030,903,443
Genome annotation		
Tandem repeats	468,796	72,964,816
Transposable elements	2,273,150	655,921,134
Gene models	20,353	27,622,289

Note: Each scaffold represents a natural chromosome.

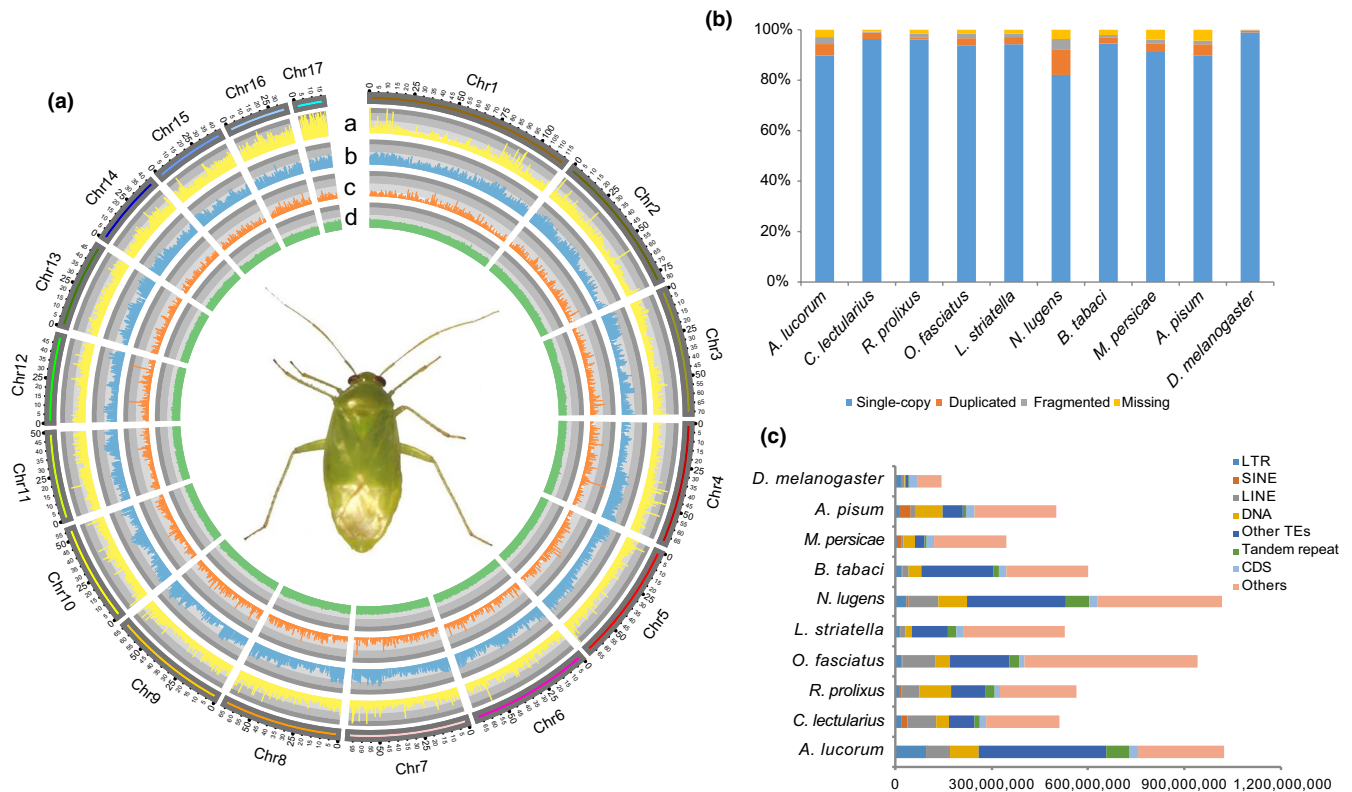


FIGURE 1 The genome landscape of *Apolygus lucorum*. (a) Circular representation of the chromosomes. Tracks a–d represents the distribution of tandem repeats density, transposable elements (TEs) density, gene density and GC density, respectively, with densities calculated in 500 kb windows. (b) BUSCO (Benchmarking Universal Single-Copy Orthologues) assessment of genomes of *A. lucorum* and other insects, using ORTHODB version 9 of insecta ($n = 1,658$). (c) The genomic composition of *A. lucorum* and other insects. TEs include LTR (long terminal repeats), SINE (short interspersed nuclear elements), LINE (long interspersed nuclear elements), DNA (DNA transposons), and Other TEs

C. lectularius, *Rhodnius prolixus*, *O. fasciatus*, *L. striatella*, *N. lugens*, and *D. melanogaster* as the outgroup (Table S9). All protein-coding genes from these ten species were used to construct gene orthologous groups, resulting in 13,933 orthologous groups. The phylogeny showed that *A. lucorum* diverged from *C. lectularius* about 168 million years ago (Ma) and from *A. pisum* about 275 MYA (Figure 2a).

Compared to the other eight hemipteran species, 1,086 expanded orthologous groups (rank sum test) containing 2,872 genes were identified in *A. lucorum*. Gene ontology analyses observed significant enriched GO terms (Fisher's exact test, p -value < .01) involved in odorant recognition, including sensory perception of smell (GO: 0007608) and sensory perception of chemical stimulus (GO: 0007606; Figure S9), which provided clues for the extremely broad host plant ranges of *A. lucorum*, as host plant selection mainly relies on odorant recognition for many herbivorous species. In addition, enriched GO terms associated with digestion in *A. lucorum* were also observed, such as hydrolase activity, acting on glycosyl bonds (GO: 001698), hydrolase activity, hydrolysing O-glycosyl compounds (GO: 0004553) and polygalacturonase (PG) activity (GO: 0004650). Particularly, PG is an essential enzyme for digestion, which hydrolyses pectin substances and then destroys plant cell walls (Marković & Janeček, 2001). In summary, these expanded genes could play an important role in the severe damage on a wide range of plants, as

PGs can hydrolyse the pectin substances and then destroy the plant cell walls and ORs could promote the pest search for host.

Novel genes were mostly generated from gene duplication, which is recognized as a driving force of evolution. Using a within-genome reciprocal best blast hit, 2,609 paralogue pairs were identified, and distribution of synonymous distances (K_s values) showed that 1,502 (58%) paralogue pairs had a K_s value smaller than 0.3, suggesting that most gene duplications possibly occurred in a recent period (Figure 2b). Among all the recent duplicated pairs, 89%, 8%, and 2% were classified as dispersed duplication, tandem duplication, and proximal duplication, respectively, indicating that recent duplicated genes in *A. lucorum* are mostly derived from small local scale gene duplications, instead of whole genome duplication (Figure 2c). Notably, 642 (42%) recent duplicated paralogue pairs are included in 675 (62%) expanded orthologous groups, so recent gene duplication events have made considerable contribution to the gene expansion in *A. lucorum*.

3.3 | Expansion of digestive enzyme genes promotes processing of diverse foods

Our genome-wide analysis showed that *A. lucorum* has a comprehensive digestive enzyme spectrum including 55 PGs, 49

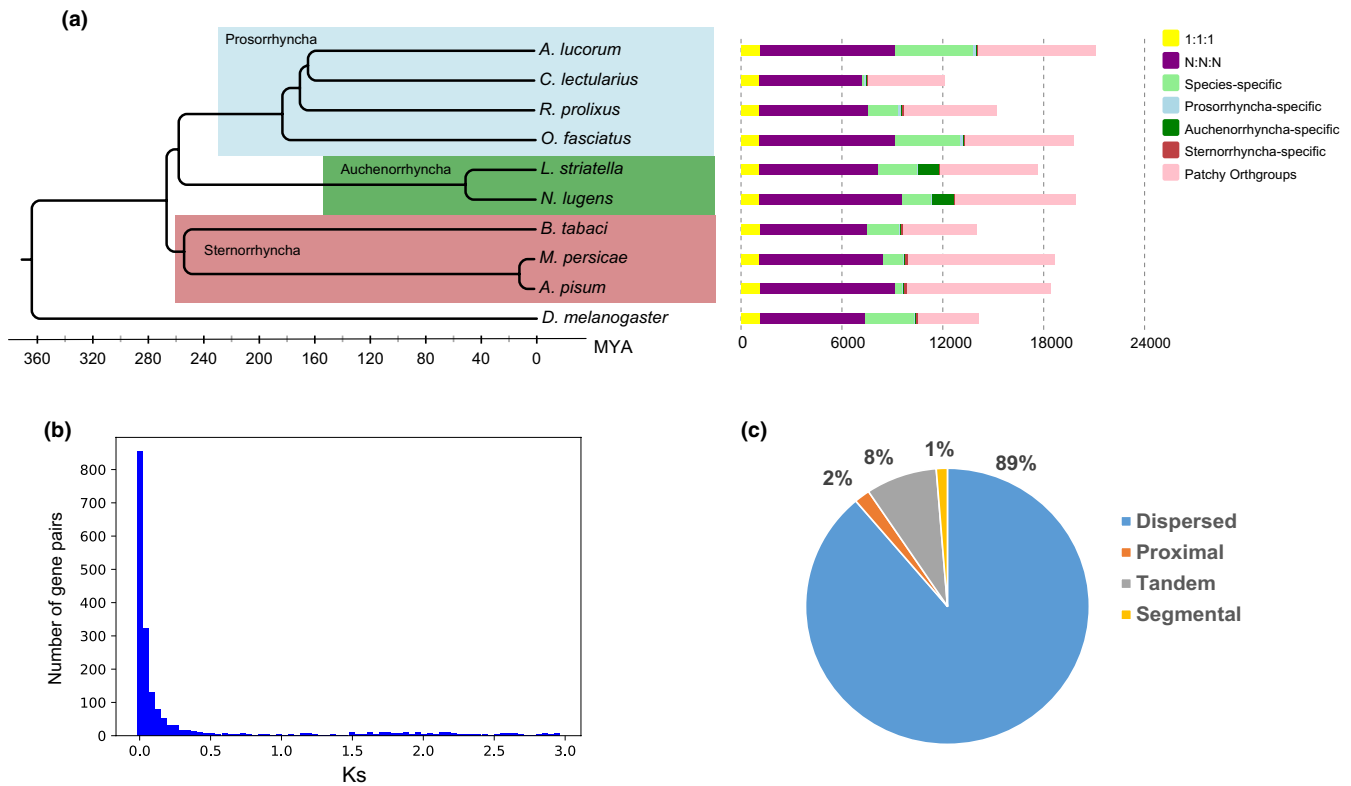


FIGURE 2 Genome evolution of *Apolygus lucorum*. (a) Phylogenetic relationships and gene orthology of *A. lucorum* with other insects. The maximum likelihood phylogenomic tree was calculated based on 1,252 single-copy universal genes. The coloured histogram indicates category of orthology as follows: '1:1:1', single-copy universal genes; 'N:N:N', multicopy universal genes; 'Prosorrhyncha-specific', genes specific to Prosorrhyncha-lineage; 'Auchenorrhyncha-specific', genes specific to Auchenorrhyncha lineage; 'Sternorrhyncha-specific', genes specific to Sternorrhyncha lineage; 'Species-specific', genes without an orthologue in any other species; 'Patchy', orthologues exist in part of species. (b) Distribution of synonymous mutation rate (Ks) values for paralogous gene pairs in *A. lucorum*, defined by reciprocal best blast hit. (c) Distribution of different types of gene duplication in *A. lucorum*

carboxypeptidases, 39 cathepsins, seven alpha-amylases, 224 serine proteases (SPs), 44 aminopeptidases, 48 phospholipases, 73 lipases, and 12 glucosidases (Table S10). Compared with the other eight hemipteran insects, *A. lucorum* had a comprehensive digestive enzyme spectrum, with a unique group of PGs and a significantly expanded group of SPs (Figure 3a).

PG is a group of plant cell wall-degrading enzyme, ubiquitous in fungi, bacteria, and plants. It is also found in Hemiptera and Coleoptera, predicted to be horizontally transferred from fungi (Wybouw et al., 2016; Xu et al., 2019). PG activity has been detected in the salivary glands of Hemiptera including some aphid species, but is especially common in mirid bugs (Zhang et al., 2015). Most notably, 55 unique PGs were detected in *A. lucorum*, but that were not found in the other eight Hemiptera including two aphids. Many PGs were arranged in tandem on the genome with high identity, which suggests recent replication (Figure S10). The duplication of PGs in *A. lucorum* might contribute to expanding the host plant range of this insect (Wybouw et al., 2016). The expression profile showed that 55 PGs were specifically expressed in salivary gland with high expression levels, indicating that the salivary gland of *A. lucorum* has a very high ability to synthesize PGs (Figure 3b). A large amount of multi-form PGs are secreted by salivary glands and injected into the plant to destroy the tissues, elicit

feeding-damage symptoms and induce much larger lesions than other piercing-sucking insects (Wheeler, 2001).

SPs are involved in various physiological processes of insects, such as digestion, development and innate immunity (Di Cera, 2009). The digestive SPs are essential in degrading proteins into free amino acids and inactivating toxic compounds from food sources. The expansion of SPs is the most obvious among the digestive enzyme genes of *A. lucorum*, and a total of 224 SPs were identified in the genome, the most of all nine Hemiptera species (Figure 3a), the phylogenetic tree of SPs within four Prosorrhyncha species exhibited most subclades are presented in *A. lucorum* (Figure S11). The type and abundance of digestive enzymes is related to the nature of the food source that an insect can assimilate (Agusti & Cohen, 2000). Carnivorous insects should secrete greater amounts of proteases than herbivorous insects (Zeng & Cohen, 2000). Some strictly phytophagous mirid bugs even lack detectable digestive proteases (Cohen & Wheeler, 1998). The expansion of SPs in *A. lucorum* can improve its digestive capacity and may contribute to its omnivorous feeding habit, mainly phytophagous with prey to complement (Li, et al., 2019; Xu et al., 2019). The slight expansion of SPs was also observed in bed bug, which is consistent with its carnivorous feeding habit. Transcriptome analysis showed that 201 (90%) SPs were

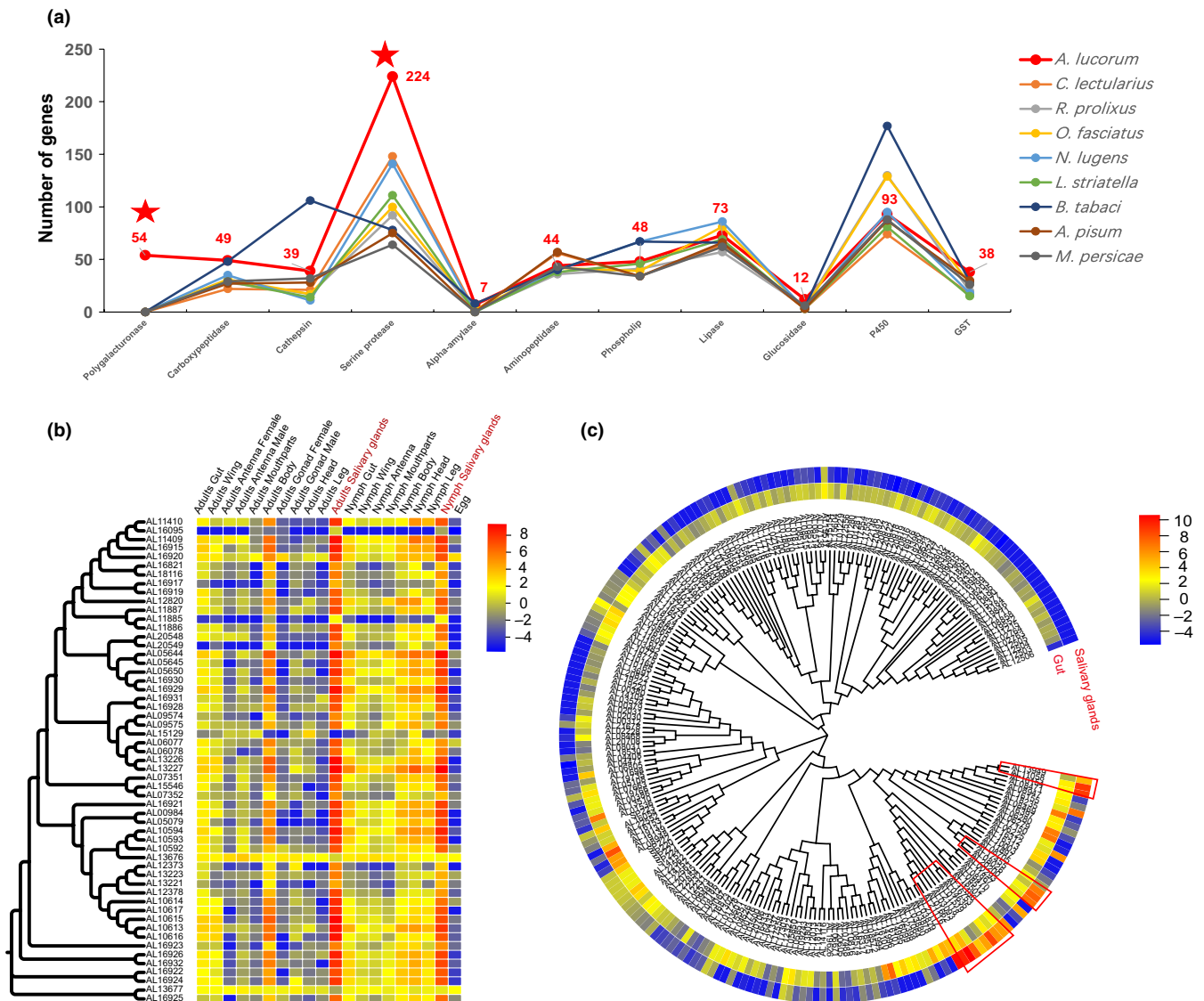


FIGURE 3 Miridae-specific polygalacturonases (PGs) and expansion of serine proteases (SPs) elucidate omnivorousness of *Apolygus lucorum*. (a) Gene number comparison of digestive enzymes between *A. lucorum* and other eight Hemiptera species. PGs and SPs are marked with red star. (b) Phylogenetic gene tree and the expression profile of PGs in *A. lucorum*. Each data block shows the base 10 logarithm of TPM (\log_{10} TPM) value of the corresponding tissue organ. Salivary gland of adult and nymph with high expression level are marked with red colour. (c) Phylogenetic gene tree and expression profile of SPs in *A. lucorum*. The three clades containing the SPs highly expressed in salivary gland are marked with red square frame

specifically expressed in the gut, in contrast, the other 21 SPs that clustered in three clades (Figure 3c) were highly expressed in the salivary gland and the expression levels were much higher than the SPs expressed in the gut, indicating that these 21 SPs may function in digestion and degrading toxic compounds from the plants in the mesophyll feeding process of *A. lucorum*.

3.4 | Rapid evolution of chemosensory receptors expands the range of host plants

Insects have developed a highly efficient chemosensory recognition system to recognize and distinguish chemical signals that are essential to host plant selection (Blomquist & Vogt, 2003). Reception of

chemical signals in insects is mediated by three families of chemosensory receptors, including ORs, GRs, and IRs. The size of chemosensory receptor families is largely related to the complexity of chemical signals the insect detects. A large number of chemosensory receptors containing 135 ORs, 57 GRs and 33 IRs were identified in the *A. lucorum* genome, indicating that this insect has a sensitive and specific perception system to distinguish complicated chemical signals in the environment (Figure 4a).

OR is the first chemosensory receptor gene family identified in insects, which locates in the dendritic membrane of odorant receptor neurons and function as a heterodimer with a highly conserved noncanonical OR coreceptor (Orco). The number of ORs varies considerably in different insects, from ten in the body louse to up to 400 in the ants, and protein sequence identities of ORs are also widely

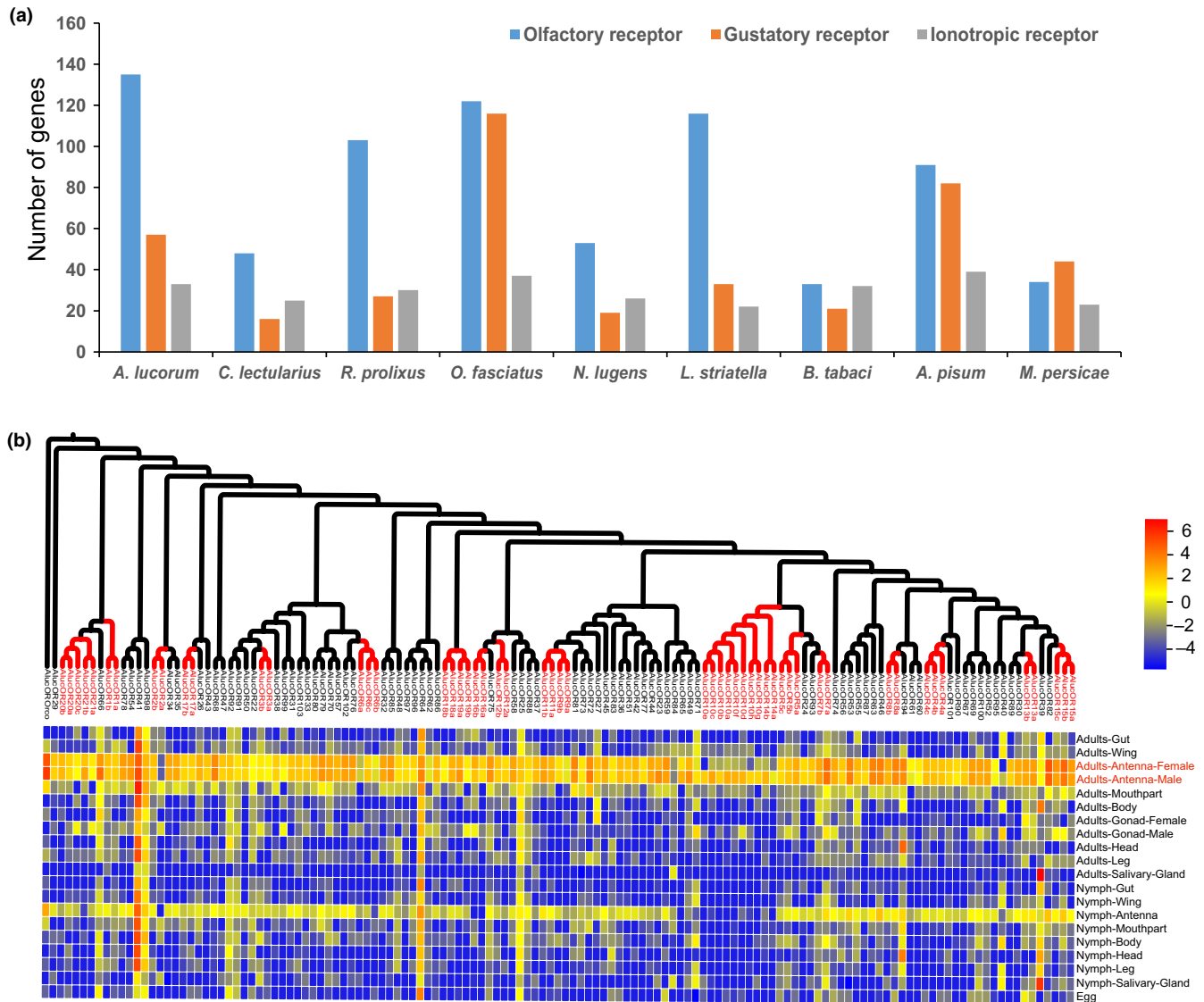


FIGURE 4 Rapid evolution of ORs in *Apolygus lucorum*. (a) Gene number comparison of chemosensory receptor genes between *A. lucorum* and other eight species in Hemiptera. (b) Phylogenetic gene tree and expression profile of olfactory receptors in *A. lucorum*. Each data block shows the base 10 logarithm of TPM (log₁₀ TPM) value of the corresponded tissue organ. High sequence identity (>80%) clades are marked with red colour. Most ORs showed high expression in Adults-Antenna-Female and Adults-Antenna-Male which marked with red colour

divergent, suggesting ORs are minimally conserved genes in insects (Kirkness et al., 2010; Zhou et al., 2012). We identified 135 ORs in *A. lucorum*, the most among the nine Hemiptera species. ORs expanded three-fold in *A. lucorum* compared with the closely related species *C. lectularius*. The phylogenetic analysis that showed 40% of the OR genes (55) were contained in several clades with high protein sequence identity (>80%), indicating that OR has experienced recent gene replication (Figure 4b). Transcriptome analysis showed that 99% of ORs were highly expressed in antenna, as antenna is the major olfactory perception organ in insects. The expansion of ORs, mostly due to the recent replication in *A. lucorum*, may be related to its extensive hosts and host-switching behaviour.

GRs, another kind of chemosensory receptor gene, are expressed in gustatory receptor neurons in taste organs (Clyne et al., 2000).

Unlike ORs, which mainly sense volatile compounds, these GRs sense nonvolatile compounds that the insects directly contact with gustatory sensilla, including carbon dioxide (CO₂), sugars, bitter compounds, and salts (Vosshall & Stocker, 2007). Similar to ORs, the number of GRs also greatly varies in different insects, and 237 GRs were reported in *Spodoptera litura* from Lepidoptera (Cheng et al., 2017). In contrast, only 10 GRs were found in the honey bee *Apis mellifera* from Hymenoptera (Robertson & Wanner, 2006), associated with their different feeding habits. Our results show that the number of GRs varies greatly in different Hemiptera, from 21 in *L. striatella* and *N. lugens* to 116 in *O. fasciatus*. GRs showed significant expansion in *A. lucorum* compared to the two related carnivorous species *C. lectularius* (16) and *R. prolixus* (27), indicating that phytophagous insects have more GRs than carnivorous insects, which is

consistent with the adaptation of phytophagous insects to multiple hosts. Expression profile analysis showed that GRs exhibit different expression patterns and most GRs are expressed in various tissues (Figure S12), which is consistent with the previous reports in *C. lectularius* and *O. fasciatus*, indicating the diversity of GR functions.

IRs, which evolved from the ionotropic glutamate receptor superfamily (iGluRs), were recently identified as involved in chemosensory reception (Benton et al., 2007). Different from ORs or GRs, IRs exhibit diverse functions and participate in olfactory, gustatory and auditory sensation. The sequences of IRs are more conserved than OR and GR, and the numbers of IRs are also more stable in different insect species. We identified 33 IRs in *A. lucorum*, comparable to the other eight Hemiptera species. Expression profile studies showed that 88% of IRs are expressed in antennae, suggesting that most IRs have olfactory functions. However, some IRs were highly expressed in tissues other than antennae, reflecting the diversity of IR functions (Figure S13).

3.5 | Expansion of detoxification enzymes contributes to degrading toxin

Agricultural pests usually employ an efficient detoxification system containing various enzymes to overcome numerous toxins in food sources or the environment. We performed a genome-wide analysis of the two important detoxification enzymes GST and P450 to study the mechanism of detoxification in *A. lucorum*.

GST is a superfamily of multifunctional isoenzymes involved in the cellular detoxification of various physiological and xenobiotic

substances, which is highly related to insecticide resistance in insects, as it can directly detoxify the insecticides (Fang, 2012). A total of 38 GSTs were identified in *A. lucorum*, which is the highest of the nine Hemiptera species (Figure 5a, Table S11). Expression profile analysis showed that most GST genes were expressed in various tissues, in line with their detoxification function (Figure S14). The number of GSTs was more than twice as high as the other three closely related true bug species *C. lectularius* (17), *R. prolixus* (20), and *O. fasciatus* (27), indicating a significant expansion in *A. lucorum*. Phylogenetic analysis of the GSTs of four true bugs showed three *A. lucorum*-specific branches, which contained 58% GST genes (22; Figure 5b), suggesting the GSTs experienced a recent species-specific expansion in *A. lucorum*, enabling better detoxification of toxic substances and adaptation to the environment.

P450s constitute the largest and most functionally diverse class of insect detoxification enzymes, including four distinct clades CYP2, CYP3, CYP4 and CYPMit0 (Li et al., 2007). A total of 93 P450s containing seven CYP2, 47 CYP3, 32 CYP4, and seven CYPMit0 have been identified in the genome of *A. lucorum*. Most P450s also showed similar expression patterns as GSTs and were simultaneously expressed in multiple tissues (Table S11, Figure S15). Members of the CYP3 clade have been implicated in the oxidative detoxification of plant secondary metabolites and synthetic insecticides. CYP3s were significantly expanded in *A. lucorum* compared with the closely related species *C. lectularius*, but were comparable to the gene numbers of other two true bugs *R. prolixus* and *O. fasciatus*, which have different feeding habits. The phylogenetic tree of P450s exhibited

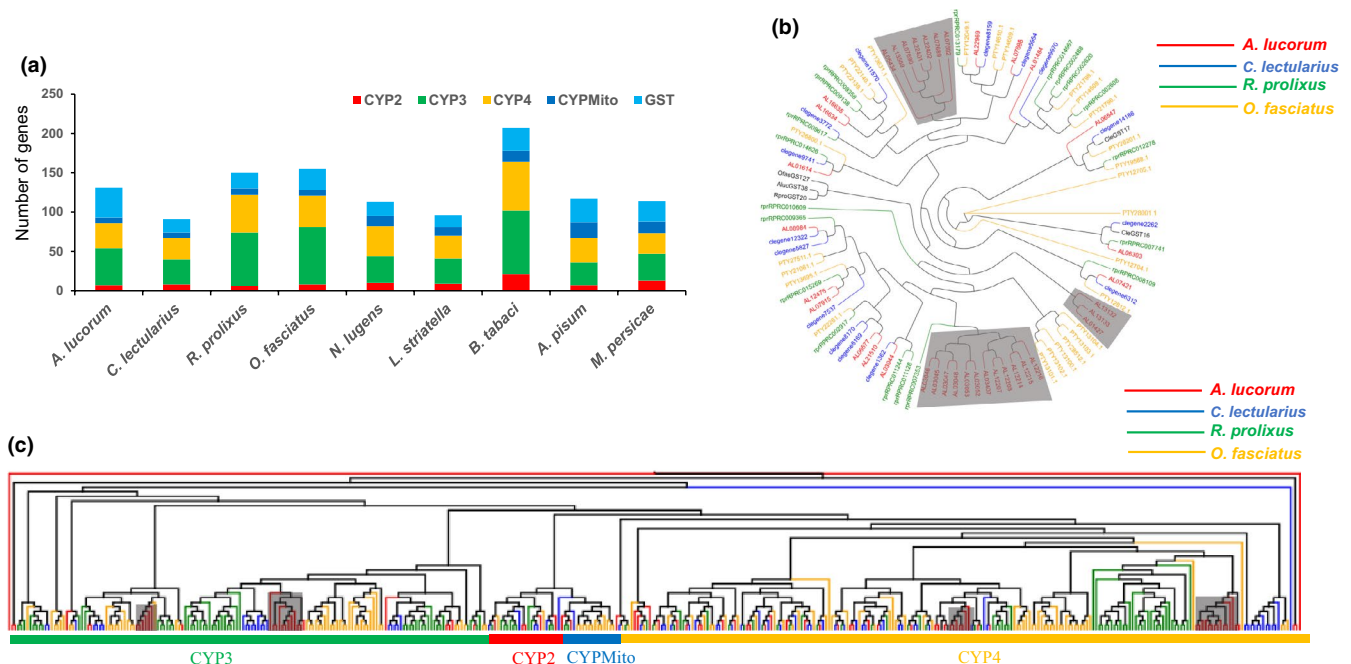


FIGURE 5 Expansion of P450 and GST in *Apolygus lucorum*. (a) Gene number comparison of P450 and GST between *A. lucorum* and other eight species in Hemiptera. (b) Phylogenetic tree of GSTs in Prosorrhyncha. Species-specific expanded clades in *A. lucorum* are emphasized with shadow. (c) Phylogenetic tree of four subfamily of P450 in Prosorrhyncha. Species-specific expanded clades in *A. lucorum* are emphasized with shadow

four *A. lucorum*-specific branches (Figure 5c), suggesting that P450s experienced similar species-specific expansion as GSTs, enhancing the detoxification activity in *A. lucorum*.

4 | DISCUSSION

Although long and ultra-long reads from the third-generation sequencing technologies have been widely adopted, large animal genomes with high heterozygosity still pose great challenges for genome assembly. Here, we present a 1.02 Gb chromosome-scale reference genome of *Apolygus lucorum* accomplished using PacBio and Hi-C data. Due to the limitations in available sequencing and assembly technologies, the current genome assembly is still far from perfect. Future sequencing that can decrease the required amount of DNA to achieve single individual sequencing, and improvements in the assembly algorithms on resolving heterozygosity problem, will contribute to improving assembly quality. With the genome sequence of *A. lucorum*, we are able to comprehensively characterize the TEs and find a recent explosion of DNA and LINE TEs. In addition, we identified a wave of recent gene duplications, which may be responsible for the damage capacity of *A. lucorum*. As one of a few species in Hemiptera with published genomes and the first sequenced mirid bug species, *A. lucorum* has the potential to become a model species of Miridae, and the reference genome of *A. lucorum* not only facilitates biological studies of Hemiptera, but also contributes to in-depth studies of damage mechanisms of agricultural pests.

The *A. lucorum* genome can be used to uncover the molecular and genetic mechanism of specific biological problems in Miridae. At the genomic level, we clarified that PG is unique in mirid bugs, which is the reason for the serious damage of *A. lucorum* caused by its mesophyll feeding; moreover, these PG genes are also specific targets for the control of this important pest. Differences in the living environments and eating habits exhibited a great influence on the abundance of related genes, as many genes associated to digestion, chemosensory perception, and detoxification expanded in *A. lucorum* compared to other Hemiptera species. Consistent with its phytozoophagy, the numbers of plant-feeding associated genes identified in *A. lucorum* are more similar to distantly related phytophagous species than the two closely related carnivorous species. In contrast, the numbers of the carnivorous-related genes in *A. lucorum* are similar to those of carnivorous species, which reflects the environmental plasticity of the insect genome. As the number of sequenced genomes in Miridae increases, the environmental adaption and damage mechanism will be further clarified in the future.

In recent years, with the worldwide expansion of Bt cotton planting area, the damage caused by mirid bugs gradually increased (Wu et al., 2002). Various ecological and biological strategies have been used to control *A. lucorum*, with some successes (Lu & Wu, 2011a; Lu, et al., 2007). But these management strategies have not reached the optimal effect, so the control of *A. lucorum* still relies

on the application of synthetic insecticides (Lu & Wu, 2011b; Zhen et al., 2018). Large-scale and long-term use of insecticides results in several ecological and environmental problems, especially increased resistance to insecticides, and *A. lucorum* was reported to have different levels of resistance to many insecticides (Zhen et al., 2018). There is an urgent need for developing new strategies to control this pest, and the genome data provide us abundant gene targets associated with specific biological processes in *A. lucorum*. By targeting these key genes, RNA interference (RNAi), genetic regulation, behavioural regulation and other environmentally friendly strategies can be introduced to control *A. lucorum*. Therefore, the genome and transcriptome data as well as the candidate drug targets provide a basis to develop efficient control technologies for mirid bug pests, protecting global agricultural safety.

ACKNOWLEDGEMENTS

This work was funded by National Natural Science Foundation of China (31621064 to K.W., and 31672095 to Y.L.), National Key R&D Program of China (2017YFD0200400 to YL, and 2016YFC1200600 to W.F.), Agricultural Science and Technology Innovation Program & The Elite Young Scientists Program of CAAS, Fundamental Research Funds for Central Non-profit Scientific Institution (No. Y2017JC01), the Agricultural Science and Technology Innovation Program Cooperation and Innovation Mission (CAAS-XTX2016), and the Fund of Key Laboratory of Shenzhen (ZDSYS20141118170111640). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

K.W., W.F., and G.W. designed and led the research. Yang. L., B.Y., Bi.L., S.Z., X.Z., B.W., and Yan. L. prepared DNA and RNA for sequencing. H.W., Bo.L., Y.Z., F.J., Y.R., S.W., and W.F. assembled the genome and generated the gene set. Yang. L., H.L., T.H., and B.Y. annotated the gene families. Yang. L. and H.L. analysed the transcriptome. Yang. L., H.L., H.W., G.W., W.F., and K.W. wrote the manuscript.

DATA AVAILABILITY STATEMENT

All the raw sequencing data and genome data in this study have been deposited at NCBI as a BioProject under accession PRJNA526332. Genomic sequence reads have been deposited in the SRA database with BioSample: SAMN11095929. Transcriptome sequence reads have been deposited in the SRA database with BioSample: SAMN13219540. Genome assembly and gene annotation has been deposited at DDBJ/ENA/GenBank under the whole genome shotgun project accession WIXP00000000. The version described in this paper is version WIXP02000000. The genome assemblies and annotation files are also available at the website ftp://ftp.agis.org.cn/~fanwei/Apolygus_lucorum_Genome/.

ORCID

Wei Fan  <https://orcid.org/0000-0001-5036-8733>

Guirong Wang  <https://orcid.org/0000-0002-3794-0106>

REFERENCES

- Agusti, N., & Cohen, A. C. (2000). *Lygus hesperus* and *L. lineolaris* (Hemiptera : Miridae), phytophages, zoophages, or omnivores: Evidence of feeding adaptations suggested by the salivary and mid-gut digestive enzymes. *Journal of Entomological Science*, 35(2), 176–186. <https://doi.org/10.18474/0749-8004-35.2.176>
- Benoit, J. B., Adelman, Z. N., Reinhardt, K., Dolan, A., Poelchau, M., Jennings, E. C., Szuter, E. M., Hagan, R. W., Gujar, H., Shukla, J. N., Zhu, F., Mohan, M., Nelson, D. R., Rosendale, A. J., Derst, C., Resnik, V., Wernig, S., Menegazzi, P., Wegener, C., ... Richards, S. (2016). Unique features of a global human ectoparasite identified through sequencing of the bed bug genome. *Nature Communications*, 7, 10165. <https://doi.org/10.1038/Ncomms10165>
- Benson, G. (1999). Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Research*, 27(2), 573–580. <https://doi.org/10.1093/nar/27.2.573>
- Benton, R., Vannice, K. S., & Vossell, L. B. (2007). An essential role for a CD36-related receptor in pheromone detection in *Drosophila*. *Nature*, 450(7167), 289–293. <https://doi.org/10.1038/nature06328>
- Blomquist, G. J., & Vogt, R. G. (2003). In G. J. Blomquist & R. G. Vogt (Eds.), *Biosynthesis and detection of pheromones and plant volatiles-introduction and overview. Insect pheromone biochemistry and molecular biology: The biosynthesis and detection of pheromones and plant volatiles*. Academic Press.
- 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), 1119–1125. <https://doi.org/10.1038/nbt.2727>
- Cassis, G., & Schuh, R. T. (2012). Systematics, biodiversity, biogeography, and host associations of the Miridae (Insecta: Hemiptera: Heteroptera: Cimicomorpha). *Annual Review of Entomology*, 57, 377–404. <https://doi.org/10.1146/annurev-ento-121510-133533>
- Cheng, T., Wu, J., Wu, Y., Chilukuri, R. V., Huang, L., Yamamoto, K., Feng, L. I., Li, W., Chen, Z., Guo, H., Liu, J., Li, S., Wang, X., Peng, L. I., Liu, D., Guo, Y., Fu, B., Li, Z., Liu, C., ... Mita, K. (2017). Genomic adaptation to polyphagy and insecticides in a major East Asian noctuid pest. *Nature Ecology & Evolution*, 1, 1747–1756. <https://doi.org/10.1038/s41559-017-0314-4>
- Clyne, P. J., Warr, C. G., & Carlson, J. R. (2000). Candidate taste receptors in *Drosophila*. *Science*, 287(5459), 1830–1834. <https://doi.org/10.1126/science.287.5459.1830>
- Cohen, A. C., & Wheeler, A. G. (1998). Role of saliva in the highly destructive fourlined plant bug (Hemiptera : Miridae : Mirinae). *Annals of the Entomological Society of America*, 91(1), 94–100. <https://doi.org/10.1093/aesa/91.1.94>
- Di Cera, E. (2009). Serine proteases. *IUBMB Life*, 61(5), 510–515. <https://doi.org/10.1002/iub.186>
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Emms, D. M., & Kelly, S. (2015). OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, 16, 157. <https://doi.org/10.1186/s13059-015-0721-2>
- Fang, S. (2012). Insect glutathione S-transferase: A review of comparative genomic studies and response to xenobiotics. *Bulletin of Insectology*, 65(2), 265–271.
- Frati, F., Galletti, R., De Lorenzo, G., Salerno, G., & Conti, E. (2006). Activity of endo-polygalacturonases in mirid bugs (Heteroptera : Miridae) and their inhibition by plant cell wall proteins (PGIPs). *European Journal of Entomology*, 103(3), 515–522. <https://doi.org/10.14411/Eje.2006.067>
- Guan, D., McCarthy, S. A., Wood, J., Howe, K., Wang, Y., & Durbin, R. (2020). Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*, 36(9), 2896–2898. <https://doi.org/10.1093/bioinformatics/btaa025>
- Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, 52(5), 696–704. <https://doi.org/10.1080/10635150390235520>
- He, Z., Zhang, H., Gao, S., Lercher, M. J., Chen, W.-H., & Hu, S. (2016). Evolvview v2: An online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Research*, 44(W1), W236–W241. <https://doi.org/10.1093/nar/gkh340>
- Jiang, Y., Lu, Y., & Zeng, J. (2015). *Forecast and management of mirid bugs in multiple agroecosystems of China*. China Agriculture Press.
- Jung, S., & Lee, S. (2012). Molecular phylogeny of the plant bugs (Heteroptera: Miridae) and the evolution of feeding habits. *Cladistics*, 28(1), 50–79. <https://doi.org/10.1111/j.1096-0031.2011.00365.x>
- Kennedy, G. G., & Storer, N. P. (2000). Life systems of polyphagous arthropod pests in temporally unstable cropping systems. *Annual Review of Entomology*, 45, 467–493. <https://doi.org/10.1146/annurev.ento.45.1.467>
- Kirkness, E. F., Haas, B. J., Sun, W., Braig, H. R., Perotti, M. A., Clark, J. M., Lee, S. H., Robertson, H. M., Kennedy, R. C., Elhaik, E., Gerlach, D., Kriventseva, E. V., Elisk, C. G., Graur, D., Hill, C. A., Veenstra, J. A., Walenz, B., Tubio, J. M. C., Ribeiro, J. M. C., ... Pittendrigh, B. R. (2010). Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proceedings of the National Academy of Sciences of the United States of America*, 107(27), 12168–12173. <https://doi.org/10.1073/pnas.1003379107>
- Knyaz, C., Stecher, G., Li, M., Kumar, S., & 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>
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*, 27(5), 722–736. <https://doi.org/10.1101/gr.215087.116>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
- Letunic, I., & Bork, P. (2016). Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research*, 44(W1), W242–W245. <https://doi.org/10.1093/nar/gkw290>
- 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. <https://doi.org/10.1186/1471-2105-12-323>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, W., Wang, L., Jaworski, C. C., Yang, F., Liu, B., Jiang, Y., Lu, Y., Wu, K., & Desneux, N. (2019). The outbreaks of nontarget mirid bugs promote arthropod pest suppression in Bt cotton agroecosystems. *Plant Biotechnology Journal*, 18, 322–324. <https://doi.org/10.1111/pbi.13233>
- Li, W., Wyckhuys, K. A., & Wu, K. (2016). Does feeding behavior of a zoophagous mirid differ between host plant and insect prey items? *Arthropod-Plant Interactions*, 10(1), 79–86. <https://doi.org/10.1007/s11829-015-9410-z>
- Li, X., Schuler, M. A., & Berenbaum, M. R. (2007). Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. *Annual*

- Review of Entomology, 52, 231–253. <https://doi.org/10.1146/annurev.ento.51.110104.151104>
- Li, Y., Park, H., Smith, T. E., & Moran, N. A. (2019). Gene family evolution in the pea aphid based on chromosome-level genome assembly. *Molecular Biology and Evolution*, 36(10), 2143–2156. <https://doi.org/10.1093/molbev/msz138>
- Liu, C., Zhang, Y., Ren, Y., Wang, H., Li, S., Jiang, F., Yin, L., Qiao, X. I., Zhang, G., Qian, W., Liu, B. O., & Fan, W. (2018). The genome of the golden apple snail *Pomacea canaliculata* provides insight into stress tolerance and invasive adaptation. *Gigascience*, 7(9). <https://doi.org/10.1093/gigascience/giy101>
- Lu, Y., Jiao, Z., & Wu, K. (2012). Early season host plants of *Apolygus lucorum* (Heteroptera: Miridae) in northern China. *Journal of Economic Entomology*, 105(5), 1603–1611. <https://doi.org/10.1603/ec12003>
- Lu, Y., Liang, G., & Wu, K. (2007). Advances in integrated management of cotton mirids. *Plant Protection*, 33(6), 10–15.
- Lu, Y., Qiu, F., Feng, H., Li, H., Yang, Z., Wyckhuys, K., & Wu, K. (2008). Species composition and seasonal abundance of pestiferous plant bugs (Hemiptera: Miridae) on Bt cotton in China. *Crop Protection*, 27(3–5), 465–472. <https://doi.org/10.1016/j.cropro.2007.07.017>
- Lu, Y., & Wu, K. (2008). *Biology and control of cotton mirids*. Golden Shield.
- Lu, Y., & Wu, K. (2011a). Effect of relative humidity on population growth of *Apolygus lucorum* (Heteroptera: Miridae). *Applied Entomology and Zoology*, 46(3), 421–427. <https://doi.org/10.1007/s13355-011-0058-6>
- Lu, Y., & Wu, K. (2011b). Mirid bugs in China: Pest status and management strategies. *Outlooks on Pest Management*, 22(6), 248–252. <https://doi.org/10.1564/22dec02>
- Lu, Y., Wu, K., & Guo, Y. (2007). Flight potential of *Lygus lucorum* (Meyer-Dur) (Heteroptera: Miridae). *Environmental Entomology*, 36(5), 1007–1013. <https://doi.org/10.1093/ee/36.5.1007>
- Lu, Y., Wu, K., Jiang, Y., Xia, B., Li, P., Feng, H., Wyckhuys, K. A. G., & Guo, Y. (2010). Mirid bug outbreaks in multiple crops correlated with wide-scale adoption of Bt cotton in China. *Science*, 328(5982), 1151–1154. <https://doi.org/10.1126/science.1187881>
- Lu, Y., Wu, K., Wyckhuys, K. A., & Guo, Y. (2009). Comparative flight performance of three important pest *Adelphocoris* species of Bt cotton in China. *Bulletin of Entomological Research*, 99(6), 543–550. <https://doi.org/10.1017/S000748530800655X>
- Marković, O., & Janeček, Š. (2001). Pectin degrading glycoside hydrolases of family 28: Sequence-structural features, specificities and evolution. *Protein Engineering, Design and Selection*, 14(9), 615–631. <https://doi.org/10.1093/protein/14.9.615>
- Mathers, T. C., Chen, Y., Kaithakottil, G., Legeai, F., Mugford, S. T., Baa-Puyoulet, P., Bretaudeau, A., Clavijo, B., Colella, S., Collin, O., Dalmay, T., Derrien, T., Feng, H., Gabaldón, T., Jordan, A., Julca, I., Kettles, G. J., Kowitzanich, K., Lavenier, D., ... Hogenhout, S. A. (2017). Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species. *Genome Biology*, 18(1), 27. <https://doi.org/10.1186/s13059-016-1145-3>
- Mesquita, R. D., Vionette-Amaral, R. J., Lowenberger, C., Rivera-Pomar, R., Monteiro, F. A., Minx, P., & Oliveira, P. L. (2015). Genome of *Rhodnius prolixus*, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. *Proceedings of the National Academy of Sciences of the United States of America*, 112(48), 14936–14941. <https://doi.org/10.1073/pnas.1506226112>
- Misof, B., Liu, S., Meusemann, K., Peters, R. S., Donath, A., Mayer, C., Frandsen, P. B., Ware, J., Flouri, T., Beutel, R. G., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J., Aberer, A. J., Aspöck, U., Aspöck, H., Bartel, D., ... Zhou, X. (2014). Phylogenomics resolves the timing and pattern of insect evolution. *Science*, 346(6210), 763–767. <https://doi.org/10.1126/science.1257570>
- Pan, H., Liu, B., Lu, Y., & Wyckhuys, K. A. (2015). Seasonal alterations in host range and fidelity in the polyphagous mirid bug, *Apolygus lucorum* (Heteroptera: Miridae). *PLoS One*, 10(2), e0117153. <https://doi.org/10.1371/journal.pone.0117153>
- Panfilio, K. A., Vargas Jentzsch, I. M., Benoit, J. B., Erezylmaz, D., Suzuki, Y., Colella, S., Robertson, H. M., Poelchau, M. F., Waterhouse, R. M., Ioannidis, P., Weirauch, M. T., Hughes, D. S. T., Murali, S. C., Werren, J. H., Jacobs, C. G. C., Duncan, E. J., Armisen, D., Vreede, B. M. I., Baa-Puyoulet, P., ... Richards, S. (2019). Molecular evolutionary trends and feeding ecology diversification in the Hemiptera, anchored by the milkweed bug genome. *Genome Biology*, 20(1), 64. <https://doi.org/10.1186/s13059-019-1660-0>
- Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., & Lopez, R. (2005). InterProScan: Protein domains identifier. *Nucleic Acids Research*, 33, W116–W120. <https://doi.org/10.1093/nar/gki442>
- Richards, S., Gibbs, R. A., Gerardo, N. M., Moran, N., Nakabachi, A., Richards, S., Stern, D., Tagu, D., Wilson, A. C. C., Muzny, D., Kovar, C., Cree, A., Chacko, J., Chandrabose, M. N., Dinh, H. H., Gabis, R. A., Hines, S., Hume, J., Jhangian, S. N., ... Hunter, W. (2010). Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLOS Biology*, 8(2), e3000029. <https://doi.org/10.1371/journal.pbio.1000313>
- Robertson, H. M., & Wanner, K. W. (2006). The chemoreceptor superfamily in the honey bee, *Apis mellifera*: Expansion of the odorant, but not gustatory, receptor family. *Genome Research*, 16(11), 1395–1403. <https://doi.org/10.1101/gr.5057506>
- Rosenfeld, J. A., Reeves, D., Brugler, M. R., Narechania, A., Simon, S., Durrett, R., Foox, J., Shianna, K., Schatz, M. C., Gandara, J., Afshinnikoo, E., Lam, E. T., Hastie, A. R., Chan, S., Cao, H., Saghibini, M., Kentsis, A., Planet, P. J., Kholodovych, V., ... Mason, C. E. (2016). Genome assembly and geospatial phylogenomics of the bed bug *Cimex lectularius*. *Nature Communications*, 7, 10164. <https://doi.org/10.1038/ncomms10164>
- 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, 259. <https://doi.org/10.1186/s13059-015-0831-x>
- Sharma, A., Khan, A. N., Subrahmanyam, S., Raman, A., Taylor, G. S., & Fletcher, M. J. (2014). Salivary proteins of plant-feeding hemipteroids: Implication in phytophagy. *Bulletin of Entomological Research*, 104(2), 117–136. <https://doi.org/10.1017/S0007485313000618>
- 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>
- Sloan, D. B., Nakabachi, A., Richards, S., Qu, J., Murali, S. C., Gibbs, R. A., & Moran, N. A. (2014). Parallel histories of horizontal gene transfer facilitated extreme reduction of endosymbiont genomes in sap-feeding insects. *Molecular Biology and Evolution*, 31(4), 857–871. <https://doi.org/10.1093/molbev/msu004>
- Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006). AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Research*, 34, W435–W439. <https://doi.org/10.1093/nar/gkl200>
- Vosshall, L. B., & Stocker, R. E. (2007). Molecular architecture of smell and taste in *Drosophila*. *Annual Review of Neuroscience*, 30, 505–533. <https://doi.org/10.1146/annurev.neuro.30.051606.094306>
- 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. <https://doi.org/10.1371/journal.pone.0112963>
- Wang, D., Zhang, Y., Zhang, Z., Zhu, J., & Yu, J. (2010). KaKs_Calculator 2.0: A toolkit incorporating gamma-series methods and sliding window strategies. *Genomics, Proteomics & Bioinformatics*, 8(1), 77–80. [https://doi.org/10.1016/s1672-0229\(10\)60008-3](https://doi.org/10.1016/s1672-0229(10)60008-3)

- Wang, L., Tang, N., Gao, X., Chang, Z., Zhang, L., Zhou, G., Guo, D., Zeng, Z., Li, W., Akinyemi, I. A., Yang, H., & Wu, Q. (2017). Genome sequence of a rice pest, the white-backed planthopper (*Sogatella furcifera*). *Gigascience*, 6(1), 1–9. <https://doi.org/10.1093/gigascience/giw004>
- Wang, P., Li, H., Wang, Y., Zhang, J., Dai, X., Chang, J., & Cai, W. (2014). The mitochondrial genome of the plant bug *Apolygus lucorum* (Hemiptera: Miridae): Presently known as the smallest in Heteroptera. *Insect Science*, 21(2), 159–173. <https://doi.org/10.1111/1744-7917.12029>
- Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., Lee, T.-H., Jin, H., Marler, B., Guo, H., Kissinger, J. C., & Paterson, A. H. (2012). MCSanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, 40(7), e49. <https://doi.org/10.1093/nar/gkr1293>
- Wenger, J. A., Cassone, B. J., Legeai, F., Johnston, J. S., Bansal, R., Yates, A. D., Coates, B. S., Pavinato, V. A. C., & Michel, A. (2017). Whole genome sequence of the soybean aphid, *Aphis glycines*. *Insect Biochemistry and Molecular Biology*, 123, 102917. <https://doi.org/10.1016/j.ibmb.2017.01.005>
- Wheeler, A. G. (2001). *Biology of the plant bugs (Hemiptera: Miridae): Pests, predators, opportunists*. Cornell University Press.
- Wu, K., & Guo, Y. (2005). The evolution of cotton pest management practices in China. *Annual Review of Entomology*, 50, 31–52. <https://doi.org/10.1146/annurev.ento.50.071803.130349>
- Wu, K., Li, W., Feng, H., & Guo, Y. (2002). Seasonal abundance of the mirids, *Lygus lucorum* and *Adelphocoris* spp. (Hemiptera Miridae) on Bt cotton in northern China. *Crop Protection*, 21, 997–1002. [https://doi.org/10.1016/S0261-2194\(02\)00080-7](https://doi.org/10.1016/S0261-2194(02)00080-7)
- Wybouw, N., Pauchet, Y., Heckel, D. G., & Van Leeuwen, T. (2016). Horizontal gene transfer contributes to the evolution of arthropod herbivory. *Genome Biology and Evolution*, 8(6), 1785–1801. <https://doi.org/10.1093/gbe/evw119>
- Xu, P., Lu, B., Liu, J., Chao, J., Donkersley, P., Holdbrook, R., & Lu, Y. (2019). Duplication and expression of horizontally transferred polygalacturonase genes is associated with host range expansion of mirid bugs. *BMC Evolutionary Biology*, 19, 12. <https://doi.org/10.1186/s12862-019-1351-1>
- Xue, J., Zhou, X., Zhang, C.-X., Yu, L.-L., Fan, H.-W., Wang, Z., Xu, H.-J., Xi, Y. U., Zhu, Z.-R., Zhou, W.-W., Pan, P.-L., Li, B.-L., Colbourne, J. K., Noda, H., Suetsugu, Y., Kobayashi, T., Zheng, Y., Liu, S., Zhang, R., ... Cheng, J.-A. (2014). Genomes of the rice pest brown planthopper and its endosymbionts reveal complex complementary contributions for host adaptation. *Genome Biology*, 15(12), 521. <https://doi.org/10.1186/s13059-014-0521-0>
- Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24(8), 1586–1591. <https://doi.org/10.1093/molbev/msm088>
- Yuan, W., Li, W., Li, Y., & Wu, K. (2013). Combination of plant and insect eggs as food sources facilitates ovarian development in an omnivorous bug *Apolygus lucorum* (Hemiptera: Miridae). *Journal of Economic Entomology*, 106(3), 1200–1208. <https://doi.org/10.1603/EC13016>
- Zeng, F., & Cohen, A. C. (2000). Comparison of alpha-amylase and protease activities of a zoophytophagous and two phytozoophagous Heteroptera. *Comparative Biochemistry and Physiology Part A: Molecular and Integrative Physiology*, 126(1), 101–106. [https://doi.org/10.1016/S1095-6433\(00\)00193-8](https://doi.org/10.1016/S1095-6433(00)00193-8)
- Zhang, L., Lu, Y., & Liang, G. (2013). A method for field assessment of plant injury elicited by the salivary proteins of *Apolygus lucorum*. *Entomologia Experimentalis Et Applicata*, 149(3), 292–297. <https://doi.org/10.1111/eea.12125>
- Zhang, L., Xu, P., Xiao, H., Lu, Y., Liang, G., Zhang, Y., & Wu, K. (2015). Molecular characterization and expression profiles of polygalacturonase genes in *Apolygus lucorum* (Hemiptera: Miridae). *PLoS One*, 10(5), e0126391. <https://doi.org/10.1371/journal.pone.0126391>
- Zhao, J. H., Ho, P., & Azadi, H. (2011). Benefits of Bt cotton counterbalanced by secondary pests? Perceptions of ecological change in China. *Environmental Monitoring and Assessment*, 173(1–4), 985–994. <https://doi.org/10.1007/s10661-010-1439-y>
- Zhen, C., Tan, Y., Miao, L., Wu, J., & Gao, X. W. (2018). Overexpression of cytochrome P450s in a lambda-cyhalothrin resistant population of *Apolygus lucorum* (Meyer-Dur). *PLoS One*, 13(6), e0198671. <https://doi.org/10.1371/journal.pone.0198671>
- Zhou, X., Slone, J. D., Rokas, A., Berger, S. L., Liebig, J., Ray, A., Reinberg, D., & Zwiebel, L. J. (2012). Phylogenetic and transcriptomic analysis of chemosensory receptors in a pair of divergent ant species reveals sex-specific signatures of odor coding. *PLoS Genetics*, 8(8), e1002930. <https://doi.org/10.1371/journal.pgen.1002930>
- Zhu, J., Jiang, F., Wang, X., Yang, P., Bao, Y., Zhao, W., Wang, W., Lu, H., Wang, Q., Cui, N. A., Li, J., Chen, X., Luo, L., Yu, J., Kang, L. E., & Cui, F. (2017). Genome sequence of the small brown planthopper, *Laodelphax striatellus*. *Gigascience*, 6(12), 1–12. <https://doi.org/10.1093/gigascience/gix109>

SUPPORTING INFORMATION

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

How to cite this article: Liu Y, Liu H, Wang H, et al. *Apolygus lucorum* genome provides insights into omnivorousness and mesophyll feeding. *Mol Ecol Resour* 2020;00:1–14. <https://doi.org/10.1111/1755-0998.13253>