frontiers in Genetics

ORIGINAL RESEARCH
published: 29 October 2021
doi: 10.3389/fgene.2021.728418

Chromosome-Level Genome Assembly Reveals Significant Gene Expansion in the Toll and IMD Signaling Pathways of *Dendrolimus kikuchii*

*Jielong Zhou[1], Peifu Wu[2], Zhongping Xiong[1], Naiyong Liu[1], Ning Zhao[2], Mei Ji[3], Yu Qiu[2] and Bin Yang[1]**

[1]Key Laboratory of Forest Disaster Warning and Control of Yunnan Province, Southwest Forestry University, Kunming, China, [2]College of Life Science, Southwest Forestry University, Kunming, China, [3]Yunnan Academy of Forestry and Grassland, Kunming, China

OPEN ACCESS

Edited by:
Ben-Yang Liao,
National Health Research Institutes,
Taiwan

Reviewed by:
Mei-Yeh Jade Lu,
Academia Sinica, Taiwan
Shu-Dan Yeh,
National Central University, Taiwan
Huabin Zhao,
Wuhan University, China

***Correspondence:**
Bin Yang
yangbin48053@swfu.edu.cn

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: 21 June 2021
Accepted: 28 September 2021
Published: 29 October 2021

Citation:
Zhou J, Wu P, Xiong Z, Liu N, Zhao N, Ji M, Qiu Y and Yang B (2021) Chromosome-Level Genome Assembly Reveals Significant Gene Expansion in the Toll and IMD Signaling Pathways of Dendrolimus kikuchii. Front. Genet. 12:728418. doi: 10.3389/fgene.2021.728418

A high-quality genome is of significant value when seeking to control forest pests such as *Dendrolimus kikuchii*, a destructive member of the order Lepidoptera that is widespread in China. Herein, a high quality, chromosome-level reference genome for *D. kikuchii* based on Nanopore, Pacbio HiFi sequencing and the Hi-C capture system is presented. Overall, a final genome assembly of 705.51 Mb with contig and scaffold N50 values of 20.89 and 24.73 Mb, respectively, was obtained. Of these contigs, 95.89% had unique locations on 29 chromosomes. *In silico* analysis revealed that the genome contained 15,323 protein-coding genes and 63.44% repetitive sequences. Phylogenetic analyses indicated that *D. kikuchii* may diverged from the common ancestor of *Thaumetopoea. Pityocampa*, *Thaumetopoea ni, Heliothis virescens*, *Hyphantria armigera*, *Spodoptera frugiperda*, and *Spodoptera litura* approximately 122.05 million years ago. Many gene families were expanded in the *D. kikuchii* genome, particularly those of the Toll and IMD signaling pathway, which included 10 genes in *peptidoglycan recognition protein*, 19 genes in *MODSP*, and 11 genes in *Toll*. The findings from this study will help to elucidate the mechanisms involved in protection of *D. kikuchii* against foreign substances and pathogens, and may highlight a potential channel to control this pest.

Keywords: lepidoptera, Dendrolimus kikuchii, nanopore, Hi-C, chromosome-level genome, gene expansion, toll and imd pathways

INTRODUCTION

Dendrolimus kikuchii (Matsumura, 1927), a member of the genus *Dendrolimus* (Lepidoptera: Lasiocampidae), is an economically significant pest of coniferous forests in southern China (Kong et al., 2007) (**Figures 1A, B**). Approximately 30 species of *Dendrolimus* have been reported as occurring in Eurasia (Mikkola and Stahls, 2008), and six of these species—*D. kikuchii* Matsumura, *D. houi* Lajonquiere, *D. punctatus* Walker, *D. superans* Butler, *D. spectabilis* Butler and *D. tabulaeformis* Tsai & Liu—are dangerous and widespread in China (Hou, 1987; Chen, 1990). *D. kikuchii* and *D. houi* are grouped together and nested in the core groups, based on the mitochondrial phylogeny (Kononov et al., 2016; Wang et al., 2019).

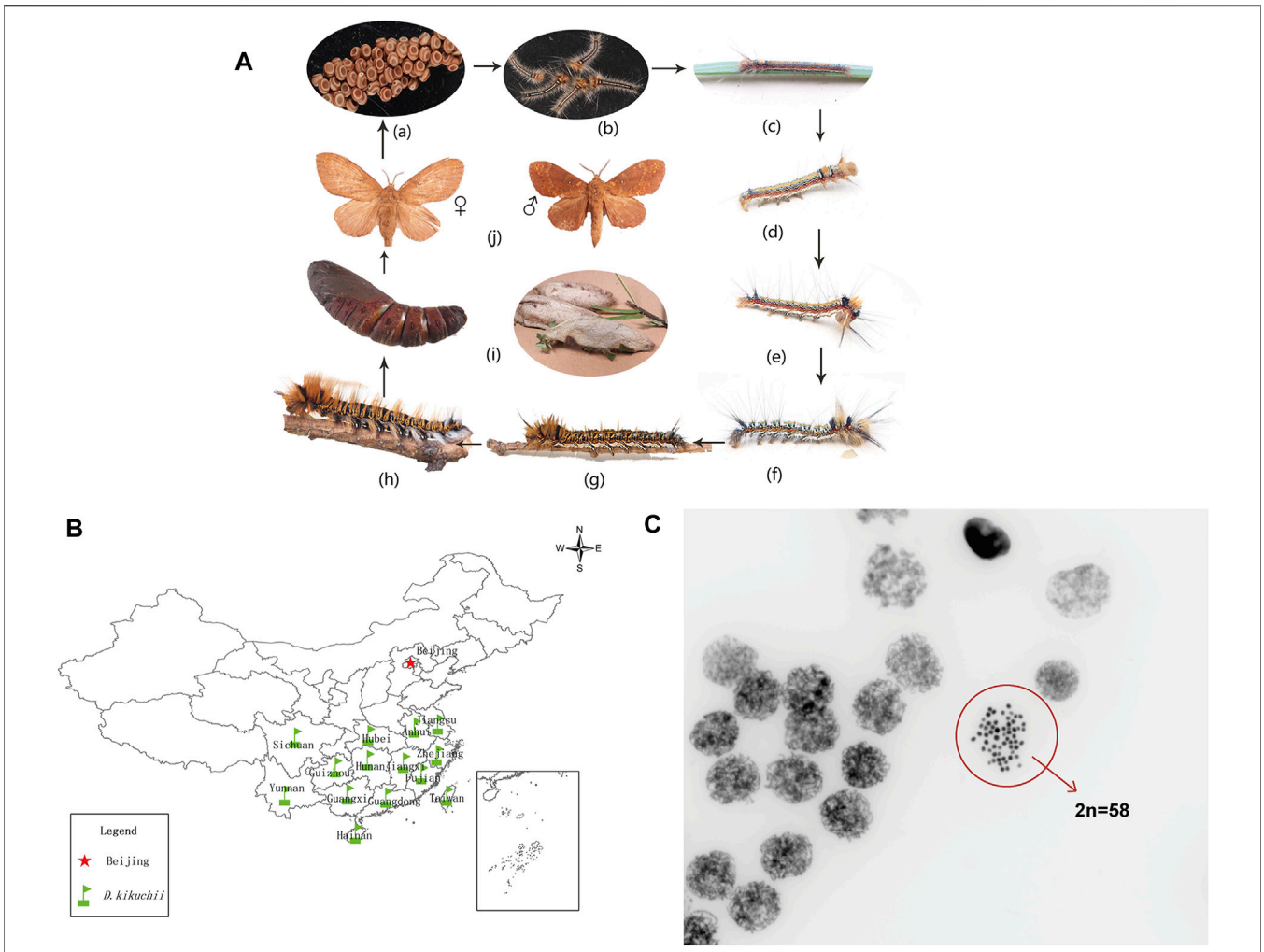

FIGURE 1 | (A) Life cycle of *Dendrolimus kikuchii*: (a) eggs (b–h) first to seventh instar larvae, respectively; (i) pupa (j) ♀ - adult female, ♂ - adult male. Photos by Mr. Zhongping XIONG **(B)** Distribution of *D. kikuchii* in southern China. Designed by Dr. Xun ZHAO **(C)** Chromosomes of gonadal cells of *D. kikuchii* in mitotic metaphase (2n = 58, 630 X). Photo by Kunming Cell Bank, Chinese Academy of Sciences.

D. kikuchii is widely distributed across southern China and Vietnam (Kishida and Wang, 2011) and has caused serious damage to coniferophytes in this region. The larvae of *D. kikuchii* endanger various coniferous trees by feeding extensively on conifer needles. A study on the food consumption of these larvae revealed that they consume approximately 7,486.6 cm of pine needles of *Pinus kesiya* var. *langbianensis* (A. Chev) to complete their growth and development (Tong and He, 2009). The large infestations of *D. kikuchii* larvae harm the growth rate of pines, causing heavy defoliation, dieback, and even tree death, and thereby reducing the yield of cones, timber, and resin (Hou, 1987; Dai et al., 2012; Men et al., 2017). Previous studies have shown that local epidemics of pine caterpillar disease in humans have been accompanied by an outbreak of *D. kikuchii* larvae, and that direct contact with either living or dead caterpillars, or their pupae, will cause a poisoning reaction known as caterpillar arthritis, which has serious consequences for human health (Chen, 1990; Xiao, 1992; Wang et al., 1999).

Pest management of *D. kikuchii* mostly involves routine technologies, for instance, manual, physical, chemical, and biocontrol methods as well as forestry management. Biocontrol of the genus *Dendrolimus* with organisms such as *Trichogramma dendrolimi*, *Beauveria bassiana*, and *Bacillus thuringiensis* is safe, environmentally friendly, and effective long-term (Hou, 1987; Hou, 1993; Kunimi, 2007; Konecka et al., 2019). However, despite the broad prospect of utilizing pathogens in the biocontrol of *D. kikuchii*, the molecular mechanisms of interaction between *D. kikuchii* and such pathogens are not well understood. A deeper understanding of the genomics of *D. kikuchii* is urgently required to provide new strategies and methods for targeting biocontrol and regulation.

The explosive development of bioinformatics and high-throughput sequencing technologies, particularly the rise of the

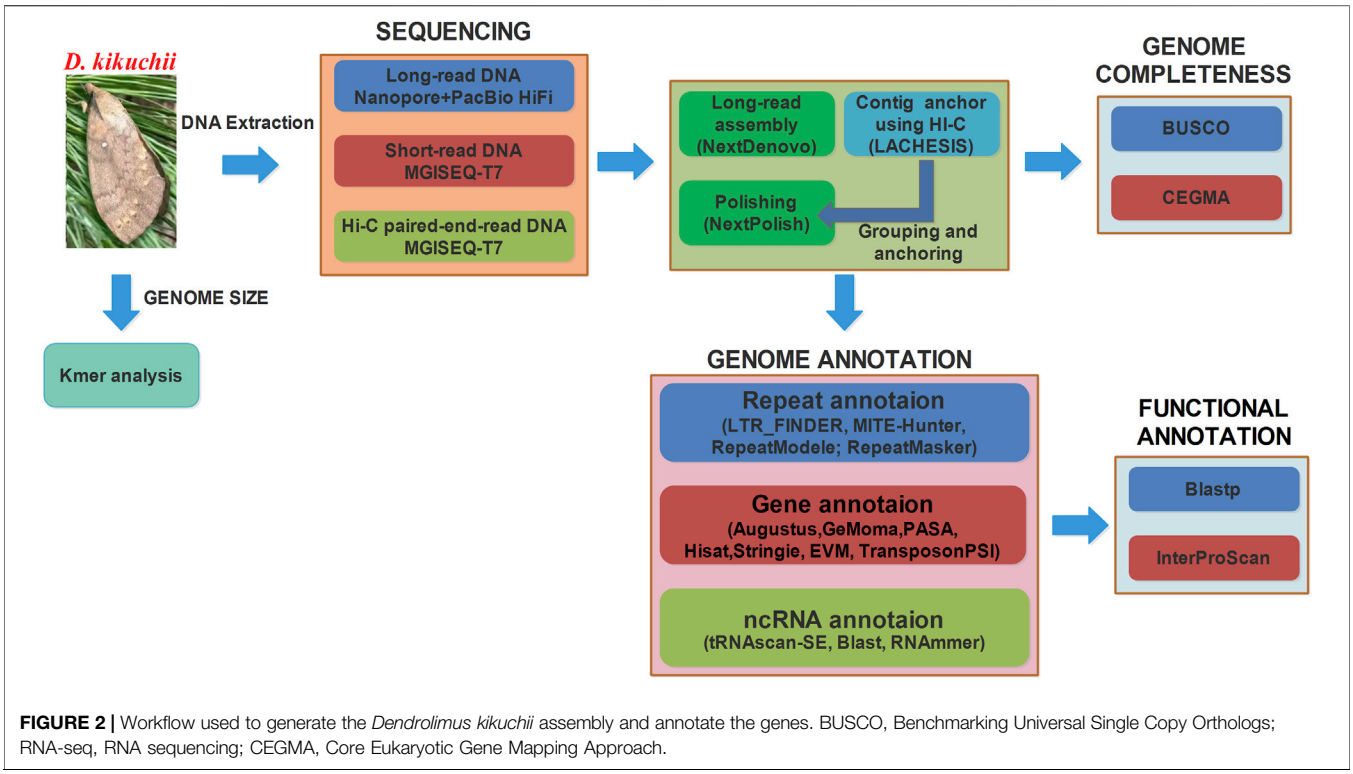

FIGURE 2 | Workflow used to generate the *Dendrolimus kikuchii* assembly and annotate the genes. BUSCO, Benchmarking Universal Single Copy Orthologs; RNA-seq, RNA sequencing; CEGMA, Core Eukaryotic Gene Mapping Approach.

Oxford Nanopore Technology (ONT) and PacBio third-generation sequencing platforms (Senol Cali et al., 2019; Wick et al., 2019) and Hi-C technology (Servant et al., 2015; Zhuang et al., 2019), have facilitated the resolution of the challenges of high repetition and high heterozygosity in insect genome assembly in the past few years. Consequently, chromosome-level genome assemblies of many insects have been published (Harrop et al., 2020; Biello et al., 2021), providing abundant information and the foundations for research in areas such as fundamental insect biology, insect-plant interactions and co-evolution, chemical ecology and insecticide resistance, comparative genomics and phylogenomics, detoxification metabolism, and ecological adaptations of the insects. Furthermore, the genome assemblies may illuminate potential targets for the development of next-generation control strategies and monitoring of potential resistance to chemical control.

To date, the genomes of more than 100 species of Lepidoptera have been sequenced and published in the NCBI database. Based on genomics and transcriptomics, the application of gene editing and interference technology could revolutionize pest control and the utilization of economic insects (Hou et al., 2017). Using a mix of the PacBio and Illumina platforms, Zhang *et al.* (Zhang et al., 2020) first reported a chromosome-level genome assembly of a species of the genus *Dendrolimus* with the sequence of *Dendrolimus punctatus*. However, obvious differences between genomes exist among species of the genus *Dendrolimus*. For example, the genome size of *D. punctatus* was 563.36 ± 7.26 Mb, but that of *D. kikuchii* was 719.30 ± 9.70 Mb as measured by flow cytometry (Zhang et al., 2014). In the present study, a higher quality chromosome-level genome assembly and annotation was obtained for *D. kikuchii* using Oxford Nanopore PromethION, PacBioHiFi, MGISEQ-T7 platform and Hi-C (**Figure 2**). This reference genome provides a foundation for genome-based investigations of the unique ecological and evolutionary characteristics of *D. kikuchii* and helps illuminate the genetic basis of gene selection and immune resistance of the species, such as Toll and IMD signaling pathways (Kim and Kim, 2005; Buchon et al., 2009), for protection against foreign substances and pathogens. Elucidating the molecular mechanism of immune resistance of *D. kikuchii* could identify potential gene targets for developing novel environmentally friendly approaches to manage this dangerous and widespread pest.

MATERIALS AND METHODS

Samples and Genomic Survey

Pupae of *D. kikuchii* were collected in Anning County (24°31′–25°6′ N, 102°8′–102°37′ E), Kunming City, Yunnan Province, China in June 2020 from host yunnanensis pine trees (*Pinus yunnanensis*). The pupae were reared at 27.5 ± 2°C and 75 ± 3% relative humidity, with a 16-h light/8-h dark photoperiod. Upon emergence, adults were immediately frozen in liquid nitrogen and preserved at −80°C until DNA extraction.

High quality genomic DNA was purified from a female adult using the QIAGEN® Genomic kit. After quality testing of extracted DNA, the resulting genomic DNA was used and sequenced based on the three platforms (Nanopore, Pacbio HiFi and Hi-C) to ensure the quality and accuracy of genome

assembly. The sequence data resulted from Pacbio HiFi and Hi-C capture system were used for genome correction.

For Nanopore sequencing, the DNA was randomly fragmented, size-selected. The ends of fragments were repaired, A-linked, ligated. Finally, Sequencing was performed on a PromethION sequencer (Oxford Nanopore Technologies, United Kingdom) instrument (**Supplementary information S2**, Protocols for genome sequencing and assembly of *D. kikuchii*).

For PacBio HiFi equencing, the DNA was fragmented, damage repaired, end polished and ligated with the stem-loop adaptor for PacBio sequencing. The SMRTbell library was purified and sequenced on a PacBio Sequel II instrument with Sequel II Sequencing Kit 2.0 (**Supplementary information S2**, Protocols for genome sequencing and assembly of *D. kikuchii*).

To ensure reads were reliable, sequenced raw reads were filtered (Chen et al., 2018). The genome of *D. kikuchii* was characterized using k-mer analysis. Briefly, quality-filtered reads were subjected to 17-mer frequency distribution analysis using the Jellyfish tool (Marcais and Kingsford, 2011). Through analysis of the 17-mer depth distribution from the 350-bp clean library sequencing reads using GenomeScope (Vurture et al., 2017) and FindGSE (Sun et al., 2018), the genome size of *D. kikuchii* was estimated via the following equation: G = K-num/K-depth (where K-num is the total number of 17-mers, K-depth represents the k-mer depth and G is the genome size).

Genome Assembly and Polish

After quality control of raw reads, the pass reads were used for *de novo* genome assembly of using an OLC (overlap layout-consensus)/string graph method with NextDenovo (v2.3.0) with reads_cutoff:1 k and seed_cutoff:30 k. Firstly, self-correction of the original subreads was finished by NextCorrect to obtain consistent sequences (CNS reads). Then, CNS reads were used to obtain preliminary assembly through NextGraph (default parameter). The ONT, CCS and Hi-C data were used to correct the preliminary assembly using Racon (v1.3.1, default, CCS data) (Vaser et al., 2017)and Nextpolish (v1.2.4, default, ONT and Hi-C data) (Hu et al., 2020). BlastN was used to check the genome contamination (**Supplementary information S2**).

Completeness of the genome assembly was assessed using BUSCO v4.0.5 (Benchmarking universal Single-Copy Orthologs) (Simao et al., 2015) and CEGMA (Core Eukaryotic Gene Mapping Approach) (Parra et al., 2007). To evaluate the accuracy of the assembly, all paired-end reads were mapped to the assembled genome using BWA (Burrows–Wheeler Aligner) (Li and Durbin, 2010) and the mapping rate and the genome coverage of sequencing reads were both assessed using SAMtools v0.1.1855 (Li et al., 2009). In addition, the base accuracy of the assembly was calculated using bcftools (Danecek and Mccarthy, 2017). Coverage of the expressed genes of the assembly was examined by aligning all the RNA-seq reads against the assembly using HISAT with default parameters. To ensure that mitochondrial sequences were not included in the assembly, the draft genome assembly was submitted to the NT library and matching sequences were eliminated.

Genome Anchoring to Chromosome

Based on Hi-C libraries, hybrid scaffolds were anchored onto the chromosomes of *D. kikuchii*. First, chromosome numbers (2n) from gonads of the fifth instar of *D. kikuchii* were counted following the method of Gautam and Paul (Gautam and Paul, 2013), and then the Hi-C library was constructed and sequenced. In brief, freshly harvested thoraxes of adult insects were cut into pieces and nuclei were purified. The purified nuclei were digested with 100 units of DpnII, and nuclear DNA was marked with biotin-14-dCTP and sheared into 300–600 bp fragments. The fragments were blunt-end repaired, A-tailed, and purified through biotin-streptavidin-mediated pull down. Lastly, the Hi-C libraries were quantified and sequenced using the MGISEQ-T7 platform (**Supplementary information S2**).

The read quality (370 million paired-end reads) was controlled using Hi-C-Pro. Firstly low-quality sequences (quality scores <20), adaptor sequences, and sequences shorter than 30 bp were filtered out using fastp (Chen et al., 2018). Next, clean reads were mapped to the draft assembled sequence using bowtie2 (v2.3.2) (-end-to-end --very-sensitive -L 30) (Langmead and Salzberg, 2012) to obtain the unique mapped paired-end reads. Invalid read pairs were filtered using HiC-Pro (v2.8.1) (Servant et al., 2015). The scaffolds were further clustered, ordered, and oriented onto chromosomes by LACHESIS (https://github.com/shendurelab/LACHESIS), with parameters CLUSTER_MIN_RE_SITES = 100, CLUSTER_MAX_LINK_DENSITY = 2.5, CLUSTER NONINFORMATIVE RATIO = 1.4, ORDER MIN N RES IN TRUNK = 60, ORDER MIN N RES IN SHREDS = 60. Lastly, placement and orientation errors exhibiting obvious discrete chromatin interaction patterns were manually adjusted.

Synteny of the *D. kikuchii* genome with the *D. punctatus* genomes was analyzed using Minimap2 and dotPlotly to identify chromosome structural changes among the two species.

Genome Annotation

The software GMATA (Wang and Wang, 2016) and Tandem Repeats Finder (TRF) (Benson, 1999) were used to respectively identify the simple or tandem repeat elements. RepeatMasker (Bedell et al., 2000) was applied to search for known and novel transposable elements (TEs) by mapping sequences against the *de novo* repeat library and the Repbase TE library.

The transcriptome of *D. kikuchii* was obtained using samples of critical developmental stages and representative tissues for genome annotation on an MGISEQ-T7 platform. The samples of *D. kikuchii* at different developmental stages included eggs (~1–2 days, 50 eggs), larvae (20 insects at 1–2 instar, 10 insects at 3–4 instar, three insects at 5–7 instar, respectively), pupae (~5 days, three males and three females), and adults (3 males and three females). The different tissue samples included adult heads, adults except the heads, testes and ovaries of adults (from 20 males and 20 females, respectively), hemolymph, epidermis, midgut, silk gland, and fat body. The samples reared and collected in the lab for RNA-seq. Clean reads of the transcriptome were mapped to the assembly genome of *D. kikuchii* with TopHat, specifying "-no-novel-juncs." The uniquely mapped reads were used for subsequent analysis,

including transcripts construction, quantification of gene and transcript expression. Gene expression profiles were determined as fragments per kilobase of transcript per million mapped reads (FPKM) using RSEM version 1.3.0 (Liu et al., 2016). R language software (ver 3.6.3) was used for gene expression visualization and to generate heatmaps.

Three independent approaches—ab initio prediction, homology search, and reference guided transcriptome assembly—were employed for gene prediction in a repeat-masked genome. In detail, GeMoMa (Birney and Durbin, 2000) was used to align the homologous peptides from seven related species (*Spodoptera litura*, *Bombyx mori*, *Thaumetopoea pityocampa*, *Drosophila melanogaster*, *Plutella xylostella*, *Operophtera brumata*, and *Stenopsyche tienmushanensis*) to the assembly of *D. kikuchii* and then obtain the gene structure information. For RNA-seq gene prediction, filtered mRNA-seq reads were aligned to the reference genome using STAR (default). Transcripts were then assembled using StringTie and open reading frames (ORFs) were predicted using PASA to produce a training set (Haas et al., 2008). AUGUSTUS, with default parameters, was then used for *ab initio* gene prediction with the training set (Alioto et al., 2018; Majoros et al., 2004; Stanke and Waack, 2003). Finally, EVidenceModeler (EVM) was employed to produce an integrated gene set, from which genes with TEs were removed using the TransposonPSI package (http://transposonpsi. sourceforge.net/) and miscoded genes were further filtered (Haas et al., 2008). Untranslated regions (UTRs) and alternative splicing regions were determined using PASA based on RNA-seq assemblies. The longest transcripts for each locus were retained and regions outside of the ORFs were designated as UTRs.

Gene function information, motifs, and domains of their proteins were assigned through comparison with public databases, including SwissProt, Non-Redundant Protein database (NR), Kyoto Encyclopedia of Genes and Genomes (KEGG), Eukaryotic Orthologous Groups of protein (KOG), and Gene Ontology (GO). Blastp was used to compare the EVM-integrated protein sequences against the four well-known public protein databases with an E-value cutoff of 1e−05. The results were concatenated from the five database searches.

Database searching and prediction were employed to obtain noncoding RNAs (ncRNAs). Transfer RNAs (tRNAs) were predicted using tRNAscan-SE with eukaryote parameters (Lowe and Eddy, 1997); microRNA, rRNA, small nuclear RNA, and small nucleolar RNA were detected using Infernal cmscan to search the Rfam database; and rRNAs and their subunits were predicted using RNAmmer (Lagesen et al., 2007).

Phylogenetic Analyses

Protein sequences obtained from *D. kikuchii* and 15 published species (*Drosophila melanogaster*, *Stenopsyche tienmushanensis*, *Plutella xylostella*, *Danaus plexippus*, *Papilio xuthus*, *O. brumata*, *T. pityocampa*, *Thaumetopoea ni*, *Heliothis virescens*, *Hyphantria armigera*, *Spodoptera frugiperda*, *Spodoptera litura*, *B. mori*, *Manduca sexta*, and *Dendroctonus ponderosae*) (**Supplementary Table S12**), were aligned using OrthMCL to obtain orthologous gene sets. Molecular phylogenetic analysis using the shared single-copy genes was then conducted through Mafft (Katoh et al., 2002). Poorly aligned sequences were eliminated using Gblocks (Castresana, 2000), and the GTRGAMMA substitution model of RAxML (Stamatakis, 2014) was used for phylogenetic tree reconstruction with 1,000 bootstrap replicates. Based on the phylogenetic tree, RelTime of MEGA-CC was employed to compute the mean substitution rates along each branch and estimate the species divergence time. Fossil calibration times were obtained from the TimeTree database (http://www.timetree.org/) as the time control. The date of the node between *Papillo xuthus* and *Danaus plexippus* was constrained to 76–146 million years ago (Ma) and that of the node between *Drosophila* and *Lepidoptera* to 217–314 Ma according to the divergence times from TimeTree (You et al., 2013; Cheng et al., 2017; Kawahara et al., 2019).

Significant expansion or contraction of specific gene families, which is frequently associated with adaptive divergence of closely related species, was identified through comparing the *D. kikuchii* genome with those of *Drosophila melanogaster*, *Stenopsyche tienmushanensis*, *Plutella xylostella*, *Danaus plexippus*, *Papilio xuthus*, *O. brumata*, *T. pityocampa*, *T. ni*, *Heliothis virescens*, *Hyphantria armigera*, *Spodoptera frugiperda*, *Spodoptera litura*, *B. mori*, *M. sexta* and *Dendroctonus ponderosae* using OrthoMCL (Li et al., 2003). Expansions and contractions of orthologous gene families were determined using CAFE 3 (Han et al., 2013), which employs a birth and death process to model gene gain and loss over a phylogeny.

Genes Under Positive Selection

The ratio of the nonsynonymous substitution rate (Ka) and the synonymous substitution rate (Ks) of protein-coding genes were used to identify positively selected genes in the *D. kikuchii* lineage following the branch-site likelihood ratio test using Codeml implemented in the PAML package (Yang, 2007). Genes with a p value < 0.05 under the branch-site model were considered to be positively selected genes.

RESULTS AND DISCUSSION

De Novo Assembly Genome

After filtering adapter and low-quality reads, 75.95 Gb clean data was used for genomic survey based on ONT, CCS and Hi-C data. A quality check detected no exogenous contamination (**Supplementary Figure S1**). K-mer analyses of the DNA data revealed the *D. kikuchii* genome to be 687.3 Mb with a heterozygosity of 1.2% following the distribution frequency of 17-mers (**Supplementary Figure S2**). This genomic heterozygosity of *D. kikuchii* is similar to that of non-model insects with published genomes (Luo et al., 2018; Zhang et al., 2020).

For long-read sequencing, the genome of an adult female *D. kikuchii* was sequenced on an ONT PromethION platform and 3,336,618 reads were obtained from 64.01 Gb of clean data with N50 and average length of long sub-reads of 29.86 and 19.18 kb, respectively (**Supplementary Table S1**, **Table 1**).

TABLE 1 | Comparison of genome assemblies of *Dendrolimus kikuchii* and *Dendrolimus punctatus*.

Species	*D. kikuchii*	*D. punctatus*
Platform	Nanopore + PacBio HiFi + MGI	PacBio + Illumina
Sequencing depth	90.66× + 13.95× + 107.57×	123× + 44.01×
Contig N50 (bp)	20,894,489	1,388,385
Contig N90 (bp)	6,875,281	486,050
Longest contig (bp)	39,754,656	8,686,113
Total assembly size of contigs (bp)	705,512,268	614,029,055
Number of contigs	86	724
Number of scaffolds after Hi-C	49	107
N50 scaffold length (bp) after Hi-C	24,734,651	22,146,069
N90 scaffold length (bp) after Hi-C	16,859,236	13,408,297
Total assembly size of scaffolds (bp)	705,515,768	614,070,971

NN50, shortest sequence length at 50% of the genome; N90, shortest sequence length at 90% of the genome.

TABLE 2 | Information for predicted protein-coding genes.

Gene set		Total no. genes	Average gene length (bp)	Average CDS length (bp)	Average no. exons per gene	Average exon length (bp)	Average intron length (bp)
RNA-seq	PASA	11,521	31,494	4,500	7	621	4,327
Homology	*Thaumetopoea pityocampa*	33,257	22,861	486	3	173	12,379
	Spodoptera litura	24,825	25,173	1,440	5	263	5,321
	Drosophila melanogaster	17,956	33,820	1,187	5	249	8,685
	Bombyx mori	18,848	25,864	1,288	6	228	5,306
	Stenopsyche tienmushanensis	21,158	52,001	1,351	6	217	9,697
	Plutella xylostella	25,660	31,347	1,323	6	230	6,338
	Operophtera brumata	20,817	19,769	985	4	280	7,495
	GeMoMa	16,025	25,024	1,538	6	249	4,555
De novo	AUGUSTUS	13,935	22,716	1,574	8	209	3,244
Final set	EVM	15,323	24,961	1,543	7	225	4,009

A *de novo* assembly was performed using Racon and NextPolish following MGISEQ paired-ended, CCS and Nanopore clean data. Finally, a 705.51 Mb assembly with contig N50 of 20.89 Mb was obtained for *D. kikuchii* (**Table 1**), which was bigger than the genome survey of 687.3 Mb obtained with the k-mer estimate. A continuous length for maximum contig size showed a high-quality genome assembly for *D. kikuchii* (**Supplementary Figure S3**). The genome of *D. kikuchii* is longer than that of *D. punctatus*, which has a 614-Mb assembly with contig N50 of 1.39 Mb (Zhang et al., 2020).

The accuracy of the *D. kikuchii* assembly was assessed, based on the Orthologs database insecta_odb10, 1,319 (96.49%), and complete, highly conserved, insect orthologs genes in the assembly were identified with BUSCO (Supplementary S1_**Table 2**). Moreover, 232 core genes (93.55%) were found in the assembly following CEGMA (**Supplementary Table S3**). BWA was used to remap the paired-end reads to the assembled genome, revealing a mapping rate of 99.33%, an average sequence depth of 104.73×, and single-base accuracy of 99.997533% in the genome assembly ((**Supplementary Table S4**). In addition, a 99.03% mapping rate and 83.27 × average sequence depth was obtained for nanopore sequences. The distributions of GC depth of the genome sequences focus on 30–40% (**Supplementary Figure S4**). Together, these findings demonstrate that the assembled *D. kikuchii* genome sequence was complete and had a markedly high accuracy ratio.

The Genome at Chromosome Level by Hi-C Data

Chromosomes of *D. kikuchii* were observed through an optical microscope and the diploid chromosome numbers of *D. kikuchii* were determined as 2n = 58 (**Figure 1C**).

After filtering adapter sequences and low-quality paired-end reads, 74.94 Gb clean data were mapped onto the genome assembly for chromosome construction with bowtie2 (Langmead and Salzberg, 2012). All assembled contigs were anchored, ordered, and orientated to the 29 chromosomes of *D. kikuchii* that were 12–39 Mb in length, with more than 95.89% of assembled bases located on the chromosomes (**Figure 3A**, (**Supplementary Table S5**). The final genome size and N50 were 705.51 and 24.73 Mb, respectively (**Table 1**).

Syntenic relationships between the newly assembled *D. kikuchii* genome and the genomes of another lepidopteran insect, *D. punctatus* (Zhang et al., 2020) were compared. *D. kikuchii* had 29 chromosomes compared with 30 chromosomes in *D. punctatus*. The alignments of the *D. kikuchii* and *D. punctatus* genomes indicated high levels of gene collinearity (**Figure 3B**); the separate chromosomes of

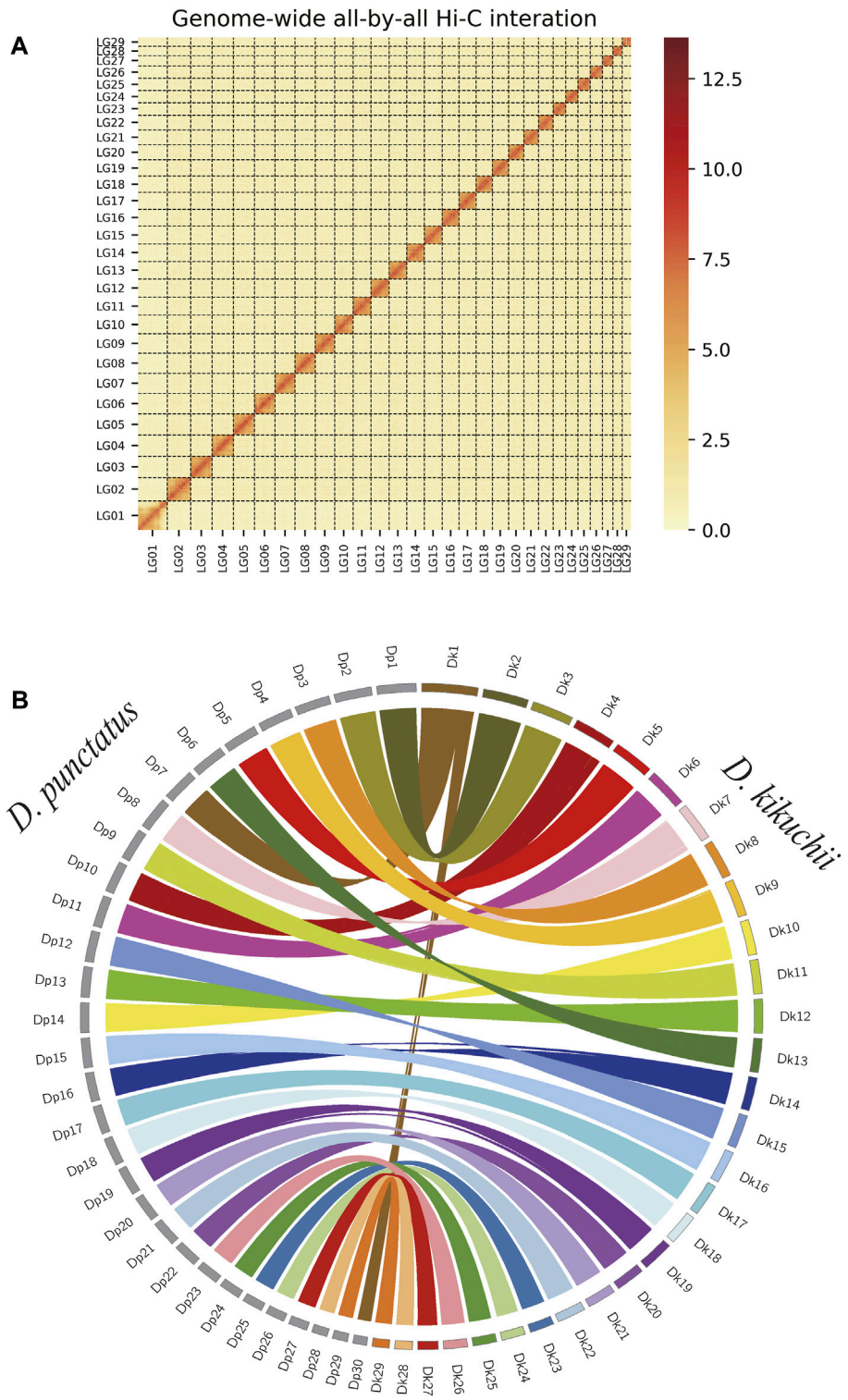

FIGURE 3 | Chromosome-level assembly of *Dendrolimus kikuchii* **(A)** Genome-wide all-by-all Hi-C interaction map of the *D. kikuchii* genome. Calculated interaction frequency distribution of Hi-C links between and within chromosomes **(B)** Comparative analysis of the synteny between *D. kikuchii* (Dk) and *D. punctatus* (Dp) chromosomes. Each colored arc represents a best match between the two species.

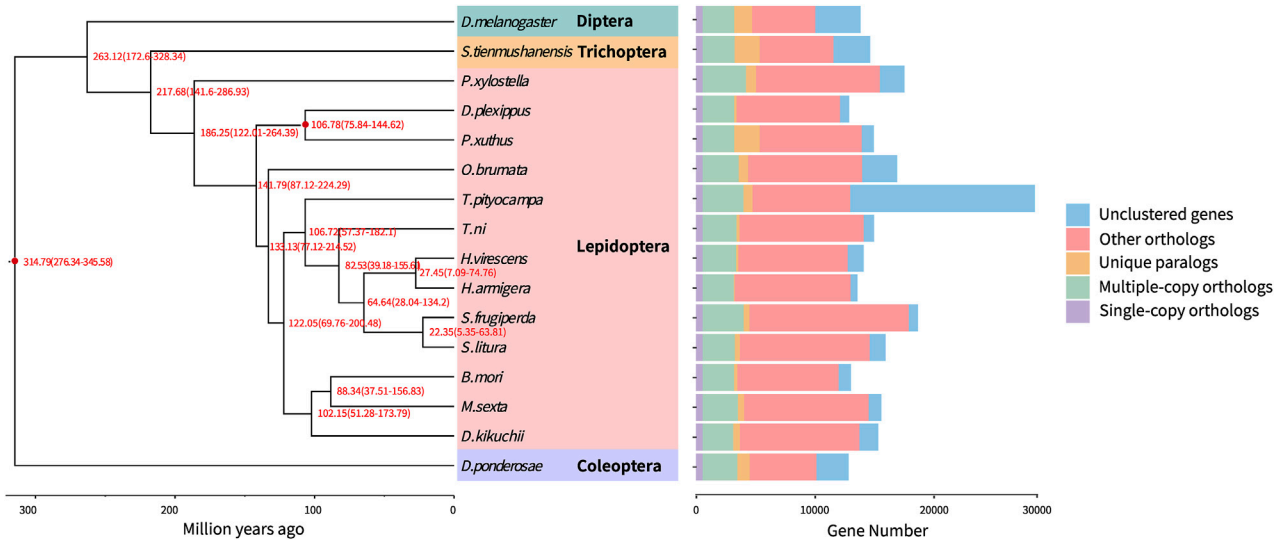

FIGURE 4 | Phylogenetic tree and gene orthology of *D. kikuchii* with 15 other insect genomes. The phylogeny was inferred from 565 strict single-copy genes by RAxML maximum-likelihood methods employing a LG + G model and 1,000 bootstrap replicates. Numbers at nodes represent divergence times (Ma) and red nodes indicate calibration times. Divergences were estimated by the PhyloBayes Bayesian method using a relaxed clock with nodes' calibration: mean age is given for each node with 95% posterior densities. Bars showing gene counts are subdivided to represent classes of orthologys.

the *D. punctatus* genome (including Dp7 and Dp30) were fused and corresponded to Dk1 of *D. kikuchii* (**Figure 3B**), thus proving the reliability and completeness of the genome assembly of *D. kikuchii*.

Genome Annotation

A total of 2,833,714 repeat sequences, spanning ~447.6 Mb and constituting 63.44% of the *D. kikuchii* genome (**Supplementary Table S6**), were identified following the prediction with RepeatMasker (Bedell et al., 2000), TRF (Benson, 1999), and GMATA (Wang and Wang, 2016). Protein-coding genes were annotated using PASA (Haas et al., 2008). By integrating the expression evidence from RNA-Seq samples, 11,521 protein-coding genes were detected in the *D. kikuchii* genome. GeMoMa (Keilwagen et al., 2016) identified 16,025 protein-coding genes by homological searching with other species, while 13,935 protein-coding genes were obtained through AUGUSTUS (Stanke et al., 2008) (**Table 2**). After removing redundancy and errors, a set of 15,323 protein-coding genes were identified in *D. kikuchii* (**Table 2**) based on EVM (Haas et al., 2008) and TransposonPSI (Urasaki et al., 2017). Out of the 15,323 protein-coding genes, 11,521 genes were supported by RNA-seq reads. The average transcript length, average length of protein-coding sequences, exon number per gene, average exon length, and average intron length of the *D. kikuchii* gene set were similar to those of other lepidopteran genes (**Table 2**). The completeness of the gene set of *D. kikuchii* was determined to be 95.61% of insect single-copy orthologs using BUSCO, and 96.83% for 25 transcriptome analysis (**Supplementary Table S8**). The high level of completeness of the assembly of *D. kikuchii* is likely due to deep long-read sequencing, which allows the assembly of long and complex regions of the genome. Next, different types of ncRNAs, including 152 small nucleolar RNAs, 683 tRNAs, 181 rRNAs, and 172 regulatory RNAs, were identified in the genome of *D. kikuchii* (**Supplementary Table S7**).

Gene functions were assigned based on the best match of the predicted proteins to SwissProt using Blastp (with E-value ≤ $1e^{-5}$), GO using InterProScan, KEGG, KOG, and NR. Of the annotated 15,323 genes, 10,879 (70.97%), 6,652 (43.39%), 9,061 (59.11%), 7,971 (52.00%), and 13,978 (91.19%) had significant hits with genes catalogued in SwissProt, KEGG, KOG, GO, and NR databases, respectively. In summary, 14,199 annotated genes were assigned with at least one related function, accounting for 92.63% of the total genes identified in *D. kikuchii*, and 4,642 genes were assigned functions with all five databases (**Supplementary Figure S5**).

Phylogenetic Analysis of the *D. kikuchii* Genome

OrthoMCL (Li et al., 2003) was employed to identify orthologous genes in *D. kikuchii* and 15 other insect species covering four insect orders (Lepidoptera, Diptera, Coleoptera, and Trichoptera), and 565 single-copy orthologous genes and 2,580 multiple-copy genes were identified in *D. kikuchii* (**Figure 4**, **Supplementary Table S9**). A phylogeny and divergence estimate was inferred using the 565 single-copy orthologs concatenated using Gblocks (Castresana, 2000) with default parameters.

Phylogenetic relationships based on the whole-genome sequence of *D. kikuchii* and published whole-genome sequences of 15 other insect species—*Drosophila melanogaster*, *Stenopsyche tienmushanensis*, *Plutella xylostella*, *Danaus plexippus*, *Papilio xuthus*, *O. brumata*, *T. pityocampa*, *T. ni*,

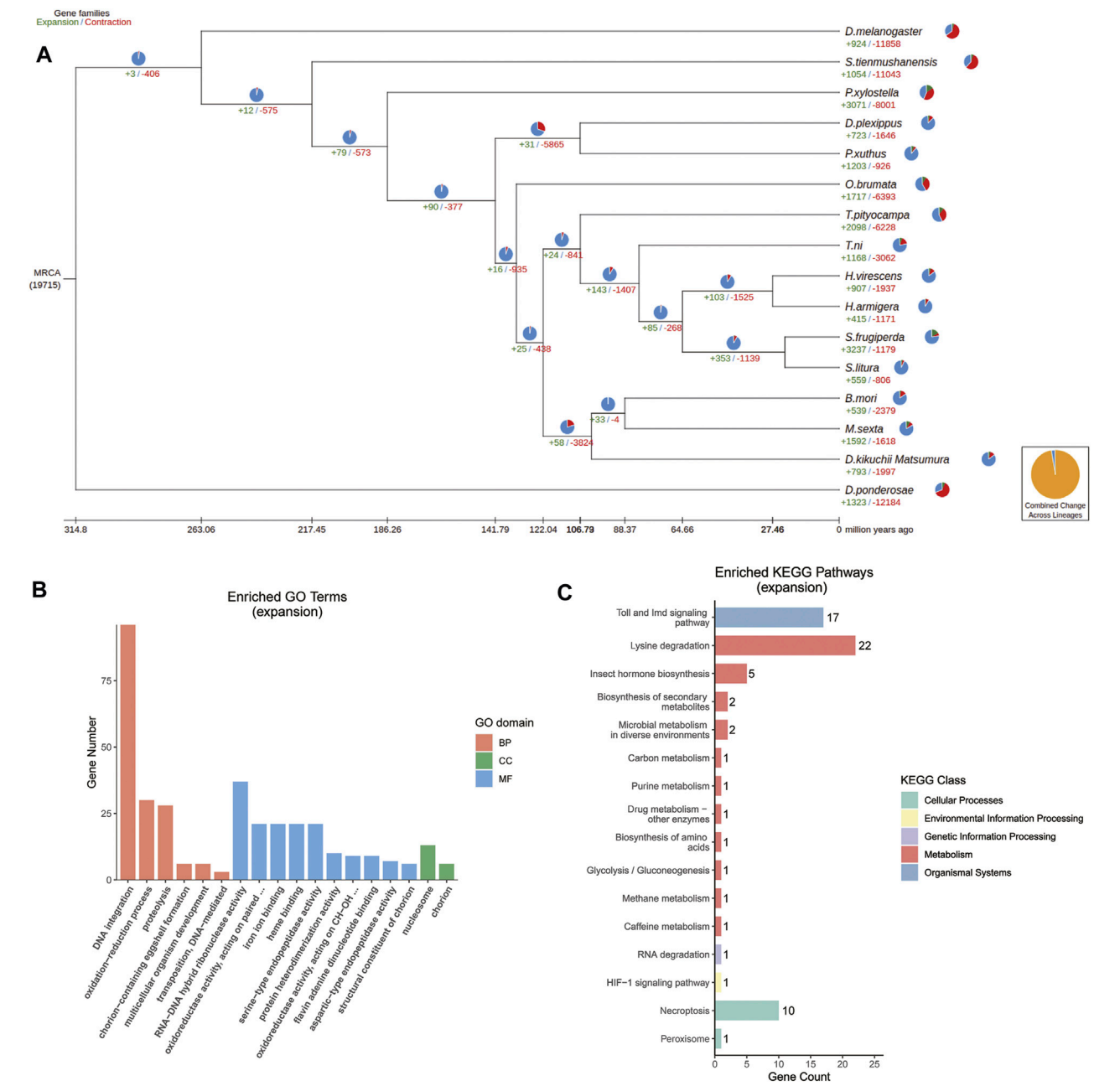

FIGURE 5 | (A) Gene family evolution in *D. kikuchii* and 15 other insect species. Trees show gene family expansions and contractions. Pie charts represent proportions of gene family expansions, contractions, or no changes. Expanded gene families are marked in green, contracted gene families are marked in red, and a gene family with no changes is indicated in blue. Yellow charts indicate the proportion of total expansions and contractions of gene families. MRCA, most recent common ancestor. The number below MRCA is the total group number from the OrthoMCL analysis. Only some of the gene expansions/contractions are significant (B) Gene ontology (GO) and (C) KEGG pathway enrichment analysis ($p < 0.05$) was performed for the expansion gene family of *D. kikuchii*. BP, biological process; CC, cellular component; MF, molecular function.

Heliothis virescens, *Hyphantria armigera*, *Spodoptera frugiperda*, *Spodoptera litura*, *B. mori*, *M. sexta*, and *Dendroctonus ponderosae*—suggested that *P. xylostella* was a basal lepidopteran species comparing to the rest of the species included in this study (**Figure 4**), which is in accordance with the findings for *D. punctatus* (Zhang et al., 2020). The divergence time of *P. xylostella*, when butterflies diverged from moths, reported by Zhang et al. (Zhang et al., 2020) is similar to the results of the present study. Phylogenomics revealing the evolutionary timing and pattern of butterflies and moths

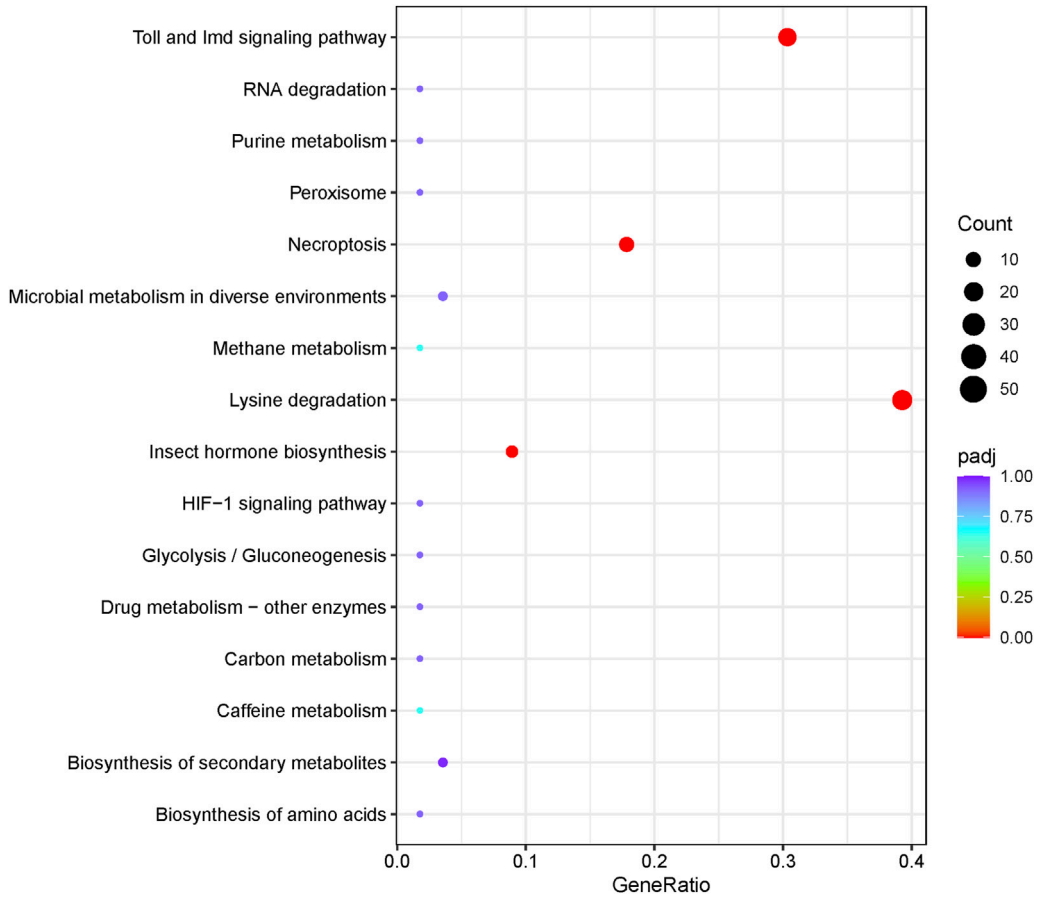

FIGURE 6 | Statistics of KEGG pathway enrichment in the genome.

(Lepidoptera), comprehensively analyzed the phylogeny of Lepidoptera using 34 superfamilies, in which Lepidoptera evolved the tube-like proboscis in the Middle Triassic (241 Ma), and the genus *Dendrolimus* nested into Bombycoidea and then grouped with a clade comprising *Artace* and *Tolype* (Kawahara et al., 2019). In the present study, *D. kikuchii* groups into Lepidoptera and shares a closer relationship with *B. mori* and *M. sexta*, diverged from the common ancestor of both taxa approximately 102.15 Ma (**Figure 4**), which is similar to the result of previous study (Kawahara et al., 2019).

Considering the branch species containing *Heliothis virescens*, *Hyphantria armigera*, *Spodoptera frugiperda*, *Spodoptera litura*, *T. ni*, or *T. pityocampa*, it could be concluded that *D. kikuchii* may have diverged from the common ancestor of these six species approximately 122.05 Ma (**Figure 4**), which is in accordance with other findings for *D. punctatus* (Zhang et al., 2020).

Expansion and Selection of Genes

Contractions and expansions of gene families were identified through comparing the *D. kikuchii* genome with the published genomes of the 15 species of insects that were used for the phylogenetic analysis. There were 793 and 1,997 gene families that had expanded and contracted, respectively, after diverging from the ancestor of *T. pityocampa*, *T. ni*, *Heliothis virescens*, *Hyphantria armigera*, *Spodoptera frugiperda*, and *Spodoptera litura* (**Figure 5A**). This finding suggested that more gene families in *D. kikuchii* contracted than expanded during adaptive evolution.

GO analysis showed that the expanded orthogroups were enriched significantly in DNA integration, RNA–DNA hybrid ribonuclease activity, oxidoreductase activity, iron ion binding, heme binding, nucleosome, serine-type endopeptidase activity, aspartic-type endopeptidase activity, protein heterodimerization activity, structural constituents of chorion, chorion-containing eggshell formation, the oxidation–reduction process, proteolysis, transposition, flavin adenine dinucleotide binding, and multicellular organism development (**Figure 5B**). KEGG annotations indicated that the expanded genes were enriched significantly in the Toll and IMD signaling pathway (17/68), lysine degradation (22/68), necroptosis (10/68), and insect hormone biosynthesis (5/68) (**Figure 5C**).

The ratio of Ka and Ks of protein-coding genes revealed six genes under positive selection (**Supplementary Table S10**). GO analysis revealed that the six genes were enriched in RNA processing, the ubiquitin-dependent protein catabolic process, metal ion binding, and calcium ion binding, while KEGG

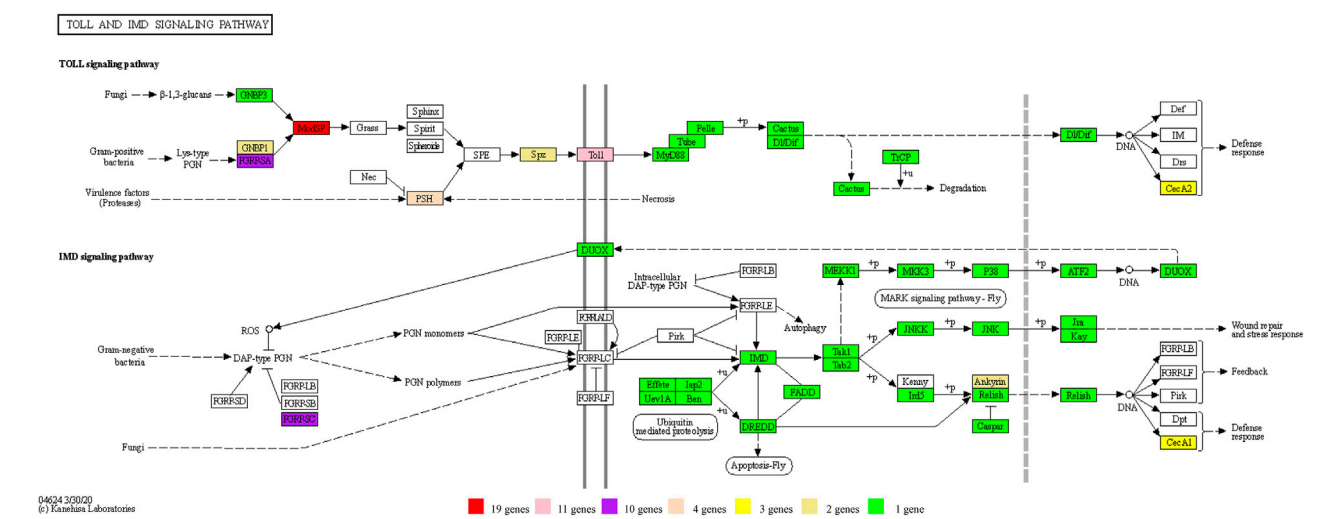

FIGURE 7 | The genes of the Toll and IMD signaling pathways identified in the *D. kikuchii* genome. The gene numbers for the corresponding KO are indicated by the color footnotes.

annotations indicated that the six enriched genes were in the Wnt signaling pathway, the MAPK signaling pathway, and protein processing in the endoplasmic reticulum.

In addition, detoxification pathways are commonly employed by insect herbivores to overcome plant defense compounds (You et al., 2013), which would help to express the reason of broad distribution of *D. kikuchii*. Thus, we carefully identified detoxification-related gene families in the *D. kikuchii* genome and found that 101 cytochrome P450 (P450) genes, six glutathione-S-transferase (GST) genes, 16 carboxylesterase (COE) genes, and 49 ATP-binding cassette (ABC) genes were annotated in this species genome. The numbers of the four detoxification-related gene families of *D. kikuchii* are less than that of *D. punctatus* reported by Zhang et al. (2020) who found 132 genes in P450, 30 genes in GST, 52 genes in COE, and 50 genes in ABC in this species. Out of these genes, we found the significant expansion in P450, which enable its extraordinary ability to detect and detoxify terpenes of pine needles (Li et al., 2007; Feyereisen. 2012) and speculate that this expansion may play a role in the adaptation of *D. kikuchii* to a wide range, although further investigations are needed to prove this hypothesis.

Expansion of Genes in the Toll and IMD Signaling Pathways

Insects have evolved an innate immune system to defend themselves against infection and survive in hostile environments. Research indicates that natural selection may drive the evolution of proteins related to the immune system. The Toll and IMD pathways are two well-studied immune signaling pathways that ultimately lead to melanization of encapsulated parasitoid eggs and bacteria-laden nodules, and synthesis of antimicrobial peptides (AMPs) (Zou et al., 2010; Li et al., 2012). In the current study, gene enrichment analysis revealed that the Toll and IMD signaling pathways had the highest gene ratio among immunity pathways (**Figure 6**), with significant gene expansion (**Figure 5C**).

The Toll pathway defends against Gram-positive bacteria or fungi, while the IMD pathway controls resistance to infections with Gram-negative bacteria (Valanne et al., 2011). A total of 81 genes directly involved in the Toll and IMD signaling pathways of *D. kikuchii* were identified. These 81 genes belong to 36 KEGG Orthologies (Kos), of which those with gene counts more than one included K01446 *PGRP* (10 genes), K10380 *ANK* (2 genes), K18809 *Toll* (11 genes), K20671 *PSH* (4 genes), K20674 *MODSP* (19 genes), K20694 *SPZ* (2 genes), K20696 *CEC* (3 genes), and K20697 *GNBP1* (2 genes) (**Figure 7**. Toll protein is the predominant gene product involved in the Toll pathway (Kim and Kim, 2005). The number of Toll genes in the genome of *D. kikuchii* (11 genes) is greater than in *Pteromalus puparum* (6), *Apis mellifera* (4), *B. mori* (5), *Anopheles gambiae* (7), and *Drosophila melanogaster* (5) (Yang et al., 2019). In the peptidoglycan recognition protein (*PGRP*) family, the IMD signaling pathway has two pattern-recognition receptors: *PGRP-LC* and *PGRP-LE* (Bosco-Drayon et al., 2012; Lu et al., 2020; Neyen et al., 2012). Through these two receptors, the IMD signaling pathway recognizes the diaminopimelic acid-type peptidoglycan of Gram-negative bacteria and some Gram-positive bacteria (*Bacillus*) and activates the downstream transcription factor *Relish*, which transfers to the nucleus and mediates expression of antibacterial peptide genes (Leulier et al., 2003; Yu et al., 2010). Ten *PGRP* genes were identified in the genome of *D. kikuchii* (**Figure 7**). In addition, 19 *MODSP* genes were found in the *D. kikuchii* genome (**Figure 7**). *MODSP* is a modular serine protease that can activate another set of serine proteases including Grass, spirit, Spheroide, and Sphinx1/2 (Kambris et al., 2006). These findings demonstrated gene expansions in the Toll and IMD signaling pathways of *D. kikuchii*.

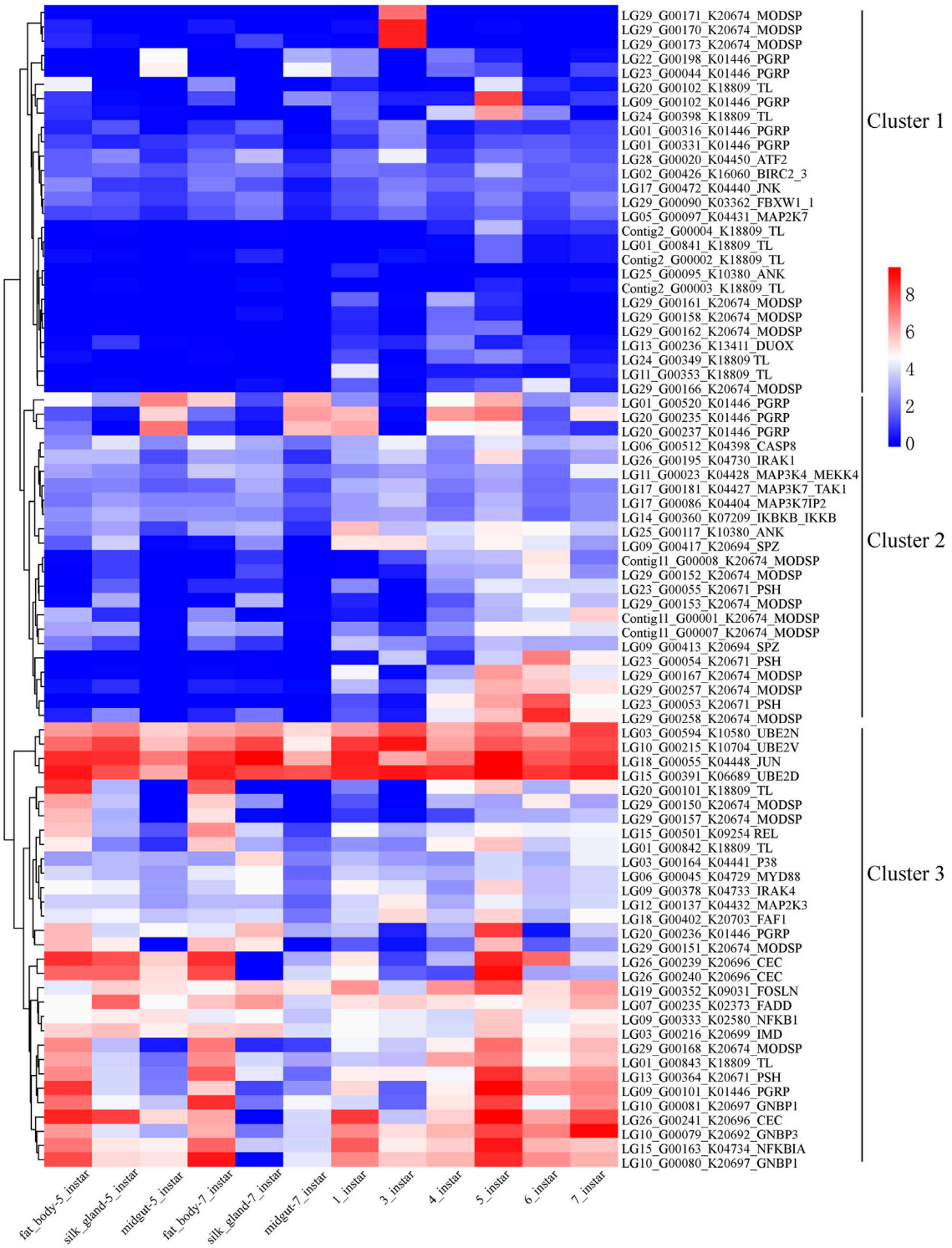

FIGURE 8 | Clustering gene expression pattern of the significantly expanded genes of the Toll and IMD signaling pathways. Note: Each column represents one sample, each row represents one gene. All the FPKM values of genes in the pathway were transformed with $\log_2$, and then normalized into Z-scores along the rows. The $\log_2$ values were color-coded as shown in the color bar.

Expression of Genes in the Toll and IMD Signaling Pathways

Gene expression levels may affect the evolution of genes within networks and pathways (Drummond et al., 2005; Duret and Mouchiroud, 2000). Three clusters could be found following the expression, cluster one had the lowest expression and cluster 3 with the highest expression (**Figure 8**). Gene LG18_G00055 in *D. kikuchii* belongs to K04448 (JUN, transcription factor AP-1) and gene LG15_G00391 belongs to K06689 (UBE2D, UBC4, UBC5; ubiquitin-conjugating enzyme E2 D [EC:2.3.2.23]). The FPKM values of the two genes are greater than 78 in the 12 tissue **Supplementary Table S11** and obviously higher than that of the other 79 genes of the Toll and IMD signaling pathways (**Figure 8**). The two genes belonged to cluster 3 (**Figure 8**). *Drosophila* STAT forms a complex with transcription factor AP-1 and chromatin modifying proteins (Dsp1 and HDAC) to compete for Relish binding sites, thus regulating NF-κB immune responses (Kim et al., 2007). The E2-ubiquitin-conjugating enzymes UEV1a, Bendless (Ubc13), and Effete (Ubc5) help activate Dredd, which cleaves IMD, removing a 30-amino acid N-terminal fragment and creating a novel binding site for Iap2, which can, in turn, mediate ubiquitination for K63-linked IMD (Zhou et al., 2005; Meinander et al., 2012). Thus, the high level of expression of genes LG18_G00055 and LG15_G00391 may help increase innate immunity in *D. kikuchii*.

In addition, some genes showing distinct differential regulation at a certain stage, such as LG29_G00170/171/173 belonged to cluster one and were highly expressed only at three instar larvae (**Figure 8**, **Supplementary Table S11**). The three genes belongs to *MODSP* who is a modular serine protease and can integrate signals originating from the circulating recognition molecules *GNBP3* and *PGRP-SA* and connect them to the Grass-SPE-Spatzle extracellular pathway upstream of the Toll receptor (Buchon et al., 2009). A total of 19 genes in the genome of *D. kikuchii* belongs to *MODSP*. Just three out of the 19 genes showed significantly higher expression in only one sample than the remaining 11 samples (**Figure 8**, **Supplementary Table S11**). These results suggested that these three genes might have special functions in the activation of the Toll pathway by Gram-positive bacteria and fungi at larvae with three instar.

CONCLUSION

This study employed the most mainstream technology available to assemble a chromosome-level genome of *D. kikuchii* with high-quality genomic data, such as contig N50 of 20.89 Mb and scaffold N50 of 24.73 Mb. The contractions and expansions of gene families identified in this study provide ideas for future work in *D. kikuchii*; for example, immune gene family and insect hormone biosynthesis warrant further attention. A preliminary investigation of gene expansion and expression in the Toll and IMD signaling pathway suggests that *D. kikuchii* has a strong immune system that defends this pathogen against infection. The high-quality genomic data generated in this study provide the foundations to study chromosome evolution and immune mechanisms in *D. kikuchii*.

DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the NCBI repository, the genome sequence and annotation accession number is JAHHIN010000000; the transcriptome analysis accession numbers are SRR15334172-SRR15334183 and SRR15927891-SRR15927903.

AUTHOR CONTRIBUTIONS

JZ and BY designed the research. NZ and MJ authenticated the insect samples. JZ and PW analyzed the data using bioinformatics tools. YQ, JZ, and ZX collected samples. JZ drafted the manuscript. PW and NL further reviewed and decorated the manuscript. BY was acquired funding and was the project supervisor. All authors have read and agreed to the submitted version of the manuscript.

FUNDING

This work was funded by National Key R & D Program of China (2018YFC1200400) and Natural Science Foundation of Southwest Forestry University (SWFU-18200132).

ACKNOWLEDGMENTS

The authors thank China Nextomics Biosciences Co., Ltd. (Wuhan) for genome sequencing; College of Life Sciences, Southwest Forestry University for the work platform; and Kunming Cell Bank, Chinese Academy of Sciences for the chromosomes photo.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.728418/full#supplementary-material.

REFERENCES

Alioto, T., Blanco, E., Parra, G., and Guigó, R. (2018). Using Geneid to Identify Genes. *Curr. Protoc. Bioinformatics* 64, e56. doi:10.1002/cpbi.56

Bedell, J. A., Korf, I., and Gish, W. (2000). MaskerAid : a Performance Enhancement to RepeatMasker. *Bioinformatics* 16, 1040–1041. doi:10.1093/bioinformatics/16.11.1040

Benson, G. (1999). Tandem Repeats Finder: a Program to Analyze DNA Sequences. *Nucleic Acids Res.* 27, 573–580. doi:10.1093/nar/27.2.573

Biello, R., Singh, A., Godfrey, C. J., Fernández, F. F., Mugford, S. T., Powell, G., et al. (2021). A Chromosome-level Genome Assembly of the Woolly Apple Aphid, Eriosoma Lanigerum Hausmann (Hemiptera: Aphididae). *Mol. Ecol. Resour.* 21, 316–326. doi:10.1111/1755-0998.13258

Birney, E., and Durbin, R. (2000). Using GeneWise in the *Drosophila* Annotation experiment. *Genome Res.* 10, 547–548. doi:10.1101/gr.10.4.547

Bosco-Drayon, V., Poidevin, M., Boneca, I. G., Narbonne-Reveau, K., Royet, J., Charroux, B., et al. (2012). Peptidoglycan Sensing by the Receptor PGRP-LE in the *Drosophila* Gut Induces Immune Responses to Infectious Bacteria and Tolerance to Microbiota. *Cell Host & Microbe* 12, 153–165. doi:10.1016/j.chom.2012.06.002

Buchon, N., Poidevin, M., Kwon, H.-M., Guillou, A., Sottas, V., Lee, B.-L., et al. (2009). A Single Modular Serine Protease Integrates Signals from Pattern-Recognition Receptors Upstream of the *Drosophila* Toll Pathway. *Proc. Natl. Acad. Sci.* 106, 12442–12447. doi:10.1073/pnas.0901924106

Castresana, J. (2000). Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Mol. Biol. Evol.* 17, 540–552. doi:10.1093/oxfordjournals.molbev.a026334

Chen, C. (1990). *Integrated Management of Pine Caterpillar in China*. Beijing: China Forestry Press.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: an Ultra-fast All-In-One FASTQ Preprocessor. *Bioinformatics* 34, i884–i890. doi:10.1093/bioinformatics/bty560

Cheng, T., Wu, J., Wu, Y., Chilukuri, R. V., Huang, L., Yamamoto, K., et al. (2017). Genomic Adaptation to Polyphagy and Insecticides in a Major East Asian Noctuid Pest. *Nat. Ecol. Evol.* 1, 1747–1756. doi:10.1038/s41559-017-0314-4

Dai, Q.-Y., Gao, Q., Wu, C.-S., Chesters, D., Zhu, C.-D., and Zhang, A.-B. (2012). Phylogenetic Reconstruction and DNA Barcoding for Closely Related pine Moth Species (*Dendrolimus*) in China with Multiple Gene Markers. *PLoS One* 7, e32544. doi:10.1371/journal.pone.0032544

Danecek, P., and McCarthy, S. A. (2017). BCFtools/Csq: Haplotype-Aware Variant Consequences. *Bioinformatics* 33, 2037–2039. doi:10.1093/bioinformatics/btx100

Drummond, D. A., Bloom, J. D., Adami, C., Wilke, C. O., and Arnold, F. H. (2005). Why Highly Expressed Proteins Evolve Slowly. *Proc. Natl. Acad. Sci.* 102, 14338–14343. doi:10.1073/pnas.0504070102

Duret, L., and Mouchiroud, D. (2000). Determinants of Substitution Rates in Mammalian Genes: Expression Pattern Affects Selection Intensity but Not Mutation Rate. *Mol. Biol. Evol.* 17, 68–070. doi:10.1093/oxfordjournals.molbev.a026239

Feyereisen, R. (2012). *Insect Molecular Biology and Biochemistry*. Elsevier, 236–316. doi:10.1016/b978-0-12-384747-8.10008-x

Gautam, D. C., and Paul, S. (2013). Karyotype of Potato Tuber Moth, Phthorimaea Operculella (Zeller) - First Report from India. *Nucleus* 55, 171–173. doi:10.1007/s13237-012-0072-2

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., et al. (2008). Automated Eukaryotic Gene Structure Annotation Using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* 9, r7. doi:10.1186/gb-2008-9-1-r7

Han, M. V., Thomas, G. W. C., Lugo-Martinez, J., and Hahn, M. W. (2013). Estimating Gene Gain and Loss Rates in the Presence of Error in Genome Assembly and Annotation Using CAFE 3. *Mol. Biol. Evol.* 30, 1987–1997. doi:10.1093/molbev/mst100

Harrop, T. W. R., Guhlin, J., Mclaughlin, G. M., Permina, E., Stockwell, P., Gilligan, J., et al. (2020). High-quality Assemblies for Three Invasive Social Wasps from the *Vespula* Genus. *G3-genes Genom. Genet.* 10, 3479–3488. doi:10.1534/g3.120.401579

Hou, L., Zhan, S., Zhou, X., Li, F., and Wang, X. H. (2017). Advances in Research on Insect Genomics in China. *Chin. J. Appl. Entomol.* 54, 693–704.

Hou, T. (1993). Advances in the Control of Genus *Dendrolimus* (Lasiocampidae) of China. *For. Pest Dis.* 2, 40–42.

Hou, T. (1987). *The Pine Caterpillars in China*. Beijing: Science Press.

Hu, J., Fan, J., Sun, Z., and Liu, S. (2020). NextPolish: a Fast and Efficient Genome Polishing Tool for Long-Read Assembly. *Bioinformatics* 36, 2253–2255. doi:10.1093/bioinformatics/btz891

Kambris, Z., Brun, S., Jang, I.-H., Nam, H.-J., Romeo, Y., Takahashi, K., et al. (2006). *Drosophila* Immunity: a Large-Scale *In Vivo* RNAi Screen Identifies Five Serine Proteases Required for Toll Activation. *Curr. Biol.* 16, 808–813. doi:10.1016/j.cub.2006.03.020

Katoh, K., Misawa, K., Ki, K., and Miyata, T. (2002). MAFFT: a Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform. *Nucleic Acids Res.* 30, 3059–3066. doi:10.1093/nar/gkf436

Kawahara, A. Y., Plotkin, D., Espeland, M., Meusemann, K., Toussaint, E. F. A., Donath, A., et al. (2019). Phylogenomics Reveals the Evolutionary Timing and Pattern of Butterflies and Moths. *Proc. Natl. Acad. Sci. U.S.A.* 116, 22657–22663. doi:10.1073/pnas.1907847116

Keilwagen, J., Wenk, M., Erickson, J. L., Schattat, M. H., Grau, J., and Hartung, F. (2016). Using Intron Position Conservation for Homology-Based Gene Prediction. *Nucleic Acids Res.* 44, e89. doi:10.1093/nar/gkw092

Kim, L. K., Choi, U. Y., Cho, H. S., Lee, J. S., Lee, W.-b., Kim, J., et al. (2007). Down-Regulation of NF-Kb Target Genes by the AP-1 and STAT Complex during the Innate Immune Response in Drosophila. *Plos Biol.* 5, e238. doi:10.1371/journal.pbio.0050238

Kim, T.-I., and Kim, Y.-J. (2005). Overview of Innate Immunity in *Drosophila*. *BMB Rep.* 38, 121–127. doi:10.5483/bmbrep.2005.38.2.121

Kishida, Y., and Wang, M. (2011). "Lasiocampidae," in *Moths of Guangdong Nanling National Nature Reserve*. Editors M. Wang and Y. Kishida (Keltern, Germany: Keltern: Goecke & Evers), 140–145.

Konecka, E., Kaznowski, A., Stachowiak, M., and Macig, M. (2019). Activity of Spore-crystal Mixtures of New *Bacillus Thuringiensis* Strains against *Dendrolimus Pini* (Lepidoptera: Lasiocampidae) and *Spodoptera Exigua* (Lepidoptera: Noctuidae). *Folia For. Pol. Ser. A. For.* 60, 91–98.

Kong, X. B., Zhang, Z., Zhao, C. H., and Wang, H. B. (2007). Female Sex Pheromone of the Yunnan pine Caterpillar Moth *Dendrolimus Houi*: First (E,Z)-isomers in Pheromone Components of *Dendrolimus* Spp. *J. Chem. Ecol.* 33, 1316–1327. doi:10.1007/s10886-007-9313-2

Kononov, A., Ustyantsev, K., Wang, B., Mastro, V. C., Fet, V., Blinov, A., et al. (2016). Genetic Diversity Among Eight *Dendrolimus* Species in Eurasia (Lepidoptera: Lasiocampidae) Inferred from Mitochondrial COI and COII, and Nuclear ITS2 Markers. *BMC Genet.* 17, 157. doi:10.1186/s12863-016-0463-5

Kunimi, Y. (2007). Current Status and Prospects on Microbial Control in Japan. *J. Invertebr. Pathol.* 95, 181–186. doi:10.1016/j.jip.2007.03.007

Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H.-H., Rognes, T., and Ussery, D. W. (2007). RNAmmer: Consistent and Rapid Annotation of Ribosomal RNA Genes. *Nucleic Acids Res.* 35, 3100–3108. doi:10.1093/nar/gkm160

Langmead, B., and Salzberg, S. L. (2012). Fast Gapped-Read Alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi:10.1038/nmeth.1923

Leulier, F., Parquet, C., Pili-Floury, S., Ryu, J.-H., Caroff, M., Lee, W.-J., et al. (2003). The *Drosophila* Immune System Detects Bacteria through Specific Peptidoglycan Recognition. *Nat. Immunol.* 4, 478–484. doi:10.1038/ni922

Li, H., and Durbin, R. (2010). Fast and Accurate Long-Read Alignment with Burrows-Wheeler Transform. *Bioinformatics* 26, 589–595. doi:10.1093/bioinformatics/btp698

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/bioinformatics/btp352

Li, L., Stoeckert, C. J., and Roos, D. (2003). OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Res.* 13, 2178–2189. doi:10.1101/gr.1224503

Li, X., Schuler, M. A., and Berenbaum, M. R. (2007). Molecular Mechanisms of Metabolic Resistance to Synthetic and Natural Xenobiotics. *Annu. Rev. Entomol.* 52, 231–253. doi:10.1146/annurev.ento.51.110104.151104

Li, Y., Xiang, Q., Zhang, Q., Huang, Y., and Su, Z. (2012). Overview on the Recent Study of Antimicrobial Peptides: Origins, Functions, Relative Mechanisms and Application. *Peptides* 37, 207–215. doi:10.1016/j.peptides.2012.07.001

Liu, P., Sanalkumar, R., Bresnick, E. H., Keleş, S., and Dewey, C. N. (2016). Integrative Analysis with ChIP-Seq Advances the Limits of Transcript Quantification from RNA-Seq. *Genome Res.* 26, 1124–1133. doi:10.1101/gr.199174.115

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. *Nucleic Acids Res.* 25, 955–964. doi:10.1093/nar/25.5.955

Lu, Y., Su, F., Li, Q., Zhang, J., Li, Y., Tang, T., et al. (2020). Pattern Recognition Receptors in *Drosophila* Immune Responses. *Dev. Comp. Immunol.* 102, 103468. doi:10.1016/j.dci.2019.103468

Luo, S., Tang, M., Frandsen, P. B., Stewart, R. J., and Zhou, X. (2018). The Genome of an Underwater Architect, the Caddisfly *Stenopsyche Tienmushanensis* Hwang (Insecta: Trichoptera). *Gigascience* 7, 1–12. doi:10.1093/gigascience/giy143

Majoros, W. H., Pertea, M., and Salzberg, S. L. (2004). TigrScan and GlimmerHMM: Two Open Source *Ab Initio* Eukaryotic Gene-Finders. *Bioinformatics* 20, 2878–2879. doi:10.1093/bioinformatics/bth315

Marçais, G., and Kingsford, C. (2011). A Fast, Lock-free Approach for Efficient Parallel Counting of Occurrences of K-Mers. *Bioinformatics* 27, 764–770. doi:10.1093/bioinformatics/btr011

Matsumura, S. (1927). New Species and Subspecies of Moths from the Japanese empire. *J. Assoc. Physicians India* 19, 1–91.

Meinander, A., Runchel, C., Tenev, T., Chen, L., Kim, C.-H., Ribeiro, P. S., et al. (2012). Ubiquitylation of the Initiator Caspase DREDD Is Required for Innate Immune Signalling. *EMBO J.* 31, 2770–2783. doi:10.1038/emboj.2012.121

Men, Q., Xue, G., Mu, D., Hu, Q., and Huang, M. (2017). Mitochondrial DNA Markers Reveal High Genetic Diversity and strong Genetic Differentiation in Populations of *Dendrolimus Kikuchii* Matsumura (Lepidoptera: Lasiocampidae). *PLoS One* 12, e0179706. doi:10.1371/journal.pone.0179706

Mikkola, K., and Ståhls, G. (2008). Morphological and Molecular Taxonomy of *Dendrolimus Sibiricus* Chetverikov stat.Rev. And Allied Lappet Moths (Lepidoptera: Lasiocampidae), with Description of a New Species. *Entomol. Fennica* 19, 65–85. doi:10.33338/ef.84417

Neyen, C., Poidevin, M., Roussel, A., and Lemaitre, B. (2012). Tissue- and Ligand-specific Sensing of Gram-Negative Infection in *Drosophila* by PGRP-LC Isoforms and PGRP-LE. *J.I.* 189, 1886–1897. doi:10.4049/jimmunol.1201022

Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a Pipeline to Accurately Annotate Core Genes in Eukaryotic Genomes. *Bioinformatics* 23, 1061–1067. doi:10.1093/bioinformatics/btm071

Senol Cali, D., Kim, J. S., Ghose, S., Alkan, C., and Mutlu, O. (2019). Nanopore Sequencing Technology and Tools for Genome Assembly: Computational Analysis of the Current State, Bottlenecks and Future Directions. *Brief. Bioinform.* 20, 1542–1559. doi:10.1093/bib/bby017

Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C.-J., Vert, J.-P., et al. (2015). HiC-Pro: an Optimized and Flexible Pipeline for Hi-C Data Processing. *Genome Biol.* 16, 259. doi:10.1186/s13059-015-0831-x

Simão, 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, 3210–3212. doi:10.1093/bioinformatics/btv351

Stamatakis, A. (2014). RAxML Version 8: A Tool for Phylogenetic Analysis and post-analysis of Large Phylogenies. *Bioinformatics* 30, 1312–1313. doi:10.1093/bioinformatics/btu033

Stanke, M., and Waack, S. (2003). Gene Prediction with a Hidden Markov Model and a New Intron Submodel. *Bioinformatics* 19 Suppl 2, ii215–25. doi:10.1093/bioinformatics/btg1080

Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using Native and Syntenically Mapped cDNA Alignments to Improve De Novo Gene Finding. *Bioinformatics* 24, 637–644. doi:10.1093/bioinformatics/btn013

Sun, H., Ding, J., Piednoël, M., and Schneeberger, K. (2018). findGSE: Estimating Genome Size Variation within Human and Arabidopsis Using K-Mer Frequencies. *Bioinformatics* 34, 550–557. doi:10.1093/bioinformatics/btx637

Tong, Q., and He, J. (2009). Study on Biological Characters of *Dendrolimus Kikuchii* Maxsumura and Food Consumption of its Larva. *J. Anhui Agric. Sci.* 37, 13122–13124.

Urasaki, N., Takagi, H., Natsume, S., Uemura, A., Taniai, N., Miyagi, N., et al. (2017). Draft Genome Sequence of Bitter Gourd (*Momordica Charantia*), a Vegetable and Medicinal Plant in Tropical and Subtropical Regions. *DNA Res.* 24, 51–58. doi:10.1093/dnares/dsw047

Valanne, S., Wang, J.-H., and Rämet, M. (2011). The *Drosophila* Toll Signaling Pathway. *J.I.* 186, 649–656. doi:10.4049/jimmunol.1002302

Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and Accurate De Novo Genome Assembly from Long Uncorrected Reads. *Genome Res.* 27, 737–746. doi:10.1101/gr.214270.116

Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., et al. (2017). GenomeScope: Fast Reference-free Genome Profiling from Short Reads. *Bioinformatics* 33, 2202–2204. doi:10.1093/bioinformatics/btx153

Wang, X., and Wang, L. (2016). GMATA: An Integrated Software Package for Genome-Scale SSR Mining, Marker Development and Viewing. *Front. Plant Sci.* 7, 1350. doi:10.3389/fpls.2016.01350

Wang, Y., Jiang, S. Z., Zheng, C. X., Zhou, Q. X., Zhang, Y. L., and Jiang, F. Z. (1999). Study on Eye Injury Caused by *Dendrolimus* Spp. *Chin. J. Ocul. Trauma Occup. Eye Dis.* 21, 187–188.

Wang, Y., Kong, X. B., Zhang, S. F., Liu, F., Zhang, Z., and Yan, S. C. (2019). Sequencing and Analysis of the Complete Mitochondrial Genome of *Dendrolimus Houi*. *For. Res* 32, 11–20.

Wick, R. R., Judd, L. M., and Holt, K. E. (2019). Performance of Neural Network Basecalling Tools for Oxford Nanopore Sequencing. *Genome Biol.* 20, 129. doi:10.1186/s13059-019-1727-y

Xiao, G. (1992). *Forest Insects of china*. 2nd Edition. Beijing: China Forestry Press.

Yang, L., Wang, J., Jin, H., Fang, Q., Yan, Z., Lin, Z., et al. (2019). Immune Signaling Pathways in the Endoparasitoid, *Pteromalus Puparum*. *Arch. Insect Biochem. Physiol.* 103, e21629. doi:10.1002/arch.21629

Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi:10.1093/molbev/msm088

You, M., Yue, Z., He, W., Yang, X., Yang, G., Xie, M., et al. (2013). A Heterozygous Moth Genome Provides Insights into Herbivory and Detoxification. *Nat. Genet.* 45, 220–225. doi:10.1038/ng.2524

Yu, Y., Park, J.-W., Kwon, H.-M., Hwang, H.-O., Jang, I.-H., Masuda, A., et al. (2010). Diversity of Innate Immune Recognition Mechanism for Bacterial Polymeric Meso-Diaminopimelic Acid-type Peptidoglycan in Insects. *J. Biol. Chem.* 285, 32937–32945. doi:10.1074/jbc.m110.144014

Zhang, S. F., Zhang, Z., Wang, H. B., Kong, X. B., Luo, J. T., Yang, Z. W., et al. (2014). Genome Size Determination of Several *Dendrolimus* Species in China. *For. Res* 27, 583–589.

Zhang, S., Shen, S., Peng, J., Zhou, X., Kong, X., Ren, P., et al. (2020). Chromosome-level Genome Assembly of an Important pine Defoliator, Dendrolimus Punctatus (Lepidoptera; Lasiocampidae). *Mol. Ecol. Resour.* 20, 1023–1037. doi:10.1111/1755-0998.13169

Zhou, R., Silverman, N., Hong, M., Liao, D. S., Chung, Y., Chen, Z. J., et al. (2005). The Role of Ubiquitination in *Drosophila* Innate Immunity. *J. Biol. Chem.* 280, 34048–34055. doi:10.1074/jbc.m506655200

Zhuang, W., Chen, H., Yang, M., Wang, J., Pandey, M. K., Zhang, C., et al. (2019). The Genome of Cultivated Peanut Provides Insight into Legume Karyotypes, Polyploid Evolution and Crop Domestication. *Nat. Genet.* 51, 865–876. doi:10.1038/s41588-019-0402-2

Zou, Z., Shin, S. W., Alvarez, K. S., Kokoza, V., and Raikhel, A. S. (2010). Distinct Melanization Pathways in the Mosquito *Aedes aegypti*. *Immunity* 32, 41–53. doi:10.1016/j.immuni.2009.11.011

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Zhou, Wu, Xiong, Liu, Zhao, Ji, Qiu and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.