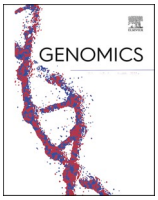
Genomics 114 (2022) 110440

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno



Chromosomal-level reference genome of the moth *Heortia vitessoides* (Lepidoptera: Crambidae), a major pest of agarwood-producing trees

Sean T.S. Law [a,*,1], Wenyan Nong [a,1], Wai Lok So [a], Tobias Baril [b], Thomas Swale [c], Chi Bun Chan [d], Stephen S. Tobe [e], Zhen-Peng Kai [f], William G. Bendena [g], Alexander Hayward [b], Jerome H.L. Hui [a,*]

[a] School of Life Sciences, Simon F.S. Li Marine Science Laboratory, State Key Laboratory of Agrobiotechnology, The Chinese University of Hong Kong, Hong Kong, China
[b] University of Exeter, United Kingdom
[c] Dovetail Genomics, United States of America
[d] School of Biological Science, The University of Hong Kong, Hong Kong, China
[e] Department of Cell and Systems Biology, University of Toronto, Canada
[f] School of Chemical and Environmental Engineering, Shanghai Institute of Technology, Shanghai, China
[g] Department of Biology, Queen's University, Canada

ARTICLE INFO

Keywords:
Aquilaria
Sesquiterpenoid
microRNA
Lepidopteran

ABSTRACT

The moth *Heortia vitessoides* Moore (Lepidoptera: Crambidae) is a major pest of ecologically, commercially and culturally important agarwood-producing trees in the genus *Aquilaria*. In particular, *H. vitessoides* is one of the most destructive defoliating pests of the incense tree *Aquilaria sinesis*, which produces a valuable fragrant wood used as incense and in traditional Chinese medicine [33]. Nevertheless, a genomic resource for *H. vitessoides* is lacking. Here, we present a chromosomal-level assembly for *H. vitessoides*, consisting of a 517 megabase (Mb) genome assembly with high physical contiguity (scaffold N50 of 18.2 Mb) and high completeness (97.9% complete BUSCO score). To aid gene annotation, 8 messenger RNA transcriptomes from different developmental stages were generated, and a total of 16,421 gene models were predicted. Expansion of gene families involved in xenobiotic metabolism and development were detected, including duplications of cytosolic sulfotransferase (SULT) genes shared among lepidopterans. In addition, small RNA sequencing of 5 developmental stages of *H. vitessoides* facilitated the identification of 85 lepidopteran conserved microRNAs, 94 lineage-specific microRNAs, as well as several microRNA clusters. A large proportion of the *H. vitessoides* genome consists of repeats, with a 29.12% total genomic contribution from transposable elements, of which long interspersed nuclear elements (LINEs) are the dominant component (17.41%). A sharp decrease in the genome-wide percentage of LINEs with lower levels of genetic distance to family consensus sequences suggests that LINE activity has peaked in *H. vitessoides*. In contrast, opposing patterns suggest a substantial recent increase in DNA and LTR element activity. Together with annotations of essential sesquiterpenoid hormonal pathways, neuropeptides, microRNAs and transposable elements, the high-quality genomic and transcriptomic resources we provide for the economically important moth *H. vitessoides* provide a platform for the development of genomic approaches to pest management, and contribute to addressing fundamental research questions in Lepidoptera.

1. Introduction

Among the ~160,000 described moth species, the family Crambidae contains >10,000 species which are commonly known as grass moths. Several crambids are well-known crop pests with stem borer larvae that affect agricultural products such as bean, corn, millet, rice, and sugarcane. The crambid moth *Heortia vitessoides* Moore is widely distributed in Southeast Asia, including southern China, Fiji, India, Malaysia, and Thailand, and can also be found in Australia. *H. vitessoides* larvae feed on trees in the genus *Aquilaria* (family Thymelaeaceae), which are also commonly known as "lign aloes", that produce valuable agarwood which is used as incense and in traditional Chinese medicine

* Corresponding authors.
E-mail addresses: seantslaw@link.cuhk.edu.hk (S.T.S. Law), jeromehui@cuhk.edu.hk (J.H.L. Hui).
[1] Contributed equally.

https://doi.org/10.1016/j.ygeno.2022.110440
Received 20 May 2022; Received in revised form 13 July 2022; Accepted 24 July 2022
Available online 26 July 2022

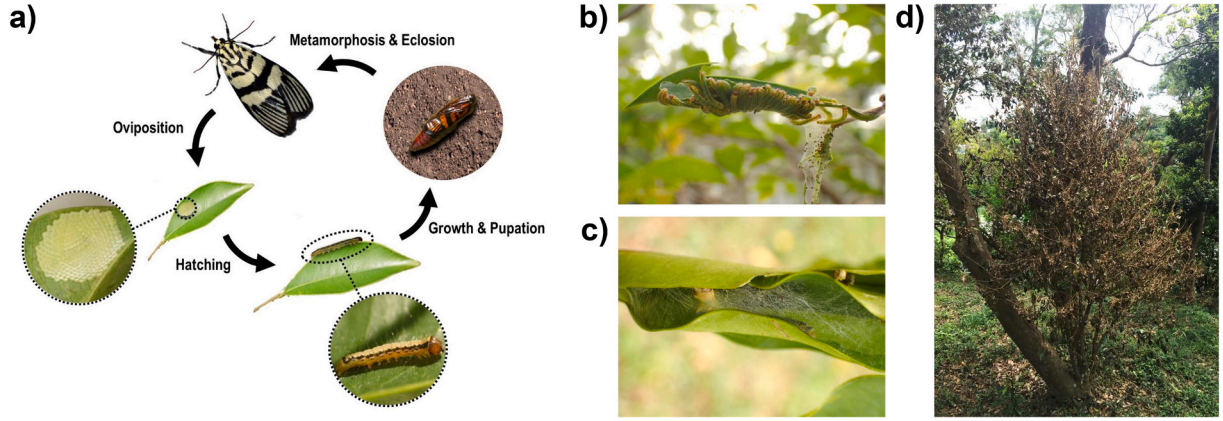


Fig. 1. a) Life cycle of *Heortia vitessoides*; b) Aggregated feeding behavior of third instar larvae; c) Solitary feeding of fifth instar larva with leaf-folding behavior; d) an incense tree *Aquilaria sinensis* completely defoliated by *H. vitessoides* at the countryside in Hong Kong.

[35,81,94].

H. vitessoides is recognised as the most destructive pest of the lign aloes *A. malaccensis* and *A. sinensis* in Malaysia and southern China, respectively [50,81]. During its larval stages, the first to third instars aggregate and feed on leaves of *Aquilaria* species, while the older instars tend to be solitary ([50], Fig. 1). Infestations of these leaf defoliators have been reported in agarwood plantations with multiple detrimental effects, including lowering agarwood production, reducing the survival of seedlings, and increasing the costs of pest management [75].

Previous studies on *H. vitessoides* have focused on the mechanisms underlying its xenobiotic detoxification [13,14], development [53,54,86], ecological relationships [69], and its potential spread under climate change [94]. With the advancement of sequencing technologies, molecular resources including well-assembled genomes, transcriptomes and small RNA-seq data are becoming increasingly available for lepidopterans. These resources can be highly beneficial for elucidating the evolutionary relationships between taxa [39], deepening the understanding of detoxification mechanisms [57], and elucidating genome biology more generally, such as revealing the burst of microRNA innovation in lepidopteran lineages [73]. For Crambidae, to date, genomic resources are available for five subfamilies including Acentropinae (*Parapoynx stratiotata*), Crambinae (*Chilo suppressalis*, *Diatraea saccharalis* and *Chrysoteuchia culmella*), Pyraustinae (*Ostrinia scapulalis*, *Ostrinia nubilalis* and *Leucinodes orbonalis*), Schoenobiinae (*Scirpophaga incertulas*), and Spilomelinae (*Cnaphalocrocis medinalis* and *Cnaphalocrocis exigua*) [17,22,36,38,43,55,95,98]. *H. vitessoides* belongs to the subfamily Odontiinae, and despite its ecological importance, genomic resources for this moth lineage remain lacking.

Here, we provide genomic and transcriptomic (both mRNA and small RNAs) resources for the damaging pest species *H. vitessoides,* to better understand the evolutionary genomics of moths, and provide a platform for further research and development into its control.

2. Materials & methods

2.1. Sample collection and genome sequencing

Specimens of the moth *H. vitessoides* were collected from The Chinese University of Hong Kong. Genomic DNA (gDNA) was extracted from a 5th instar larva using the PureLink Genomic DNA Mini Kit (Invitrogen) following the manufacturer's protocol. Extracted gDNA was subjected to quality control using a Nanodrop spectrophotometer (Thermo Scientific) and gel electrophoresis. The qualifying sample was sent to Novogene (Hong Kong, China) Co. Ltd. and Dovetail Genomics LLC (Santa Cruz, CA, USA) for library preparation and sequencing. A 10× Genomics Chromium linked-read library was prepared and sequenced using the Illumina HiSeq X platform to produce 160 million 2 × 150 bp paired-end sequences. The length-weighted mean molecule length is 16,819.15 bp, and the raw data can be found at NCBI's Small Read Archive (SRR12799705).

2.2. Chicago library preparation and sequencing

A Chicago library was prepared as described previously [67]. Briefly, ~500 ng of high molecular weight (HMW) gDNA (mean fragment length = 85 kbp) was reconstituted into chromatin *in vitro* and fixed with formaldehyde. Fixed chromatin was digested with *Dpn*II, the 5′ overhangs filled in with biotinylated nucleotides, and free blunt ends were ligated. After ligation, crosslinks were reversed and the DNA was purified from protein. Purified DNA was treated to remove biotin that was not internal to ligated fragments. The DNA was then sheared to ~350 bp mean fragment size and sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. The libraries were sequenced on an Illumina HiSeq X to produce 170 million 2 × 150 bp paired-end reads (1–100 kb pairs). The raw data have been submitted to NCBI's Small Read Archive (SRR13205889).

2.3. Dovetail library preparation and sequencing

A Dovetail HiC library was prepared in a similar manner to that of the Chicago library as described previously [51]. The libraries were sequenced on an Illumina HiSeq X sequencer to produce 132 million 2 × 150 bp paired-end reads (100–10,000 kb pairs). The raw data have been submitted to NCBI's Small Read Archive (SRR13205888).

2.4. mRNA and sRNA transcriptome sequencing

Total RNA was extracted from different developmental stages, namely egg, day 1 and 3 of each instar (from 1st to 5th) larvae, pre-pupa, and male and female adult using the mirVana miRNA Isolation Kit (Ambion), following the manufacturer's protocol. Pools of 30, 10, 5, and 1 individual(s) were used during RNA extraction from egg, 1st instar larvae, 2nd instar larvae, and other developmental stages respectively. The quality of extracted total RNA was inspected using a Nanodrop spectrophotometer (Thermo Scientific), gel electrophoresis, and an Agilent 2100 Bioanalyser (Agilent RNA 6000 Nano Kit). Qualified samples were sent to Novogene (Hong Kong, China) Co. Ltd. for library construction and sequencing. PolyA-selected RNA-Sequencing libraries were prepared using the TruSeq RNA Sample Prep Kit v2. Insert sizes and library concentrations of final libraries were examined using an Agilent 2100 Bioanalyser instrument (Agilent DNA 1000 Reagents). Similarly, isolated small RNA was subjected to Novogene (Hong Kong,

China) Co. Ltd. for HiSeq Small RNA library construction and 50 bp single-end (SE) sequencing. Details of the sequencing data can be found in (Supplementary Table 1).

2.5. Genome assembly

The contamination reads were removed by Kraken2 [91]. Processed reads were used to make a *de novo* assembly in Supernova (v 2.1.1) with default parameters (raw coverage = 56.55×). The Supernova output pseudohaplotype assembly, shotgun reads, Chicago library reads, and Dovetail HiC library reads were used as input data for HiRise, a software pipeline designed specifically for scaffolding genome assemblies using proximity ligation data [67]. An iterative analysis was conducted as follows. First, Shotgun and Chicago library sequences were aligned to the draft input assembly using a modified SNAP read mapper (http://snap.cs.berkeley.edu). Separation of Chicago read pairs mapped within draft scaffolds was analyzed in HiRise to produce a likelihood model of genomic distance between read pairs, which was used to identify and break putative misjoins, score prospective joins, and make joins above a threshold. After aligning and scaffolding Chicago data, Dovetail HiC library sequences were aligned and scaffolded following the same method. After scaffolding, shotgun sequences were used to close gaps between contigs. Genome size estimation was analyzed using a k-mer-based statistical approach in the GenomeScope2 webtool [84]. Completeness of genome assembly was examined by BUSCO (v5.1.3, metazoa_odb10 created by 2021–02–17, [76]). Syntenic relationships among *H. vitessoides* and four other lepidopterans with different numbers of chromosomes, namely tobacco cutworm *Spodoptera litura* (GCA_002706865.2; [15]), monarch butterfly *Danaus plexippus plexippus* (GCF_009731565.1; [23]), silkmoth *Bombyx mori* (SilkBase, [40]), and rice stem borer *C. suppressalis* (InsectBase, [55]), were also analyzed using MCScanX [88] and visualized with TBtools version 1.055 [12].

2.6. Gene model prediction

Gene models were predicted as previously described [48,59]. Briefly, gene models were trained and predicted using funannotate (v1.7.4, https://github.com/nextgenusfs/funannotate) [61] specifying the parameters "–repeats2evm –protein_evidence uniprot_sprot.fasta –genemark_mode ET –busco_seed_species fly –optimize_augustus –busco_db arthropoda –organism other –max_intronlen 350000". Gene models from several prediction sources including GeneMark, high-quality Augustus predictions (HiQ), pasa, Augustus, GlimmerHM and snap were passed to Evidence Modeler and used to generate gene model annotation files, and applied in PASA to update EVM consensus predictions, add UTR annotations and models for alternatively spliced isoforms. Protein-coding gene models were then blasted (blastp) to the NCBI nr and swissprot databases using diamond (v0.9.24) specifying the following parameters "–more-sensitive –evalue 1e-3", and mapped by HISAT2 (version 2.1.0) [41] with transcriptome reads. Gene models with no homology to any known protein in the nr and swissprot databases and no mRNA support were removed from the final version. The expression profile of gene models from different developmental stages was generated by extracting transcript per million (TPM) values from mapped reads using StringTie [64] and visualized with row Z score in Heatmapper [1].

2.7. microRNA annotation

The obtained small RNA sequencing data were trimmed and processed using the mirDeep2 package [20]. The microRNA annotation and novel microRNA characterisation as previous described [60,72]. Briefly, known microRNAs were identified by comparing predicted microRNA hairpins to metazoan microRNA precursor sequences from miRBase with BLASTn (e-value <0.01) [42]. Other predicted microRNAs with no significant sequence similarity to any of the microRNAs in miRBase were manually checked. Novel microRNAs were characterized according to the listed criteria for microRNAs in MirGeneDB 2.0 [21].

2.8. Comparative analyses for gene family evolution

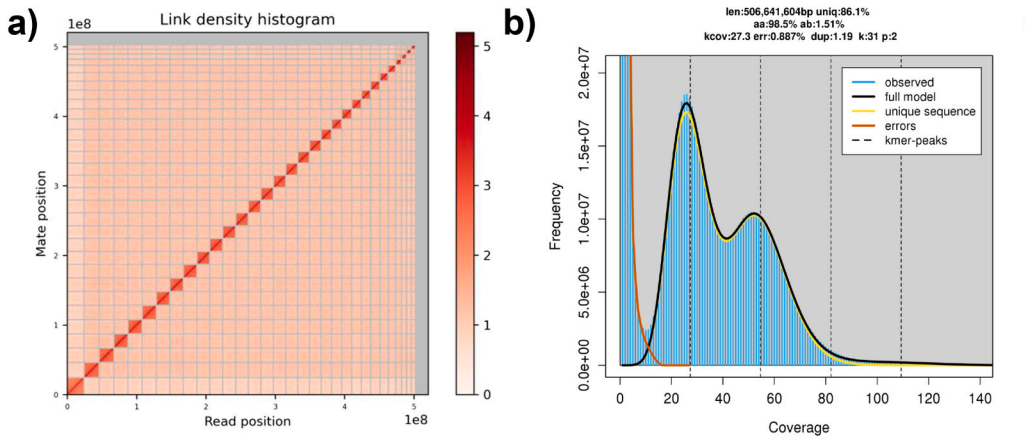
The longest transcript of whole protein gene sets from *H. vitessoides* genome and 25 other Lepidoptera genomes across 15 families together with 7 hexapod outgroup species were used for construction of a putative orthologous gene family classification using OrthoMCL (version 2.0.9) [46]. Total universal single-copy orthologs were extracted and aligned using MAFFT v7.271 [37] and trimmed with trimAl v1.4.rev15 [11]. From the concatenated alignment, a maximum likelihood phylogenetic tree was constructed using RAxML v8.2.9 [78] with PROTGAMMAILGF model and 1000 bootstrap replicates, with the springtail *S. curviseta* specified as the outgroup. The maximum likelihood tree was used as an input tree to infer divergence times using BEAST v2.6.2 [8]. Strict clock rate and a calibrated Yule model with gamma prior distribution were employed to estimate divergences time. Five divergence time points were used for calibration according to the TimeTree web database (http://timetree.org/): *Cryptotermes secundus* - *Nilaparvata lugens* (mean: 358 million years ago (MYA), sigma: 15), *Photinus pyralis* - *Tribolium castaneum* (mean: 286 MYA, sigma: 10), *Papilio bianor* - *Danaus plexippus plexippus* (mean: 106 MYA, sigma: 10), *Arctia plantaginis* - *Spodoptera frugiperda* (mean: 69, sigma: 4), and *Operophtera brumata* - *Ectropis grisescens* (mean: 16.9 MYA, sigma: 0.5).

2.9. Gene family expansion and contraction analyses

The putative orthologous genes from OrthoMCL and phylogenetic tree topology inferred from universal orthologs were used as inputs in CAFE (Computational Analysis of gene Family Evolution, v4.2.1) [26]. The global gene birth-death parameter (λ) was first estimated from the input tree and gene family counts to describe the probability of gain or loss of any gene and a family-wise *p*-value of threshold 0.01 (as default) was employed with resampling (×1000) to indicate if significant expansion or contraction occurred in each gene family across species. Gene families that are undergoing expansion and with relation to detoxification and developmental processes were further investigated with protein family searches using HMMER (version 3.3.1; cut-off *E*-value $<10^{-5}$) [18]. Validation of protein domains of extracted sequences were processed with InterproScan (version 5.31–70.0) [56] and CD-Search (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) to remove false positive domain hits. Gene family trees were constructed to identify if an expansion occurred in specific groups within certain gene families. The validated protein sequences were aligned using MAFFT (--maxiterate 1000 –genafpair –ep 0) [37] and manually adjusted in MEGA7 [44], followed by tree construction using FastTree [66] with the JTT model and visualized in Evolview v3 [79]. For the identification of cytochrome P450, the P450 Blast server (http://drnelson.uthsc.edu/CytochromeP450.html; [58]) was used for nomenclature.

2.10. Gene family and microsynteny analyses of cytosolic sulfotransferases

Cytosolic sulfotransferases (SULTs) were identified as one of the rapid expanding gene families in *H. vitessoides* and several other lepidopterans. To further investigate the evolutionary process for this gene family, a gene family tree was constructed using SULTs identified from the protein search employed in this study together with mammalian and plant SULTs [6,29], and inclusion of 8 SULTs from blastp searches of the *Aquilaria sinensis* genome [59], the host of *H. vitessoides*. Pairwise sequence identity was calculated using Clustal Omega (v1.2.4, [77]). Microsynteny was assessed by comparing *H. vitessoides* and other 31 hexapod genomes (*Operophtera brumata* was excluded due to genome quality) using MCScanX with default settings (tolerating maximum gene



c)

Heortia vitessoides genome assembly	
No. of scaffolds	5,313
Assembly size	517,196,383
Gap content (N)	1.21678%
Contig N50	882,731
Contig L50	160
Scaffold N50	18,244,730
Scaffold L50	13
No. of gene models	16,421
Complete BUSCOs	97.90%

Fig. 2. Genomic features of *Heortia vitessoides*. a) Hi-C interaction of 31 linkage groups; b) GenomeScope profile of *H. vitessoides* using k-mer (K = 31); c) Summary of genome assembly statistics of *H. vitessoides*.

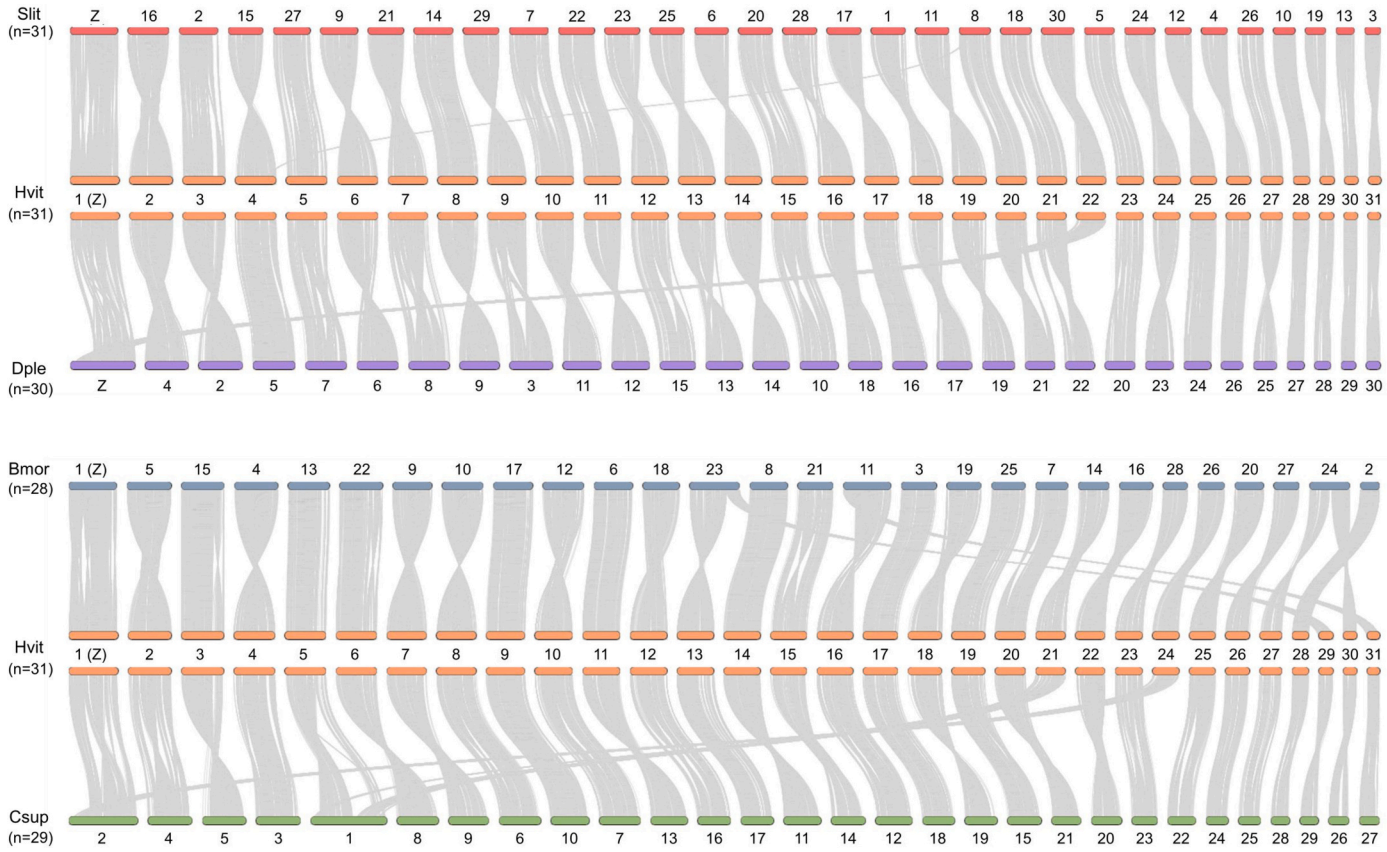


Fig. 3. Genomic synteny between *Heortia vitessoides*, *Spodoptera litura*, *Danaus plexippus plexippus*, *Bombyx mori*, and *Chilo suppressalis*.

gap of 25) [88]. The syntenic blocks with expansion of SULTs were visualized in gggenomes [25]. Jalview (v2.11.2.2) was used to visualize sequence alignments and SWISS-MODEL was used for modelling protein structures of selected SULTs [89,90].

2.11. Annotation of gene pathways and neuropeptides

Gene family sequences involved in biosynthesis pathways of juvenile hormone and edcysone were first retrieved from the KEGG database and used for homology searches against the gene models in the genome manually using BLASTP. Putatively identified genes were tested by reciprocal blasts in the NCBI nr database using BLASTP. For neuropeptides, amino acid sequences were analyzed by MEGA 7.0 [44], by BLAST algorithm at the online NCBI database and by SignalP 3.0 [5].

2.12. Repeat annotation

Transposable elements (TEs) were annotated using the automated Earl Grey TE annotation pipeline (version 1.2, https://github.com/TobyBaril/EarlGrey) as previously described [3,4]. Briefly, this pipeline first identified known TEs from the *lepidoptera* subset of Dfam (release 3.4) and RepBase (release 20,181,026) [30,34]. Following this, *de novo* TEs were identified and consensus boundaries were extended using an automated "BLAST, Extract, Extend" process [65]. Redundant

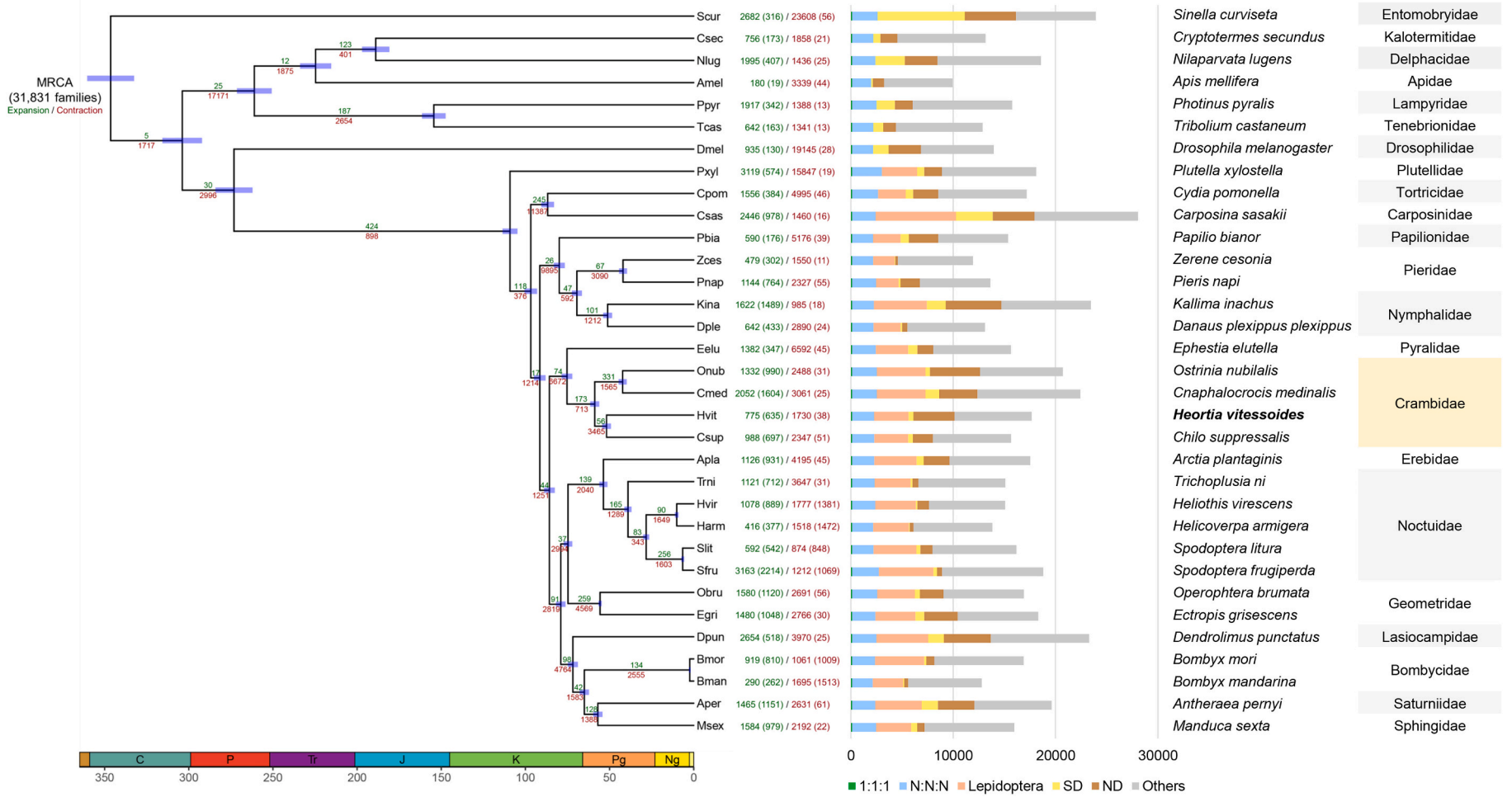


Fig. 4. Orthologues and phylogenetic analyses. The right panel shows the bar chart of orthologous grouping among the 33 taxa. 1:1:1 indicates single-copy orthologues; N:N:N indicates multicopy orthologous genes; "Lepidoptera" indicates orthologous genes shared between at least two lepidopteran species; SD indicates single copy of species-specific genes; ND indicates multiple copies of species-specific genes; "others" indicates any other orthologous genes shared between at least any two taxa. The left panel shows the maximum likelihood species phylogenetic tree constructed from the concatenation of 125 single-copy orthologues. Blue bars at the internodes indicate the 95% highest posterior density (HPD) of divergence times estimated using BEAST; the gene families undergone expansion and contraction are shown in green and red at the branches and tips respectively, with the number of rapid evolving gene families shown in brackets; MRCA indicates the most recent common ancestor. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Number of genes in selected expanding gene families across 33 hexapod taxa.

Family	Species	UGT	SULT	abhydrolipase	SDR	CYP	C-type lectin	Chorion
		PF00201	PF00685	PF04083	PF00106	PF00067	PF00059	PF01723
Entomobryidae	*Sinella curviseta*	80	6	49	141	349	92	0
Kalotermitidae	*Cryptotermes secundus*	26	6	12	74	99	23	0
Delphacidae	*Nilaparvata lugens*	24	6	13	61	72	13	0
Apidae	*Apis mellifera*	12	3	4	43	46	9	0
Lampyridae	*Photinus pyralis*	65	14	19	103	167	9	0
Tenebrionidae	*Tribolium castaneum*	40	6	22	78	126	11	0
Drosophilidae	*Drosophila melanogaster*	35	4	46	57	89	35	0
Plutellidae	*Plutella xylostella*	28	4	35	100	97	16	17
Tortricidae	*Cydia pomonella*	30	15	36	103	137	37	21
Carposinidae	*Carposina sasakii*	37	4	25	97	75	21	22
Papilionidae	*Papilio bianor*	31	4	27	79	127	26	46
Pieridae	*Zerene cesonia*	52	8	13	87	84	16	7
Pieridae	*Pieris napi*	44	17	9	56	96	13	14
Nymphalidae	*Kallima inachus*	62	5	28	108	155	28	32
	Danaus plexippus plexippus	35	7	15	90	76	14	22
Pyralidae	*Ephestia elutella*	25	8	45	98	110	16	12
Crambidae	*Ostrinia nubilalis*	35	10	37	76	94	25	7
	Cnaphalocrocis medinalis	32	6	27	92	80	23	9
	Heortia vitessoides	34	13	27	88	67	37	26
	Chilo suppressalis	26	10	7	47	73	34	42
Erebidae	*Arctia plantaginis*	26	14	20	78	102	39	10
Noctuidae	*Trichoplusia ni*	55	8	23	120	114	13	3
	Heliothis virescens	27	9	26	100	101	28	6
	Helicoverpa armigera	46	7	24	95	106	36	20
	Spodoptera litura	52	11	23	151	122	32	54
	Spodoptera frugiperda	48	16	26	144	181	28	66
Geometridae	*Operophtera brumata*	37	7	25	91	127	33	1
	Ectropis grisescens	43	8	22	114	96	25	7
Lasiocampidae	*Dendrolimus punctatus*	30	9	32	120	131	34	30
Bombycidae	*Bombyx mori*	37	8	26	95	82	18	22
	Bombyx mandarina	41	6	20	88	70	21	91
Saturniidae	*Antheraea pernyi*	27	19	32	165	104	19	67
Sphingidae	*Manduca sexta*	40	6	26	128	118	30	72

CYP, Cytochrome P450; SDR, Short-Chain Dehydrogenase/Reductase; SULT, cytosolic sulfotransferase; UGT, UDP-glucuronosyltransferase.

sequences were removed from the consensus library before the genome assembly was annotated with the combined known and *de novo* TE libraries. Annotations were processed to remove overlaps and defragment annotations before final TE quantification.

3. Results

3.1. High quality moth genome

Here, we present a high-quality chromosome-level genome assembly of the moth *H. vitessoides* (2n = 62), for which 97.02% of the genomic sequences are contained on 31 pseudomolecules (~517 Mb, scaffold N50 = 18.2 Mb) (Fig. 2a, Supplementary Table 1). The assembled genome size is compatible with the estimated genome size (499.5–506.6 Mb estimated by k-mer sizes 21 and 31 respectively)(Fig. 2b). From 8 transcriptomes sequenced from different development stages, a total of 16,421 predicted gene models including 13,867 protein coding genes and 2554 tRNA genes, were annotated, and genome completeness estimated by BUSCO was 97.90% (Fig. 2c; Supplementary Table 2).

Synteny analysis showed that the *H. vitessoides* genome shares a highly conserved gene order with other lepidopterans (Fig. 3). Chromosome 1 of *H. vitessoides* was recognised as the Z chromosome according to its high syntenic conservation to the Z chromosomes of *S. litura* and *B. mori* [55]. There were three chromosome fusion/fission events observed between *H. vitessoides* and *B. mori*, and two others between *C. suppressalis* and *H. vitessoides*.

In the family Crambidae, it was reported that the assembled chromosome number varies between 31 in *Cnaphalocrocis medinalis* and 32 in *Cnaphalocrocis exigua* (including Z and W sex chromosomes) [95,98]. While the pattern of fusion events of Chr11 and Chr24 of *B. mori* from *H. vitessoides* was similar to that from *C. medinalis*, Chr23 of *B. mori* was reported to be the fusion of Chr12 and a portion of the W chromosome in *C. medinalis* [98]. Our synteny dotplots showed that Chr29 of *H. vitessoides* was not assigned to any assembled chromosome in *C. medinalis*, but to unscaffolded contigs in *C. medinalis* and Chr32 in *C. exigua* (Supplementary Fig. 1). Our analyses suggest that the chromosome number of *H. vitessoides* is 32 (Z and W sex chromosomes and 30 autosomes), and the genome assembly we present here consists of both the Z chromosome and a full set of autosomes.

3.2. Orthology and phylogenetic analysis of H. vitessoides and other lepidopterans

By sampling protein-coding gene sequences from *H. vitessoides*, 25 other lepidopteran species and 7 non-lepidopteran hexapods, a total of 22,679 orthologous groups including 125 single copy orthologous genes were identified by OrthoMCL (Fig. 4). These single copy genes contain 10,683 sites and were used to construct the phylogenetic tree (Fig. 4). The tree topology is consistent with recent phylogenomic analyses [52,68].

3.3. Expansion of gene families in H. vitessoides

Among the 31,831 hexapod gene families inferred from 33 species, 173 genes were found to be expanded among the crambids. In *H. vitessoides*, 635 gene families were identified as rapid expanding (Fig. 4). These include gene families related to detoxification, metabolism, and development including UDP-glucuronosyltransferases (UGTs, Pfam: PF00201), abhydrolipases (Pfam: PF04083), short-chain dehydrogenases/reductases (SDRs, Pfam: PF00106), cytochromes P450 (CYPs, Pfam: PF00067), C-type lectins (Pfam: PF00059), and lepidopteran-specific chorion gene family (Pfam: PF01723), and

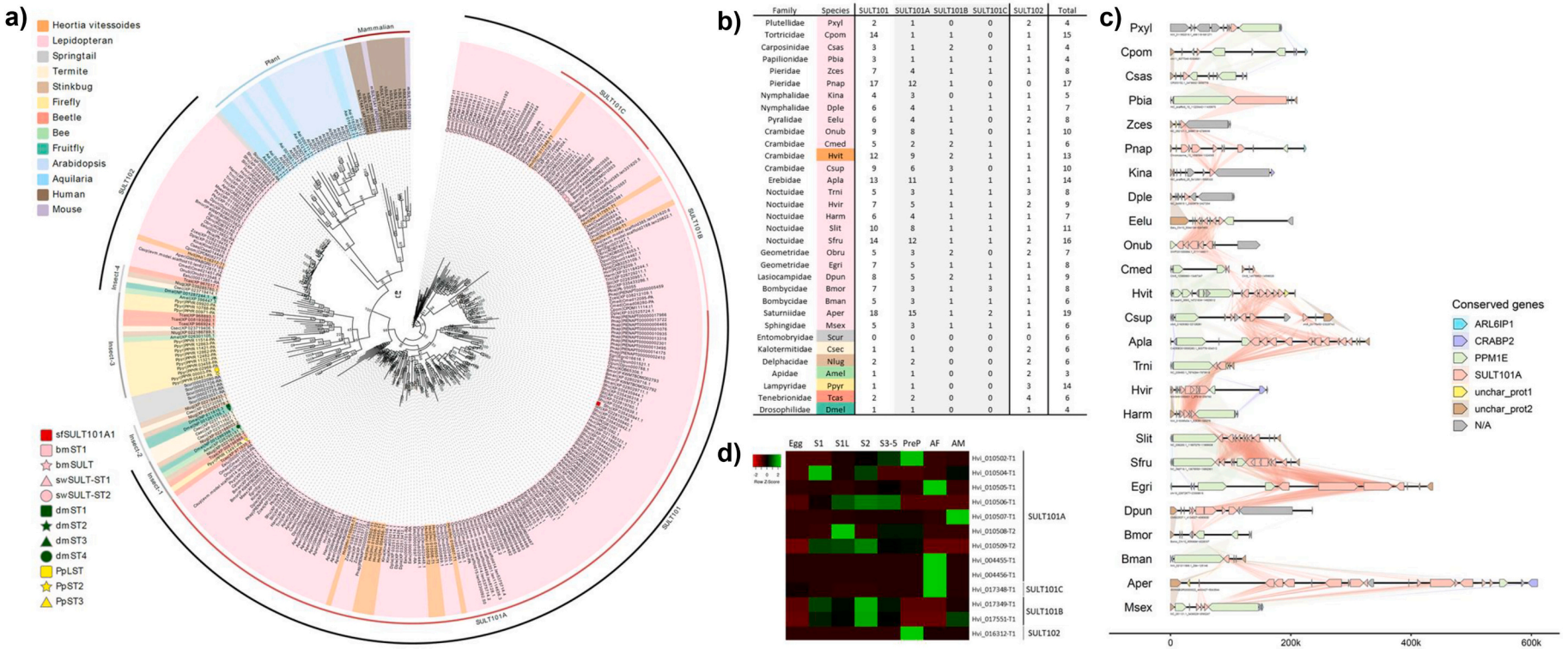


Fig. 5. a) Gene family tree of cytosolic sulfotransferases (SULTs). Taxonomic groups are indicated by corresponding leaf background colors; previously characterized insect SULTs are indicated with leaf decorations, including *sfSULT101A1* [83], *bmST1* [27], *bmSULT* [96], *swSULT-ST1–2* [2], *dmST1–4* [28], and *PpLST*, *PpST2–3* [19]; b) Table summarizing the gene numbers of SULT families across 33 hexapod taxa; c) Microsynteny of subfamily SULT101A in 25 lepidopteran species, indicated by red linkage groups; d) Heatmap representing the row Z score of transcript per million (TPM) normalized counts for the SULT expression levels in different developmental stages of *H. vitessoides*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

cytosolic sulfotransferases (SULTs) (Table 1; Supplementary Table 3).

A total of 67 CYPs could be identified in the *H. vitessoides* genome assembly, which contains an additional 21 CYPs to those identified in a previous transcriptomic study [13] (Supplementary Table 3). Moreover, copy number expansion was identified in two CYP families (CYP6AB and CYP9G) that function in xenobiotic metabolism and plant allelochemical detoxification [9] (Supplementary Fig. 2). It is worth noting that the expansion of CYP6AB was also reported in another closely related crambid, *C. suppressalis* [85].

Apart from detoxification genes, gene expansion was also detected in C-type lectins, which contain carbohydrate-recognition domains and mediate innate immunity [93]. In addition, a lepidopteran-specific chorion gene family that encodes essential proteins for eggshell formation was also expanded in Crambidae, with *O. nubilalis*, *C. medinalis*, *H. vitessoides*, and *C. suppressalis* containing 7, 9, 26, and 42 chorion genes respectively. How this variation contributes to the adaptation and biology of cambids remain to be functionally tested.

3.4. Cytosolic sulfotransferases (SULT) in H. vitessoides and other lepidopterans

SULT genes encode enzymes that catalyse sulfate conjugation in animals ranging from 3′-phosphoadenosine-5′-phosphosulfate (PAPS) to substrates containing hydroxyl or amino groups, such as phenols, enols, alcohols, and amines [6]. In insects, SULT has been reported to convert retinol to retro-retinoid anhydroretinol in the armyworm *Spodoptera frugiperda*, and is also involved in the formation of sulfoluciferin in the firefly *Pyrophorus luminosus* [19,83,92]. While the classification SULTs in mammals and plants has been robustly resolved [6,29], the evolution of SULT family genes in lepidopterans remains poorly studied. In our analyses, SULT family gene numbers have unexpectedly found to expand in lepidopterans In the gene tree, two groups of lepidopteran SULTs can be identified, SULT101 and SULT102 (Fig. 5a).

The SULT101 family contains *dmST4* of the fruit fly *Drosophila melanogaster* and *PpST3* of the firefly *Photinus pyralis*, which can further be subdivided into subfamilies SULT101A, SULT101B and SULT101C [2,6,83,96] (Fig. 5b; Supplementary Fig. 3). Microsynteny analysis revealed that SULT101As are located in a conserved genomic region across lepidopterans containing the cellular retinoic acid-binding protein 2 (*CRABP2*)(Fig. 5c).

The SULT102 family contains *dmST2* of *D. melanogaster* [28](Supplementary Fig. 3). SULT102 sequences share a unique serine catalytic residue at the PAPS-binding site in region I (aligned position 153) (Supplementary Fig. 4). For the RX_7S 3′-phosphate binding site in region II, lepidopteran SULT101s contain a $RDPRDX_3S$ motif, while SULT102s contain a $RNPKDX_3S$ motif similar to plants and mammals (Supplementary Fig. 3 and 4).

Previous studies identified that the lid domain of retinol dehydratase (*sfSULT101A1*) in the fall armyworm *Spodoptera frugiperda* is responsible for enclosing the β-ionone ring of substrate retinol to promote its dehydratase activity, and loss of the lid domain restricts the enzyme to function only as a sulfotransferase that sulfonates phenolic compounds [62,63]. The three tyrosine residues that interact with the ring of the retinol in the lid domain were nevertheless not conserved in SULT101A (Supplementary Fig. 5), suggesting that the cation–π interactions could differ among lepidopterans.

In *H. vitessoides*, SULT genes show diverse expression patterns (Fig. 5d), with SULT101A gene family members Hvi_010504, Hvi_010506, Hvi_010508 and Hvi_010509 primarily expressed in larval stages, Hvi_010502 in prepupae, Hvi_010505, Hvi_004455 and Hvi_004456 in adult female, and Hvi_010507 in adult male. Moreover, higher expression levels of SULT101B, SULT101C, and SULT102 were found in the 2nd instar larval, adult female and prepupal stages respectively. Considering that the *Aquilaria* host plant produces phenolics and volatile compounds in leaves after the first day of herbivory [31,69], SULT101A genes that are expressed in *H. vitessoides* larval stages could be involved in the metabolism of phenolic compounds. Further investigations into the functional roles of SULTs will be required to reveal their specific roles in this insect-plant interaction.

Table 2
Conserved miRNAs identified in *H. vitessoides*.

Family	Precursor
mir-1	mir-1
mir-2	mir-2-1, mir-2-2, mir-13
mir-8	mir-8
mir-9	mir-9-1, mir-9-2, mir-79–1, mir-79–2
mir-10	mir-10
mir-11	mir-11-1, mir-11–2
mir-14	mir-14
mir-25	mir-92–1, mir-92–2, mir-92–3
mir-29	mir-285
mir-31	mir-31
mir-33	mir-33
mir-34	mir-34
mir-67	mir-307
mir-71	mir-71
mir-87	mir-87
mir-124	mir-124
mir-132	mir-212
mir-133	mir-133
mir-137	mir-137
mir-184	mir-184
mir-193	mir-193
mir-210	mir-210, mir-3286
mir-252	mir-252
mir-263	mir-263-1, mir-263-2
mir-274	mir-274
mir-275	mir-275
mir-276	mir-276
mir-278	mir-278, mir-281, mir-279–1, mir-279–2, mir-279-3, mir-279-4
mir-282	mir-282
mir-305	mir-305
mir-308	mir-308
mir-316	mir-316
mir-317	mir-317
mir-750	mir-750
mir-927	mir-927-1, mir-927-2
mir-929	mir-929
mir-932	mir-932
mir-970	mir-970
mir-971	mir-971
mir-981	mir-981
mir-988	mir-988
mir-989	mir-989
mir-998	mir-998
mir-1000	mir-1000
mir-1175	mir-1175
mir-2733	mir-2733-1, mir-2733-2, mir-2733-3, mir-2733-4, mir-2733-5, mir-2733-6, mir-2733-7, mir-2733-8, mir-2733-9
mir-2756	mir-2756
mir-2766	mir-2766
mir-2767	mir-2767
mir-2788	mir-2788
mir-2796	mir-2796
mir-3338	mir-3338
bantam	bantam
let-7	let-7
n.a.	mir-6094
n.a.	mir-6307
n.a.	mir-8498-1, mir-8498-2
n.a.	mir-8506-1, mir-8506-2
n.a.	mir-8522

3.5. Sesquiterpenoid hormone biosynthesis genes and neuropeptides

Sesquiterpenoids, ecdysteroids, and neuropeptides are key hormonal regulators in insect development, physiology, and reproduction, and have been considered targets for pest management [7,47,54,82]. In the *H. vitessoides* genome assembly, the whole set of genes involved in the biosynthetic pathways of sesquiterpenoids and ecdysteroids are

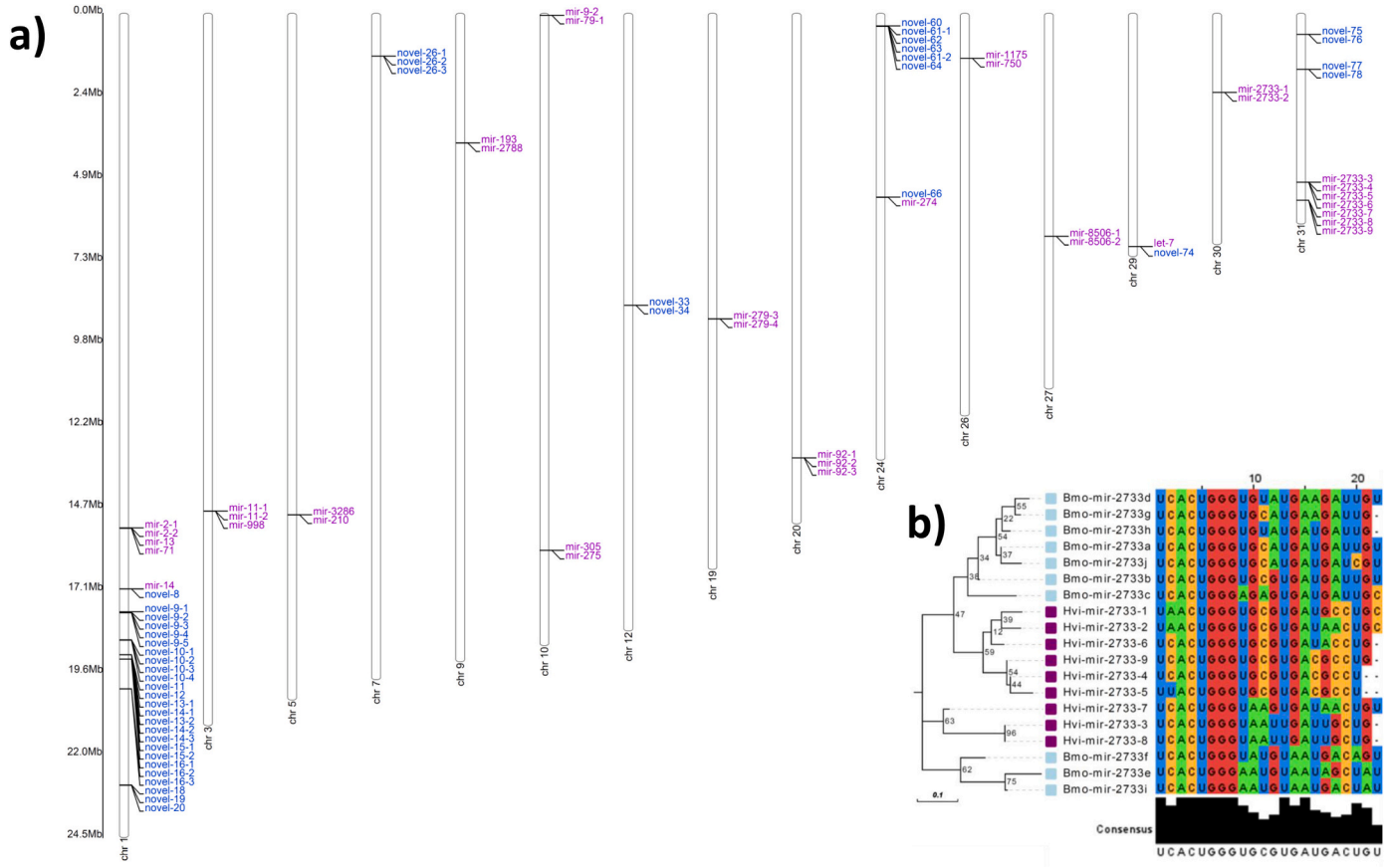


Fig. 6. a) miRNA clusters within 10 kb window visualized on chromosome map. miRNA colored in purple and blue indicates conserved and novel miRNA, respectively; b) Sequence alignment of the 3′ arm of mir-2733 of *H. vitessoides* and *B. mori*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

revealed (Supplementary Table 4). In addition, arthropod conserved neuropeptides could also be identified, with most displaying higher expression levels in the early stages ([45,74]; Supplementary Fig. 6,7; Supplementary Table 5). These identified genes could serve as a useful resource for future development in pest management strategies of *H. vitessoides*.

3.6. Conserved and novel microRNA clusters

MicroRNAs are important post-transcriptional regulators in insects [70,71], and a previous study sequencing small RNAs of lepidopterans revealed a burst of microRNA evolution in these animals [73]. Here, we sequenced small RNA transcriptomes from 1st and 2nd instar larvae, and male and female adults of *H. vitessoides*, and identified a total of 85 conserved microRNAs and 94 novel microRNAs (Table 2; Supplementary Table 6; Supplementary Fig. 8). Similar to previous study, our results identified a high birth rate of lineage-specific microRNAs.

In addition, we identified microRNA clusters as defined by inter-microRNA genomic distance within 10 kb (Fig. 6a), which include the miR-2/13/71 cluster [24] and the lepidopteran-specific miR-2733 cluster [32]. Sequence alignment showed that the 3′ arms of miR-2733 were not conserved between *H. vitessoides* and the silk moth *B. mori* (Fig. 6b). Furthermore, novel microRNA clusters were also found on Chr1, in which 4 novel miRNAs (novel-10-1–4) were located at the intronic region of an otoferlin gene, and another 9 novel miRNAs (novel-11/12/13–1–2/14–1–3/15–1–2) were found in the intergenic regions of two neighbour E3 ubiquitin-protein ligase MYCBP2 (*Mycbp2*) genes. How these different novel microRNAs and microRNA clusters contribute to the development and reproduction of *H. vitessoides* warrants further investigation.

Table 3
Summary of transposable elements in the genome of *H. vitessoides*.

Repeat Class	No. of elements	Total length (Mb)	Percentage sequence (%)	No. of distinct classifications
Retroelement	4,29,349	108.14	20.91%	504
SINE	65,490	14.45	2.79%	18
LINE	3,54,399	90.03	17.41%	428
Penelope	3908	0.74	0.14%	4
LTR element	5552	2.92	0.56%	54
DNA transposon	28,454	13.31	2.57%	248
Rolling-circle	74,313	18.46	3.57%	112
Unclassified	40,570	10.71	2.07%	145
Other	0	0.00	0.00%	0
Total repeats	5,72,686	150.61	29.12%	1009

3.7. Repeat content

We identify a repeat content of 29.12% in the *H. vitessoides* genome (Table 3, Fig. 7a), which is close to midrange compared to figures reported for other lepidopterans [97]. Retroelements dominate repeat content in *H. vitessoides*, accounting for 20.91% of all repeats, of which 17.41% is composed of LINEs, and 2.79% is composed of SINEs (Table 3, Fig. 7a). In contrast, DNA elements account for just 6.14% of repeat content, of which 3.57% is composed of rolling-circle elements, and 2.57% is composed of DNA transposons (*i.e.* terminal inverted repeat transposons) (Table 3, Fig. 7a). Only 2.07% of total repeat content

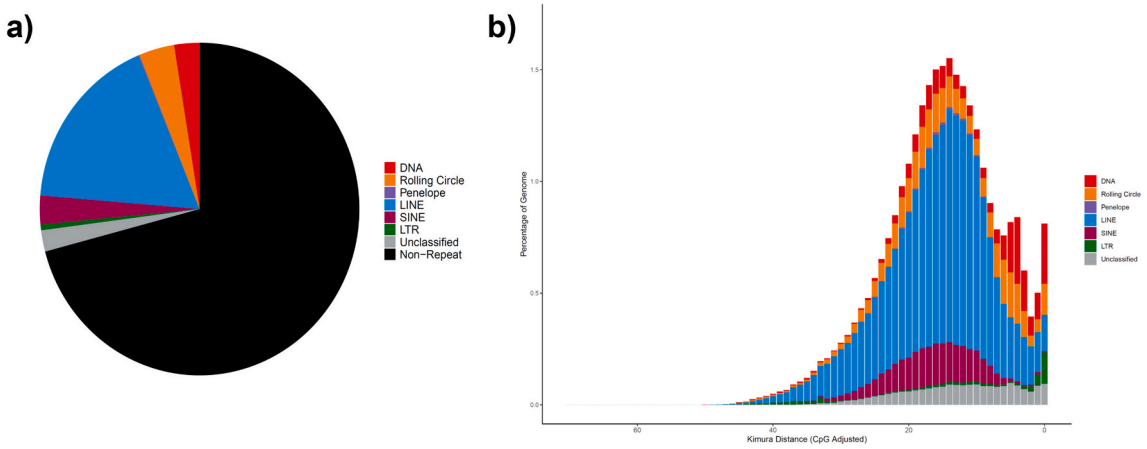


Fig. 7. a) Pie chart showing the proportions of repeat types present in the genome of *H. vitessoides*; b) Repeat landscape plot based on a Kimura distance-based copy divergence analysis of transposable elements.

remained unclassified. A dominance of LINEs in not unusual among moth genomes (*e.g.* winter moth, *Operophtera brumata* [16] and peach fruit moth, *Carposina sasakii,* [10]. Inspection of a repeat landscape plot (Fig. 7b) reveals a sharp decrease in the genome-wide proportion of LINEs at lower levels of genetic distance to family consensus sequences, suggesting that LINE activity has peaked and is now decreasing in *H. vitessoides*. Also, the pattern for SINEs suggests that activity has almost completely halted for this group. In contrast, opposing patterns suggest a substantial recent increase in DNA and LTR element activity. Consequently, it appears that there has been a recent switch in transposon activity, from a pattern dominated by LINEs and SINEs, to a pattern dominated by terminal inverted repeat DNA elements (Fig. 7b).

4. Conclusion

This study reports a chromosome-level genome assembly and transcriptomic (both mRNA and small RNAs) resources of the agarwood pest moth *Heortia vitessoides*. The gene family and microRNA analyses carried out in this study shed new light on the biology and evolution of *H. vitessoides* and other lepidopterans, and establish a platform for further research into understanding of insect-plant interactions between *H. vitessoides* and *Aquilaria* species and advances in pest management.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2022.110440.

Authors contributions

JHLH conceived and supervised the study. SLTS, WN, WLS, TB, TS, AH, JHLH performed the genome assembly, gene model prediction, gene annotation, and analyses. STSL, WN, TB, TS, WGB, AH, JHLH wrote the manuscript.

Availability of data and materials

The final chromosome assembly was submitted to NCBI Assembly under accession number JACJUM000000000 in NCBI. The raw reads generated in this study have been deposited to the NCBI database under the BioProject accessions: PRJNA654728, the genome annotation files were deposited in the Figshare (https://doi.org/10.6084/m9.figshare.19633668). The microRNA sequences of known species were obtained from both miRbase [42] and MirGeneDB [21].

CRediT authorship contribution statement

Sean T.S. Law: Conceptualization, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Wenyan Nong:** Methodology, Software, Investigation, Writing – original draft, Writing – review & editing. **Wai Lok So:** Conceptualization, Methodology, Visualization, Writing – original draft. **Tobias Baril:** Conceptualization, Methodology, Visualization, Writing – original draft. **Thomas Swale:** Methodology. **Chi Bun Chan:** Conceptualization, Investigation. **Stephen S. Tobe:** Conceptualization, Investigation. **Zhen-Peng Kai:** Conceptualization, Investigation. **William G. Bendena:** Conceptualization, Investigation. **Alexander Hayward:** Conceptualization, Methodology, Writing – original draft. **Jerome H.L. Hui:** Conceptualization, Resources, Supervision, Project administration, Writing – original draft, Writing – review & editing.

Data availability

Data will be made available on request.

Acknowledgements

This study was supported by the Hong Kong Research Grant Council Collaborative Research Fund (C4015-20EF), General Research Fund (14100919, 14100420), and The Chinese University of Hong Kong Direct Grant (4053489, 4053547). STSL was supported by PhD studentship provided by The Chinese University of Hong Kong.

References

[1] S. Babicki, D. Arndt, A. Marcu, Y. Liang, J.R. Grant, A. Maciejewski, D.S. Wishart, Heatmapper: web-enabled heat mapping for all, Nucleic Acids Res. 44 (W1) (2016) W147–W153.
[2] A.F. Bairam, Z.W. Kermasha, M.C. Liu, K. Kurogi, K. Yamamoto, Functional analysis of novel sulfotransferases in the silkworm *Bombyx mori*, Arch. Insect Biochem. Physiol. 104 (3) (2020), e21671.
[3] T. Baril, A. Hayward, Migrators within migrators: exploring transposable element dynamics in the monarch butterfly, *Danaus plexippus*, Mob. DNA 13 (1) (2022) 1–19.
[4] T. Baril, R. Imrie, A. Hayward, Earl Grey, 2021, https://doi.org/10.5281/zenodo.5654616.
[5] J.D. Bendtsen, H. Nielsen, G. Von Heijne, S. Brunak, Improved prediction of signal peptides: SignalP 3.0, J. Mol. Biol. 340 (4) (2004) 783–795.
[6] R.L. Blanchard, R.R. Freimuth, J. Buck, R.M. Weinshilboum, M.W. Coughtrie, A proposed nomenclature system for the cytosolic sulfotransferase (SULT) superfamily, Pharmacogenet. Genomics 14 (3) (2004) 199–211.
[7] D.A.T. Boncan, S.S. Tsang, C. Li, I.H. Lee, H.M. Lam, T.F. Chan, J.H. Hui, Terpenes and terpenoids in plants: interactions with environment and insects, Int. J. Mol. Sci. 21 (19) (2020) 7382.
[8] R. Bouckaert, J. Heled, D. Kühnert, T. Vaughan, C.H. Wu, D. Xie, M.A. Suchard, A. Rambaut, A.J. Drummond, BEAST 2: a software platform for Bayesian evolutionary analysis, PLoS Comput. Biol. 10 (4) (2014), e1003537.
[9] B. Calla, K. Noble, R.M. Johnson, K.K. Walden, M.A. Schuler, H.M. Robertson, M. R. Berenbaum, Cytochrome P450 diversification and hostplant utilization patterns in specialist and generalist moths: birth, death and adaptation, Mol. Ecol. 26 (21) (2017) 6021–6035.

[10] L.J. Cao, W. Song, L. Yue, S.K. Guo, J.C. Chen, Y.J. Gong, A.A. Hoffmann, S.J. Wei, Chromosome-level genome of the peach fruit moth *Carposina sasakii* (Lepidoptera: Carposinidae) provides a resource for evolutionary studies on moths, Mol. Ecol. Resour. 21 (3) (2021) 834–848.
[11] S. Capella-Gutiérrez, J.M. Silla-Martínez, T. Gabaldón, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses, Bioinformatics 25 (15) (2009) 1972–1973.
[12] C. Chen, H. Chen, Y. Zhang, H.R. Thomas, M.H. Frank, Y. He, R. Xia, TBtools: an integrative toolkit developed for interactive analyses of big biological data, Mol. Plant 13 (8) (2020) 1194–1202.
[13] J. Cheng, J. Chen, T. Lin, De novo assembly and analysis of the *Heortia vitessoides* transcriptome via high-throughput Illumina sequencing, J. Asia Pac. Entomol. 20 (4) (2017) 1241–1248.
[14] J. Cheng, C.Y. Wang, Z.H. Lyu, T. Lin, Multiple glutathione S-transferase genes in *Heortia vitessoides* (Lepidoptera: Crambidae): identification and expression patterns, J. Insect Sci. 18 (3) (2018) 23.
[15] T. Cheng, J. Wu, Y. Wu, R.V. Chilukuri, L. Huang, K. Yamamoto, L. Feng, W. Li, Z. Chen, H. Guo, J. Liu, S. Li, X. Wang, L. Peng, D. Liu, Y. Guo, B. Fu, Z. Li, C. Liu, Y. Chen, A. Tomar, F. Hilliou, N. Montagné, E. Jacquin-Joly, E. d'Alençon, R. K. Seth, R.K. Bhatnagar, A. Jouraku, T. Shiotsuki, K. Kadono-Okuda, A. Promboon, G. Smagghe, K.P. Arunkumar, H. Kishino, M.R. Goldsmith, Q. Feng, Q. Xia, K. Mita, Genomic adaptation to polyphagy and insecticides in a major East Asian noctuid pest, Nat. Ecol. Evol. 1 (11) (2017) 1747–1756.
[16] M.F. Derks, S. Smit, L. Salis, E. Schijlen, A. Bossers, C. Mateman, A.S. Pijl, D. de Ridder, M.A.M. Groenen, M.E. Visser, H.J. Megens, The genome of winter moth (*Operophtera brumata*) provides a genomic perspective on sexual dimorphism and phenology, Genom. Biol. Evol. 7 (8) (2015) 2321–2332.
[17] L.B. dos Santos, J.P.G. Viana, F.J.B. Francischini, S.V. Fogliata, A.L. Joyce, A.P. de Souza, M.G. Murúa, S.J. Clough, M.I. Zucchi, A first draft genome of the sugarcane borer, *Diatraea saccharalis*, F1000Research 9 (1269) (2020) 1269.
[18] S.R. Eddy, Accelerated profile HMM searches, PLoS Comput. Biol. 7 (10) (2011), e1002195.
[19] T.R. Fallon, F.S. Li, M.A. Vicent, J.K. Weng, Sulfoluciferin is biosynthesized by a specialized luciferin sulfotransferase in fireflies, Biochemistry 55 (24) (2016) 3341–3344.
[20] M.R. Friedländer, S.D. Mackowiak, N. Li, W. Chen, N. Rajewsky, miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades, Nucleic Acids Res. 40 (1) (2012) 37–52.
[21] B. Fromm, D. Domanska, E. Høye, V. Ovchinnikov, W. Kang, E. Aparicio-Puerta, M. Johansen, K. Flatmark, A. Mathelier, E. Hovig, M. Hackenberg, M. R. Friedländer, K.J. Peterson, MirGeneDB 2.0: the metazoan microRNA complement, Nucleic Acids Res. 48 (D1) (2020) D132–D141.
[22] B. Gschloessl, F. Dorkeld, P. Audiot, A. Bretaudeau, C. Kerdelhue, R. Streiff, De novo genome and transcriptome resources of the Adzuki bean borer *Ostrinia scapulalis* (Lepidoptera: Crambidae), Data Brief 17 (2018) 781–787.
[23] L. Gu, P.F. Reilly, J.J. Lewis, R.D. Reed, P. Andolfatto, J.R. Walters, Dichotomy of dosage compensation along the neo Z chromosome of the monarch butterfly, Curr. Biol. 29 (23) (2019) 4071–4077.
[24] Q. Guo, Y. Huang, F. Zou, B. Liu, M. Tian, W. Ye, J. Guo, X. Sun, D. Zhou, Y. Sun, L. Ma, B. Shen, C. Zhu, The role of miR-2~ 13~ 71 cluster in resistance to deltamethrin in *Culex pipiens pallens*, Insect Biochem. Mol. Biol. 84 (2017) 15–22.
[25] T. Hackl, S. Duponchel, K. Barenhoff, A. Weinmann, M.G. Fischer, Virophages and retrotransposons colonize the genomes of a heterotrophic flagellate, Elife 10 (2021), e72674.
[26] M.V. Han, G.W. Thomas, J. Lugo-Martinez, M.W. Hahn, Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3, Mol. Biol. Evol. 30 (8) (2013) 1987–1997.
[27] K. Hattori, M. Hirayama, H. Suzuki, H. Hamamoto, K. Sekimizu, H.O. Tamura, Cloning and expression of a novel sulfotransferase with unique substrate specificity from *Bombyx mori*, Biosci. Biotechnol. Biochem. 71 (4) (2007) 1044–1051.
[28] K. Hattori, N. Motohashi, I. Kobayashi, T. Tohya, M. Oikawa, H.O. Tamura, Cloning, expression, and characterization of cytosolic sulfotransferase isozymes from *Drosophila melanogaster*, Biosci. Biotechnol. Biochem. 72 (2) (2008) 540–547.
[29] F. Hirschmann, F. Krause, J. Papenbrock, The multi-protein family of sulfotransferases in plants: composition, occurrence, substrate specificity, and functions, Front. Plant Sci. 5 (2014) 556.
[30] R. Hubley, R.D. Finn, J. Clements, S.R. Eddy, T.A. Jones, W. Bao, A.F.A. Smit, T. J. Wheeler, The Dfam database of repetitive DNA families, Nucleic Acids Res. 44 (D1) (2016) D81–D89.
[31] T. Ito, M. Kakino, S. Tazawa, M. Oyama, H. Maruyama, Y. Araki, H. Hara, M. Iinuma, Identification of phenolic compounds in *Aquilaria crassna* leaves via liquid chromatography-electrospray ionization mass spectroscopy, Food Sci. Technol. Res. 18 (2) (2012) 259–262.
[32] G. Jagadeeswaran, Y. Zheng, N. Sumathipala, H. Jiang, E.L. Arrese, J.L. Soulages, W. Zhang, R. Sunkar, Deep sequencing of small RNA libraries reveals dynamic regulation of conserved and novel microRNAs and microRNA-stars during silkworm development, BMC Genomics 11 (1) (2010) 1–18.
[34] J. Jurka, V.V. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany, J. Walichiewicz, Repbase update, a database of eukaryotic repetitive elements, Cytogenet. Genom. Res. 110 (1–4) (2005) 462–467.
[35] J. Kalita, P.R. Bhattacharyya, S.C. Nath, *Heortia vitessoides* Moore (Lepidoptera: Pyralidae): a serious pest of agarwood plant (*Aquilaria malaccensis* Lamk.), Geobios 29 (1) (2001) 13–16.
[36] B. Kariyanna, A. Prabhuraj, R. Asokan, A. Agrawal, R. Gandhi Gracy, P. Jyoti, T. Venkatesan, M. Bheemanna, B. Kalmath, J.R. Diwan, Y. Pampanna, M. Mohan, Genome mining and expression analysis of carboxylesterase and glutathione S-transferase genes involved in insecticide resistance in eggplant shoot and fruit borer, *Leucinodes orbonalis* (Lepidoptera: Crambidae), Front. Physiol. (2020) 1479.
[37] K. Katoh, D.M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, Mol. Biol. Evol. 30 (4) (2013) 772–780.
[38] D. Kattupalli, K.M. Barbadikar, V. Balija, S. Ballichatla, A.P. Padmakumari, S. Saxena, K. Gaikwad, S. Yerram, P. Kokku, M.S. Madhav, The draft genome of yellow stem borer, an agriculturally important pest, provides molecular insights into its biology, development and specificity towards rice for infestation, Insects 12 (6) (2021) 563.
[39] A.Y. Kawahara, D. Plotkin, M. Espeland, K. Meusemann, E.F. Toussaint, A. Donath, F. Gimnich, P.B. Frandsen, A. Zwick, M. dos Reis, J.R. Barber, R.S. Peters, S. Liu, X. Zhou, C. Mayer, L. Podsiadlowski, C. Storer, J.E. Yack, B. Misof, J.W. Breinholt, Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths, Proc. Natl. Acad. Sci. 116 (45) (2019) 22657–22663.
[40] M. Kawamoto, A. Jouraku, A. Toyoda, K. Yokoi, Y. Minakuchi, S. Katsuma, A. Fujiyama, T. Kiuchi, K. Yamamoto, T. Shimada, High-quality genome assembly of the silkworm, *Bombyx mori*, Insect Biochem. Mol. Biol. 107 (2019) 53–62.
[41] D. Kim, J.M. Paggi, C. Park, C. Bennett, S.L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype, Nat. Biotechnol. 37 (8) (2019) 907–915.
[42] A. Kozomara, S. Griffiths-Jones, miRBase: annotating high confidence microRNAs using deep sequencing data, Nucleic Acids Res. 42 (D1) (2014) D68–D73.
[43] G.M. Kozak, C.B. Wadsworth, S.C. Kahne, S.M. Bogdanowicz, R.G. Harrison, B. S. Coates, E.B. Dopman, Genomic basis of circannual rhythm in the European corn borer moth, Curr. Biol. 29 (20) (2019) 3501–3509.
[44] S. Kumar, G. Stecher, K. Tamura, MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets, Mol. Biol. Evol. 33 (7) (2016) 1870–1874.
[45] B. Li, R. Predel, S. Neupert, F. Hauser, Y. Tanaka, G. Cazzamali, M. Williamson, Y. Arakane, P. Verleyen, L. Schoofs, J. Schachtner, C.J.P. Grimmelikhuijzen, Y. Park, Genomics, transcriptomics, and peptidomics of neuropeptides and protein hormones in the red flour beetle *Tribolium castaneum*, Genome Res. 18 (1) (2008) 113–122.
[46] L. Li, C.J. Stoeckert, D.S. Roos, OrthoMCL: identification of ortholog groups for eukaryotic genomes, Genome Res. 13 (9) (2003) 2178–2189.
[47] Z. Li, Z. Lyu, Q. Ye, J. Cheng, C. Wang, T. Lin, Cloning, expression analysis, 20-hydroxyecdysone induction, and RNA interference study of autophagy-related gene 8 from *Heortia vitessoides* Moore, Insects 11 (4) (2020) 245.
[48] Y. Li, W. Nong, T. Baril, H.Y. Yip, T. Swale, A. Hayward, D.E.K. Ferrier, J.H. Hui, Reconstruction of ancient homeobox gene linkages inferred from a new high-quality assembly of the Hong Kong oyster (*Magallana hongkongensis*) genome, BMC Genomics 21 (1) (2020) 1–17.
[50] S. Liang, J. Cai, X. Chen, Z. Jin, J. Zhang, Z. Huang, L. Tang, Z. Sun, X. Wen, C. Wang, Larval aggregation of *Heortia vitessoides* Moore (Lepidoptera: Crambidae) and evidence of horizontal transfer of avermectin, Forests 10 (4) (2019) 331.
[51] E. Lieberman-Aiden, N.L. Van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, J. Dekker, Comprehensive mapping of long-range interactions reveals folding principles of the human genome, science 326 (5950) (2009) 289–293.
[52] T. Léger, R. Mally, C. Neinhuis, M. Nuss, Refining the phylogeny of Crambidae with complete sampling of subfamilies (Lepidoptera, Pyraloidea), Zool. Scr. 50 (1) (2021) 84–99.
[53] Z. Lyu, Z. Li, J. Cheng, C. Wang, J. Chen, T. Lin, Suppression of gene juvenile hormone diol kinase delays pupation in *Heortia vitessoides* Moore, Insects 10 (9) (2019) 278.
[54] Z. Lyu, J. Chen, Z. Li, J. Cheng, C. Wang, T. Lin, Knockdown of β-N-acetylglucosaminidase gene disrupts molting process in *Heortia vitessoides* Moore, Arch. Insect Biochem. Physiol. 101 (4) (2019), e21561.
[55] W. Ma, X. Zhao, C. Yin, F. Jiang, X. Du, T. Chen, Q. Zhang, L. Qiu, H. Xu, J.J. Hull, G. Li, W.K. Sung, F. Li, Y. Lin, A chromosome-level genome assembly reveals the genetic basis of cold tolerance in a notorious rice insect pest, Chilo suppressalis, Mol. Ecol. Resour. 20 (1) (2020) 268–282.
[56] A.L. Mitchell, T.K. Attwood, P.C. Babbitt, M. Blum, P. Bork, A. Bridge, S.D. Brown, H.Y. Chang, S. El-Gebali, M.I. Fraser, J. Gough, D.R. Haft, H. Huang, I. Letunic, R. Lopez, A. Luciani, F. Madeira, A. Marchler-Bauer, H. Mi, D.A. Natale, M. Necci, G. Nuka, C. Orengo, A.P. Pandurangan, T. Paysan-Lafosse, S. Pesseat, S.C. Potter, M.A. Qureshi, N.D. Rawlings, N. Redaschi, L.J. Richardson, C. Rivoire, G. A. Salazar, A. Sangrador-Vegas, C.J.A. Sigrist, I. Sillitoe, G.G. Sutton, N. Thanki, P. D. Thomas, S.C.E. Tosatto, S.Y. Yong, R.D. Finn, InterPro in 2019: improving coverage, classification and access to protein sequence annotations, Nucleic Acids Res. 47 (D1) (2019) D351–D360.
[57] S. Nallu, J.A. Hill, K. Don, C. Sahagun, W. Zhang, C. Meslin, E. Snell-Rood, N. L. Clark, M.I. Morehouse, J. Bergelson, C.W. Wheat, M.R. Kronforst, The molecular genetic basis of herbivory between butterflies and their host plants, Nat. Ecol. Evol. 2 (9) (2018) 1418–1427.
[58] D.R. Nelson, The cytochrome p450 homepage, Human Genom. 4 (1) (2009) 1–7.
[59] W. Nong, S.T. Law, A.Y. Wong, T. Baril, T. Swale, L.M. Chu, A. Hayward, D.T. W. Lau, J.H. Hui, Chromosomal-level reference genome of the incense tree *Aquilaria sinensis*, Mol. Ecol. Resour. 20 (4) (2020) 971–979.
[60] W. Nong, Z. Qu, Y. Li, T. Barton-Owen, A.Y. Wong, H.Y. Yip, H.T. Lee, S. Narayana, T. Baril, T. Swale, J. Cao, T.F. Chan, H.S. Kwan, S.M. Ngai, G. Panagiotou, P. Y. Qian, J.W. Qiu, K.Y. Yip, K. Ismail, S. Pati, A. John, S.S. Tobe, W.G. Bendena, S. G. Cheung, A. Hayward, J.H. Hui, Horseshoe crab genomes reveal the evolution of genes and microRNAs after three rounds of whole genome duplication, Commun. Biol. 4 (1) (2021) 1–11.
[61] J.M. Palmer, J. Stajich, nextgenusfs/Funannotate: Funannotate v1.7.4 (1.7.4), Zenodo, 2020, https://doi.org/10.5281/zenodo.3679386.

[62] S. Pakhomova, J. Buck, M.E. Newcomer, The structures of the unique sulfotransferase retinol dehydratase with product and inhibitors provide insight into enzyme mechanism and inhibition, Protein Sci. 14 (1) (2005) 176–182.
[63] S. Pakhomova, M. Kobayashi, J. Buck, M.E. Newcomer, A helical lid converts a sulfotransferase to a dehydratase, Nat. Struct. Biol. 8 (5) (2001) 447–451.
[64] M. Pertea, G.M. Pertea, C.M. Antonescu, T.C. Chang, J.T. Mendell, S.L. Salzberg, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads, Nat. Biotechnol. 33 (3) (2015) 290–295.
[65] R.N. Platt, L. Blanco-Berdugo, D.A. Ray, Accurate transposable element annotation is vital when analyzing new genome assemblies, Genom. Biol. Evol. 8 (2) (2016) 403–410.
[66] M.N. Price, P.S. Dehal, A.P. Arkin, FastTree 2–approximately maximum-likelihood trees for large alignments, PLoS One 5 (3) (2010), e9490.
[67] N.H. Putnam, B.L. O'Connell, J.C. Stites, B.J. Rice, M. Blanchette, R. Calef, C. J. Troll, A. Fields, P.D. Hartley, C.W. Sugnet, D. Haussler, D.S. Rokhsar, R.E. Green, Chromosome-scale shotgun assembly using an in vitro method for long-range linkage, Genome Res. 26 (3) (2016) 342–350.
[68] M. Qi, H. Zhao, F. Yu, A. Zhang, H. Li, The first mitogenomes of the subfamily Odontiinae (Lepidoptera, Crambidae) and phylogenetic analysis of Pyraloidea, Insects 12 (6) (2021) 486.
[69] H. Qiao, P. Lu, S. Liu, C. Xu, K. Guo, R. Xu, J. Chen, Volatiles from *Aquilaria sinensis* damaged by *Heortia vitessoides* larvae deter the conspecific gravid adults and attract its predator *Cantheconidea concinna*, Sci. Rep. 8 (1) (2018) 1–15.
[70] Z. Qu, W.G. Bendena, W. Nong, K.W. Siggens, F.G. Noriega, Z.P. Kai, Y.Y. Zang, A. C. Koon, H.Y.E. Chan, T.F. Chan, K.H. Chu, H.M. Lam, M. Akam, S.S. Tobe, J. H. Lam Hui, MicroRNAs regulate the sesquiterpenoid hormonal pathway in *Drosophila* and other arthropods, Proc. R. Soc. B Biol. Sci. 284 (1869) (2017) 20171827.
[71] Z. Qu, W.G. Bendena, S.S. Tobe, J.H. Hui, Juvenile hormone and sesquiterpenoids in arthropods: biosynthesis, signaling, and role of MicroRNA, J. Steroid Biochem. Mol. Biol. 184 (2018) 69–76.
[72] Z. Qu, W. Nong, W.L. So, T. Barton-Owen, Y. Li, T.C. Leung, C. Li, T. Baril, A.Y. P. Wong, T. Swale, T.F. Chan, A. Hayward, S.M. Ngai, J.H. Hui, Millipede genomes reveal unique adaptations during myriapod evolution, PLoS Biol. 18 (9) (2020), e3000636.
[73] S. Quah, J.H. Hui, P.W. Holland, A burst of miRNA innovation in the early evolution of butterflies and moths, Mol. Biol. Evol. 32 (5) (2015) 1161–1174.
[74] L. Roller, N. Yamanaka, K. Watanabe, I. Daubnerová, D. Žitňan, H. Kataoka, Y. Tanaka, The unique evolution of neuropeptide genes in the silkworm *Bombyx mori*, Insect Biochem. Mol. Biol. 38 (12) (2008) 1147–1157.
[75] M. Saikia, K. Shrivastava, Biological control of *Heortia vitessoides* Moore, the most serious insect defoliator of *Aquilaria malaccensis* Lamk., a commercially important tree species of Northeast India, Ann. Biol. Res. 6 (2015) 26–32.
[76] M. Seppey, M. Manni, E.M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness, in: Gene Prediction, Humana, New York, NY, 2019, pp. 227–245.
[77] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J.D. Thompson, D.G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, Mol. Syst. Biol. 7 (1) (2011) 539.
[78] A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, Bioinformatics 30 (9) (2014) 1312–1313.
[79] B. Subramanian, S. Gao, M.J. Lercher, S. Hu, W.H. Chen, Evolview v3: a webserver for visualization, annotation, and management of phylogenetic trees, Nucleic Acids Res. 47 (W1) (2019) W270–W275.
[81] S.A. Syazwan, S.Y. Lee, S.P. Ong, R.O.Z.I. Mohamed, Damaging insect pests and diseases and their threats to agarwood tree plantations, Sains Malaysiana 48 (3) (2019) 497–507.
[82] S.S. Tsang, S.T. Law, C. Li, Z. Qu, W.G. Bendena, S.S. Tobe, J.H. Hui, Diversity of insect sesquiterpenoid regulation, Front. Genet. 11 (2020) 1027.
[83] E. Vakiani, J.G. Luz, J. Buck, Substrate specificity and kinetic mechanism of the insect sulfotransferase, retinol dehydratase, J. Biol. Chem. 273 (52) (1998) 35381–35387.
[84] G.W. Vurture, F.J. Sedlazeck, M. Nattestad, C.J. Underwood, H. Fang, J. Gurtowski, M.C. Schatz, GenomeScope: fast reference-free genome profiling from short reads, Bioinformatics 33 (14) (2017) 2202–2204.
[85] B. Wang, M.F. Shahzad, Z. Zhang, H. Sun, P. Han, F. Li, Z. Han, Genome-wide analysis reveals the expansion of cytochrome P450 genes associated with xenobiotic metabolism in rice striped stem borer, Chilo suppressalis, Biochem. Biophys. Res. Commun. 443 (2) (2014) 756–760.
[86] C.Y. Wang, J. Cheng, Z.H. Lyu, Z.X. Li, J.X. Chen, T. Lin, Chitin deacetylase 1 and 2 are indispensable for larval–pupal and pupal–adult molts in *Heortia vitessoides* (Lepidoptera: Crambidae), Comp. Biochem. Physiol. B: Biochem. Mol. Biol. 237 (2019), 110325.
[88] Y. Wang, H. Tang, J.D. DeBarry, X. Tan, J. Li, X. Wang, T.H. Lee, H. Jin, B. Marler, H. Guo, J.C. Kissinger, A.H. Paterson, MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity, Nucleic Acids Res. 40 (7) (2012) e49.
[89] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T.A.P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, SWISS-MODEL: homology modelling of protein structures and complexes, Nucleic Acids Res. 46 (W1) (2018) W296–W303.
[90] A.M. Waterhouse, J.B. Procter, D.M. Martin, M. Clamp, G.J. Barton, Jalview version 2—a multiple sequence alignment editor and analysis workbench, Bioinformatics 25 (9) (2009) 1189–1191.
[91] D.E. Wood, J. Lu, B. Langmead, Improved metagenomic analysis with Kraken 2, Genome Biol. 20 (1) (2019) 1–13.
[92] R.S. Yang, C.F. Wilkinson, Enzymic sulphation of p-nitrophenol and steroids by larval gut tissues of the southern armyworm (*Prodenia eridania* Cramer), Biochem. J. 130 (2) (1972) 487–493.
[93] X. Xia, M. You, X.J. Rao, X.Q. Yu, Insect C-type lectins in innate immunity, Dev. Comp. Immunol. 83 (2018) 70–79.
[94] D. Xu, X. Li, Y. Jin, Z. Zhuo, H. Yang, J. Hu, R. Wang, Influence of climatic factors on the potential distribution of pest *Heortia vitessoides* Moore in China, Glob. Ecol. Conserv. 23 (2020), e01107.
[95] H. Xu, X. Zhao, Y. Yang, X. Chen, Y. Mei, K. He, L. Xu, X. Ye, Y. Liu, Z. Lu, Chromosome-level genome assembly of an agricultural pest, the rice leaffolder *Cnaphalocrocis exigua* (Crambidae, Lepidoptera), Mol. Ecol. Resour. 22 (1) (2022) 307–318.
[96] K. Yamamoto, M.C. Liu, Identification and characterization of a new type of cytosolic sulfotransferase in the silkworm *Bombyx mori*, J. Insect Biotechnol. Sericol. 84 (3) (2015), 3_063-3_068.
[97] J. Zhang, Q. Cong, E.A. Rex, W. Hallwachs, D.H. Janzen, N.V. Grishin, D. B. Gammon, Gypsy moth genome provides insights into flight capability and virus–host interactions, Proc. Natl. Acad. Sci. 116 (5) (2019) 1669–1678.
[98] X. Zhao, H. Xu, K. He, Z. Shi, X. Chen, X. Ye, Y. Mei, Y. Yang, M. Li, L. Gao, L. Xu, H. Xiao, Y. Liu, Z. Lu, F. Li, A chromosome-level genome assembly of rice leaffolder, Cnaphalocrocis medinalis, Mol. Ecol. Resour. 21 (2) (2021) 561–572.