A Chromosome-Level Genome Assembly of *Ephestia elutella* (Hübner, 1796) (Lepidoptera: Pyralidae)

Bin Yan [1], Houding Ou[1], Lan Wei[1], Xiuqin Wang[1], Xiaofei Yu[2], Jianfeng Liu[1,*], and Maofa Yang[1,2,*]

[1]Institute of Entomology, Guizhou University, Guizhou Provincial Key Laboratory for Agricultural Pest Management of the Mountainous Region, Guiyang, China

[2]College of Tobacco Science, Guizhou University, Guiyang, China

*Corresponding authors: E-mails: jianfengliu25@126.com; gdgdly@126.com.

Accepted: 17 May 2021

Abstract

The moth *Ephestia elutella* (Hübner), is a storage pest that feeds on tobacco, cacao beans, cereals, dried fruits, and nuts. We generated a chromosome-level genome assembly containing 576.94 Mb using Nanopore long reads (approximately 130×) and Hi-C data (approximately 134×). The final assembly contained 804 scaffolds, with an N50 length of 19.00 Mb, and 94.96% (547.89 Mb) of the assembly was anchored into 31 pseudochromosomes. We masked 58.12% (335.32 Mb) of the genome as repetitive elements, identified 727 noncoding RNAs, and predicted 15,637 protein-coding genes. Gene family evolution and functional enrichment analyses revealed significantly expanded gene families primarily involved in digestion, detoxification, and chemosensation. Strong chromosomal syntenic relationships were also observed among *E. elutella*, silkworm, and tobacco cutworm. This study could provide a valuable genomic basis for better understanding the biology of *E. elutella*.

Key words: Phycitinae, genome annotation, comparative genomics, gene family evolution, synteny.

Significance

Genomic resources for the important storage pest *Ephestia elutella* are essentially nonexistent at present, making it difficult to understand the genomic architecture underlying this species' behavior, food preferences, and potential susceptibilities to various forms of management. In this study, we generated a high-quality, chromosome-level genome sequence for *E. elutella*, and performed an analysis of gene family evolution and synteny, respectively. Our work provides fundamental and valuable data for better understanding the tobacco moth and can contribute to improved moth management.

Introduction

Moths are one of the major super-radiations of Lepidoptera, comprising near 160,000 extant species in the world, which play key roles in many terrestrial systems. A total of 149 lepidopteran genomes were reported in NCBI database. *Ephestia elutella* (Hübner) (Lepidoptera: Pyralidae), the cacao moth, tobacco moth, or warehouse moth, is an important storage pest worldwide that preferring to feed on dried materials of plant origin, such as cereal products, cacao beans, dried fruits, and nuts (Athanassiou et al. 2018; Trematerra 2020). Current methods used to control *E. elutella* infestation in tobacco storage primarily include fumigation, using phosphine (PH_3), and contact insecticides (Ou et al. 2021). Until now, the whole genome of *E. elutella* has not been sequenced. High-quality moth genomes are important genetic resources for the study of pest biology, evolution, and pest control. Thus far, only four pyralid genomes have been published on NCBI (accessed April 25, 2021), including *Endotricha flammealis*, *Galleria mellonella*, *Plodia interpunctella*, and *Amyelois transitella*. Their genome sizes range from 382 Mb to 483 Mb. However, none of the chromosome-level assemblies for these species are available to the public, making this the first chromosome-level

assembly for the entire subfamily Phycitinae. Here, we provided a de novo chromosome-level genome assembly of *E. elutella* using ONT (Nanopore) long reads and Hi-C sequencing. We annotated the protein-coding genes, repetitive elements, and noncoding RNAs (ncRNAs) within the genome, analyzed gene family evolution across the main lepidopteran lineages, and investigated chromosomal syntenic relationships among three economically important moth species with better genomic resources available, *E. elutella*, *Bombyx mori*, and *Spodoptera litura*.

Results and Discussion

Genome Assembly

We generated 93.62 Gb (approximately 162×) Illumina and 75.20 Gb (approximately 130×) ONT reads for the genome assembly and 7.17 Gb transcriptome data. The N50 and the mean length of the long reads were 28.93 kb and 14.26 kb, respectively. After quality control, 75.15 Gb short reads were retained for the subsequent genome polishing step.

NextDenovo generated 63.38 Gb (approximately 110×) corrected ONT reads with an N50/mean length value reaching 31.75/26.82 kb. Using the corrected reads, NextDenovo produced a 2.28 Gb assembly, which was much larger than publicly available lepidopteran genomes; BUSCO assessment ($n = 1,367$) identified 992 (72.6%) complete and duplicated single-copy genes, indicating that the NextDenovo assembly was highly redundant due to the high heterozygosity of the sampled *Ephestia elutella* strain. After polishing, removing redundancy and contaminants, and Hi-C scaffolding, our final assembly (table 1) had a length of 576.94 Mb, comprising 804 scaffolds and 3,121 contigs, with a scaffold/contig N50 length of 19.00/0.43 Mb, a GC content of 37.89%, and BUSCO completeness of 92.9% (1.0% complete and duplicated, 2.2% fragmented, 4.9% missing); 94.96% (547.89 Mb) of the assembly was anchored into 31 pseudochromosomes (supplementary fig. S1, Supplementary Material online). Based on the comparable size of this genome with those from other lepidopterans, the low ratio of duplicates (1%), and the Hi-C contact heatmap (supplementary fig. S1, Supplementary Material online), no obvious heterogeneous regions were observed within the assembly. The genome size of *Ephestia elutella* was comparable to those of the four publicly Pyralidae species, only slightly smaller than one assembly version (GCA_002589825.1) of *G. mellonella*. The high mapping rates of the ONT (99.79%) and Illumina (95.42%) reads confirmed the integrity of our assembly.

Table 1
Genome Assembly and Annotation Statistics for *Ephestia elutella*

Elements	Current Version
Genome assembly	
Assembly size (Mb)	576.94
Number of scaffolds/contigs	804/3,121
Longest scaffold/contig (Mb)	23.24/2.51
N50 scaffold/contig length (Mb)	19.00/0.43
GC content (%)	37.89
Gaps (%)	0.04
BUSCO completeness (%)	92.9
Protein-coding genes	
Numbers	15,637
Mean gene length (bp)	9,619.59
Exons/introns per gene	7.44/6.25
Exon/intron ratio (%)	5.54/20.56
Mean exon/intron length	274.87/1,212.06
Repetitive elements	335.32 Mb (58.12%)
Number of ncRNAs	727

Genome Annotation

We masked 58.12% (335.32 Mb) of the genome as repetitive elements. The top five abundant repeat categories were LTR (19.25%), unclassified (11.59%), LINE (10.81%), rolling-circles (RCs, 9.83%), and DNA elements (4.24%) (fig 1*a*, supplementary table S1, Supplementary Material online). LTR and LINE retrotransposons and RCs were the primary contributors to the expansion of repetitive elements, particularly the LTR families Pao (8.71%), Gypsy (4.74%), and Copia (1.70%), the LINE families L2 (2.47%), RTE-BovB (1.65%), RTE-RTE (1.60%), and CR1-Zenon (1.15%), and the RC family helitron (9.83%). The significant expansion of repeat content may explain the larger genome size of *Ephestia elutella* compared with those of other pyralid species.

We annotated 727 ncRNAs: 60 ribosomal RNAs (rRNAs), 71 micro RNAs (miRNAs), 79 small nuclear RNAs (snRNAs), 4 long noncoding RNAs (lncRNAs), 318 tRNAs (20 isotypes, Supres and SelCys lacking), 3 ribozymes, and 192 other ncRNAs (supplementary table S2, Supplementary Material online). The snRNAs were classified as 59 spliceosomal RNAs (U1, U2, U4, U5, U6, and U11), seven minor spliceosomal RNAs (U4atac, U6atac, and U12), 11C/D box small nucleolar RNAs (snoRNAs), and two H/ACA box snoRNA.

MAKER pipeline predicted 15,637 protein-coding gene models (table 1); among them, 15,578 (94.38%), 12,698 (80.56%), and 14,822 (94.79%) genes matched the UniprotKB, InterProScan, and eggNOG records, respectively. After combining the InterProScan and eggNOG results, GO terms, KEGG pathways, Reactome pathways, Enzyme Codes, and COG categories were assigned to 11,005, 9,896, 10,873, 3,116, and 13,236 genes, respectively.

Gene Family Evolution

Thirteen species were used to analyze orthogroups using OrthoFinder v2.3.8. A total of 184,352 (94.70%) genes were clustered into 14,940 orthogroups (gene families). Among all defined orthogroups, 5,082 were multicopy groups with all species present, whereas 1,691 were single-

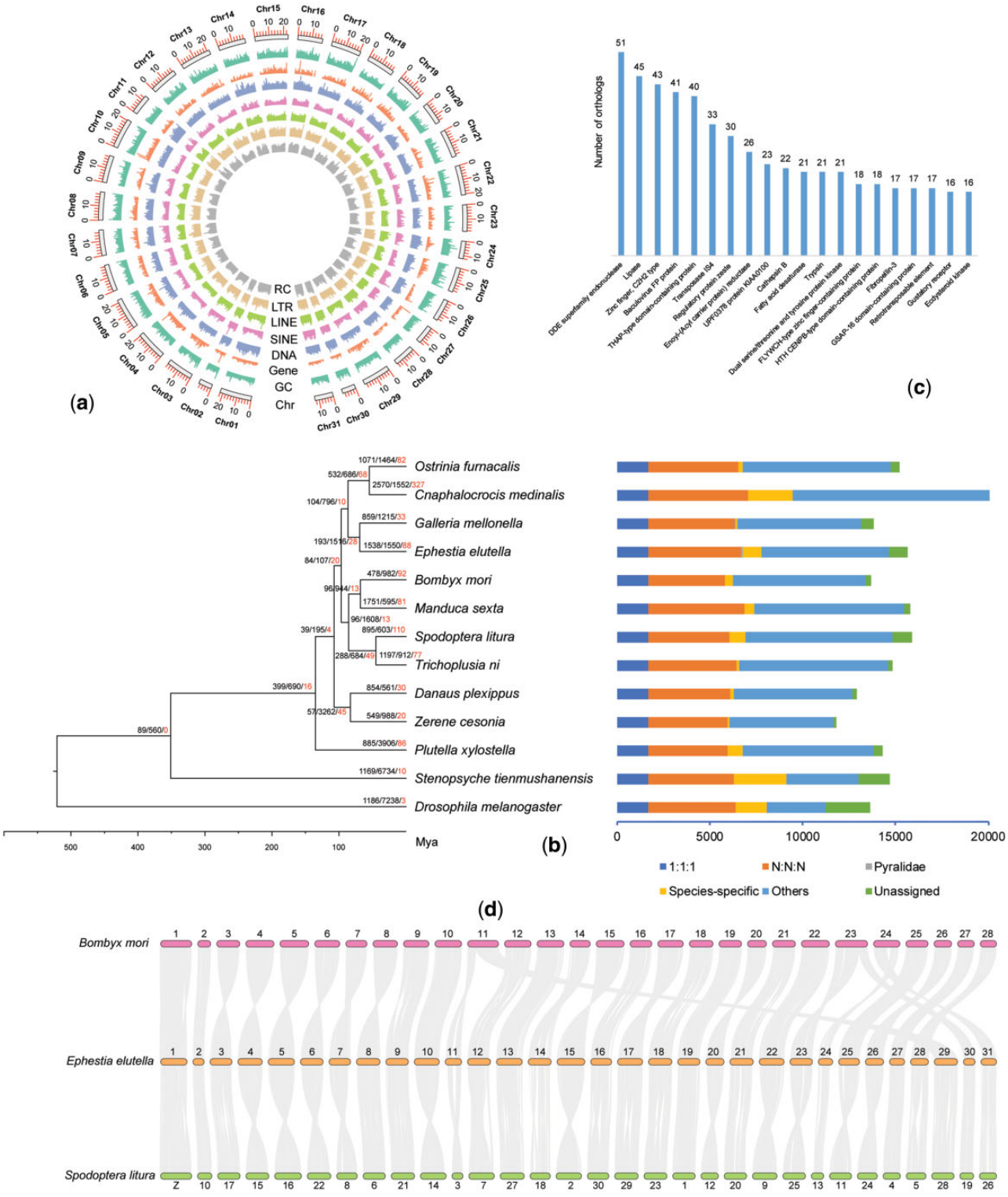

Fig. 1.—Genome characteristics, gene family evolution, and synteny. (a) Circos tracks showing element distributions in 100 kb sliding windows from outer to inner: chromosome length, GC content, density of protein-coding genes, DNA transposons, SINE/LINE/LTR retrotransposons, and rolling-circles. (b) Gene family evolution and statistics of orthologs. Node values representing the number of expanded, contracted, and rapidly evolving families, respectively; "1:1:1" represents shared single-copy genes, "N: N: N" represents multicopy genes shared by all species, "Pyralidae" represents orthologs unique to Pyralidae, "Others" represents unclassified orthologs, "Unassigned" represents orthologs that cannot be assigned to any orthogroups. (c) Top twenty significantly expanded families. (d) Synteny between *Ephestia elutella* and *Bombyx mori*/*Spodoptera litura*.

copy groups (fig. 1*b*). For the *Ephestia elutella*, 14,635 (93.58%) genes were assigned to 10,254 orthogroups, including 236 orthogroups that contained 1,003 species-specific genes.

After "symtest" loci filtering, IQ-TREE inferred the phylogenetic tree based on 1,497 single-copy genes and 651,642 amino acid sites (fig. 1*b*). All nodes were fully resolved, with all node supports as 100/100. Pyraloidea was identified as a sister clade for Bombycoidea + Noctuoidea. The Pyraloidea originated from a transition period during the late or early Cretaceous period (95.06–102.23 Mya). Crambidae and Pyralidae diverged at the beginning of the Late Cretaceous period (85.01–91.66 Mya). Two Pyralidae species diverged during the last portion of the Late Cretaceous period (67.61–73.47 Mya). The tree topology (classification) and divergence estimation were largely consistent with those reported in a recent phylogenomic study (Kawahara et al. 2019).

Gene family evolution analyses revealed that 1,538 and 1,550 gene families experienced expansions and contractions, respectively, including 88 gene families (79 expansions and 9 contractions) that were recognized as rapidly evolving gene families (fig. 1*b*). The significantly expanded families were primarily associated with digestion, chemosensation, and detoxification (fig. 1*c*, supplementary table S3, Supplementary Material online). The digestion-related families included lipase (45), trypsin (21), enoyl-(Acyl carrier protein) reductase (26), and fatty acid desaturase (21). These strongest expansions may reflect the possible mechanisms necessary to feed on dry storage foods. Families are associated with chemosensation-, such as gustatory and odorant receptors, and detoxification, such as ecdysteroid kinase, ABC transporter, and glutathione S-transferase (supplementary table S3, Supplementary Material online). Further GO (supplementary fig. S2, Supplementary Material online) and KEGG (supplementary fig. S3, Supplementary Material online) enrichment analyses for 79 significantly expanded gene families emphasized the high representation of digestion- and detoxification-related categories. Large expansions of digestion-, detoxification-, and chemosensation-related genes are crucial for feeding, foraging, and adapting to harsh environments.

Chromosomal Synteny

We recovered 232 syntenic blocks (16,448 genes) between *Ephestia elutella* and *B. mori* and 370 syntenic blocks (15,613 genes) between *Ephestia elutella* and *S. litura*. Strong syntenic relationships among all three species indicated conserved chromosome-level gene collinearity (fig. 1*d*). Chromosome 1 of both *Ephestia elutella* and *B. mori* corresponded to the Z chromosome of *S. litura*. Chromosomal pairs 11/29, 23/30, 24/31 in *Ephestia elutella* were related to chromosomes 11, 23, and 24 in silkworms. The numbers of syntenic blocks and chromosomal correspondence showed that chromosomes were more conserved between *Ephestia elutella* and *S. litura*.

Materials and Methods

Sample Collection and Sequencing

The species of *E. elutella* strain used for sequencing was originally collected in May 2016 in a warehouse (26°52′N, 106°73′E; 1120 m altitude) of the Guiyang Branch Tobacco Company, Guizhou, China and has been maintained by feeding on a previously developed artificial diet in the laboratory, without exposure to any insecticides (Ou et al. 2019). Female pupae were collected for sequencing: 35 for Illumina and ONT whole-genome sequencing, 10 for transcriptome sequencing, and 5 for Hi-C sequencing, respectively. Genomic DNA and RNA were extracted using the Qiagen Blood & Cell Culture DNA Mini Kit and TRIzol™ Reagent, and libraries with a 350 bp insert size were constructed using the TruSeq DNA PCR-Free LT Library Preparation Kit and TruSeq RNA v2 Kit. An ONT library with a 40 kb insert size was prepared using a Ligation Sequencing Kit (SQK-LSK109). The restriction enzyme MboI was used to digest DNA for the Hi-C assay. Short-read and long-read libraries were sequenced on the HiSeq NovaSeq 6000 and PromethION platforms, respectively. All library construction and sequencing were procedures at BENAGEN (Wuhan, China).

Genome Assembly

Raw ONT reads longer than 10 kb were selfcorrected and assembled using NextDenovo v2.3.1 (https://github.com/Nextomics/NextDenovo). Preliminary assembly was polished with one round of long reads and two rounds of short reads using NextPolish v1.3.0 (Hu et al. 2020). Prior to polishing, quality control for short reads was conducted using BBTools suite v38.82 (Bushnell 2014), including the following steps: the removal of duplicates using "clumpify.sh"; quality trimming (>Q20); length filtering (>15 bp); polymer trimming (>10 bp for poly-A/G/C tails); and the correction of overlapping paired reads using "bbduk.sh." Redundant heterozygous regions were removed based on read depth using three rounds of Purge_Dups v1.0.1 (Guan et al. 2020) with a minimum alignment score of 50 and a minimum chaining score of 3,000 for a match ("-a 50 -l 3000"). Minimap2 v2.17 (Li 2018) was used as a sequence mapper for short-read polishing and redundancy removal.

Hi-C data quality control, which included mapping, duplicate removal, and Hi-C contact extraction, was performed using Juicer v1.6.2 (Durand et al. 2016). Contigs were anchored to pseudochromosomes using two rounds of 3D-DNA v180922 (Dudchenko et al. 2017). Possible errors (misjoins, translocations, inversions, and chromosome boundaries) were manually corrected using the Assembly Tools module

within Juicebox (Durand et al. 2016), and the resulting assembly was further refined in a second 3D-DNA round.

Potential contaminants were detected using blastn (BLAST+ v2.9.1) (Camacho et al. 2009) against the NCBI nucleotide (nt) and UniVec databases. Scaffolds greater than 10 kb were retained in the final assembly. We assessed the assembly quality in terms of genome completeness and raw read mapping rate. Genome completeness was assessed using BUSCO v3.0.2 pipeline (Waterhouse et al. 2018) against the insecta_odb10 gene set ($n = 1{,}367$). The mapping rate was estimated by aligning raw long and short reads with the genome assembly using Minimap2.

Genome Annotation

We annotated three essential genomic elements: protein-coding genes, repetitive elements, and ncRNAs. A de novo repeat library was constructed using RepeatModeler v2.0.1 (Flynn et al. 2020) with the LTR discovery pipeline included ("-LTRStruct") and combined with Dfam 3.1 (Hubley et al. 2016) and RepBase-20181026 databases (Bao et al. 2015) to generate a custom library. Repeats were masked using RepeatMasker v4.1.0 (Smit et al. 2013–2015) on the custom library. All ncRNAs were annotated using Infernal v1.1.3 (Nawrocki and Eddy 2013) and tRNAscan-SE v2.0.7 (Chan and Lowe 2019). Only high-confidence tRNAs were retained, using the tRNAscan-SE script "EukHighConfidenceFilter."

We employed the MAKER v3.01.03 pipeline (Holt and Yandell 2011) to predict protein-coding gene models by integrating ab initio, transcriptome, and protein homology-based evidence. We generated ab initio predictions using the BRAKER v2.1.5 pipeline (Hoff et al. 2016), which automatically trained the predictors Augustus v3.3.4 (Stanke et al. 2004) and GeneMark-ES/ET/EP 4.59_lic (Brůna et al. 2020) and simultaneously incorporated evidence from transcriptome and protein homology information; transcriptome evidence in BAM alignments was produced using HISAT2 v2.2.0 (Kim et al. 2019), and the protein source was mined from the OrthoDB10 v1 database (Kriventseva et al. 2019). Transcripts passed to MAKER were assembled using the genome-guided assembler StringTie v2.1.4 (Kovaka et al. 2019). Protein sequences for *Apis mellifera*, *B. mori*, *Danaus plexippus*, *Drosophila melanogaster*, *S. litura*, and *Tribolium castaneum* were downloaded from NCBI and passed to MAKER as evidence of protein homology. Weights 8, 2, and 1 were respectively assigned to transcript, protein, and ab initio evidence for the EVidenceModeler (EVM) module built into MAKER3. Gene functions were annotated by searching the UniProtKB database using Diamond v0.9.24 (Buchfink et al. 2015) using the sensitive mode "–more-sensitive -e 1e−5." Protein domains, Gene Ontology (GO), and pathways [Kyoto Encyclopedia of Genes and Genomes (KEGG), Reactome] were assigned using eggNOG-mapper v2.0.1 (Huerta-Cepas et al. 2017) against the eggNOG v5.0 database (Huerta-Cepas et al. 2019) and InterProScan 5.47–82.0 (Finn et al. 2017) against Pfam (El-Gebali et al. 2019), Gene3D (Lewis et al. 2018), Superfamily (Wilson et al. 2009), SMART (Letunic and Bork 2018), and CDD (Marchler-Bauer et al. 2017) databases.

Gene Family Evolution

We inferred orthology using OrthoFinder v2.3.8 (Emms and Kelly 2019) with Diamond as the sequence aligner. In addition to *E. elutella*, high-quality protein annotation sequences of one Diptera (*Dr. melanogaster*), one Trichoptera (*Stenopsyche tienmushanensis*), ten Lepidoptera (*B. mori*, *Cnaphalocrocis medinalis*, *D. plexippus*, *G. mellonella*, *Manduca sexta*, *Ostrinia furnacalis*, *Plutella xylostella*, *S. litura*, *Trichoplusia ni*, and *Zerene cesonia*) were downloaded from the NCBI for analyses except for *St. tienmushanensis* (doi: 10.5524/100538).

Protein sequences of single-copy orthologs were aligned using MAFFT v7.450 (Katoh and Standley 2013) with the L-INS-I mode. Unreliable homologous sites within alignments were trimmed using BMGE v1.12 (Criscuolo and Gribaldo 2010), based on the stringent parameters of "-m BLOSUM90 -h 0.4." The phylogenetic tree was reconstructed using IQ-TREE v2.0-rc1 (Minh et al. 2020), using the parameters "-m MFP –mset LG –msub nuclear –rclusterf 10 -B 1000 –alrt 1000 –symtest-remove-bad –symtest-pval 0.10." Divergence times were estimated using MCMCTree within the PAML v4.9j package (Yang 2007). Six fossils from the PBDB database (https://www.paleobiodb.org/navigator/) were used for node calibration: root (Trichoptera <358.9 Mya), Lepidoptera (201.3–252.2 Mya), Noctuoidea (>28.1 Mya), Bombycoidea (>33.9 Mya), Pyraloidea (>54 Mya), and Papilionoidea (>54 Mya).

Expansions and contractions of gene families were estimated using CAFÉ v4.2.1 (Han et al. 2013); the model of single birth–death parameter lambda was used with the significance level of 0.01. For significantly expanded families, we further performed GO and KEGG functional enrichment analyses using R package clusterProfiler v3.14.3 (Yu et al. 2012) with the default parameters (P value = 0.01 and q-value = 0.05).

Synteny

Intergenomic chromosomal synteny between *E. elutella* (Pyralidae) and Bombycoidea/Noctuoidea species (*B. mori* and *S. litura*) were inferred using TBtools v1.0692 (Chen et al. 2020). The input genome and annotation files were downloaded from NCBI. Initial blastp search parameters were 1e−10 for the e-value and 5 for the number of blast hits. At least five genes were required to define a collinear block. Synteny plot was also visualized by TBtools.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

This work was financially supported by the Science and Technology Project of Guiyang Branch Company of Guizhou Tobacco Company (Grant No. 2019-03), the Program of Excellent Innovation Talents, Guizhou Province (Grant No. [2016]-4022) and the Natural Science Special Project of Guizhou University (Special post,[2020]-02).

Data Availability

Genome assembly and raw sequencing data have been deposited at NCBI under the accessions numbers JAEMBS000000000 and SRR13340377–SRR13340380, respectively. Genome annotations are available at Figshare and can be accessed at https://doi.org/10.6084/m9.figshare.14216660.

Literature cited

Athanassiou C, Bray DP, Hall DR, Phillips C, Vassilakos TN. 2018. Factors affecting field performance of pheromone traps for tobacco beetle, *Lasioderma serricorne*, and tobacco moth, *Ephestia elutella*. J Pest Sci. 91(4):1381–1391.

Bao W, Kojima KK, Kohany O. 2015. Repbase update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 6:11–11–6.

Brůna T, Lomsadze A, Borodovsky M. 2020. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. NAR Genom Bioinform. 2(2):lqaa026.

Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 12(1):59–60.

Bushnell B. 2014. BBtools. Available from: https://sourceforge.net/projects/bbmap/.

Camacho C, et al. 2009. BLAST+: architecture and applications. BMC Bioinformatics. 10(1):421.

Chan PP, Lowe TM. 2019. tRNAscan-SE: searching for tRNA genes in genomic sequences. Methods Mol Biol. 1962:1–14.

Chen C, et al. 2020. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant. 13(8):1194–1202.

Criscuolo A, Gribaldo S. 2010. BMGE (Block Mapping and Gathering with Entropy): selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol Biol. 10(1):210.

Dudchenko O, et al. 2017. *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science 356(6333):92–95.

Durand NC, et al. 2016. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 3(1):95–98.

El-Gebali S, et al. 2019. The Pfam protein families database in 2019. Nucleic Acids Res. 47:D427–D432.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20(1):238.

Finn RD, et al. 2017. InterPro in 2017-beyond protein family and domain annotations. Nucleic Acids Res. 45(D1):D190–D199.

Flynn J, et al. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 117(17):9451–9457.

Guan D, et al. 2020. Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics 36(9):2896–2898.

Han MV, Thomas G, Lugo-Martinez J, Hah MW. 2013. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. Mol Biol Evol. 30(8):1987–1997.

Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics. 32(5):767–769.

Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics. 12(1):491.

Hu J, et al. 2020. NextPolish: a fast and efficient genome polishing tool for long read assembly. Bioinformatics 36(7):2253–2255.

Hubley R, et al. 2016. The Dfam database of repetitive DNA families. Nucleic Acids Res. 44(D1):D81–D89.

Huerta-Cepas J, et al. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol Biol Evol. 34(8):2115–2122.

Huerta-Cepas J, et al. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 47(D1):D309–D314.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.

Kawahara AY, et al. 2019. Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. Proc Natl Acad Sci U S A. 116(45):22657–22663.

Kim D, et al. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 37(8):907–915.

Kovaka S, et al. 2019. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. 20(1):278.

Kriventseva EV, et al. 2019. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic Acids Res. 47(D1):D807–D811.

Letunic L, Bork P. 2018. 20 years of the SMART protein domain annotation resource. Nucleic Acids Res. 46(D1):D493–D496.

Lewis T, et al. 2018. Gene3D: extensive prediction of globular domains in proteins. Nucleic Acids Res. 46(D1):D435–D439.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34(18):3094–3100.

Marchler-Bauer A, et al. 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. 45(D1):D200–D203.

Minh BQ, et al. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 37(5):1530–1534.

Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 29(22):2933–2935.

Ou HD, et al. 2019. Control efficiency of *Bracon hebetor* Say against *Ephestia elutella* (Hübner). Chin Tob Sci. 40(5):44–51.

Ou HD, et al. 2021. Host deprivation effects on population performance and paralysis rates of *Habrobracon hebetor* (Hemenoptera: braconidae). Pest Manag Sci. 77(4):1851–1863.

Smit AFA, Hubley R, Green P. 2013–2015. Repeat Masker Open-4.0. Available from: http://www.repeatmasker.org.

Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res. 32(Web Server):W309–W312.

Trematerra P. 2020. Combined control of *Lasioderma serricorne* (F.) and *Ephestia elutella* (Hbn.) in a tobacco processing facility by attracticide method. J Appl Entomol. 144(7):598–604.

Waterhouse RM, et al. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Bio Evol. 35(3):543–548.
Wilson D, et al. 2009. SUPERFAMILY—sophisticated comparative genomics, data mining, visualization and phylogeny. Nucleic Acids Res. 37(Database issue):D380–D386.
Yang ZH. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24(8):1586–1591.
Yu G, Wang LG, Han Y, He QY. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 16(5):284–287.

Associate editor: Dennis Lavrov