A chromosome-level genome assembly for the beet armyworm (*Spodoptera exigua*) using PacBio and Hi-C sequencing

Feng Zhang, Jianpeng Zhang, Yihua Yang, Yidong Wu*

Department of Entomology, College of Plant Protection, Nanjing Agricultural University, Nanjing 210095, China
*Correspoding author

Abstract

The beet armyworm, *Spodoptera exigua* (Hübner), is a worldwide, polyphagous agricultural pest feeding on vegetable, field, and flower crops. However, the lacking of genome architecture severely limits our understanding of its rapid adaptation and the development of efficient pest managements. We report a chromosome-level genome assembly using single-molecule real-time PacBio sequencing and Hi-C data. The final genome assembly was 446.80 Mb with a scaffold N50 of 14.36 Mb, and captured 97.9% complete arthropod Benchmarking Universal Single-Copy Orthologs (n=1,658). A total of 367 contigs were anchored to 32 pseudo-chromosomes, covering 96.18% (429.74 Mb) of the total genome length. We predicted 17,707 protein-coding genes, of which 81.69% were supported by transcriptome evidence and 97.94% matched the UniProt protein records. We also identified 867,102 (147.97 Mb/33.12%) repeats and 1,609 noncoding RNAs. Synteny analyses indicated a strong collinearity between three lepidopteran species. Our high-quality genome information provides a valuable resource for better understanding and management of the beet armyworm.

Key words Noctuidae, agricultural pest, genome annotation, phylogenomics

Introduction

Spodoptera exigua (Lepidoptera, Noctuidae), known as beet armyworm or BAW, is one of the most polyphagous pests feeding on over 90 agricultural and horticultural plants, and heavily threats agricultural economy and food security (Mehrkhou et al. 2012; Saeed et al. 2019; Tay & Gordon 2019). The beet armyworm has been distributed throughout the tropical and temperate regions of Asia, Europe, Africa and North America although it was native to Southeast Asia (Han et al. 2009; Zhou et al. 2016; Fu et al. 2017). Severe outbreaks of this pest have been recorded in some regions duo to complex factors, i.e. weather (rainfall, temperature), high-level resistance and increase of host planting (Che et al. 2013; Zhou et al. 2017). Hence, *S. exigua* is thought as an ideal model to study the invasion events and population dynamics for agricultural pests (Han et al. 2009; Tay & Gordon 2019).

Currently, chemical insecticides are the most widely adopted approach to suppress beet armyworm populations. Whereas, surge in resistance to traditional insecticides (e.g. organophosphates, carbamates and pyrethroids), as well as newer insecticides with novel modes of action (e.g. emamectin benzoate, spinosad, abamectin), severely obstructs the further

applications of insecticides (Che et al. 2013; Ishtiaq & Saleem 2011). The molecular mechanisms of resistance are still gapped in beet armyworm. Owing to the advancement of modern molecular techniques, in particular CRISPR/Cas9 system (Cong et al. 2013; Bassett et al. 2013), we can inspect molecular functions of conserved target loci generating resistances in relatives in *S. exigua* (Zuo et al. 2017, 2018). However, it is difficult to independently explore new molecular mechanisms based on their resistant phenotype without the knowledge of the genome context, structure and functions. Serving as an important pest, limited genomic resources has been accumulated for beet armyworm.

Here, we present a de novo chromosome-level genome assembly of *S. exigua* from a Chinese susceptible strain using single-molecule real-time (SMRT) Pacific Bioscience (PacBio) long reads and Hi-C data. Mitochondrial genome and transcripts were also assembled. We annotated the essential genomic elements, i.e. repetitive sequences, protein-coding genes, and non-coding RNAs (ncRNAs). These valuable genomic data help to reveal the downstream molecular mechanism and further facilitate the future pest management.

Materials and Methods

Sample collection and sequencing

The strain WH-S, originally obtained from Wuhan Institute of Vegetables (Hubei, China) in 2009, has been maintained in laboratory without exposure to any insecticides since its initial collection in 1998 from Wuhan. Five female pupae were collected for sequencing experiments: one for Illumina and PacBio whole genome, one for Hi-C, and three for transcriptome, respectively. Genomic DNA/RNA extraction, library preparation and sequencing were carried out at BGI (Shenzhen, China) except for Hi-C data produced at Frasergen Co. Ltd. (Wuhan, China). Libraries were constructed with an insert size of 15 kb for the PacBio Sequel platform and 270 bp on the Illumina HiSeq X Ten platform, respectively. Restriction Enzyme MboI was used to digest DNA for the Hi-C assay.

Genome assembly

Quality control was performed for the raw Illumina short reads using BBTools v38.49 (Bushnell): duplicates removal with clumpy.sh and low-quality base trimming with bbduk.sh. Preliminary genome survey was investigated employing the strategy of short-read k-mer distributions using GenomeScope v1.0.0 (Vurture et al. 2017). The histogram of k-mer frequencies was estimated with 21-mers using khist.sh (BBTools). We selected 1,000 and 10,000 as the maximum k-mer coverage cutoffs.

The raw PacBio long reads were assembled into contigs using Flye v2.5 (Kolmogorov et al. 2019) and Falcon v1.3.0 (Chin et al. 2016); both assemblies were self-polished with a round of Flye polishing module "--polish-target". To improve contig contiguity, Flye and Falcon assemblies were merged with three rounds of quickmerge (Chakraborty et al. 2016) following USAGE 2 (https://github.com/mahulchak/quickmerge/wiki; last accessed September 1, 2019). We removed redundant heterozygous contigs with Purge Haplotigs v1.1.0 (Roach et al. 2018); percent cutoff for identifying a contig as a haplotig was set as 50 (-a 50) with other parameters as the default. Non-redundant assembly was further polished with Illumina short reads using two

rounds of Pilon v1.22 (Walker et al. 2014). To performed the chromosome-level concatenation, we aligned Hi-C read pairs to the genome, removed duplicates, and extracted the Hi-C contacts using Juicer v1.6.2 (Durand et al. 2016). Pseudo-chromosomes were generated using 3D-DNA v180922 (Dudchenko et al. 2017) in combination with a review step in Juicebox Assembly Tools module within Juicebox (Durand et al. 2016). Sex chromosomes with female heterogamety were determined according to male-to-female comparison approach (Tomaszkiewicz et al. 2017). Chromosomal nomenclature order and orientation were assigned in comparison with *Bombyx mori* genome (Kawamoto et al. 2019). Contaminants were removed against the nt and UniVec databases using HS-BLASTN (Chen et al. 2015) and BLAST+ (i.e. blastn) v2.7.1 (Camacho et al. 2009). We assessed the assembly completeness using Benchmarking Universal Single-Copy Orthologs (BUSCO, Waterhouse et al. 2018) analyses against insect data set (n= 1658). We also mapped PacBio long reads and Illumina short reads to the final genome assembly with Minimap2 (Li et al. 2018). In addition, we assembled the mitochondrial genome based on Illumina short reads with NOVOPlasy v2.7.2 (Dierckxsens et al. 2017) using the seed (JN290214).

Genome annotation

We constructed a de novo species specific repeat library using RepeatModeler v1.0.11 (Smit and Hubley 2008–2015), and then combined it with Dfam_3.0 (Hubley et al. 2016) and RepBase-20181026 databases (Bao et al. 2015) to generate a custom library. We then used RepeatMasker v4.0.9 (Smit et al. 2013–2015) to mask repeats in the genome assembly with the combined custom library. Non-coding RNAs were identified with Infernal v1.1.2 (Nawrocki and Eddy 2013) and tRNAscan-SE v2.0 (Chan & Lowe 2019).

We conducted gene prediction with the MAKER v2.31.10 pipeline (Holt & Yandell 2011) by integrating ab initio, transcriptome-based and protein homology-based evidence. For the ab initio gene predictions, we used Augustus v3.3 (Stanke et al. 2004) and GeneMark-ET v4.38 (Lomsadze et al. 2005), which were trained using BRAKER v2.1.0 (Hoff et al. 2016) with RNA-seq data. For the transcriptome-based evidence, a genome-guided method was applied to transcriptome assembly. RNA-seq reads were mapped to the genome with HISAT2 v2.1.0 (Kim et al. 2015) and assembled with StringTie v1.3.4 (Pertea et al. 2015). Redundant isoforms were removed with CD-HIT v4.8.1 (Fu et al. 2012) with an identity cutoff of 0.9. For the protein homology-based evidence, protein sequences of *Acyrthosiphon pisum*, *Apis mellifera*, *Bombyx mori*, *Drosophila melanogaster*, *Helicoverpa armigera* and *Tribolium castaneum* were downloaded from Ensembl (Flicek et al. 2014). Gene functions were assigned using Diamond v0.9.24 (Buchfink et al. 2015) against UniProtKB (SwissProt + TrEMBL) database with a sensitive mode and an e-value threshold of 1e-5 (--more-sensitive -e 1e-5). We also search protein domains, Gene Ontology (GO) and pathways (KEGG, MetaCyc, Reactome) with InterProScan 5.34-73.0 (Finn et al. 2017) against Pfam (El-Gebali et al. 2019), PANTHER (Mi et al. 2019), Gene3D (Lewis et al. 2018), Superfamily (Wilson et al. 2009), and CDD (Marchler-Bauer et al. 2017) databases (-dp -f TSV, GFF3 -goterms -iprlookup -pa -t p -appl Pfam, PANTHER, Gene3D, Superfamily, CDD).

Phylogenomic analyses

Orthogroups of 16 species (*Amyelois transitella*, *Bombyx mori*, *Drosophila melanogaster*,

Heliconius melpomene, *Helicoverpa armigera*, *Helicoverpa zea*, *Heliothis virescens*, *Manduca sexta*, *Operophtera brumata*, *Phoebis sennae*, *Plutella xylostella*, *Spodoptera exigua*, *Spodoptera frugiperda*, *Spodoptera litura*, *Stenopsyche tienmushanensis*, *Trichoplusia ni*) were inferred using OrthoFinder v2.3.7 (Emms & Kelly 2019) with the Diamond as the sequence aligner. Single-copy orthologous were selected for the phylogenomic reconstructions. Protein sequences of each orthogroup were aligned using MAFFT v7.394 (Katoh & Standley 2013) with the L-INS-I strategy, trimmed using trimAl v1.4.1 (Capella-Gutiérrez et al. 2009) with the heuristic method 'automated1', and concatenated using FASconCAT-G v1.04 (Kück & Longo 2014). Phylogenetic tree was reconstructed using IQ-TREE v1.6.10 (Nguyen et al. 2015) with the unpartitioned GHOST model (Crotty et al. 2019). Node support values were calculated using 1,000 SH-aLRT replicates (Guindon et al. 2010) and 1,000 ultrafast bootstraps (Hoang et al. 2018). Divergence time was inferred using r8s (Sanderson 2003) with calibration nodes extracted from TimeTree database (Kumar et al. 2017).

Synteny analysis

To search for syntenic blocks in whole genome, we extracted the coding sequences (CDSs) of *S. exigua* (except for W chromosome) and other two species (*B. mori*, *Spodoptera litura*) using custom python script, then we searched pairwise synteny using LAST (Kiełbasa et al. 2011) and removed tandem duplications and weak hits by the module McScan (Python version) in jcvi (https://github.com/tanghaibao/jcvi). We extracted subset of blocks with following options "--minspan=30 --minsize=15 --simple". Finally, the high-quality syntenic blocks were visualized by Circos (Krzywinski et al. 2009).

Results and Discussion

Genome assembly

We generated 3,919,515 PacBio subreads of 38.98 Gb (87X) with the mean and N50 length reaching 9.95 and 16.01 kb, respectively. A total of 34.62 Gb (77X), 65.21 Gb and 56.32 Gb (126X) were produced in the Illumina X Ten platform for whole genome, transcriptome and Hi-C sequencing, respectively. Estimated genome size ranged from 408.58 Mb to 448.90 Mb (Table 1). Genome repeat length increased with the maximum k-mer coverage cutoff, ranging from 64.96 Mb to 104.93 Mb. Heterozygosity rate (0.585–0.594) and a small peak on the left of main peak (Fig. 1) implied that the genome may have a moderate rate of heterozygous regions, which should be carefully addressed when assembling genome.

Table 1. Genome survey for *Spodoptera exigua*.

K-mer	**Max k-mer coverage**	**Heterozygosity (%)**		**Repeat Length (Mb)**		**Unique Length (Mb)**		**Genome size (Mb)**	
		Mix	Max	Mix	Max	Mix	Max	Mix	Max
21	1000	0.585	0.594	64.96	65.07	343.62	344.16	408.58	409.22
21	10000	0.588	0.591	104.88	104.93	343.80	343.97	448.68	448.90

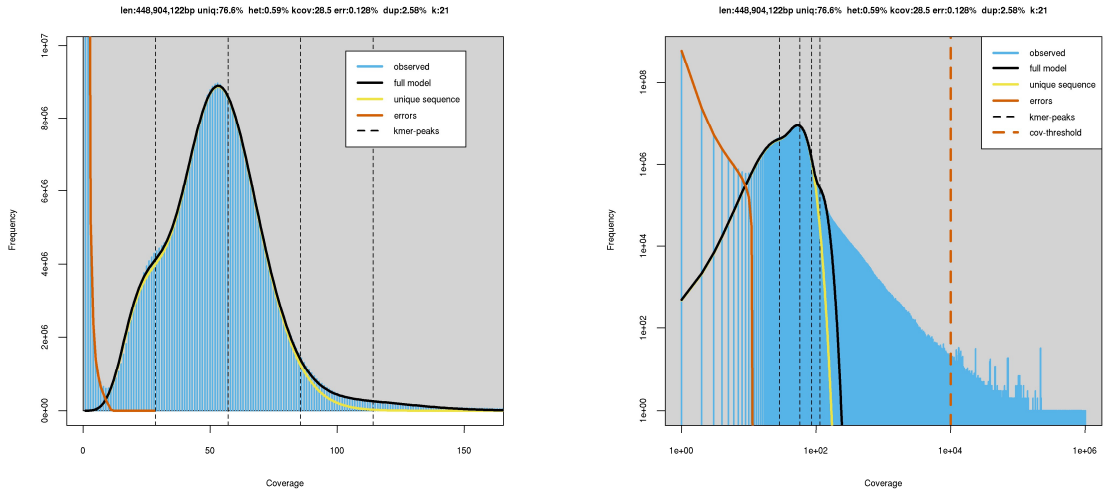

Figure 1. GenomeScope profile plots of k-mer frequency (left) and log-transformed k-mer coverage (right) at a k-mer length of 21 and a maximum k-mer coverage of 10,000.

Both Flye and Falcon assemblies (Table 2) were ca. 100 Mb larger than the estimated size due to a moderate rate of heterozygosity (duplicated BUSCOs > 10%), and merged into a 561.68 Mb-assembly of 934 contigs (N50 length 3.47 Mb). We removed 668 redundant heterozygous sequences using Purge Haplotigs, resulting an obvious improvement of N50 length (5.53 Mb). After polishing, chromosome concatenation and contaminant removal, the final genome assembly had 301 scaffolds/667 contigs, a total length of 446.80 Mb and scaffold/contig N50 length of 14.36/3.47 Mb; 367 contigs were concatenated into 32 pseudo-chromosomes (Fig. 2), accounting for 96.18% (429.74 Mb) of the total genome length. Through 3D-DNA pipeline, partial assembled scaffolds or contigs in previous assembly process were split into shorter sequences due to the misjoints. We successfully determined 30 autosomes, Z and W chromosome with their nomenclature order and orientation homogeneous to *B. mori*. The assembled size was very close to the estimated size (Table 1). BUSCO completeness analyses against insect dataset (n=1,658) (Table 2) showed our assembled versions covered 97.8–98.6% complete, 0.2–0.4% fragmented, and 1.1–1.9 missing BUSCO genes. We mapped 97.62% and 93.78% Illumina short and PacBio long reads to our final assembly. All these assessments indicated the accuracy and the high integrity of our final genome assembly with low redundancy (2.1 % duplicated).

A circularized mitochondrial genome of 15,383 bp was assembled with NOVOPlasy. It consists of 13 protein-coding genes, 2 rRNA genes and 22 tRNA genes and a control region. The mitochondrial gene number and order are similar to most lepidopteran species. The overall nucleotide composition is 40.86% A, 11.40% C, 7.68% G,40.05% T, indicating a very high A+T bias (80.91%).

Genome annotation

By combing *de novo* and homology-based evidence, we identified 33.12% (147.97 Mb) repetitive elements in the beet armyworm genome. The top five abundant repeat families were LINE (14.81%), Unknown (5.89%), RC (3.95%), LTR (3.10%), and SINE (2.53%) (Table S1). We identified 172 rRNAs (5S, 5.8S, LSU, SSU), 61 miRNAs (37 families), 118 small nuclear RNAs (snRNAs, 19 families), 2 long non-coding RNAs (LncRNA-Sphinx), 1,135 tRNAs, 3 ribozymes (2 families), 114 cis-regulatory elements, and 3 other (Metazoa_SRP) ncRNAs (Table S2). Twenty-one tRNA

isotypes were identified except for SelCys-isotype lacking. Small nuclear RNAs were classified into six spliceosomal RNAs (U1, U2, U4, U5, U6, U11), three minor spliceosomal RNAs (U12, U4atac, U6atac), and four H/ACA box and 20 C/D box snoRNAs (small nucleolar RNA).

We predicted 17,707 protein-coding genes (gene models) with a mean length of 7,294.33 bp in the beet armyworm genome. Exons and introns had a mean count of 5.88 and 4.88 per gene and a mean length of 254.65 bp and 1043.82 bp, respectively. Among predicted genes, 81.69% were supported by transcriptome evidence and 97.94% matched the UniProt protein records. InterProScan identified protein domains of 13,234 (74.74%) genes; among them, 8,814 were assigned with GO terms, and 732, 622 and 3,151 genes matched KEGG, MetaCyc, and Reactome pathway databases, respectively. BUSCO completeness assessment identified 1,528 (92.1%) complete, 45 (2.7%) duplicated, 39 (2.4%) fragmented, and 91 (5.5%) missing BUSCO genes.

Table 2. Summary of each assembly version for *Spodoptera exigua*. C, complete BUSCOs; D, complete and duplicated BUSCOs; F, fragmented BUSCOs; M, missing BUSCOs.

Assembly	**Total length (Mb)**	**No. scaffolds**	**N50 length (kb)**	**Longest scaffold (Mb)**	**GC (%)**	**BUSCO (n=1,658) (%)**			
						C	D	F	M
Flye	544.64	2,339	810	14.84	36.81	98.6	10.9	0.3	1.1
Falcon	559.54	1,061	1,600	14.32	36.71	98.4	16.6	0.2	1.4
quickmerge	561.68	934	3,474	22.01	36.71	98.1	16.9	0.4	1.5
purge_haplotigs	447.05	266	5,529	22.01	36.70	97.8	2.4	0.3	1.9
Pilon	446.52	266	5,526	22.01	36.67	98.0	2.2	0.2	1.8
Final genome assembly	**446.80**	**301**	**14,363**	**19.74**	**36.67**	**97.9**	**2.1**	**0.2**	**1.9**
Transcript assembly	**51.07**	**18,696**	**4.32**	**0.05**	**41.12**	**93.6**	**11.2**	**2.4**	**4.0**

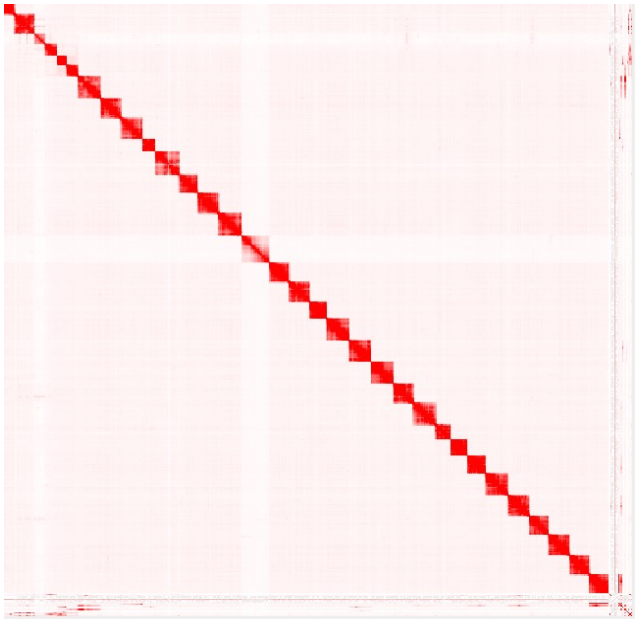

Figure 2. Hi-C Contact map of the *Spodoptera exigua* genome assembly indicating 32 pseudo-chromosomes.

Phylogeny

OrthoFinder assigned 250,040 (91.16%) genes into 14,637 gene families with a mean orthogroup size of 17.08. of 3,767 families shared by all 16 species, 595 are single-copy ones (Table S3). Phylogenetic tree clustered *Spodoptera exigua* within Noctuidae with absolutely high support values. Divergence time estimation indicated that Noctuidae and beet armyworm arised 31 and 10 millions years ago, respectively (Fig. 3).

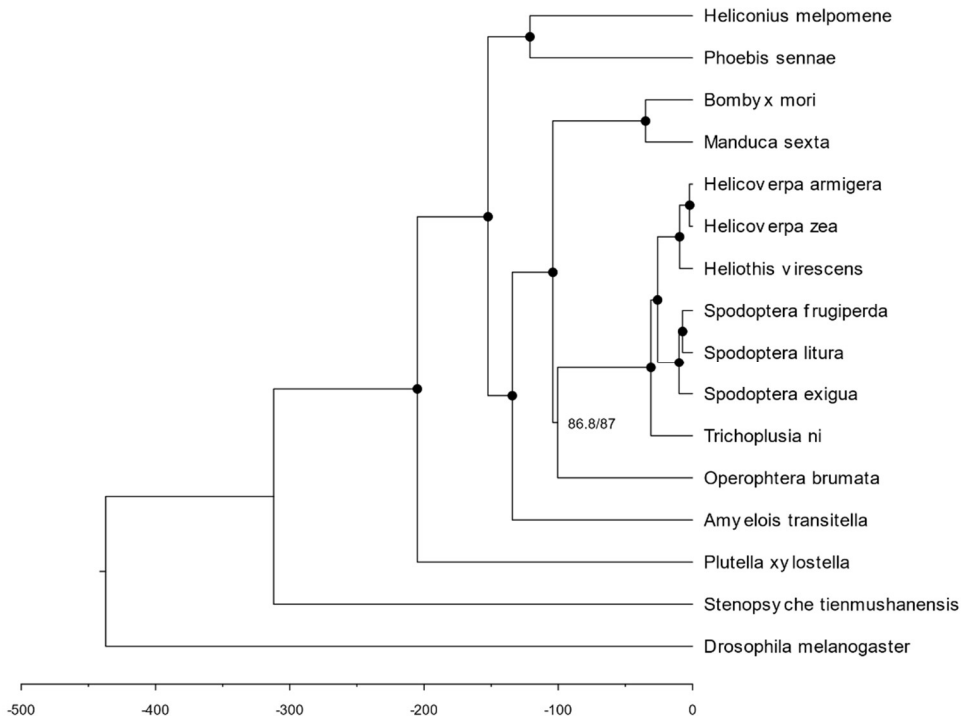

Figure 3. Phylogeny and divergence of Lepidoptera based on 595 single-copy genes. Node support values below 100 (SH-aLRT, UFBoot) are given on nodes. Circles represent the best support values of 100/100.

Synteny

Syntenic relationships showed that 100 syntenic blocks between *S. exigua* and *B. mori*, and 148 syntenic blocks between *S. exigua* and *S. litura* occurred. Chromosomes chr11 and chr29, chr23 and chr30, chr24 and chr31 in *S. exigua* were fused to chr11, chr23 and chr24 in *B. mori* (Fig. 4a). Strong syntenic relationships between *S. exigua* and *S. litura* indicated the conserved chromosome-level gene collinearity, which was also reported previously (d'Alençon et al. 2010; Chen et al. 2017; Hill et al. 2019).

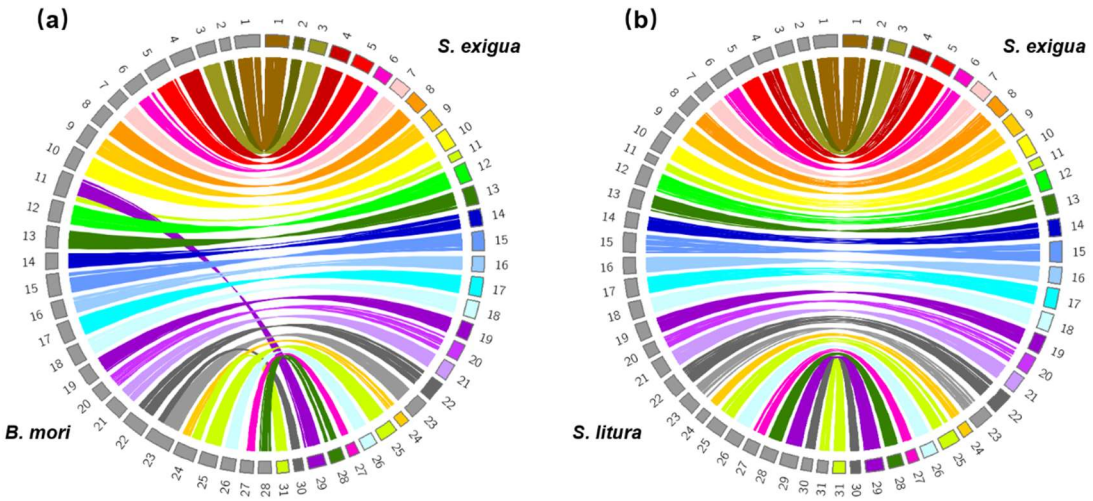

Figure 4. Circos plot of the synteny between the genomes of *S. exigua* and *B. mori* (a), *S. exigua* and *S. litura* (b).

Supplementary material

Table S1. Repeat annotation in *Spodoptera exigua*.
Table S2. Summary statistics for non-coding RNAs in *Spodoptera exigua*.
Table S3. Summary statistics for gene family inference.

Acknowledgments

This work was supported by a grant (no. 31572030) from the National Natural Science Foundation of China.

Data accessibility

This project has been deposited at the NCBI under the accessions WNNL00000000, MN735183 and SRR10441878–SRR10441889.

References

Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* 6: 11.

Bassett AR, Tibbit C, Ponting CP, Liu JL. 2013. Highly efficient targeted mutagenesis of *Drosophila* with the CRISPR/Cas9 system. Cell Rep 4: 220–228.

Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12: 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: 421.
Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25: 1972–1973.
Chakraborty M, Baldwin-Brown JG, Long AD, Emerson JJ. 2016. Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. *Nucleic Acids Res* 44: e147.
Chan PP, Lowe TM. 2019. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. *Methods Mol Biol* 1962: 1–14.
Che W, Shi T, Wu Y, Yang Y. 2013. Insecticide resistance status of field populations of *Spodoptera exigua* (Lepidoptera: Noctuidae) from China. *J Econ Entomol* 106: 1855–1862.
Chen Y, Ye W, Zhang Y, Xu Y. 2015. High speed BLASTN: an accelerated MegaBLAST search tool. *Nucleic Acids Res* 43: 7762–7768.
Cheng T, et al. 2017. Genomic adaptation to polyphagy and insecticides in a major East Asian noctuid pest. *Nat Ecol Evol* 1: 1747–1756.
Chin CS, et al. 2016. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods* 13: 1050–1054.
Cong L, et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* 339: 819–823.
Crotty SM, et al. 2019. GHOST: Recovering historical signal from heterotachously-evolved sequence alignments. *Syst Biol*. Doi: 10.1093/sysbio/syz051
d'Alençon E, et al. 2010. Extensive synteny conservation of holocentric chromosomes in Lepidoptera despite high rates of local genome rearrangements. Proc Natl Acad Sci U S A 107: 7680–7685.
Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res* 45: e18.
Dudchenko O, et al. 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* 356: 92–95.
Durand NC, et al. 2016. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* 3: 95–98.
El-Gebali S, et al. 2019. The Pfam protein families database in 2019. *Nucleic Acids Res* 47(D1): D427–D432.
Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 20: 238.
Flicek P, et al. 2014. Ensembl 2014. *Nucleic Acids Res* 42: D749–D755.
Finn RD, et al. 2017. InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res* 45(D1): D190–D199.
Fu X, Feng H, Liu Z, Wu K. 2017. Trans-regional migration of the beet armyworm, *Spodoptera exigua* (Lepidoptera: Noctuidae), in North-East Asia. PLoS One. 12: e0183582.
Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* 59: 307–321.
Han LZ, Gu HN, Zhai BP, Zhang XX. 2009. Genetic effects on flight capacity in the beet armyworm, *Spodoptera exigua* (Lep., Noctuidae). *J Appl Entomol* 133: 262–271.
Hill J, et al. 2019. Unprecedented reorganization of holocentric chromosomes provides insights into the enigma of lepidopteran chromosome evolution. *Sci Adv* 5: eaau3648.
Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the

ultrafast bootstrap approximation. *Mol Biol Evol* 35: 518–522.
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: 767–769.
Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12: 491.
Hubley R, et al. 2016. The Dfam database of repetitive DNA families. *Nucleic Acids Res* 44(D1): D81–D89.
Ishtiaq M, Saleem MA. 2011. Generating susceptible strain and resistance status of field populations of *Spodoptera exigua* (Lepidoptera: Noctuidae) against some conventional and new chemistry insecticides in Pakistan. *J Econ Entomol* 104: 1343–1348.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 30: 772–780.
Kawamoto M, et al. 2019. High-quality genome assembly of the silkworm, *Bombyx mori*. *Insect Biochem Mol Biol* 107: 53–62.
Kiełbasa SM, Wan R, Sato K, Horton P, Frith MC. 2011. Adaptive seeds tame genomic sequence comparison. *Genome Res* 21: 487–493.
Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 12: 357–360.
Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 37: 540–546.
Krzywinski M, et al. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res* 19: 1639–1645.
Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol Biol Evol* 34: 1812–1819.
Kück P, Longo GC. 2014. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Front Zool* 11: 81.
Lewis TE, 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. *Bioinformatics* 34: 3094–3100.
Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. 2005. Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res* 33: 6494–6506.
Marchler-Bauer A, et al. 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* 45(D1): D200–D203.
Mehrkhou F, Talebi AA, Moharramipour S, Naveh VH. 2012 Demographic parameters of *Spodoptera exigua* (Lepidoptera: Noctuidae) on different soybean cultivars. *Environ Entomol* 41: 326–332.
Mi H, Muruganujan A, Ebert D, Huang X, Thoas PD. 2019. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res* 47: D419–D426.
Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA ho- mology searches. *Bioinformatics* 29: 2933–2935.
Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol Biol Evol* 32: 268–274.

Osorio A, et al. 2008. Monitoring of beet armyworm resistance to spinosad and methoxyfenozide in Mexico. *Pest Manag Sci*. 64: 1001–1007.

Pertea M, et al. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 33: 290–295.

Roach MJ, Schmidt SA, Borneman AR. 2018. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* 19: 460.

Saeed Q, Ahmad F, Iqbal N, Zaka SM. 2019. Chemical control of polyphagous pests on their auxiliary hosts can minimize insecticide resistance: A case study of *Spodoptera exigua* Hübner (Lepidoptera: Noctuidae) in cotton agroecosystem. *Ecotoxicol Environ Saf* 171: 721–727.

Sanderson MJ. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics. 19: 301–302.

Smit AFA, Hubley R. 2008–2015. RepeatModeler Open-1.0. Available from: http://www.repeatmasker.org.

Smit AFA, Hubley R, Green P. 2013–2015. RepeatMasker 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: W309–W312.

Tay WT, Gordon KHJ. 2019. Going global - genomic insights into insect invasions. *Curr Opin Insect Sci* 31: 123–130.

Tomaszkiewicz M, Medvedev P, Makova KD. 2017. Y and W Chromosome Assemblies: Approaches and Discoveries. *Trends Genet* 33: 266–282.

Vurture GW, et al. 2017. GenomeScope: fast reference-free genome profling from short reads. *Bioinformatics* 33: 2202–2204.

Walker BJ, et al. 2014. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS ONE* 9: e112963.

Waterhouse RM, et al. 2018. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol Biol Evol* 35: 543–548.

Wilson D, et al. 2009. SUPERFAMILY—sophisticated comparative genomics, data mining, visualization and phylogeny. *Nucleic Acids Res* 37: D380–D386.

Zhou LH, Wang XY, Xu GQ, Lei JJ. 2016. Mitochondrial DNA phylogeography of *Spodoptera exigua* across a broad geographic area in China. *J Appl Entomol* 141: 527–539.

Zuo Y, et al. 2017. CRISPR/Cas9 mediated G4946E substitution in the ryanodine receptor of *Spodoptera exigua* confers high levels of resistance to diamide insecticides. *Insect Biochem Mol Biol* 89: 79–85.

Zuo YY, et al. 2018. Knockout of a P-glycoprotein gene increases susceptibility to abamectin and emamectin benzoate in *Spodoptera exigua*. *Insect Mol Biol* 27: 36–45.