



OPEN

# A chromosome-level genome assembly of *Sesamia inferens*

DATA DESCRIPTOR

Hongran Li<sup>1,2</sup>, Yan Peng<sup>1,2</sup>, Chao Wu<sup>1,2</sup>, Chess-Kadouste Vigan<sup>1,2</sup>, Kaikai Mao<sup>1</sup>, Jingyun Zhu<sup>1</sup>, Luming Zou<sup>1</sup>, Minghui Jin<sup>1</sup>, Lei Zhang<sup>1</sup> & Yutao Xiao<sup>1</sup>✉

The pink stem borer, *Sesamia inferens* (Walker), is a significant polyphagous pest historically restricted to regions south of N34° latitude. However, with changes in global climate and farming practices, the distribution of this moth has progressively exceeded its traditional limit of 34° N and encompassed most regions in North China. The genetic adaptations of *S. inferens* remain incompletely understood due to the lack of high-quality genome resources. Here, we sequenced the genome of *S. inferens* using PacBio and Hi-C technology, yielding a genome assembly of 865.04 Mb with contig N50 of 1.23 Mb. BUSCO analysis demonstrated this genome assembly has a high-level completeness of 96.1% gene coverage. In total, 459.72 Mb repeat sequences (53.14% of the assembled genome) and 20858 protein-coding genes were identified. We used the Hi-C technique to anchor 1135 contigs to 31 chromosomes, yielding a chromosome-level genome assembly with a scaffold N50 of 29.99 Mb. In conclusion, our high-quality genome assembly provided valuable resource that exploring the genetic characteristics of local adaptation and developing an efficient control strategy.

## Background & Summary

The pink stem borer, *Sesamia inferens* (Walker) (Lepidoptera: Noctuidae), is a significant polyphagous pest that damages a wide range of food crops including rice, maize, sorghum, and barley<sup>1,2</sup>. In recent decades, the impact of *S. inferens* has been steadily rising in most areas along the Yangtze River and coastal regions in southern China. In certain locations, the damage caused by this moth has even surpassed that of the *Chilo suppressalis*, propelling it from a secondary pest to a primary and significant threat<sup>3,4</sup>. The *S. inferens* populations likely have an abundance of ecological phenotypes to adapt to complex environmental conditions. For example, the number of generations per year decreases with increasing latitude, which is largely affected by the length of the growing season<sup>5</sup>. It consumes internal tissues of sugarcane, and the distinctive “deadheart” symptom manifests when it feeds on young plants. Infestation in older plants may lead to a decrease in growth, and stems may exhibit desiccation<sup>6</sup>. Furthermore, there have been studies concentrating on the characterization of insecticide resistance in *S. inferens*. For instance, a population resistant to Fipronil exhibited a 106.00- and 22.71-fold resistance compared to unselected and field strains<sup>7,8</sup> documented the susceptibility of three field populations of *S. inferens* to chlorantraniliprole and flubendiamide. They also conducted the cloning and characterization of the full-length cDNA of the ryanodine receptor, along with profiling its mRNA expression pattern.

The *S. inferens* is widely distributed in Asian countries such as China, Japan, India, Laos and Pakistan<sup>9–13</sup>, and was recently discovered in Hawaii and Guam (United States) according to Centre for Agriculture and Bioscience International (CABI) (<https://www.cabi.org/isc/datasheet/49751#REFDDB-202162>). In China, the natural range of this species was traditionally confined to areas south of N34° latitude<sup>14</sup>. The *S. inferens* exhibited limited ability for migratory flight, as 75.5% of the moths had a cumulative flight distance of  $\leq 5.0$  km<sup>15</sup>. Thus, during winter, the *S. inferens* predominantly overwinters as mature larvae on the roots of rice, water bamboo, and various weed species<sup>4</sup>. However, due to the change of global climate and farming system, the distribution of the moth has progressively exceeded its traditional limit of 34° N and expanded to include the majority of regions in North China<sup>4,16</sup>. Given the overwintering ability being one of the main factors restricting the distribution of the moth, it is particularly important to study the adaptable ability of the overwintering larvae of the moth in North China.

While the severe damage to rice and other crops caused by *S. inferens* has gained significant attention, the current efforts primarily focus on the field dynamics and forecast, environmental adaption, biological control

Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Gene Editing Technologies (Hainan), Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, 518116, P. R. China. <sup>2</sup>These authors contributed equally: Hongran Li, Yan Peng, Chao Wu, Chess-Kadouste Vigan. ✉e-mail: [xiaoyutao@caas.cn](mailto:xiaoyutao@caas.cn)

Category	Number of BUSCOs
C: 96.1% [S:94.7%, D:1.4%], F: 0.9%, M: 3.0%	1658
Complete BUSCOs (C)	1593
Complete and single-copy BUSCOs (S)	1570
Complete and duplicated BUSCOs (D)	23
Fragmented BUSCOs (F)	15
Missing BUSCOs (M)	50

**Table 1.** Assessment of completeness of the *S. inferens* assembly.

and pesticides resistance assessment<sup>7,8,10,17–20</sup>. Indeed, the knowledge of the population structure and genetic bases of the rapid adaptation of *S. inferens* in its expanded habitat are limited. Only a few studies have been conducted to explore the genetic diversity of *S. inferens*, using a smaller number of molecular markers<sup>21,22</sup>. These findings were based on mitochondrial DNA or microsatellite markers, which either reflected the maternal history or were limited by a lack of sufficient genetic information. Thus, the genetic adaptations of *S. inferens* in North China remain to be confirmed using whole genomic variation. Yet, the lack of a high-quality reference genome seriously limits understanding the extent to which genetic variation resulted in *S. inferens*'s expanding range and adaptations to local climatic regimes.

Here, we generated a chromosome-level genome assembly of *S. inferens* using PacBio and Hi-C technology. Phylogenetic analysis was performed to determine the relationship of *S. inferens* with other Noctuidae species. Moreover, functional enrichment showed that the rapidly expanded gene families and positively selected genes were associated with multiple metabolism-related pathways that contributed to the local adaptation in new environment. Our study provides the first genome assembly for the pink stem borer, which will facilitate studies on the genetic mechanisms of evolutionary adaptation for in large-scale habitats and agro-ecosystems across China and also significantly benefit efforts to control this important rice pest.

## Methods

**Sample collection and genomic sequencing.** The pink borer samples were collected from Guangdong Province, China in 2022 and were reared under the controlled conditions ( $27 \pm 2^\circ\text{C}$ , 16L: 8D and RH  $60 \pm 5\%$ ) in the laboratory for one generation to obtain sibling pupa for Hi-C and genome sequencing. To effectively eliminate microbial contaminants present on the surface of the pupa, the pupa sample underwent surface sterilization through a triple treatment process involving 75% alcohol and sodium hypochlorite solution.

For PacBio sequencing, genomic DNA was extracted from one male pupa and then subjected to sequencing on the Pacific Biosciences Sequel II platform to generate HiFi reads in circular consensus sequencing (CCS) mode. A total of 24 Gb ( $\sim 29.26 \times$  coverage) single molecule real-time (SMRT) long reads were generated. In parallel, Genomic DNA was extracted from another pupa of the same parent to facilitate the construction of Hi-C libraries. These libraries were generated using the MboI restriction enzyme and subsequently subjected to sequencing using the Illumina Novaseq/MGI-2000 platform. This phase of the study produced  $\sim 86.9$  Gb ( $\sim 100 \times$  coverage) of data with 150 bp paired-end sequencing raw reads.

**RNA sequencing.** For the purpose of genome annotation in *S. inferens*, total RNA was isolated from larvae, pupae, and adults employing the TRIzol reagent from Invitrogen, USA. Subsequently, a cDNA library was generated utilizing the NEBNext Ultra RNA Library Prep Kit designed for Illumina (NEB), in accordance with the provided instructions. RNA-seq libraries were constructed and sequenced using Illumina HiSeq X Ten (insert size 240 bp, 150 PE reads) at Novogene, Tianjing. As a result, a total of 61.27 Gb sequencing data was generated.

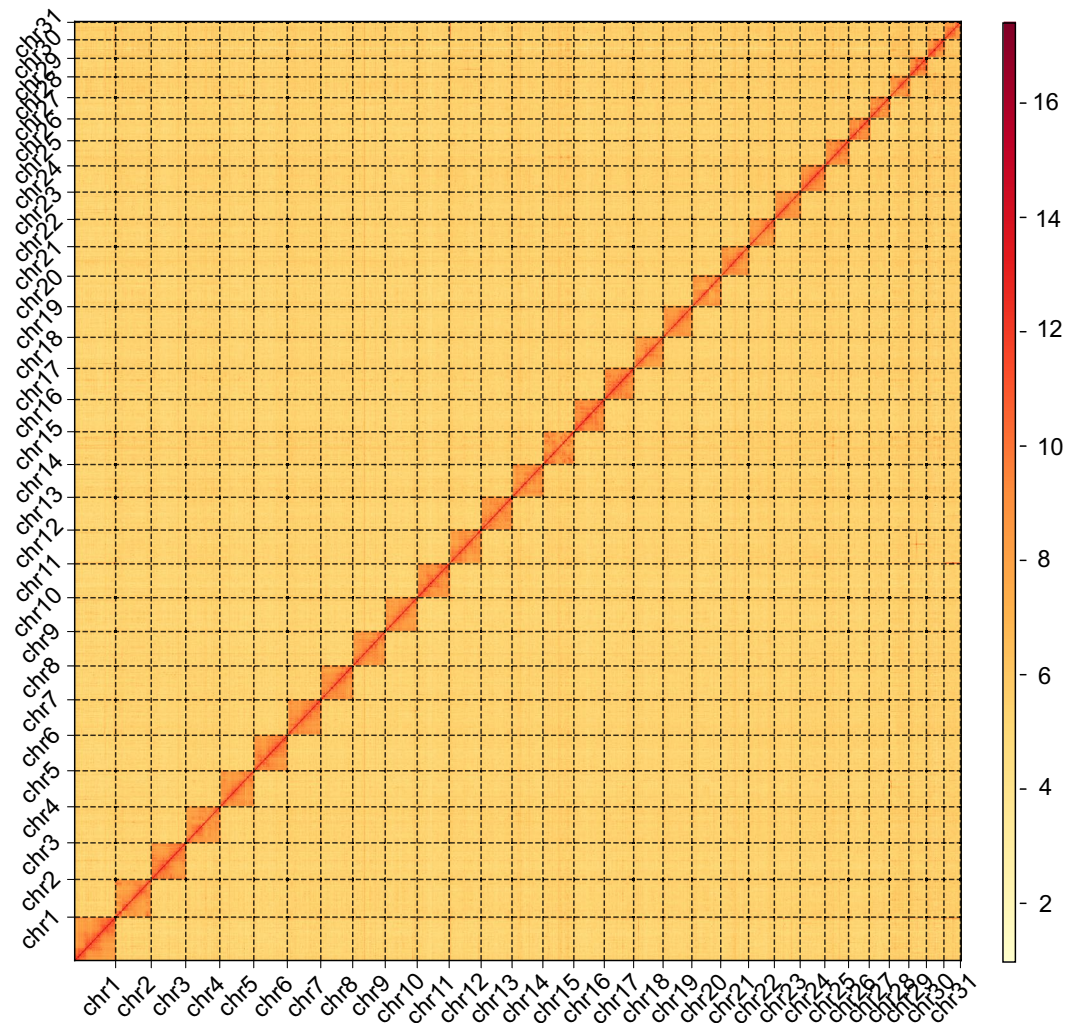
**Genome assembly.** The estimation of genome size relied on analyzing the distribution of k-mer frequencies. Here, 21-kmer analysis of Illumina paired-end sequencing reads was performed using Jellyfish v2.1.3 software<sup>23</sup>, while GenomeScope 2.0<sup>24</sup> was employed for calculating heterozygosity. As a result, the genome size of *S. inferens* was projected to be approximately 865.04 Mb (Fig. S1).

To eliminate CCS reads containing residual PacBio adapter sequences, the HiFiAdapterFilt software was applied<sup>25</sup>. The ensuing clean PacBio reads were then assembled utilizing default settings in Hicanu v2.0<sup>26</sup>. The software `purge_dups` was subsequently utilized for the removal of haplotigs and contig overlaps<sup>27</sup>. To gauge the assembly's comprehensiveness, Benchmarking Universal Single-Copy Orthologs v5.1.2 (BUSCO) was employed, using homologous genes from `insecta_odb10`. The outcomes exhibited a successful detection rate of 96.1% for BUSCO genes, of which 94.7% were identified as complete and single-copy genes and 1.4% were categorized as complete and duplicated (Table 1). Following these evaluations, a total of 1135 contigs were assembled, with a contig N50 size measuring 1.23 Mb (Table 2).

**Hi-C scaffolding.** Initial processing involved the removal of low-quality raw reads (those with a quality score  $< 20$  and shorter than 30 bp), along with adapter sequences, through the utilization of FASTP v0.20.0. Subsequently, the purified reads were aligned to the contig assembly using BOWTIE2 v2.3.2 with specific parameters (`-end-to-end-very-sensitive -L 30`)<sup>28</sup>. The Hi-C reads were used to map the draft genome using Juicer v1.5 and corrections for misjoins, ordering, and orientation were carried out using 3D-DNA v.180922 with the parameter (`-r 0`)<sup>29,30</sup>. As a result, Hi-C data were combined with the contig-level assembly to generate a chromosome-level assemble. These contigs were improved to generate 69 scaffolds (865.04 Mb) with a scaffold N50 size of 29.99 Mb (Table 2; Table S1). Finally, Hi-C data were employed for the anchoring, ordering,

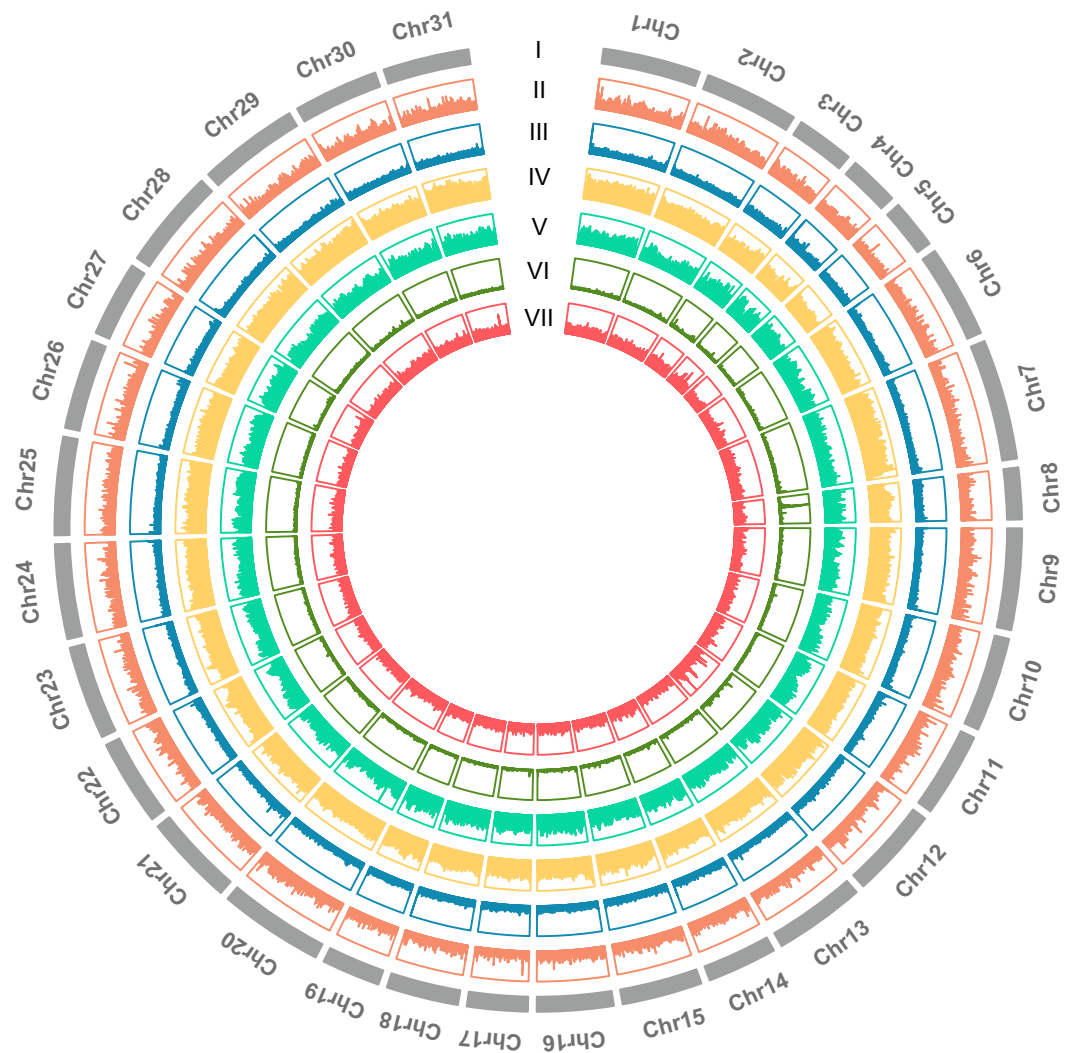
	<i>S. inferens</i>	<i>A. ipsilon</i>	<i>S. frugiperda</i>	<i>S. litura</i>	<i>H. armigera</i>	<i>T. ni</i>
Genome size (Mb)	865.04	515	390.38	438.32	356.67	333.0
Karyotype	—	—	30 + Z	30 + Z	30 + Z	30 + Z
Number of contigs	1135	274	776	13636	106	7885
Number of scaffolds	69	31	—	3597	41	1916
Max scaffold length (Mb)	38.00	—	21.4	4.4	17.88	8.92
Contig N50 (kb)	1231.9	6747	5606.9	68.4	7516.16	140.0
Scaffold N50 (Mb)	28.60	17.57	12.69	13.44	12.27	4.51
BUSCO (%)	96.1	98.1	—	99.0	97.7	98.8
G + C (%)	38.62	38	36.4	36.6	36.53	35.5
Repeat (%)	53.14	32.09	27.2	31.8	27.04	16.7
Number of genes	18937	19412	22260	15317	18668	14384

**Table 2.** Comparison of genome assemblies in six Noctuidae species. Note: *A. ipsilon*<sup>43</sup>, *S. frugiperda*<sup>44</sup>, *S. litura*<sup>45</sup>, *H. armigera*<sup>46</sup>, *T. ni*<sup>47</sup>.



**Fig. 1** Heatmap of genome-wide Hi-C data of the pink borer, *Sesamia inferens*. Linkage group contact map informed by Hi-C sequencing data in the genome. Thirty-one linkage groups were generated after the clustering of contact map. The colour bar indicates the frequency of Hi-C interaction intensity from low (yellow) to high (red) in the plot.

and orientation of these scaffolds, yielding 31 chromosomes (Chr1-Chr31), harboring >99.05% of assembled sequences (Fig. 1).



**Fig. 2** Circular representation of the chromosomes of the pink borer, *Sesamia inferens*. The genome information landscape map, from the outer ring, I: chromosome ideograms (Mb scale), II: Density of protein-coding genes; III: Density of G + C content; IV: Density of genome-wide single nucleotide polymorphism (SNP) sites; V: Density of long interspersed nuclear elements (LINEs); VI: Density of long terminal repeats (LTR); VII: Density of DNA transposons.

**Repeat annotation.** We customized a de novo repeat library of the genome using RepeatModeler open-1.0.11. RepeatMasker v2.1 (<http://www.repeatmasker.org/>)<sup>31</sup> was used to compare the genome sequence with the known repeat sequences in the reference database (Repbase). The Tandem Repeats Finder (TRF) package v4.09 was used to identify tandem repeat sequences in the *S. inferens* genomes<sup>32</sup>. In total, 459.72 Mb sequences (approximately 53.14% of the assembled genome) were identified as repeat sequences (Fig. 2). Among them, long interspersed nuclear elements (LINE) (18.01%), DNA transposons (5.12%), and short interspersed nuclear elements (SINE) (5.00%) represented the top three most abundant repeat types. Unknown repeat types accounted for 15.67% of identified repeat regions (Table S2).

**Gene prediction and functional annotation.** Three strategies including ab initio prediction, homology search and transcriptome-based approaches were integrated to predict protein-coding genes. For homology-based annotation analysis, the genome sequences of *S. inferens* were cross-referenced with the protein-coding sequences of related species (*Helicoverpa armigera*, *Manduca sexta*, *Spodoptera litura*, *Trichoplusia ni*) through BLAST<sup>33</sup> and GeneWise v2.4.1<sup>34</sup> to deduce gene structures. For transcriptome-based prediction, HISAT v2.147<sup>35</sup> was used to align the transcriptome data to the genome, and gene information was predicted using StringTie v1.3.4c<sup>36</sup>. For the ab initio method, the software packages Augustus v3.2.2<sup>37</sup> was employed with default settings. The consolidated gene models from the ab initio, homology-based, and RNA-seq-based methods were then integrated using EvidenceModeler v1.1.0 with default settings<sup>38</sup> to yield a comprehensive gene dataset. A total of 20,858 protein-coding genes within the *S. inferens* genome were generated through the merging of these approaches (Table S3).

Subsequently, the functional annotation of the protein-coding genes was performed against databases like the non-redundant protein database (NR), Kyoto Encyclopedia of Genes and Genomes (KEGG), eggNOG, and Trembl using BLASTP v2.7.1 with a threshold of  $1e^{-5}$ . This analysis revealed that functional annotations were possible for 18,937 genes, equating to 90.79% of the total (Table S3).

### Data Records

The genome assembly utilized PacBio and Hi-C sequencing techniques, and the data have been stored in the NCBI Sequence Read Archive with the accession numbers SRR25638298 and SRR25638299 under accession number SRP455073<sup>39</sup>. Additionally, the transcriptome data used for genome annotation are also available in the NCBI Sequence Read Archive with accession numbers ranging from SRR25638295 to SRR25638297 (The detailed information of the raw data in Table S4). The final genome assembly and annotation data have been deposited in the figshare database (<https://doi.org/10.6084/m9.figshare.23036267.v1>)<sup>40</sup>. The genome assembly has been also deposited at GenBank under the accession GCA\_035079275.1<sup>41</sup>.

### Technical Validation

Three validation methods were employed to assess the contiguity, accuracy and completeness of the genome assembly. Firstly, a total of 1135 contigs were assembled, with a contig N50 size measuring 1.23 Mb. We then used the Hi-C data combined with the contig-level assembly to generate a chromosome-level assemble that spanned 865.04 Mb, characterized by a scaffold N50 value of 29.99 Mb. Secondly, the Hi-C heatmap unveiled a discernible and well-structured pattern of interaction contacts along the diagonals within and around the chromosome inversion region, resulted in the encapsulation of over 99.05% of the assembled sequences within these chromosomes (Fig. 1). Finally, we employed Benchmarking Universal Single-Copy Orthologs (BUSCOv5.1.2), using homologous genes from insecta\_odb10. The results indicated that 96.1% of BUSCO genes (insecta\_odb10) were successfully detected within the genome assembly (Table 1). Among them, 94.7% were identified as single-copy genes and 1.4% were categorized as duplicated. These observations indirectly validated the precision and accuracy of the chromosome assembly.

To ensure the completeness and accuracy of the annotated gene set, the forecasted gene models underwent a comparison with multiple protein databases including nr, eggNOG, Trembl, and KEGG. The outcome demonstrated that 18,937 (90.79%) of the projected gene models exhibited notable homology with proteins found within at least one of these databases. Moreover, clean reads from three transcriptomes of larvae, pupa and adults were mapped onto the genome assembly, more than 90.5% of the RNA-Seq reads can be aligned to the coding regions of the reference genome.

### Code availability

All software and pipeline were executed following the instructions and protocols provided by the respective bioinformatic tools' publications. All the scripts for the genome assemble we have deposited in a public repository <https://doi.org/10.6084/m9.figshare.24570898.v1><sup>42</sup>. The software versions and corresponding code/parameters used are comprehensively outlined in the Methods section.

Received: 18 August 2023; Accepted: 9 January 2024;

Published online: 25 January 2024

### References

- Mahesh, P., Srikanth, J. & Chandran, K. Pattern of pink stem borer *Sesamia inferens* (Walker) incidence in different crop seasons and *Saccharum spp. J. Sugarcane. Res.* **4**, 91–95 (2014).
- Baladhiya, H., Sisodiya, D. & Pathan, N. A review on pink stem borer, *Sesamia inferens* Walker: a threat to cereals. *J. Entomol. Zool. Stud.* **6**, 1235–1239 (2018).
- Tang, X. T., Xu, J., Sun, M., Xie, F. F. & Du, Y. Z. First microsatellites from *Sesamia inferens* (Lepidoptera: Noctuidae). *Ann. Entomol. Soc. Am.* **107**, 866–871 (2014).
- Tang, X. T., Lu, M. X. & Du, Y. Z. Molecular phylogeography and evolutionary history of the pink rice borer (Lepidoptera: Noctuidae): Implications for refugia identification and pest management. *Systematic Entomology* **47**, 371–383 (2022).
- Xu, L., Li, C. C., Hu, B. J., Zhou, Z. Z. & Li, X. X. Review of History, Present Situation and Prospect of Pink Stem Borer in China. *Chin. Agr. Sci. Bull.* **27**, 244–248 (2011).
- Nagayama, A., Arakaki, N., Kishita, M. & Yamada, Y. Emergence and mating behavior of the pink borer, *Sesamia inferens* (Walker) (Lepidoptera: Noctuidae). *Appl. Entomol. Zool.* **39**, 625–629 (2004).
- Mansoor, M. M., Raza, A. B. M. & Afzal, M. B. S. Fipronil resistance in pink stem borer, *Sesamia inferens* (Walker) (Lepidoptera: Noctuidae) from Pakistan: cross-resistance, genetics and realized heritability. *Crop. Prot.* **120**, 103–108 (2019).
- Wu, S. F. *et al.* Molecular characterization and expression profiling of ryanodine receptor gene in the pink stem borer, *Sesamia inferens* (Walker). *Pestic. Biochem. Physiol.* **146**, 1–6 (2018).
- Chai, H. N. & Du, Y. Z. The complete mitochondrial genome of the pink stem borer, *Sesamia inferens*, in comparison with four other noctuid moths. *Int. J. Mol. Sci.* **13**, 10236–10256 (2012).
- Sun, M., Tang, X. T., Lu, M. X., Yan, W. F. & Du, Y. Z. Cold tolerance characteristics and overwintering strategy of *Sesamia inferens* (Lepidoptera: Noctuidae). *Fla. Entomol.* **97**, 1544–1553 (2014).
- Reddy, M. L., Babu, T. R. & Venkatesh, S. A new rating scale for *Sesamia inferens* (Walker) (Lepidoptera: Noctuidae) damage to maize. *Int. J. Trop. Insect Sci.* **23**, 293–299 (2003).
- Karim, S. & Riazuddin, S. Rice insect pests of Pakistan and their control: a lesson from past for sustainable future integrated pest management. *Pak. J. Biol. Sci.* **2**, 261–276 (1999).
- Li, C. X., Cheng, X. & Dai, S. M. Distribution and insecticide resistance of pink stem borer, *Sesamia inferens* (Lepidoptera: Noctuidae), in Taiwan. *Formosan Entomol.* **31**, 39–50 (2011).
- Zhang, S. M. Discussion the boundary line between the ancient North and the Eastern regions in the east of the Qinling Mountains in China from the distribution of some agricultural insects. *Acta. Entomol. Sin.* **14**, 411–419 (1965). (In Chinese).

15. Han, L. Z., Peng, Y. F. & Wu, K. M. Studies on larval dispersal ability in the field and flight capacity of the pink stem borer, *Sesamia inferens*. *Plant Prot.* **38**, 9–13 (2012).
16. Chen, X. J. & Lu, D. H. Study Advances in Occurrence and Control of the Polyphagous Pink Stem Borer *Sesamia inferens*. *Chin. Agric. Sci. Bull.* **31**, 171–175 (2015).
17. Jin, J. Y., Li, Z. Q., Zhang, Y. N., Liu, N. Y. & Dong, S. L. Different roles suggested by sex-biased expression and pheromone binding affinity among three pheromone binding proteins in the pink rice borer, *Sesamia inferens* (Walker) (Lepidoptera: Noctuidae). *J. Insect. Physiol.* **66**, 71–79 (2014).
18. Yang, G. Q., Du, S. G., Li, L., Jiang, L. B. & Wu, J. C. Potential positive effects of pesticides application on *Sesamia inferens* (Walker) (Lepidoptera: Insecta). *Int. J. Insect. Sci.* **6**, S16485 (2014).
19. Li, B. *et al.* *Chilo suppressalis* and *Sesamia inferens* display different susceptibility responses to Cry1A insecticidal proteins. *Pest. Manag. Sci.* **71**, 1433–1440 (2015).
20. Huang, J. *et al.* Low-temperature derived temporal change in the vertical distribution of *Sesamia inferens* larvae in winter, with links to its latitudinal distribution. *PLoS One* **15**, e0236174 (2020).
21. Yao, Y. H., Du, Y. Z., Zheng, F. S. & Wang, L. P. The variation of mtDNA COII sequences in 9 geo-populations of rice stem borer, *Sesamia inferens*. *Environ. Entomol.* **1**, 9 (2008).
22. Zhang, Z. C. *et al.* Genetic diversity of different geographical populations of *Sesamia inferens* as determined by AFLP. *J. Appl. Entomol.* **50**, 693–699 (2013).
23. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
24. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **11**, 1–10 (2020).
25. Sim, S. B., Corpuz, R. L., Simmonds, T. J. & Geib, S. M. HiFiAdapterFilt, a memory efficient read processing pipeline, prevents occurrence of adapter sequence in PacBio HiFi reads and their negative impacts on genome assembly. *BMC Genom.* **23**, 1–7 (2022).
26. Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
27. Guan, D. *et al.* Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896–2898 (2020).
28. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods.* **9**, 357–359 (2012).
29. Durand, N. C. *et al.* Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
30. Dudchenko, O. *et al.* De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
31. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **4**, 11–4 (2009).
32. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–80 (1999).
33. Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–64 (2002).
34. Birney, E., Clamp, M. & Durbin, R. GeneWise and Genomewise. *Genome Res.* **14**, 988–95 (2004).
35. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–60 (2015).
36. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, 1650 (2016).
37. Stanke, M. *et al.* AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, 435–439 (2006).
38. Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using Evidence-Modeler and the program to assemble spliced alignments. *Genome Biol.* **9**, R7 (2008).
39. *NCBI Sequence Read Archive* <https://identifiers.org/ncbi/insdc.sra:SRP455073> (2023).
40. Li, H. Genome assemble of *Sesamia inferens*. *figshare. Dataset.* <https://doi.org/10.6084/m9.figshare.23036267.v1> (2023).
41. *NCBI GenBank* [https://identifiers.org/ncbi/insdc.gca:GCA\\_035079275.1](https://identifiers.org/ncbi/insdc.gca:GCA_035079275.1) (2024).
42. Peng, Y. A chromosome-level genome assembly of the pink stem borer. *figshare. Software.* <https://doi.org/10.6084/m9.figshare.24570898.v1> (2023).
43. Jin, M. *et al.* Chromosome-level genome of black cutworm provides novel insights into polyphagy and seasonal migration in insects. *BMC Biol.* **21**, 2 (2023).
44. Zhang, L. *et al.* Genetic structure and insecticide resistance characteristics of fall armyworm populations invading China. *Mol. Ecol. Resour.* **20**, 1682–169 (2020).
45. Cheng, T. *et al.* Genomic adaptation to polyphagy and insecticides in a major East Asian noctuid pest. *Nat. Ecol. Evol.* **1**, 1747–1756 (2017).
46. Zhang, J. *et al.* Population genomics provides insights into lineage divergence and local adaptation within the cotton bollworm. *Mol. Ecol. Resour.* **22**, 1875–1891 (2022).
47. Chen, W. *et al.* A high-quality chromosome-level genome assembly of a generalist herbivore, *Trichoplusia ni*. *Mol. Ecol. Resour.* **19**, 485–496 (2019).

## Acknowledgements

This work was supported by STI 2030–Major Projects (2022ZD04021), Shenzhen Science and Technology Program (Grant No. KQTD20180411143628272) and the National Natural Science Foundation of China (32302352). The third funder had no role in study design, data collection and analysis, decision to publish or manuscript preparation.

## Author contributions

Y.X. and H.L. conceived of the study. H.L., Y.P. and C.W. prepared samples for genome sequencing and conducted bioinformatics analysis. The manuscript was written by H.L. and finalized by Y.X. with contributions from V.C., L.Z., K.M., J.Z., L.Z. and M.J. All authors read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41597-024-02937-6>.

**Correspondence** and requests for materials should be addressed to Y.X.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024