



OPEN

DATA DESCRIPTOR

A chromosome-level genome assembly of the *Hermonassa cecilia* (Lepidoptera: Noctuidae)

Chao Wu^{1,5} , Xinyue Liang^{1,5}, Chengyue Xian^{1,2,5}, Lei Zhang¹ , Zhiqiang Zhang² , Yao Tan^{2,3} , Jingang Liang⁴ & Yutao Xiao¹

Hermonassa cecilia is a Lepidoptera pest primarily distributed in East Asia, belonging to the subfamily Noctuidae, which includes species that typically target and damage the underground parts of plants. However, there is limited information available on the life history and genomic resources of *H. cecilia* to date. In this study, we present a high-quality reference genome of *H. cecilia* generated using PacBio sequencing and Hi-C methods. The assembled genome size is 626.10 Mb, with a N50 of 21.00 Mb, and the contigs were mapped onto 31 chromosomes. BUSCO analysis indicated high genome completeness, with a score of 99.40%. We identified 281.45 Mb of repetitive sequences, which account for 44.21% of the genome, and annotated 22,662 protein-coding genes, 89.29% of which had functional annotations. This study represents the first assembly and annotation of the *H. cecilia* genome, providing a valuable resource for understanding its biological characteristics and offering significant potential for comparative genomics within the Noctuidae subfamily.

Background & Summary

In the field of entomological research, exploring the genomes of specific species is critical for understanding their biological essence, ecological adaptation mechanisms, and potential applications. *Hermonassa cecilia*, a significant species within the genus *Hermonassa* of the family Noctuidae, is primarily found in East Asia¹. This genus exhibits the highest species diversity in the Sino-Himalayan mountains, straddling the border between the Palearctic and Oriental regions². Additionally, several known Lepidopteran pests that damage the underground parts of plants belong to the same subfamily, Noctuidae, as *H. cecilia*. Currently, research on *H. cecilia* mainly focuses on its morphological characteristics, geographical distribution, and preliminary ecological observations. Although existing studies have outlined its biological characteristics and life habits, in-depth genomic analysis remains an unexplored field.

Genomics has recently become an indispensable tool for studying biodiversity, tracing evolutionary history, and analyzing ecological adaptation mechanisms. Internationally, genome research on other noctuid insects (such as *Agrotis segetum*³, *Spodoptera litura*⁴ and *Spodoptera frugiperda*⁵) has made significant progress, providing valuable insights into the genetic mechanisms, ecological adaptability, and pest management strategies of noctuid insects. Similarly, studies on other Lepidopteran pests, such as *Ostrinia furnacalis*⁶ and *Conogethes punctiferalis*⁷, have expanded our understanding of pest evolution, ecological niche differentiation, and control strategies across diverse species. However, the genetic differences and ecological niche differentiation among species underscore the unique scientific value and practical significance of independent research on the *H. cecilia* genome. For example, genome research not only provides a comprehensive view of the complexity of their genetic information but also uses the powerful tool of comparative genomics to deeply analyze their

¹Shenzhen 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, 518120, China. ²Key Laboratory of Grassland Resources, Ministry of Education, Inner Mongolia Agricultural University, Hohhot, 010019, China. ³Research Center for Grassland Entomology, Inner Mongolia Agricultural University, Hohhot, 010019, China. ⁴Development Center of Science and Technology, Ministry of Agriculture and Rural Affairs, Beijing, 100176, China. ⁵These authors contributed equally: Chao Wu, Xinyue Liang, Chengyue Xian. ✉e-mail: zhangzq198909@imau.edu.cn; 850310.tanhuaf4@163.com; liangjingang@agri.gov.cn; xiaoyutao@caas.cn

Method	reads number	Base number (Gb)	Coverage (×)
HiFi	2,379,152	41.00	65.49
Hi-C	729,383,748	109.41	174.47
BGISEQ	405,555,358	60.83	97.16
RNA	196,143,596	29.42	\

Table 1. Statistics for sequencing data for *Hermonassa cecilia* genome assembly.

phylogenetic relationships with related species, gene family evolution dynamics, and potential adaptive evolutionary events.

This study focuses on a comprehensive analysis of the *H. cecilia* genome using high-throughput sequencing technology, aiming to construct a high-quality reference genome sequence and conduct systematic annotation and in-depth analysis. To achieve this goal, this study employs the advanced combination of PacBio HiFi and Hi-C sequencing technologies for high-precision sequencing of the *H. cecilia* genome. Advanced bioinformatics methods are utilized for the precise processing, assembly, annotation, and in-depth analysis of the sequencing data. As a result, we obtained a chromosome-level genome assembly for *H. cecilia*, with a genome size of 626.10 Mb and a scaffold N50 of 21.00 Mb. The genome assembly completeness was evaluated at 99.4%, showcasing its exceptional quality. Our analysis identified 44.21% repeat sequences and 22,662 protein-coding genes. Additionally, functional prediction for protein-coding genes was performed using four databases, resulting in functional annotation information for a total of 20,221 genes.

The publication of the *H. cecilia* genome will enrich the genomic data of Noctuid species, providing potential for a more comprehensive understanding of their genetic basis, evolutionary history, and complex relationships with their ecological environment. Furthermore, the results of this study will offer valuable references and insights for the genomic research of other noctuid insects, promoting the in-depth development of genomics research in entomology, and contributing new scientific knowledge and strength to ecological protection and pest management.

Methods

Sample collection, DNA and RNA extraction. The sequencing samples of *H. cecilia* were all male adults, collected through light trapping in October 2023 from the cropland of Xingping City, Xianyang, Shaanxi Province, China. Genomic DNA (gDNA) from a single male adult insect, intended for both PacBio HiFi and Hi-C sequencing, was extracted using the Qiagen Genomic DNA Kit (Cat#13323, Qiagen). The quality and quantity of the extracted gDNA were evaluated using a NanoDrop One UV-Vis spectrophotometer (Thermo Fisher Scientific) and a Qubit 3.0 Fluorometer (Invitrogen), following the manufacturers’ protocols. The total RNA from each of the four adult insects was individually extracted using the RNeasy Mini Extraction Kit (Qiagen), in accordance with the manufacturer’s instructions. The integrity and purity of the RNA were assessed using the same methods employed for gDNA quality assessment.

Genome and transcriptome sequencing. The DNA libraries were thoroughly examined and sequenced using the PacBio Sequel II platform at GrandOmics in Wuhan. The raw sequencing data underwent preprocessing with the CCS program, generating 41.00 Gb of Circular Consensus Sequencing (CCS) bases, resulting in a sequencing depth of 65.49×. Additionally, Hi-C library construction was performed using the same male moth for DNA extraction. The DNA was cross-linked *in situ*, extracted, and digested with the restriction enzyme DpnII. The resulting fragments were ligated to form chimeric junctions, purified, and the Hi-C libraries were amplified using PCR. These libraries were then sequenced on the Illumina NovaSeq 6000 platform using a 150 bp paired-end configuration, producing a total of 109.41 Gb of clean reads, representing approximately 174.47× coverage. Regarding the data used for the genome survey, short insertion fragment libraries, ranging in size from 300 to 500 base pairs (bp), were constructed at BGI (Beijing Genomics Institute) in Shenzhen. These libraries were then sequenced on the BGISEQ-500 platform using 150-bp paired-end reads, following standard protocols throughout the process. In total, 60 Gb of sequencing data was obtained. The extracted RNA was used to prepare indexed cDNA libraries with the NEBNext Ultra RNA Library Prep Kit for Illumina. These libraries, with insert sizes of 250–300 bp, were sequenced on the Illumina NovaSeq 6000 platform using a paired-end 150 bp configuration, generating a total of 29.42 Gb of filtered sequencing data (Table 1).

Genome assembly and chromosome anchoring. Before the genome assembly, the genome size was estimated using the Kmerferq v4.0 software⁸. The k-value was set to 17, resulting in an approximate size of 631.91 Mb, with repetitive sequences accounting for 44.16% (Fig. 1). The primary genome assembly was generated using HiFiasm v0.16.19 with PacBio CCS reads in --primary option, supplemented by Hi-C data (--h1/--h2 parameters) for haplotype phasing⁹. The initial assembly comprised 136 contigs, which was subsequently processed through purge_dups v1.2.5 (https://github.com/dfguan/purge_dups) to eliminate residual heterozygous sequences. Potential microbial contamination in the assembled genome was identified by Kraken2 screening against NCBI RefSeq microbial databases, with putative contaminant contigs subsequently validated using BLASTn (e-value < 1e-5). This two-step purification process ensured a contamination-free haploid assembly. The assembly’s quality was assessed using BUSCO v5.7.1¹⁰ with the lepidoptera_odb10 geneset¹¹, which indicated a completeness level of 99.4%. After performing quality control on the Hi-C reads by Trimmomatic-0.39-2¹², Juicer’s generate_site_positions.py was then used to construct possible restriction enzyme site files based on the previously assembled genome and to obtain the length information of each contig. The purified reads were

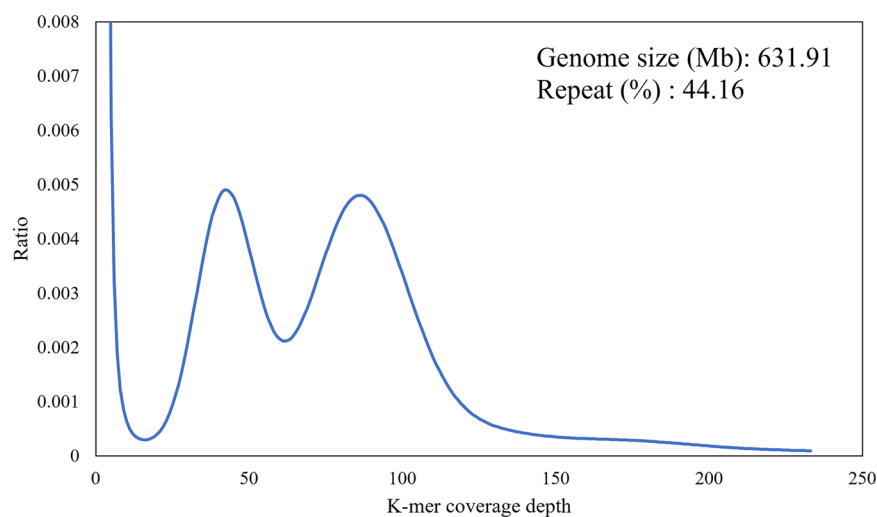


Fig. 1 K-mer analysis of *Hermonassa cecilia*. The estimated genome size is 631.91 Mb, with repetitive sequences accounting for 44.16%.

mapped to the assembled contigs using Juicer¹³ with default parameters. Subsequently, 3D de novo assembly (3D-DNA) v180922¹⁴ pipelines were employed to cluster, sort, and orient the contigs or scaffolds to achieve a chromosome-level genome assembly. Juicebox v1.11.0819¹⁵ was then used for visual inspection and correction of any errors in the order and orientation of the contigs. The final chromosome-level genome of *H. cecilia* was obtained, with a genome size of 626.10 Mb and a scaffold N50 of 21.00 Mb (Fig. 2). The GC content was 37.87%. BUSCO analysis indicated that 99.4% (98.9% single-copy genes and 0.5% duplicated genes) of the 5286 BUSCOs were complete orthologs (Table 2). The completeness of the chromosome-level genome assembly is very close to that of the initial assembly, with one more complete BUSCO and two more single-copy BUSCOs compared to the initial assembly. Moreover, its BUSCO results outperform those of several published genomes of Noctuidae species, including *Agrotis ipsilon*¹⁶, *A. segetum*³, and *Athetis lepigone*¹⁷, among others. These results suggest the assembly of a high-quality chromosome-level genome for *H. cecilia*.

Repeat element annotation. To elucidate the patterns of repetitive elements in the *H. cecilia* genome, we employed homology-based prediction methods, utilizing the established repeat sequence database Repbase. Predictions were performed using RepeatMasker v4.1.5¹⁸ and RepeatProteinMask v4.1.5. Additionally, we constructed a de novo repeat library from the genomic sequence using RepeatModeler v2.0.4¹⁹ with the -LTRStruct option to enable comprehensive LTR retrotransposon annotation, followed by repeat masking and classification using RepeatMasker. Furthermore, the Tandem Repeats Finder v4.09 (TRF)²⁰ program was used to identify tandem repeat sequences in the genome. The results from the four prediction methods were then integrated to obtain the final repeat sequence information. A total of 281.45 Mb of repetitive sequences were identified, accounting for 44.21% of the assembled genome (Fig. 1). Among these, DNA transposons accounted for 3.06%, Long terminal repeats (LTRs) for 1.93%, Long interspersed nuclear elements (LINEs) for 11.56%, and Short interspersed nuclear elements (SINEs) for 6.60% (Table 3).

Protein-coding gene annotation. Protein-coding gene prediction based on the genome with masked repetitive sequences was conducted using three methods: de novo prediction, transcriptome-based prediction, and homology-based prediction. De novo prediction was performed using Augustus v3.3²¹ with the etraining and augustus commands with the default parameters. Transcriptome data from four moths were used for gene prediction. The transcriptome sequencing data were first filtered using Trimmomatic v0.39-2¹², and then aligned to the assembled genome using Tophat v2.1.1²² to identify splice sites. Transcripts were constructed and their abundance estimated using Cufflinks v2.2.1²², and then the four transcripts were merged using Cuffmerge. Homology-based prediction was conducted using Genome-wide electronic tool for annotation Geta v2.6.1's homolog_gene-wise command <https://github.com/chenlianfu/geta/releases/tag/v2.6.1>, based on the publicly available protein sequences of four Lepidoptera noctuid insects: *A. ipsilon*¹⁶, *Xestia c-nigrum*²³, *S. frugiperda*⁵, and *Helicoverpa armigera*²⁴. Finally, the predictions from these three methods were integrated using EVidenceModeler v2.0.0²⁵, resulting in the prediction of 22,662 protein-coding sequences. The predicted gene models showed excellent BUSCO completeness (lepidoptera_odb10; n = 5,286), with 94.6% of conserved genes present (90.9% single-copy, 3.7% duplicated) and only 3.7% missing, demonstrating high annotation quality. Functional annotation of the predicted protein-coding genes was performed using Diamond v2.1.8.162's blastp command with the parameters -evalue 1e-6 -max-target-seqs. 1 -outfmt 6, based on public protein databases such as UniProt_invertebrates, non-redundant protein database (NR), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups (EggNOG)²⁶. A total of 20,221 genes were annotated with functions in at least one database, accounting for 89.29% of the predicted proteins (Fig. 3a,b, Table 4).

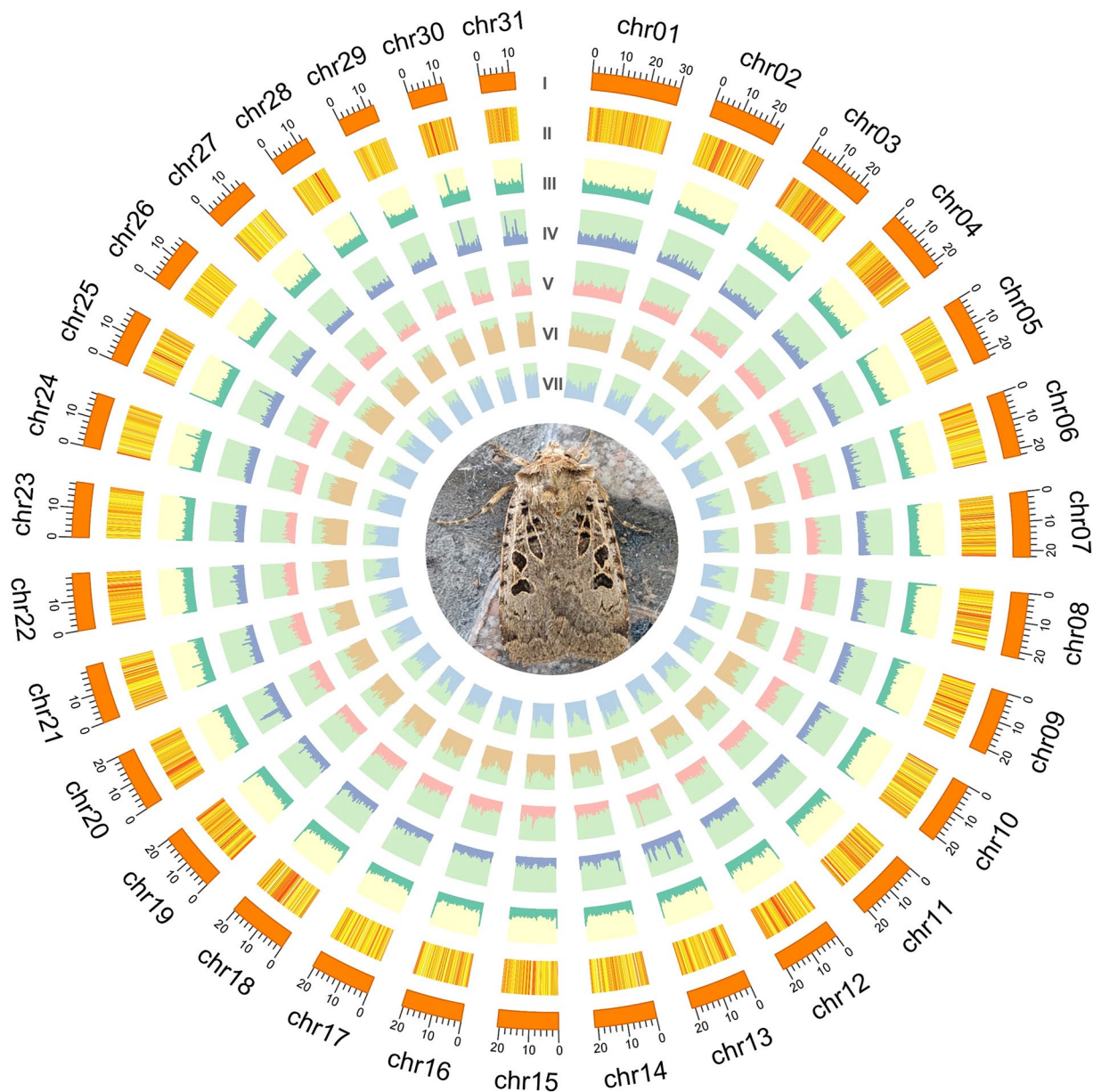


Fig. 2 Circos Diagram of the *Hermonassa cecilia* Genome This diagram offers a comprehensive visualization of the *H. cecilia* genome structure. From the outermost layer inward, the layers represent: I. Distribution of markers across 31 chromosomes, at a megabase scale. II. Gene density in each chromosomal region. III. GC ratio. IV-VII. Distribution of different repeat elements (DNA repeats, Long Terminal Repeats (LTR), Long Interspersed Nuclear Elements (LINE), and Short Interspersed Nuclear Elements (SINE)) across the chromosomes. An illustration of an adult *H. cecilia* at the core highlights the subject of this genomic study. The photo of *H. cecilia* come from <https://www.inaturalist.org/observations/187888582>.

Feature	<i>H. cecilia</i>	<i>A. ipsilon</i>	<i>A. segetum</i>	<i>A. lepigone</i>	<i>S. fruiperda</i>	<i>H. armigera</i>
level	chr	chr	scaffold	chr	chr	chr
genome size	626.10 Mb	515.32 Mb	600.00 Mb	612.49 Mb	390.40 Mb	355.23 Mb
scaffold N50	21.00 Mb	17.56 Mb	25.34 Mb	20.91 Mb	12.97 Mb	12.74 Mb
GC content	37.87%	38.04%	38.07%	37.02%	36.43%	36.57%
single-copy BUSCOs	98.9%	97.5%	97.5%	96.0%	98.3%	96.3%
Complete BUSCOs	99.4%	98.1%	98.6%	97.9%	99.4%	99.6%

Table 2. Assembly features for genomes of *Hermonassa cecilia* and other scale insects.

Types	Number	Length (bp)	Percentage (%)
ARTEFACT	11	4662	0.00073
DNA	57107	19497910	3.06265
LINE	270494	73601286	11.56110
LTR	56182	12311595	1.93387
Low_complexity	11669	539498	0.08474
Other	5	772	0.00012
PLE	1126	194894	0.03061
RC	87804	17916296	2.81423
Retroposon	48	5076	0.00080
SINE	217313	42046206	6.60448
Satellite	1312	198611	0.03120
Simple_repeat	97253	7219159	1.13397
Unknown	599983	107819534	16.93602
Smail RNA	293	99657	0.01596
Total	1400600	281454976	44.21030

Table 3. Statistics for repeat elements in the genome of *Hermonassa cecilia*.

Database	Number of annotated genes	Percentage
Uniprot	19892	87.78%
NR	19859	87.63%
KEGG	11915	52.58%
EggNOG	17414	76.84%
Total	20221	89.23%

Table 4. Functional annotation of *Hermonassa cecilia* proteins.

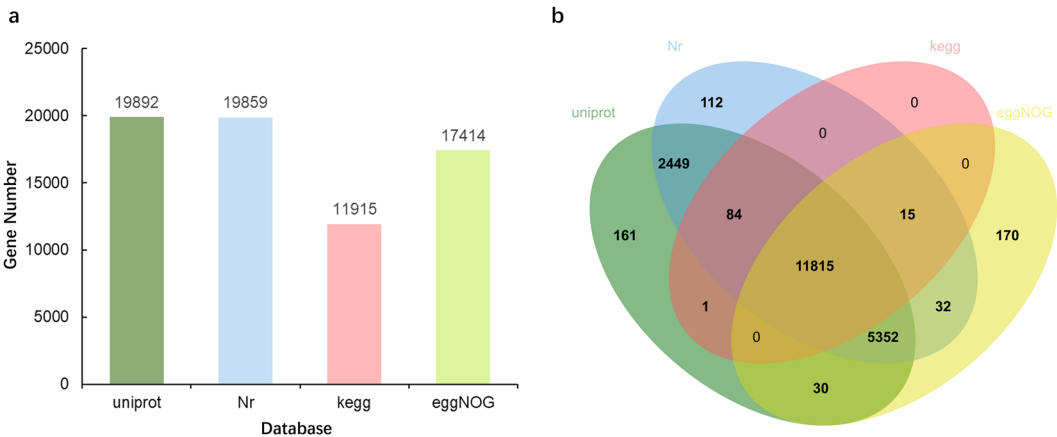


Fig. 3 Functional Annotations of *Hermonassa cecilia* Protein-Coding Genes Using Public Databases. **(a)** Bar chart showing the number of genes annotated by each of the four databases: Uniprot, NR, KEGG, and EggNOG. **(b)** Venn diagram displaying the annotation results from the four databases.

Data Records

The raw data generated by the PacBio sequencing platform has been deposited into the NCBI Sequence Read Archive (SRA) with accession number SRR31742934²⁷. The Hi-C reads are stored in the NCBI SRA with accession number SRR31742933²⁸. The RNA-seq reads are available in the NCBI SRA under accession numbers SRR31744080 to SRR31744083^{29–32}. The BGISEQ raw reads are available in the NCBI SRA under accession number SRR32082304³³. The final genome assembly is deposited in GenBank under accession number PRJNA1198691³⁴. Genome annotations are available in the FigShare repository³⁵.

Technical Validation

After extracting genomic DNA, its integrity was assessed by agarose gel electrophoresis, and DNA concentration was measured using NanoDrop and Qubit 3.0 fluorometer, with an absorbance of approximately 2.0 at 260/280. Low-quality Hi-C and transcriptome raw data were filtered using Trimmomatic-0.39, retaining only high-quality sequencing reads. The assembled genome size is 626.10 Mb, with repetitive sequences accounting

for 44.21%. The size of the assembled genome and the proportion of repetitive elements are both consistent with the estimated results of 631.91 Mb and 44.16%, respectively. The assembled genome size is also similar to the genome sizes of *A. ipsilon*¹⁶ and *Agrotis segetum*³ in the Noctuidae subfamily. The initial HiFiiasm assembly results showed only 136 contigs, indicating that the assembled genome exhibits high continuity and integrity. BUSCO analysis revealed a genome assembly completeness of 99.4%, with a low percentage of duplicated single-copy genes at just 0.5%, indicating that duplication is not a significant issue in the assembly process. This confirms the high quality of the genome assembly comparable to other recently published lepidopteran genomes^{7,36}. These results demonstrate that we have successfully obtained a high-quality genome for *H. cecilia*.

Our integrated annotation approach successfully identified 22,662 protein-coding genes, achieving exceptional genome annotation quality as evidenced by the high BUSCO completeness score (94.6%). The complementary use of de novo prediction, transcriptomic evidence, and homology-based methods ensured comprehensive detection of both evolutionarily conserved and species-specific gene content, with functional annotations assigned to 89.29% of predicted genes. While these results establish a reliable genomic framework, emerging single-nucleus RNA sequencing technologies present exciting opportunities to enhance our understanding of cell-type-specific gene expression patterns and refine current gene models³⁷. The 10.7% of genes lacking functional annotation likely include important lineage-specific adaptations and non-coding elements, highlighting valuable targets for future functional characterization. This high-confidence genome annotation provides an essential foundation for advancing Lepidoptera genomic research and facilitating downstream functional studies.

Code availability

The data analyses were conducted in accordance with the manuals and protocols provided by the developers of the respective bioinformatics tools. All software and codes utilized in this work are publicly available, with the corresponding versions specified in the Methods section. No custom code was used for the curation or validation of the dataset in this study.

Received: 17 February 2025; Accepted: 4 June 2025;

Published online: 17 June 2025

References

1. Beccaloni, G. W., Scoble, M. J., Robinson, G. S. & Pitkin, B. *The Global Lepidoptera Names Index (LepIndex)* (2003).
2. Gao, B., Han, H. L., Kononenko, V. S. & Pan, Z. H. Five new species of the genus *Hermonassa* Walker, 1865 from Xizang Autonomous Region, China (Lepidoptera, Noctuidae, Noctuinae). *Zookeys* **1179**, 35–61 (2023).
3. Wang, P. *et al.* Population genomics of *Agrotis segetum* provide insights into the local adaptive evolution of agricultural pests. *BMC Biol* **22**, 42 (2024).
4. Cheng, T. *et al.* Genomic adaptation to polyphagy and insecticides in a major East Asian noctuid pest. *Nat Ecol Evol* **1**, 1747–1756 (2017).
5. Zhang, L. *et al.* Genetic structure and insecticide resistance characteristics of fall armyworm populations invading China. *Mol Ecol Resour* **20**, 1682–1696 (2020).
6. Peng, Y. *et al.* Population Genomics Provide Insights into the Evolution and Adaptation of the Asia Corn Borer. *Mol Biol Evol* **40**, msad112 (2023).
7. Gao, B. J. *et al.* Chromosome genome assembly and whole genome sequencing of 110 individuals of *Conogethes punctiferalis* (Guenée). *Sci Data* **10**, 805 (2023).
8. Liu, B. H. *et al.* Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. *Quantitative Biology* **35**, 62–67 (2013).
9. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* **18**, 170–175 (2021).
10. Manni, M., Berkeley, M. R., Seppey, M., Simao, F. A. & Zdobnov, E. M. BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol* **38**, 4647–4654 (2021).
11. Zdobnov, E. M. *et al.* OrthoDB in 2020: evolutionary and functional annotations of orthologs. *Nucleic Acids Res* **49**, D389–D393 (2021).
12. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
13. Durand, N. C. *et al.* Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst* **3**, 95–98 (2016).
14. Dudchenko, O. *et al.* De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
15. Robinson, J. T. *et al.* Juicebox.js Provides a Cloud-Based Visualization System for Hi-C Data. *Cell Syst* **6**, 256–258.e1 (2018).
16. Jin, M. H. *et al.* Chromosome-level genome of black cutworm provides novel insights into polyphagy and seasonal migration in insects. *BMC Biol* **21**, 2 (2023).
17. Yesaya, A. *et al.* The chromosomal-scale genome sequencing and assembly of *Aethis lepigone*. *Sci Data* **11**, 338 (2024).
18. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics* **4**, 4.10.1–4.10.14 (2009).
19. Flynn, J. M. *et al.* RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci USA* **117**, 9451–9457 (2020).
20. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**, 573–580 (1999).
21. Nachtweide, S. & Stanke, M. Multi-Genome Annotation with AUGUSTUS. *Methods Mol Biol* **1962**, 139–160 (2019).
22. Ghosh, S. & Chan, C. K. Analysis of RNA-Seq Data Using TopHat and Cufflinks. *Methods Mol Biol* **1374**, 339–361 (2016).
23. Broad, G. R. *et al.* The genome sequence of the setaceous Hebrew character, *Xestia c-nigrum*, (Linnaeus, 1758). *Wellcome Open Res* **7**, 295 (2022).
24. Jin, M. H. *et al.* Adaptive evolution to the natural and anthropogenic environment in a global invasive crop pest, the cotton bollworm. *Innovation* **4**, 100454 (2023).
25. Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol* **9**, 1–22 (2008).

26. Buchfink, B., Reuter, K. & Drost, H. G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* **18**, 366–368 (2021).
27. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR31742934> (2024).
28. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR31742933> (2024).
29. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR31744080> (2024).
30. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR31744081> (2024).
31. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR31744082> (2024).
32. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR31744083> (2024).
33. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR32082304> (2025).
34. NCBI Assembly <https://identifiers.org/ncbi/insdc:JBNGIN000000000> (2025).
35. Figshare <https://doi.org/10.6084/m9.figshare.28040201.v1>.
36. Zhang, L. *et al.* A high-quality chromosome-level genome assembly for the agricultural pest *Mythimna separata*. *Sci Data* **12**, 540, <https://doi.org/10.1038/s41597-025-04855-7> (2025).
37. Sun, C., Shao, Y. & Iqbal, J. A comprehensive cell atlas of fall armyworm (*Spodoptera frugiperda*) larval gut and fat body via snRNA-Seq. *Sci Data* **12**, 250, <https://doi.org/10.1038/s41597-025-04520-z> (2025).

Acknowledgements

This work was supported by the STI 2030–Major Projects (2022ZD04021), the National Natural Science Foundation of China (3237170929), the Chinese Agrosystem Long-Term Observation Network (CALTON-DP) (Y2024JC34), and the Agricultural Science and Technology Innovation Program (ASTIP).

Author contributions

Yutao Xiao and Chao Wu conceived the study. Xinyue Liang prepared the samples for genome sequencing, HiC sequencing, and RNA sequencing. Chao Wu and Chengyue Xian performed the bioinformatics analysis. The manuscript was written by Chao Wu, proofread by Lei Zhang, and finalized by Yutao Xiao, Jingang Liang, Yao Tan and Zhiqiang Zhang. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Z.Z., Y.T., J.L. or Y.X.

Reprints and permissions information is available at 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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

© The Author(s) 2025