scientific data

Check for updates

OPEN

DATA DESCRIPTOR

A chromosome-level genome assembly for the mulberry thrips *Pseudodendrothrips mori* (Thysanoptera: Thripidae)

De-Long Guan, Yue-Mei Li, Shi-Hao Zhang, Yue Qin, Nan Wei, Lei Xin & Xiao-Dong Li ✉

The mulberry thrips, *Pseudodendrothrips mori* Niwa, is a significant agricultural pest whose genomic study has been hindered by a lack of reference data. Here, we present the first high-quality, chromosome-level genome assembly for this pest. The genome was assembled using 10.24 Gb of PacBio HiFi long reads, polished with 16.87 Gb of Illumina short reads, and scaffolded using 17.61 Gb of Hi-C data. The final assembly spans 280.93 Mb, with 98.65% of the sequence anchored into 19 pseudochromosomes. The assembly exhibits exceptional contiguity (contig N50: 1.51 Mb; scaffold N50: 14.52 Mb) and high base-level accuracy (QV: 42.82). We identified 18.32% (51.47 Mb) of the genome as repetitive elements and predicted 13,429 protein-coding genes, of which 98.68% were functionally annotated by NR and Interproscan. The quality of this assembly is further validated by a high BUSCO score (98.0% complete genes from the insecta_odb12 dataset). This genomic resource provides a critical foundation for research into pest genetics, insecticide resistance, and the evolution of Thysanoptera.

Background & Summary

Pseudodendrothrips mori Niwa (Thysanoptera: Thripidae) represents a significant agricultural pest that causes substantial economic losses through direct feeding damage on various host plants[1]. Despite its considerable impact on crop production, genomic resources for this species have remained remarkably scarce, primarily due to technical challenges associated with its minute body size[1–3]. The requirement for pooling multiple individuals to obtain sufficient nucleic acid quantities for sequencing has historically hindered molecular studies, resulting in an extreme paucity of genetic research and related literature for this economically important pest species[2,4].

Thrips species within the family Thripidae exhibit remarkable evolutionary adaptations that enable them to exploit diverse ecological niches and develop rapid resistance to pesticides[5,6]. The genus *Pseudodendrothrips* comprises specialized phytophagous insects with piercing-sucking mouthparts that cause distinctive silvering and distortion of plant tissues[1,7]. Their cryptic behavior, small size (typically 1–2 mm in length), and high reproductive potential make them particularly challenging to manage using conventional pest control strategies[4,7]. Understanding the genomic basis of these adaptations is crucial for developing sustainable management approaches and elucidating the evolutionary mechanisms underlying their success as agricultural pests.

The lack of genomic resources for *P. mori* has particularly hindered investigations into several critical biological questions. These include the genetic basis of host plant specialization, mechanisms of rapid insecticide resistance evolution, and molecular pathways underlying their remarkable reproductive plasticity[1,2,7,8]. Furthermore, the absence of reference genomes limits our ability to conduct population genomic studies essential for understanding pest dispersal patterns, invasion dynamics, and the development of molecular markers for pest monitoring programs.

Here, we present the first high-quality chromosome-level genome assembly for *P. mori*, generated through an integrated approach combining multiple sequencing technologies. Our assembly utilized 10.24 Gb of PacBio HiFi long reads, 16.87 Gb of Illumina short reads, 17.61 Gb of Hi-C chromatin interaction data, and 3.77 Gb of

Guangxi Key Laboratory of Sericulture Ecology and Intelligent Technology Application, Guangxi Collaborative Innovation Center of Modern Sericulture and Silk, Guangxi Colleges Universities Key Laboratory of Exploitation and Utilization of Microbial and Botanical Resources, School of Chemistry and Bioengineering, Hechi University, Hechi, 546300, China. ✉e-mail: lxdong_627@163.com

RNA sequencing data. This comprehensive dataset enabled the construction of a highly contiguous genome assembly spanning 280.93 Mb across 1,155 contigs, with a contig N50 of 1.51 Mb. Through Hi-C scaffolding, we successfully anchored 277.13 Mb (98.65%) of the assembly to 19 pseudochromosomes, achieving an exceptional scaffold N50 of 14.52 Mb. The assembly demonstrates remarkable completeness, with BUSCO analysis revealing 98.0% complete gene representation against the insecta_odb12 dataset of 3,114 conserved orthologous genes. This level of completeness substantially exceeds many previously published thrips genomes[5,6,8,9] and provides confidence in the comprehensive capture of the *P. mori* gene space. The high contiguity and completeness metrics of our assembly establish it as a valuable reference for comparative genomics within Thysanoptera and broader arthropod evolutionary studies.

Methods

Sample collection and sequencing. Adult individuals of *P. mori* were collected from a mulberry (*Morus alba*) garden in Liu Sanjie Town, Yizhou, Hechi City, Guangxi, China. Species identification was confirmed based on morphological characteristics by a field worker Shi-Hao Zhang. Due to the minute body size of this species, which poses challenges for extracting sufficient quantities of high-quality nucleic acids from a single individual, we employed a sample pooling strategy for all sequencing procedures. All collected specimens were flash-frozen in liquid nitrogen and stored at −80 °C until processing.

For PacBio HiFi sequencing, high-molecular-weight genomic DNA was extracted from a pool of approximately 200 adult thrips using a magnetic bead-based method. The extracted DNA was used to construct a SMRTbell library with an average fragment size of 20 kb. This library was sequenced on a PacBio Sequel IIe platform, generating 10.24 Gb of high-fidelity (HiFi) long-read data, representing approximately 36× coverage of the estimated genome. For short-read genomic sequencing, DNA from the same pooled sample was used to prepare a paired-end library with a 350 bp insert size. The library was sequenced on an Illumina NovaSeq 6000 platform, yielding 16.87 Gb of clean short-read data (approximately 60× coverage), which were used for subsequent genome polishing and heterozygosity assessment. To enable chromosome-level scaffolding, a Hi-C library was constructed from a separate pool of approximately 200 adult thrips. After proximity ligation and reverse cross-linking, the Hi-C library was sequenced on an Illumina platform, producing 17.61 Gb of valid interaction data (approximately 63 × coverage). For transcriptome sequencing to aid in gene annotation, total RNA was extracted from another pooled sample of 200 adult thrips using TRIzol reagent (Invitrogen). An mRNA-enriched library was then prepared and sequenced on an Illumina platform, generating 3.77 Gb of paired-end RNA-seq reads (Table S1).

All raw sequencing data were filtered to remove low-quality reads and adapters, yielding high-quality clean data for subsequent genome assembly and annotation. Voucher specimens were deposited at the Guangxi Key Laboratory of Sericulture Ecology and Applied Intelligent Technology, Hechi University, Hechi, China, under the accession number PMORI-GXHC01.

Genome assembly. A high-quality chromosome-level genome assembly for *P. mori* was generated using an integrated strategy combining PacBio HiFi long reads, Illumina short reads, and Hi-C chromatin interaction data. The initial *de novo* assembly was performed using 10.24 Gb of filtered PacBio HiFi reads with NextDenovo v2.5.2[10], which produced a primary contig assembly totaling 280.93 Mb in length across 1,155 contigs, achieving a contig N50 of 1.51 Mb. To enhance base-level accuracy, this preliminary assembly was subjected to two rounds of polishing using 16.87 Gb of high-quality Illumina paired-end reads with NextPolish v1.4.1[11]. For chromosome-level scaffolding, the 17.61 Gb of valid Hi-C reads were mapped to the polished contigs, and the HapHiC v1.0.5[12] pipeline was employed to order, orient, and anchor the contigs into larger scaffolds. Final manual review and correction of the scaffolding were conducted using Juicebox Assembly Tools v1.11.08[13] to resolve any misjoins and finalize the pseudochromosome structures. This comprehensive workflow resulted in 277.13 Mb (98.65%) of the assembled sequence being anchored into 19 chromosome-level scaffolds, with a final scaffold N50 of 14.52 Mb. The Hi-C interaction map is shown in Fig. 1 (Fig. 1), where the strong intra-chromosomal interactions clearly validate the scaffolding quality. The completeness of the functionally important gene space was assessed using BUSCO v6.0.0[14] against the Insecta lineage dataset (insecta_odb12, n = 3,114). The results demonstrated an exceptionally high level of completeness, with 98.0% of the expected single-copy insect genes identified as complete (97.0% single-copy and 1.0% duplicated). Only 0.4% were found to be fragmented, and a mere 1.5% were missing, confirming that the gene-rich regions of the genome are comprehensively represented.

The 19 assembled pseudochromosomes ranged in size from 10.87 Mb (PSMO_Chr19) to 23.55 Mb (PSMO_Chr01). To identify potential sex chromosomes, we analyzed the sequencing coverage of HiFi reads across all 19 pseudochromosomes using mosdepth v0.3.3[15]. The mean coverage depth across the autosomes was relatively consistent, fluctuating between approximately 18.4× and 23.2×. Unlike typical heterogametic systems (e.g., XY or X0) where a sex chromosome exhibits roughly half the coverage of autosomes, no chromosome in our assembly showed such a significant drop in coverage. The relatively uniform coverage across all chromosomes suggests that the pooled sample used for sequencing may have consisted primarily of the homogametic sex (e.g., diploid females, which is common in haplodiploid systems prevalent in Thysanoptera)[16,17], thus precluding the identification of sex chromosomes based on coverage depth alone.

The final genome assembly of *P. mori* spans a total length of 280.93 Mb (Fig. 2), a size comparable to that of *Megalurothrips usitatus*[9] (247.62 Mb) but smaller than other reported thrips genomes such as *Odontothrips loti*[5] (331.07 Mb), *Dendrothrips minowai*[6] (350.11 Mb), and *Thrips tabaci*[18] (325.07 Mb). The overall GC content of the *P. mori* genome is 51.42%, which falls within the typical range observed in other Thysanoptera species. Our assembly demonstrates exceptional long-range contiguity, characteristic of a high-quality, chromosome-level reference genome. We successfully anchored the assembly into 19 pseudochromosomes, a number similar to *D. minowai*[6] (19) and *T. tabaci*[18] (18). The resulting scaffold N50 is 14.52 Mb, a value that is highly comparable to

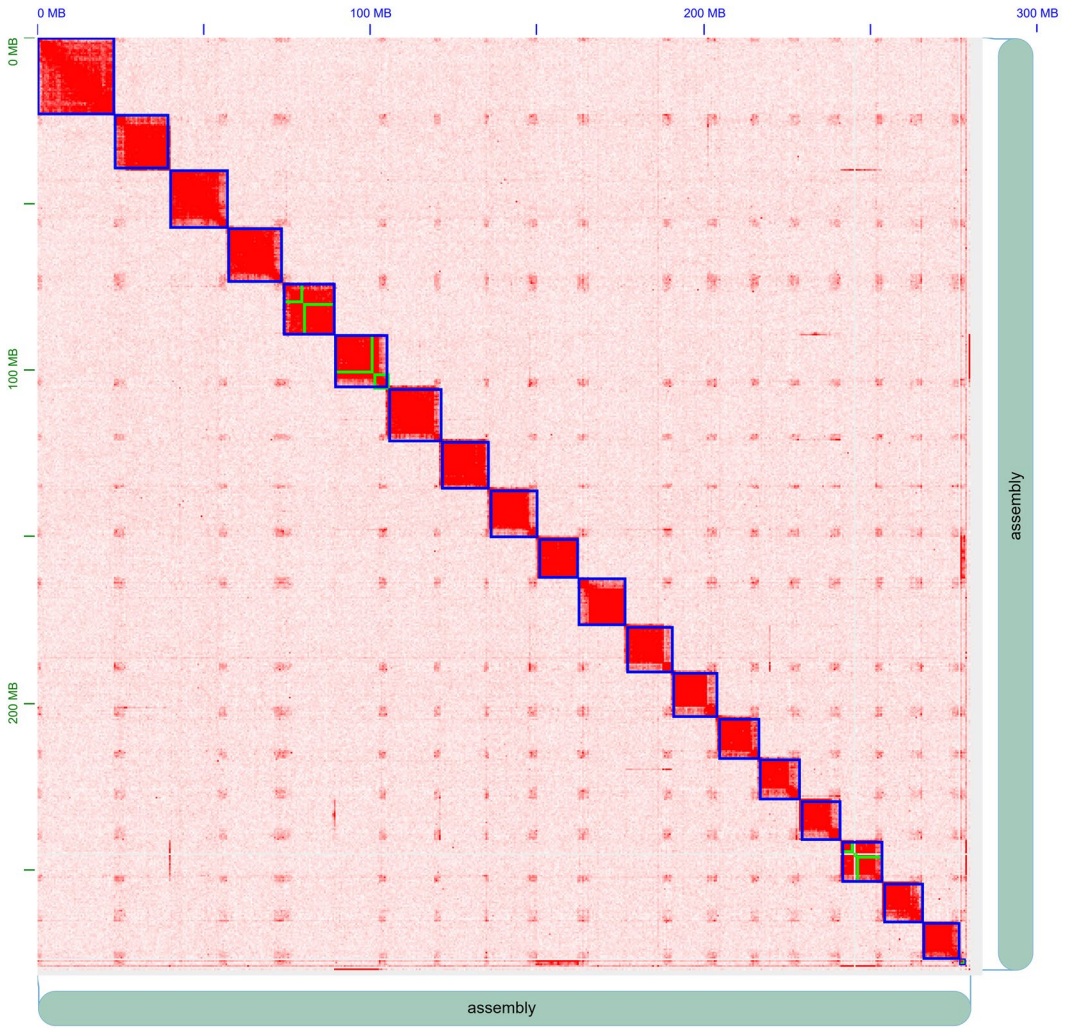

Fig. 1 Chromosome-level scaffolding of the *Pseudodendrothrips mori* genome validated by Hi-C data. The heatmap illustrates the chromatin interaction frequencies across the entire genome assembly. The 19 pseudochromosomes are concatenated and represented on both the x- and y-axes.

the scaffold N50 values of other chromosome-level thrips assemblies, including *O. loti* (18.59 Mb), *M. usitatus* (15.49 Mb), *D. minowai* (16.86 Mb), and *T. tabaci* (16.56 Mb). This confirms the high quality of the scaffolding achieved through the use of Hi-C data. At the contig level, the assembly consists of 1,155 contigs with a contig N50 of 1.51 Mb. While the number of contigs is higher than in some other thrips genomes, suggesting a degree of genomic complexity or repetitive content, the contig N50 value is robust and significantly surpasses that of *D. minowai* (0.54 Mb) and is on par with *T. tabaci* (1.58 Mb). The combination of a solid contig N50 and an excellent scaffold N50 underscores the reliability of this genomic resource for detailed comparative genomics and fine-scale gene mapping (Table 1).

Genome annotation. Comprehensive identification of repetitive elements in the *P. mori* genome was conducted using an integrated approach combining both homology-based and *de novo* methodologies. First, a *de novo* repeat library was constructed using RepeatModeler v2.0.5[19]. This custom library was then combined with known insecta repeats from the Repbase database v21.12. Finally, RepeatMasker v4.1.5[20] was used with default parameters to systematically identify, classify, and mask repetitive elements throughout the genome assembly based on this combined library.

This integrated approach revealed that repetitive elements comprise a total of 51.47 Mb, accounting for 18.32% of the *P. mori* genome assembly (Table 2). The identified repeats were broadly categorized into interspersed repeats (12.04% of the genome) and non-interspersed repeats (6.50%) (Fig. 2C). Among the non-interspersed class, simple repeats were the most dominant component, making up 5.84% of the genome. Interspersed repeats, while more abundant overall, were largely composed of elements that could not be classified into known families ("Unknown" category), which alone constituted 11.19% of the entire genome, suggesting the presence of novel or highly diverged lineage-specific repeat families in *P. mori*. Among the classified interspersed repeats, retroelements (Class I) were more prevalent (0.62% of the genome) than DNA transposons

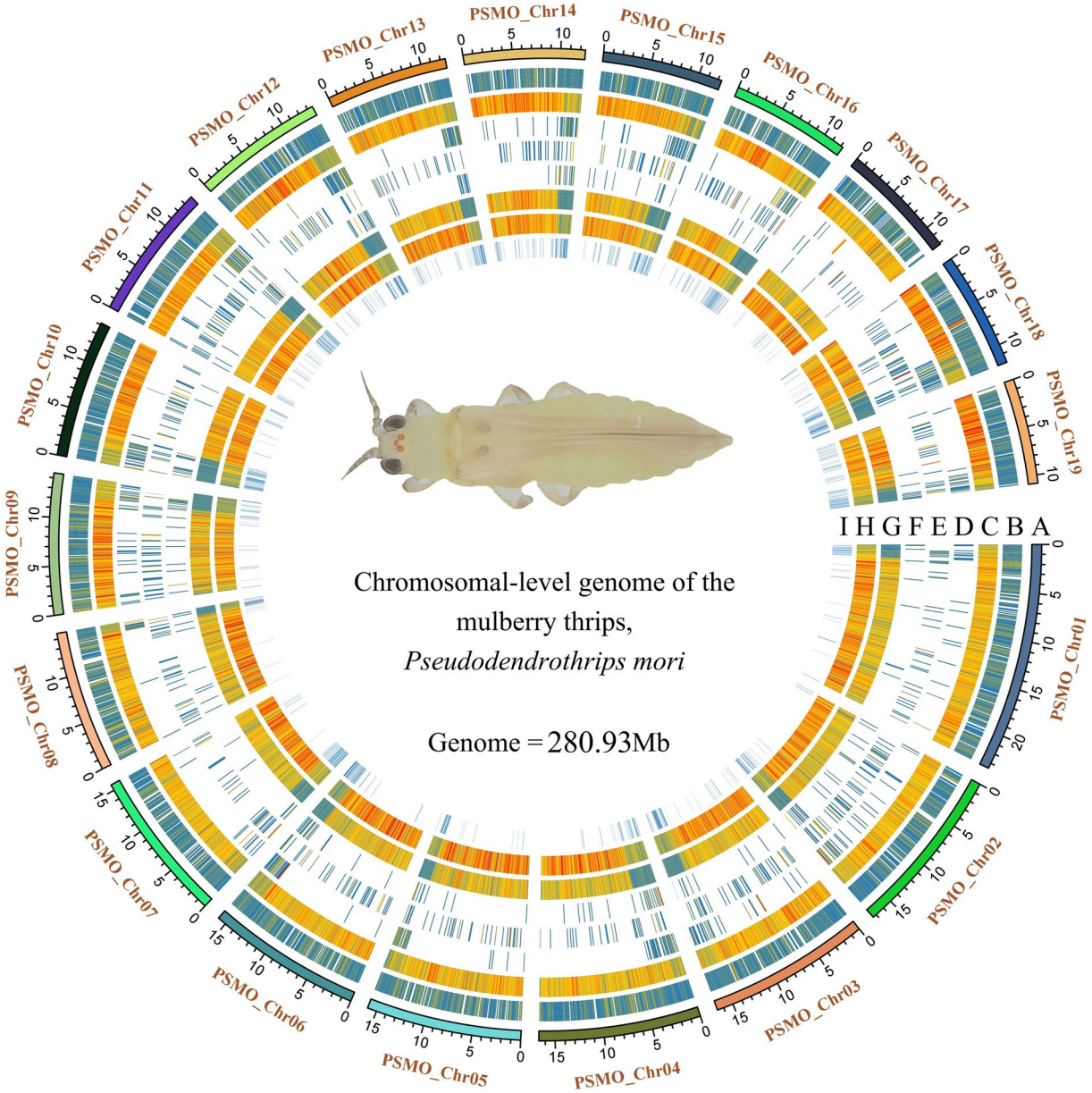

Fig. 2 Circular visualization of genomic features of the *P. mori* genome. The tracks are arranged from outermost to innermost. An image of an adult mulberry thrips is depicted in the center. The tracks represent: (A) The 19 pseudochromosomes (PSMO_Chr01 to PSMO_Chr19), with scale marks in megabases (Mb). (B) Density of protein-coding genes, where taller bars indicate higher density. (C) Density of all repetitive elements. (D) Density of LTR retrotransposons. (E) Density of LINE transposons. (F) Density of DNA transposons. (G) Density of Simple Repeats (H) Density of GC content. (I) Location of gaps among contigs. Panel B to I was calculated in non-overlapping 100 kb windows, with orange/red indicating higher GC content and blue indicating lower content.

(Class II, 0.20%). Within the retroelements, LTR elements (e.g., Gypsy and Copia) were the major subclass, occupying 0.49% of the genome, whereas LINE elements constituted a smaller fraction at 0.12% (Fig. 2D–G).

Following the masking of repetitive elements, a comprehensive strategy was employed to annotate protein-coding genes within the *P. mori* genome. This strategy integrated three complementary evidence sources: homology-based prediction, transcriptome alignments, and ab initio modeling. For homology-based prediction, protein sequences from the closely related species *D. minowai* were aligned to the assembly using Exonerate v2.4.0 (parameters: --model protein2genome --percent 50) and GenomeThreader (GTH) v1.7.1[21] (parameters: -gff3out -skipalignmentout -paralogs -prseedlength 20 -prhdist 2). Transcriptome-based evidence was generated by aligning the RNA-seq reads to the genome using HISAT2 v2.2.1[22] (parameters: --dta --max-intronlen 20000), which provided crucial experimental support for defining exon-intron boundaries. Concurrently, ab initio gene prediction was performed using AUGUSTUS v3.4.0[23] (parameters: --species = fly --gff3 = on --protein = on --codingseq = on), which was trained on high-confidence gene models derived from

Parameters	*Odontothrips loti*	*Megalurothrips usitatus*	*Dendrothrips minowai*	*Thrips tabaci*	*Pseudodendrothrips mori*
Genome size (bp)	331,074,510	247,619,268	350,114,598	325,073,080	280,930,787
Chromosomes sequence No.	16	16	19	18	19
GC content	52	55.5	46	53.5	51
Scaffold sequence No.	152	25	236	42	31
Scaffold L50	8	7	10	9	9
Scaffold N50 (bp)	18,593,632	15,492,676	16,860,160	16,564,241	14,517,657
Scaffold N90 (bp)	14,868,425	12,700,200	12,818,119	13,411,671	12,145,867
Contig sequence No.	300	168	1,269	316	1,155
Contig L50	14	9	158	61	45
Contig N50 (bp)	7,191,130	12,100,000	536,341	1,581,468	1,514,686

Table 1. Comparison of genome assembly statistics between *Pseudodendrothrips mori* and four other representative thrips.

Repeat Class	Number of Elements	Length (bp)	Percentage of Genome (%)
Total Repetitive Elements	573271	51474958	18.32
Non-interspersed	417834	18270430	6.5
Simple repeat	380537	16409466	5.84
Low complexity	37296	1894584	0.67
Satellite	1	484	0
Interspersed	155437	33840033	12.04
Retroelements (Class I)	1834	1741941	0.62
LTR elements	1095	1386939	0.49
Gypsy	336	875275	0.31
Copia	77	178809	0.06
ERVK	3	11714	0
LINE elements	699	352112	0.12
L2	534	249375	0.08
I-Jockey	107	53567	0.01
Penelope	54	36223	0.01
SINE elements	40	2890	0
DNA transposons (Class II)	376	563015	0.2
Maverick	163	277173	0.09
PiggyBac	8	85990	0.03
Others	205	200000	0.08
Rolling-circles	27	85800	0.03
Helitron	27	85800	0.03
Unknown	153155	31456172	11.19

Table 2. Classification and composition of repetitive elements in the *Pseudodendrothrips mori* genome.

the transcriptome data to enhance predictive accuracy. Finally, the evidence from all three approaches was synthesized and weighted using the GETA (Gene Eukaryotic Annotation, https://github.com/chenlianfu/geta) pipeline to generate a final, non-redundant consensus gene set.

This integrated approach resulted in the identification of 13,429 protein-coding genes (Fig. 2B). The predicted genes had a mean length of 10.56 kb and contained an average of 8.25 exons per gene, with a mean coding sequence (CDS) length of 1,854 bp. Functional annotation of the predicted gene models was conducted through sequence similarity searches against the NCBI non-redundant protein (Nr) database and protein domain identification using InterProScan v5.72-103.0[24]. This comprehensive functional annotation process successfully assigned putative functions to 13,252 genes, representing an exceptionally high annotation rate of 98.68% of the total predicted gene set. Individually, 12,988 genes were annotated based on Nr homology, and 12,315 genes were annotated via InterProScan.

To assess the completeness of the gene annotation, we performed a BUSCO[14] analysis (v6.0.0) on the predicted protein set against the Insecta lineage dataset (insecta_odb12, n = 3,114). The analysis revealed an exceptionally high level of completeness, with 97.0% complete BUSCOs (96.1% single-copy and 1.0% duplicated), 0.9% fragmented, and only 2.1% missing. This high level of completeness in the annotated gene set is highly congruent with the genome-level BUSCO assessment (98.0%), indicating that our annotation pipeline successfully captured the vast majority of the expected gene content.

Data Records

The final chromosome-level genome assembly of *P. mori* has been deposited in the National Center for Biotechnology Information (NCBI) GenBank database under the Genome accession number GCA_054486245.1[25]. All raw sequencing data generated in this study have been made publicly available through the NCBI Sequence Read Archive (SRA) under the study accession number SRP623193[26]. Specifically, the PacBio HiFi long-read sequencing data used for primary contig assembly are accessible under run accession SRR35528008. The Illumina short-read whole-genome sequencing data used for genome polishing are available under accession SRR35528007. The Hi-C sequencing data utilized for chromosome-level scaffolding can be accessed under accession SRR35528005. Additionally, the RNA-sequencing data used for transcriptome-guided gene annotation have been deposited under accession SRR35528006. Genome annotation files for Pseudodendrothrips mori are available in Figshare under https://doi.org/10.6084/m9.figshare.30256126.v2[27].

Technical Validation

The quality, completeness, and accuracy of the final *P. mori* genome assembly were rigorously evaluated through a multi-faceted approach, incorporating read-mapping concordance, reference-free k-mer analysis, and gene space completeness assessment. To validate the structural integrity and completeness, the raw sequencing reads were mapped back to the final assembly. The 16.87 Gb of Illumina short reads (NGS) were aligned using BWA-MEM v0.7.17[28], achieving a mapping rate of 91.06%. The PacBio HiFi long reads (TGS) were aligned using minimap2 v2.24[29], resulting in a mapping rate of 72.71%. For transcriptome validation, the RNA-seq reads were aligned to the genome using HISAT2 v2.2.1[22], yielding a high mapping rate of 66.20%. These consistent mapping rates across different data types confirm that the source sequencing data is well-represented in the final assembly.

The k-mer completeness of the assembly were assessed using merqury[30]. The analysis yielded a Phred-scaled quality value (QV) of 42.82, indicating exceptionally high base accuracy (less than one error per 10,000 bases). Then, we run CRAQ v1.10[31] to examine the genome errors. The results demonstrated a genome coverage rate of 99.39% with an extremely low rate of low-confidence regions (5.33×10^{-6}). The high reference-free quality scores (R-AQI: 52.24; S-AQI: 69.88) provide additional strong support for the high quality of the diploid assembly at both the read and scaffold levels. To assess the completeness of the functionally important gene space, we performed a Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis using BUSCO v6.0.0[14] against the Insecta lineage dataset (insecta_odb12, n = 3,114). The assembly was found to contain 98.0% of the expected core orthologous genes, with 97.0% identified as complete and single-copy, 1.0% as complete and duplicated, 0.4% as fragmented, and only 1.5% as missing. This high level of completeness in the gene space representation is critical for downstream functional genomics. Collectively, the high read-mapping rates, exceptional base-level accuracy, near-complete k-mer representation and comprehensive gene content validate the high quality, accuracy, and completeness of the *P. mori* reference genome.

Data availability

The final chromosome-level genome assembly of *Pseudodendrothrips mori* has been deposited in the National Center for Biotechnology Information (NCBI) GenBank database under the Genome accession number GCA_054486245.1[25]. All raw sequencing data generated in this study are publicly available through the NCBI Sequence Read Archive (SRA) under the study accession number SRP623193[26]. Genome annotation files, including repeat annotations (GFF), gene models (GFF3), coding sequences (CDS), and protein sequences (FASTA), are available in Figshare[27].

Code availability

No custom code was generated for this study. All analyses were performed using publicly available software with default or specified parameters as detailed in the Methods section. The specific software versions and key parameters are documented throughout the Methods to ensure reproducibility.

Received: 7 October 2025; Accepted: 22 January 2026;
Published online: 02 February 2026

References

1. Rasool, I., Mound, L. A., Soliman, A. M. & Aldhafer, H. M. The Dendrothripinae (Thysanoptera, Thripidae) of Saudi Arabia, with new records, one new species, and revised status of Pseudodendrothrips stuardoi (Moulton). *Zootaxa* **4999**, 219–227, https://doi.org/10.11646/zootaxa.4999.3.2 (2021).
2. Masumoto, M. & Okajima, S. Studies on Dendrothripinae (Thysanoptera, Thripidae) from Japan, with new records and one new species. *Zootaxa* **4362**, 405–420, https://doi.org/10.11646/zootaxa.4362.3.5 (2017).
3. Rebagliati, P. J. & Mola, L. M. Meiotic behavior and karyotypic variation in Acledra (Pentatomidae, Heteroptera). *Genetics and molecular research: GMR* **9**, 739–749, https://doi.org/10.4238/vol9-2gmr763 (2010).
4. Rachana, R. R., Amarendra, B. & Gracy, R. G. Studies on Pseudodendrothrips (Thysanoptera, Thripidae) from India, with one new species. *Zootaxa* **5209**, 365–372, https://doi.org/10.11646/zootaxa.5209.3.5 (2022).
5. Yingning, L. *et al.* Chromosome-level genome assembly of Odontothrips loti Haliday (Thysanoptera: Thripidae). *Scientific data* **11**, 451, https://doi.org/10.1038/s41597-024-03289-x (2024).
6. Xiu, C. *et al.* Chromosome-level genome assembly of Dendrothrips minowai and genomic analysis highlights distinct adaptations to high polyphenols in tea plants. *Pest management science* **81**, 4143–4153, https://doi.org/10.1002/ps.8781 (2025).
7. Mound, L. A. & Tree, D. J. Genera of the leaf-feeding Dendrothripinae of the world (Thysanoptera, Thripidae), with new species from Australia and Sulawesi, Indonesia. *Zootaxa* **4109**, 569–582, https://doi.org/10.11646/zootaxa.4109.5.5 (2016).
8. Zhang, Z. *et al.* Chromosome-level genome assembly of the flower thrips Frankliniella intonsa. *Scientific data* **10**, 844, https://doi.org/10.1038/s41597-023-02770-3 (2023).

9. Zhang, Z. *et al*. The Chromosome-Level Genome Assembly of Bean Blossom Thrips (Megalurothrips usitatus) Reveals an Expansion of Protein Digestion-Related Genes in Adaption to High-Protein Host Plants. *International journal of molecular sciences* **24**, https://doi.org/10.3390/ijms241411268 (2023).
10. Hu, J. *et al*. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. *Genome biology* **25**, 107, https://doi.org/10.1186/s13059-024-03252-4 (2024).
11. Hu, J., Fan, J., Sun, Z. & Liu, S. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics (Oxford, England)* **36**, 2253–2255, https://doi.org/10.1093/bioinformatics/btz891 (2020).
12. Zeng, X. *et al*. Chromosome-level scaffolding of haplotype-resolved assemblies using Hi-C data without reference genomes. *Nature Plants* **10**, 1184–1200, https://doi.org/10.1038/s41477-024-01755-3 (2024).
13. Durand, N. C. *et al*. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Systems* **3**, 99–101, https://doi.org/10.1016/j.cels.2015.07.012 (2016).
14. Seppey, M., Manni, M. & Zdobnov, E. M. BUSCO: Assessing Genome Assembly and Annotation Completeness. *Methods in molecular biology (Clifton, N.J.)* **1962**, 227–245, https://doi.org/10.1007/978-1-4939-9173-0_14 (2019).
15. Pedersen, B. S. & Quinlan, A. R. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics (Oxford, England)* **34**, 867–868, https://doi.org/10.1093/bioinformatics/btx699 (2018).
16. Zhang, Q. *et al*. Reproductive disruption induced by Beauveria bassiana in Frankliniella occidentalis as a potential long-term biocontrol strategy. *Pesticide biochemistry and physiology* **212**, 106464, https://doi.org/10.1016/j.pestbp.2025.106464 (2025).
17. Ghali, K., Ramella, E., Henry, M., Schwander, T. & van der Kooi, C. J. Ecological specialization, clonal diversity, and local adaptation explain the co-existence of sexual and asexual grass thrips. *Journal of evolutionary biology* **38**, 1016–1022, https://doi.org/10.1093/jeb/voaf076 (2025).
18. Gao, Y. *et al*. Chromosome-level genome assembly of cotton thrips Thrips tabaci (Thysanoptera: Thripidae). *Scientific data* **11**, 1003, https://doi.org/10.1038/s41597-024-03737-8 (2024).
19. Flynn, J. M. *et al*. RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 9451–9457, https://doi.org/10.1073/pnas.1921046117 (2020).
20. Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Current protocols in bioinformatics* Chapter 4, Unit 4.10, https://doi.org/10.1002/0471250953.bi0410s05 (2004).
21. Gremme, G., Brendel, V., Sparks, M. E. & Kurtz, S. Engineering a software tool for gene structure prediction in higher organisms. *Information and Software Technology* **47**, 965–978, https://doi.org/10.1016/j.infsof.2005.09.005 (2005).
22. Kim, D., Landmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* **12**, 357–U121, https://doi.org/10.1038/nmeth.3317 (2015).
23. Stanke, M. *et al*. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic acids research* **34**, W435–439, https://doi.org/10.1093/nar/gkl200 (2006).
24. Jones, P. *et al*. InterProScan 5: genome-scale protein function classification. *Bioinformatics (Oxford, England)* **30**, 1236–1240, https://doi.org/10.1093/bioinformatics/btu031 (2014).
25. Guan, D.-L. *Genbank* https://identifiers.org/ncbi/insdc.gca:GCA_054486245.1 (2025).
26. Guan, D.-L. *NCBI* https://identifiers.org/ncbi/insdc.sra:SRP623193 (2025).
27. Guan, D.-L. Genome annotation files for Pseudodendrothrips mori. *Figshare* https://doi.org/10.6084/m9.figshare.30256126.v2 (2025).
28. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* **25**, 1754–1760, https://doi.org/10.1093/bioinformatics/btp324 (2009).
29. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics (Oxford, England)* **34**, 3094–3100, https://doi.org/10.1093/bioinformatics/bty191 (2018).
30. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome biology* **21**, https://doi.org/10.1186/s13059-020-02134-9 (2020).
31. Li, K., Xu, P., Wang, J., Yi, X. & Jiao, Y. Identification of errors in draft genome assemblies at single-nucleotide resolution for quality assessment and improvement. *Nature communications* **14**, 6556, https://doi.org/10.1038/s41467-023-42336-w (2023).

Acknowledgements

This research was funded by the Special Project of Guangxi Collaborative Innovation Center of Modern Sericulture and Silk (2023GXCSSC04), Hechi University high-level talent research start-up fee project (2023GCC017). The project is supported by the Local Science and Technology Development Fund project guided by the central government (Heke ZY230301) and 2023 Annual Operation Subsidy Project of the Guangxi Key Laboratory of Sericulture Ecology and Applied Intelligent Technology (23-026-08).

Author contributions

X.-D.L. conceived and designed the project, supervised the study, and acquired funding. D.-L.G. and X.-D.L. performed the computational analyses, including genome assembly, annotation, and data curation. Y.-M.L. and S.-H.Z. collected and identified the samples. Y.Q. and N.W. conducted the laboratory experiments, including DNA and RNA extraction and library preparation. L.X. assisted with bioinformatics analysis and data visualization. D.-L.G. drafted the original manuscript. X.-D.L. and D.-L.G. reviewed and revised the manuscript. All authors read, approved, and contributed to 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/s41597-026-06697-3.

Correspondence and requests for materials should be addressed to X.-D.L.