Scientific Data

https://doi.org/10.1038/s41597-025-06348-z

Article in Press

Chromosome-level genome assembly of *Monochamus sutor* in China

Received: 4 August 2025

Accepted: 18 November 2025

Cite this article as: Yu, D., Li, M., Ren, J.-R. *et al.* Chromosome-level genome assembly of *Monochamus sutor* in China. *Sci Data* (2025). https://doi.org/10.1038/s41597-025-06348-z

Dian Yu, Meng Li, Jia-Ru Ren, Chuan-Qin Zhang & Jing Tao

Chromosome-level genome assembly of *Monochamus sutor* in China

Dian Yu[1], Meng Li[2], Jia-Ru Ren[2], Chuan-Qin Zhang[2], Jing Tao✉

Beijing Key Laboratory for Forest Pest Control, Beijing Forestry University, Beijing, 100083, China

Dian Yu, Meng Li, Jia-Ru Ren, Chuan-Qin Zhang& Jing Tao

Corresponding Author：Jing Tao (taojing1029@hotmail.com)

Abstract

Monochamus sutor (Coleoptera: Cerambycidae) is a pest that has led to significant economic and ecological losses. Herein, we created a high-quality chromosome-level reference genome for *M. sutor* by integrating MGI, Nanopore, and Hi-C sequencing technologies. We successfully assembled the scaffolds into eleven chromosomes with a total size of 762.16 MB, featuring a scaffold N50 of 84.3 Mb. In total, 381.54 Mb (50.06%) of repetitive elements were identified, and 14,404 protein-coding genes were predicted, with 83.08% of them annotated. The genome of *M. sutor* is an invaluable resource that provides data support for biological, ecological, genetic, and other research. It also offers a new basis to explore the functions of key gene families. More importantly, this genome enables data-driven risk prediction of pine wilt disease outbreaks, helping to address the vector role of *M. sutor* and support scientific surveillance strategies to reduce forest losses.

Background &Summary

The pinewood nematode *Bursaphelenchus xylophilus* is recognized as one of the most destructive quarantine pests globally, posing a severe threat to forest ecosystems worldwide. Since its introduction into China in 1982, it has caused damage to varying degrees in Jiangsu, Zhejiang, Anhui, and Jiangxi[1]. The occurrence of pine wilt disease has continued to increase in the past 30 years that has caused significant economic losses to pine forest resources in China and even global forest resources[2].

Monochamus sutor (Coleoptera: Cerambycidae) belongs to the genus *Monochamus*, according to the report of Zhang Jianjun, *M. sutor* is also a vector insect of pinewood nematode[3]. However, there is no report in China that it can transmit *B. xylophilus*, it still poses a potential risk of

transmitting pine wilt disease[11, 12]. Studies based on a well-assembled genome on the ecology and genetics of *M. sutor* are of critical significance for the control of this pest. Such a genome resource is also essential to fill the gap in risk prediction research, as it can support the exploration of genetic markers associated with vector competence.

In this research, we employed a combination of Nanopore, MGI short-read sequencing, and chromosome conformation capture Hi-C technologies to assemble a chromosome-level genome of *M. sutor*. This study provides comprehensive genomic resources to advance investigations into the ecological adaptations, genetic diversity, and evolutionary dynamics of *M. sutor*, as well as its symbiotic interactions with the pinewood nematode. Furthermore, this genome offers a key tool for developing risk prediction frameworks for pine wilt disease in China, enabling more targeted and effective control measures to mitigate economic and ecological impacts.

Methods

Sample preparation.

Samples of *M. sutor* were collected from Monochamus genus traps in the Dajianshan Forest Farm, Qinshui County, Jincheng City, Shanxi Province, China. After being captured, these samples were fed with *Pinus bungeana* for approximately two weeks[13], followed by a 48-hour starvation treatment. A single female adult was used to construct libraries of MGI short read, Oxford Nanopore Technology (ONT) long read sequencing, and Hi-C. Additionally, larvae, pupae and adults were collected through the placement of bait logs in the Jingxi Forest Farm, Beijing, China, followed by the splitting wood segments in March of the subsequent year. Subsequently, larvae, pupae, and adults of *M. sutor* were subjected to transcriptome sequencing. All samples were immediately frozen in liquid nitrogen and stored at −80 °C before being delivered to a company for sequencing[14].

Table 1. Library sequencing data and methods used in this study to assemble the *Monochamus sutor* genome.

Sequencing strategy	Sequence Read Archive (SRA)	Platform	Usage	Clean data (Gb)	Coverage (X)
Short-read	SRR34440010	MGI	Genome survey	46.84	62
Long-read	SRR34440009	Nanopore	Genome assembly	101.86	134
Hi-C	SRR34440008	MGI	Hi-C assembly	78.69	103

RNA-seq of larvae	SRR34335301	MGI	Anno-evidence	6.71	/
RNA-seq of pupae	SRR34335300	MGI	Anno-evidence	6.469	/
RNA-seq of adult	SRR34335299	MGI	Anno-evidence	7.228	/

Genomic DNA and RNA sequencing

For short-read sequencing, genomic DNA was extracted using CTAB methods. DNA purification was performed using the GrandOmics Genomic Purification Kit (GrandOmics, China) in accordance with the standard operating procedures (SOPs) provided by the manufacturer. Libraries were constructed by MGIEasy Universal DNA Library Prep Kit V1.0（CAT#1000005250, MGI）following the standard protocol. Then the DNA templates and libraries were transferred to the MGI T7 sequencing platform for high-throughput sequencing. We obtained 46.84 Gb short reads (Table 1).

For long-read sequencing, high molecular weight genomic DNA was isolated using the QIAGEN Genomic-tip extraction kit according to the standard operating procedure provided by the manufacturer[15]. Following successful sample quality control, the SQK-LSK114 library preparation kit was employed for library construction. Following library preparation, the DNA library, containing a defined concentration and volume of DNA fragments and enzyme complexes, was loaded onto a Nanopore PromethION P48 flow cell for sequencing, and 101.86 Gb long reads were generated[16, 17](Table 1).

For Hi-C sequencing, genomic DNA was isolated from tissues or cells using a magnetic bead-based purification system. Whole-genome sequencing (WGS) libraries were subsequently constructed and sequenced on the MGI T7 high-throughput sequencing platform. A total of 78.69 Gb (103 × coverage) of data was generated (Table 1).

For transcriptome sequencing, total RNA was extracted from a single *M. sutor* (larva, pupa, and adult, respectively) using the the TRIzol/CTAB method. The poly-A RNAs were enriched from total RNA using Dynabeads mRNA Purification Kit (Cat#61006, Invitrogen) and fragmented into small pieces using fragmentation reagent in MGIEasy RNA Library Prep Kit V3.1 (Cat# 1000005276, MGI). Library was analyzed on the Agilent Technologies 2100 bioanalyzer. The double stranded PCR products were heat denatured and circularized by the splint oligo sequence in MGIEasy Circularization Module (CAT#1000005260, MGI). The qualified libraries were sequenced on MGI-T7 platform. Transcriptome data was obtained (Table 1).

Sequence quality assessment

For short-read sequencing, Hi-C sequencing, the reliability of these data are evaluated using the base quality distribution plot, base content distribution plot, and GC content distribution plot.

In the base quality distribution plot, the abscissa represents the position of a base within its respective read, and the ordinate denotes the sequencing quality score. The blue line indicates the average quality score of all bases at that position. All average base quality scores fall within the green high-quality region, suggesting that the sequencing data are of high quality.

In the base content distribution plot, the blue, orange, green, and red lines represent the four bases A, T, G, and C respectively, with their values corresponding to the proportion of each base at different positions. Theoretically, due to the complementary pairing of A-T and G-C, as well as the randomness of library construction, the proportions of A-T and G-C in the sequencing data should be nearly consistent and remain relatively stable across different positions. However, due to the inherent characteristics of next-generation sequencing, the base proportions in the first 10 bp may fluctuate. The contents of A-T and G-C are basically consistent and remain relatively stable, it indicates that the sequencing data quality is acceptable.

In the GC content distribution plot, the horizontal axis represents the average GC content, and the vertical axis represents the proportion of reads corresponding to different GC contents. Theoretically, the curve shape should approximate a normal distribution. Deviations from this pattern often result from library contamination or pooled library construction with heterogeneous samples, leading to biases in subsets of reads. As shown in the plot, the GC content distribution presents a single normal distribution curve, indicating that the sequencing data quality is acceptable.

For long-reads sequencing, the abscissa represents the length of reads, the ordinate denotes the average quality score of reads, and the color indicates the number of reads. The abscissa represents the length of reads, and the ordinate represents the number of read.

Genome assembly

A draft genome at contig level was assembled using NextDenovo version 2.5.2 (https://github.com/Nextomics/NextDenovo) with parameters (genome-size = 686, read-cutoff = 1k)based on Nanopore long reads[18]. Alternative haplotypes and redundant fragments in the contig assembly were eliminated using Purge_dups[19] version 1.2.5 with default parameters (minimum matching bases for chaining = 500, maximum gap size for chaining = 20K, maximum gap size for 2nd round chaining = 50K, minimum chaining score for a match = 10K, maximum extension for contig ends = 15K). To correct base errors introduced during the assembly of Nanopore long reads, the contig-level genome assembly was polished using NextPolish[20], with the parameters of NextPolish set as follows: round = 2, threads = 64. Hi-C analysis was conducted to further anchor the assembly into chromosome-scale linkage groups. The Hi-C clean reads were processed with Fastp and mapped to the contigs via BWA[21, 22]. Chromosome-scale assembly and manual curation were performed using YaHS version 1.2.2[23]. The results showed that 95.8% of the contigs were anchored to 11 chromosomes, which was visualized in the heatmap of the chromatin contact matrix (Figure 1), the heatmap display 11 distinct "square-shaped signal clusters", with each cluster corresponding to one chromosome. The interaction intensity between chromatin regions within a cluster is significantly higher than that between clusters, which proves that the contigs have been correctly assembled onto 11 chromosomes based on the actual spatial interaction relationships of the chromosomes. Then, a circos plot of genomic features was generated using TBtools, which includes collinear blocks, GC content, gene density, and N ratio (Figure 2). Finally, genome polishing was performed using NextPolish based on ONT and MGI sequencing reads. The resulting chromosome-level genome had a size of 762.16 Mb, an N50 of 84.3 Mb, leading to the generation of a contact map that was visualized using Juicebox[24]. Busco version 5.1.2 showed that there are 94.9% single-copy genes, 0.8% multi-copy genes, and 1.1% fragmented genes (Table 3)[25,26]. Subsequently, we plotted the synteny map within this species. The results showed that chromosomes 1 through 11 each have genes that exhibit a certain degree of synteny with the remaining 10 chromosomes (Figure 3). Prove that the order and structure among the fragments after genome assembly conform to the genetic laws of the organism itself. Finally, we constructed a collinearity map between *M. sutor* and *Monochamus alternatus*[27] (Figure 4). The map exhibits continuous and intact collinear blocks with few breakpoints, and a high degree of collinearity with

M. alternatus, indicating that the genome assembly quality is high. The high collinearity with *M. alternatus* reflects that *M. sutor* and *M. alternatus* have a relatively close evolutionary relationship. The genomes of the two species have retained a large number of conserved gene arrangements and sequences during the long-term evolutionary process. By utilizing this collinearity relationship, we can refer to the genes with known functions in *M. alternatus* to quickly locate and mine genes with similar functions in the genome of *M. sutor*, accelerating the research process of genes related to its growth and development, stress resistance, pathogenicity, and providing theoretical support for its application fields such as biological control and resource utilization.

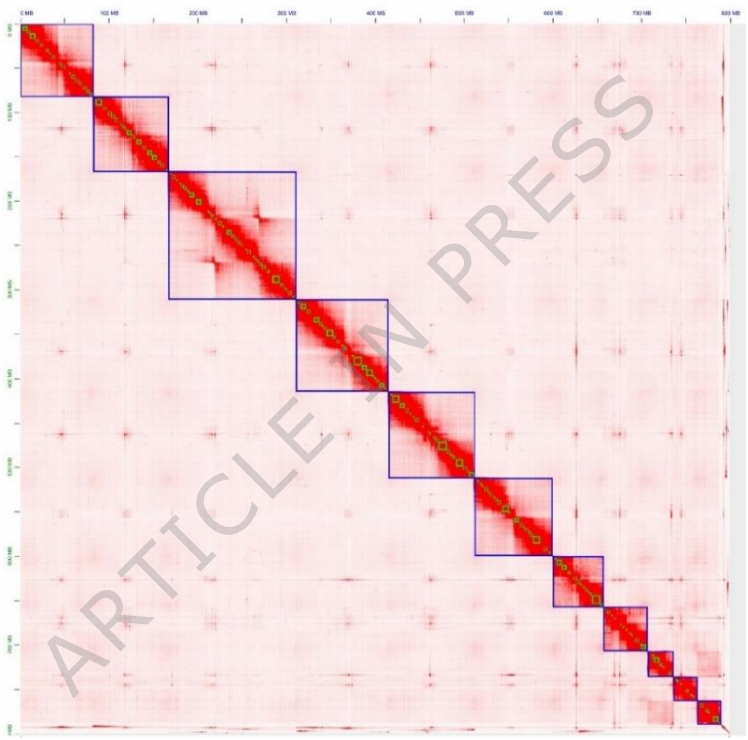

Figure 1 The genome-wide Hi-C interaction map of 11 chromosomal linkage groups in *Monochamus sutor*. Each black square represents a pseudo chromosome. The color bar indicates the interaction intensity of Hi-C contacts.

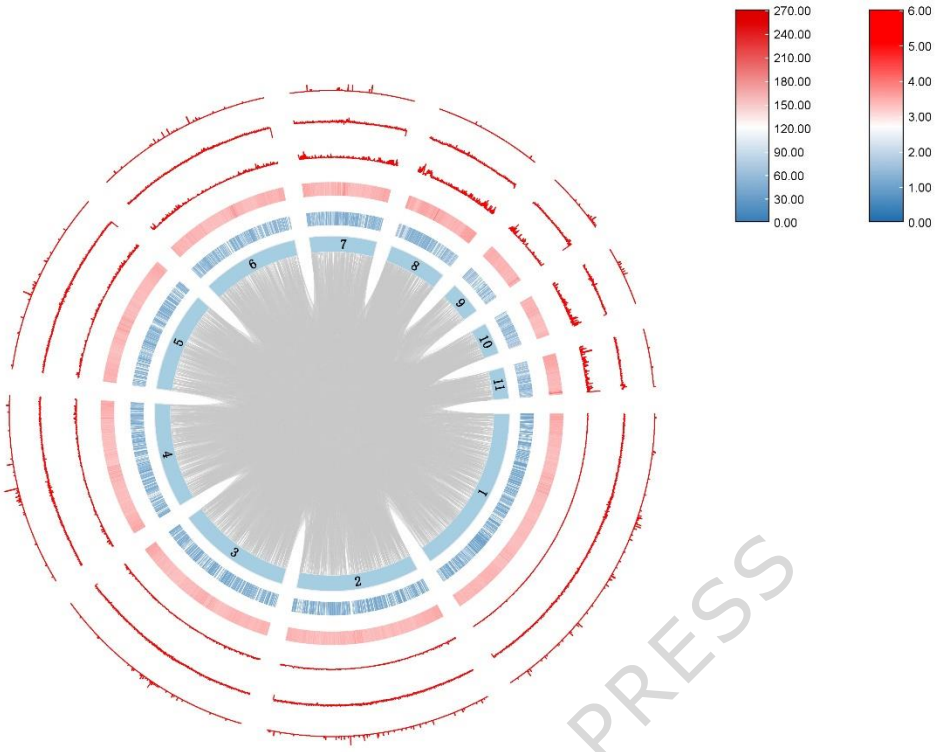

Figure 2 Circos plot of genome features. From the inside out, the order is as follows: collinear blocks (shown in gray), chromosome lengths labeled with chromosome numbers, gene density displayed in the form of a heatmap, GC content displayed in the form of a heatmap, and finally gene density, GC content, and N ratio each displayed in line format.

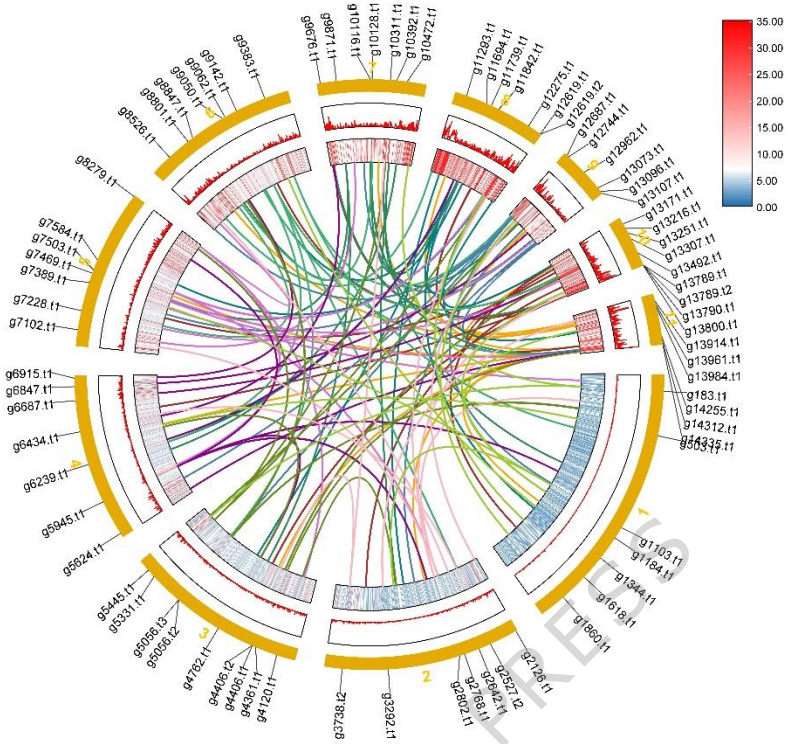

Figure 3 Intraspecific collinearity in *Monochamus sutor*. Seven genes were extracted from each chromosome, and it was found that they had collinearity with genes on other chromosomes.

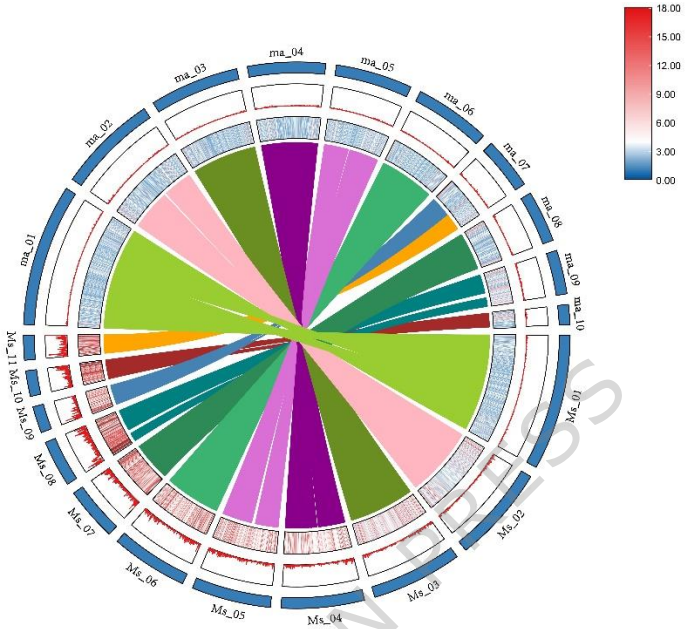

Figure 4 Interspecific collinearity between *Monochamus sutor* and *Monochamus alternatus* (The chromosome numbering of *Monochamus sutor* is designated as Ms_X, and that of *Monochamus alternatus* as ma_X.). The outer circle represents the chromosome numbers of the two species. Inside the circle, the same color connects the collinear block areas between the chromosomes of the two species: the larger the area and the smaller the gaps, the higher the collinearity.

Genome annotation

In this study, HiTE[28] version 3.1.2 was employed to construct a species-specific repetitive sequence library, and based on this library, a systematic annotation of transposable elements in the genome was performed. The parameters set for HiTE were as follows: threads = 40, domain = 1, and annotate = 1. The specific workflow is as follows: The structure prediction model and homology alignment algorithm built into HiTE were used to identify potential repetitive units in the genome. Candidate repetitive sequences were clustered by CD-HIT (with a similarity threshold of ⩾80% and a coverage threshold of ⩾80%) to merge redundant sequences and generate a non-redundant species-specific repetitive sequence library. Genome sequences were then aligned with the custom library and public databases using RepeatMasker, GeneMark-ETP and Prothint to annotate the locations and types of repetitive sequences[29, 30].

The results showed that 50.06% of the genome sequence was annotated as repetitive elements, with a total length of 400,073,441 bp. Short interspersed nuclear elements (SINEs), long interspersed nuclear elements (LINEs), long terminal repeats (LTRs), and DNA elements accounted for 0.00%, 2.08%, 26.73%, and 1.3% of the genome, respectively, while 19.95% of the repetitive sequences remained unclassified (Table 2). In *M. sutor*, SINE elements exhibited the highest proportion among known repeat types, whereas DNA elements were the least abundant.

Table 2. Annotation of repeat sequeces and small RNAs of *Monochamus sutor*.

	Number of elements	Length (bp)	Percentage (%)
Retroelements	700184	230244148	28.81
SINE	207	15056	0.00
LINE	74324	16644517	2.08
LTR	625653	213584575	26.73
DNA transposons	63572	10372672	1.30
Unclassified	555979	159456416	19.95
Total	/	/	50.06
Small RNA	0	0	0.00
Satellite	111	13277	0.00
Simple repeats	191	25508	0.00

The genome was annotated using Braker3[31] with parameters (threads=48, addUTR=on), and functionally annotated using eggNOG-mapper[32](http://eggnog-mapper.embl.de/). Then, Tbtools software was used to visualize the obtained annotation files[33]. A total of 14,404 protein-coding genes were predicted by integrating transcriptome data from larvae and adults. Among them, 7,564 genes had GO annotations, and 4,403 genes were mapped to KEGG pathways.

The Hi-C heatmap showed relatively independent Hi-C signals among the 11 pseudo-chromosomes (Figure.2). The BUSCO completeness at both the contig and chromosome levels reached over 95% (Table 3). For the genome assembly, the N50 value was 84.3 Mb, the length of the largest contig was 144,765,653 bp, and the gap number was 334, suggesting that the genome assembly is highly complete and suitable for subsequent studies on further gene function and evolution.

Table 3. BUSCO evaluation of genome assemblies and annotation.

Description	Contig-level		Chromosome-level		Genome annotation	
	Number	Percent (%)	Number	Percent (%)	Number	Percent (%)
Complete gene (C)	2033	95.7	2034	95.8	2026	95.4
Complete and single-copy gene (S)	2015	94.9	2016	94.9	2004	94.4
Complete and duplicated gene (D)	18	0.8	18	0.8	222	1.0
Fragmented gene (F)	23	1.1	24	1.1	8	0.4
Missing gene (M)	68	3.2	66	3.1	90	4.2
Total gene search	2124	/	2124	/	2124	/

Data Records

The genome project was deposited at NCBI under BioProject number PRJNA1284596. MGI second-generation sequencing data were deposited in the Sequence Read Archive at NCBI under accession number SRR34440010[34]. Hi-C sequencing data were deposited in the Sequence Read Archive at NCBI under accession number SRR34440008[35]. Nanopore sequencing raw data were deposited in the Sequence Read Archive at NCBI under accession number SRR34440009[36]. RNA-seq data were deposited in the Sequence Read Archive at NCBI under accession numbers SRR34335299- SRR34335301[37-40]. The final chromosome assembly, genome structure annotation, amino acid sequences and functional annotation results of protein-coding genes were deposited to Figshare repository under a DOI number of https://doi.org/10.6084/m9.figshare.c.7921310[41]. The

final chromosome assembly was deposited in GenBank under processing number JBPYXQ000000000[42].

Technical Validation

The Hi-C heatmap can reflect the accuracy of genome assembly (Figure.2). Relatively independent Hi-C signals are observed among the 11 pseudo-chromosomes. The Hi-C interaction heatmap shows that the assembly has low background noise, with no obvious fragmentation of cut fragments or signal conflicts.

To assess the completeness of genome assembly and genome annotation, we performed an analysis using BUSCO (version 5.1.2) based on the endopterygota_odb10 database, which contains 2124 conserved genes. The results showed that in the contig-level assembly, the first round of BUSCO analysis revealed that among 2033 single-copy genes, 95.7% were complete genes (with single-copy genes accounting for 94.9%), 1.1% were fragmented genes, and 3.2% were missing genes. For the chromosome-level assembly, BUSCO analysis indicated that among 2034 genes, 95.8% were complete genes (including 94.9% single-copy genes and 0.8% duplicated genes), 1.1% were fragmented genes, and 3.1% were missing genes. Regarding genome annotation, BUSCO analysis demonstrated a completeness of 95.4%, with single-copy genes accounting for 94.4%, fragmented genes for 0.4%, and missing genes for 4.2% (Table 3).

Data Availability

All data used in this study were generated through our own assembly. All relevant datasets have been uploaded to NCBI and Figshare and these datasets are publicly available. MGI second-generation sequencing data, Hi-C sequencing data, Nanopore sequencing raw data, and RNA-seq data were deposited in the Sequence Read Archive at NCBI (https://www.ncbi.nlm.nih.gov/sra/SRP597287). The final chromosome assembly, genome structure annotation, amino acid sequences and functional annotation results of protein-coding genes were deposited to Figshare repository under a DOI number of https://doi.org/10.6084/m9.figshare.c.7921310. The final chromosome assembly was deposited in GenBank (https://identifiers.org/insdc.gca:GCA_052757275.1).

Code Availability

There were no custom scripts or code utilized in this study.

Reference

1. Huang Yulin. Present Situation and Control Measures of Pine Wood Nematode. *World Journal of Forestry.* **11**, 193-199(2022).
2. YANG Ding, QIN Siyuan, CHI Shikuan, SUN Dongxing, FENG Zhihui, GUAN Zhenghao, YU Zhihao, LIU Zhixin, LI Xiaodong, LIU Chao& BAI Hongyan. Research Status and Hotspot Analysis of Pine Wood Nematode Disease Based on CiteSpace. *Natural Protected Areas*， **2**, 115−128(2022). DOI: 10.12335/2096-8981.2021112901
3. Zhang Jianjun, Zhang Runzhi& Chen Jingyuan. Species and their dispersal ability of Monochamus as vectors to transmit Bursaphelenchus xylophilus. *Journal of Zhejiang Forestry College*, **3**, 350-356(2007).
4. KOBAYASF, YAMANEA, IKEDAT.The Japanese pine sawyer beetle as the vector of pinewilt disease. *J. Ann Rev Entomol*, **29** ,115-135 (1984).
5. LINITMJ, KONDOE, SMITHMT.Insects associated with the pine wood nematode, *Bursaphelenchus xylophilus* (Nematoda: Aphelenchoididae), in Missouri. *J. Environ Entomol*, **12**, 457-470 (1983).
6. USDA Forest Service 1991. Pest Risk Assessment of the Importation of Larch from Siberia and the Soviet Far East. USDA Forest Service, Miscellaneous Publication No. 1495, September 1991.
7. Bakke, Alf& Torstein Kvamme. The pine sawyer (*Monochamus sutor*): distribution and life history in South Norway https://api.semanticscholar.org/CorpusID: 133813053 (1992).
8. Q.-H. Zhang, Dr. J. A. Byers& X.-D. Zhang. Influence of bark thickness, trunk diameter and height on reproduction of the longhorned beetle, *Monochamus sutor* (Col., Cerambycidae) in burned larch and pine. *Journal of Applied Entomology*. **115**, 145-154 (1993).
9. Schroeder, L.M., Weslien, J., Lindelöw, Å., & Lindhe, A. Attacks by bark- and wood-boring Coleoptera on mechanically created high stumps of Norway spruce in the two years following cutting. *Forest Ecology and Management*, **123**, 21-30 (1999).
10. Cesari, M., Marescalchi, O., Francardi, V.& Mantovani, B. Taxonomy and phylogeny of European Monochamus species: first molecular and karyological data. *Journal of Zoological Systematics and Evolutionary Research*, **43**, 1-7(2005).
11. SChroeder L. & Magnusson C., Transmission of Bursaphelenchus mucronatus (Nematoda) to

branches and bolts of Pinus sylvestris and Picea abies by the cerambycid beetle Monochamus sutor. Scandinavian Journal of Forest Research, **7**, 107-112(1992).

12. Akbulut S., Yuksel B., serin M., baYsal I. & erdeM M., Pathogenicity of Bursaphelenchus mucronatus in pine seedlings under greenhouse conditions. Turkish Journal of Agriculture Forestry, **31**, 169-173(2007).
13. Pajares, J.A., Álvarez, G., Hall, D.R., Douglas, P., Centeno, F., Ibarra, N., Schroeder, M., Teale, S.A., Wang, Z., Yan, S., Millar, J.G.& Hanks, L.M. 2-(Undecyloxy)-ethanol is a major component of the male-produced aggregation pheromone of Monochamus sutor. *Entomol Exp Appl*, **149**, 118-127(2013).
14. Andreasson A, Kiss NB, Juhlin CC, Höög A. Long-term storage of endocrine tissues at - 80°C does not adversely affect RNA quality or overall histomorphology. *Biopreserv Biobank*. **11**, 366-70(2013).
15. Zerpa-Catanho D., Zhang X., Song J., Hernandez A.G., Ming R. Ultra-long DNA molecule isolation from plant nuclei for ultra-long read genome sequencing. *STAR Protoc* **2**, 100343 (2021).
16. Wang Y., Zhao Y., Bollas A., Wang Y., Au K.F. Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol* **39**:1348-1365 (2021)
17. MacKenzie M., Argyropoulos C. An Introduction to Nanopore Sequencing: Past, Present, and Future Considerations. *Micromachines (Basel)* **14**, 459 (2023)
18. Hu, J., Wang, Z., Sun, Z. et al. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. *Genome Biol* **25**, 107 (2024).
19. Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*. **36**, 2896-2898(2020).
20. Hu, J., Fan, J. P., Sun, Z. Y. & Liu, S. L. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* **36**, 2253–2255 (2020).
21. Chen, S. F., Zhou, Y. Q., Chen, Y. R. & Gu, J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
22. De Sena B.G., Smith A.D. Falco: high-speed FastQC emulation for quality control of sequencing data. *F1000Research* **8**, 1874 (2021).
23. Zhou, C. X., McCarthy, S. A. & Durbin, R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* **39**, btac808 (2023).

24. Durand, N. C. et al. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst* **3**, 99–101(2016).
25. Seppey M, Manni M, Zdobnov EM. BUSCO: Assessing Genome Assembly and Annotation Completeness. Methods Mol Biol. 1962, 227-245(2019).
26. Manni, M., Berkeley, M. R., Seppey, M., Simão, 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. *Molecular Biology and Evolution* **38**, 4647–4654 (2021).
27. Gao, YF., Yang, FY., Song, W. et al. Chromosome-level genome assembly of the Japanese sawyer beetle Monochamus alternatus. *Sci Data* 11, 199 (2024).
28. Hu, K., Ni, P., Xu, M. *et al.* HiTE: a fast and accurate dynamic boundary adjustment approach for full-length transposable element detection and annotation. *Nat Commun* **15**, 5573 (2024).
29. Tarailo-Graovac, M. & Chen, N. S. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Protoc. *Bioinform*. **5**, 4.10.11–14.10.14 (2009).
30. Bruna T, Lomsadze A, Borodovsky M. A new gene finding tool GeneMark-ETP significantly improves the accuracy of automatic annotation of large eukaryotic genomes. *Genome Res.* **34**, 757-768(2024).
31. Gabriel L, Brůna T, Hoff KJ, Ebel M, Lomsadze A, Borodovsky M, Stanke M. BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Res*. **4**, 769-777(2024).
32. Huerta-Cepas, J. et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol*. **34**, 2115–2122 (2017).
33. Chengjie Chen, Hao Chen, Yi Zhang, Hannah R. Thomas, Margaret H. Frank, Yehua He, Rui Xia, TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data, *Molecular Plant*, **13**, 1194-1202 (2020).
34. *NCBI Sequence Read Archive*. https://identifiers.org/ncbi/insdc.sra:SRR34440010 (2025).
35. *NCBI Sequence Read Archive*. https://identifiers.org/ncbi/insdc.sra:SRR34440008 (2025).
36. *NCBI Sequence Read Archive*. https://identifiers.org/ncbi/insdc.sra:SRR34440009 (2025).
37. *NCBI Sequence Read Archive*. https://identifiers.org/ncbi/insdc.sra:SRR34335299 (2025).
38. *NCBI Sequence Read Archive*. https://identifiers.org/ncbi/insdc.sra:SRR34335300 (2025).
39. *NCBI Sequence Read Archive*. https://identifiers.org/ncbi/insdc.sra:SRR34335301 (2025).

40. *NCBI Sequence Read Archive*. https://identifiers.org/ncbi/insdc.sra:SRP597287 (2025).
41. Yu, dian. Chromosome-level genome assembly of potential vectors for pine wood nematode *Monochamus sutor*. *figshare. Collection.* https://doi.org/10.6084/m9.figshare.c.7921310(2025).
42. Yu, D., Tao, J. & Cao, L. J. *Monochamus sutor* Genome sequencing and assembly. *Genbank* https://identifiers.org/insdc.gca:GCA_052757275.1 (2025).

Author contributions

J.T. designed the study. D.Y., J.R.R.and C.Q.Z. contribute to the samples; D.Y. and M.L. established a platform for genome assembly. D.Y. contribute to the genome assembly and annotation. D.Y. wrote the draft manuscript. J.T. contributed to the revisions. The final manuscript has been read and approved by all authors.

Competing Interests

The authors declare no competing interests.

Acknowledgements

This research was supported by the Fundamental Research Funds for the Central Universities (QNTD202510).