scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Chromosome-level genome assembly of Marco Polo Blister Beetle (*Hycleus marcipoli*)

Wenhui Zhu[1], Delong Guan[2], Jiawei Wang[1], Linlin Nie[1], Binbin Yao[1] & Sheng-Quan Xu[1]✉

Hycleus marcipoli **is an agricultural pest that feeds on the flowers and leaves of leguminous plants, including *Desmodium* spp., as well as sweet potatoes. It exhibits hypermetamorphic development—an exceptionally complex life cycle shared by the genus and subfamily. Despite the availability of fragmented genomes of *Hycleus*, a high-quality, chromosome-level genome reference is not yet available for this diverse genus. To address this gap, we present the first chromosome-level genome assembly for *H. marcipoli*. The 111 Mb genome (scaffold N50: 10.22 Mb) was successfully anchored to 11 chromosomes, with repetitive elements accounting for 26.43% of the assembly. We annotated 13,357 protein-coding genes, achieving 98.80% BUSCO completeness. This high-quality genomic resource establishes a foundation for elucidating the olfactory and visual system modifications associated with metamorphic transitions, understanding the evolutionary drivers of rapid species diversification in *Hycleus*, and exploring novel targets for pest control.**

Background & Summary

Blister beetles (Coleoptera: Meloidae) are a highly diverse family within Polyphaga, with a wide global distribution[1,2]. Adult blister beetles pose a major threat to crops such as alfalfa, wheat, legumes, and nightshades, and can contaminate harvested forage, posing serious toxicity risks to horses and livestock[3,4]. Although chemical insecticides remain the primary method for controlling these pests, their overuse has led to the emergence of resistant populations, environmental contamination, and unintended impacts on beneficial organisms. RNA interference (RNAi)-based pest management strategies have emerged as a promising alternative.

The family Meloidae comprises approximately 130 genera and ~3,000 described species[5,6]. Taxonomic diversity is highly skewed, with nearly half of the species concentrated in just five genera[7,8]. Among these, *Hycleus* (Meloinae, Mylabrini) stands out as the most species-rich genus, containing over 450 described species and representing one of the most recently diverged lineages in the family[3,9].

Like all members of the Meloinae, *Hycleus* adults primarily feed on the flowers and leaves of plants in *Desmodium* spp., as well as sweet potatoes, while larvae exhibit highly specialized feeding behaviors, either feeding on locust eggs or parasitizing beehives[3,10]. These dietary transitions are associated with changes in sensory systems, particularly in olfactory and visual systems, which are critical for host recognition and foraging behaviors, and have important implications for non-chemical, biological pest control strategies[11,12]. The genus *Hycleus* exhibits distinctive ecological and evolutionary traits, including specialized host-plant interactions and complex life-history strategies, making it a valuable model for studying speciation and insect-plant dynamics. While prior research has primarily explored their taxonomic structure, ecological niches, behavioral adaptations, and geographic ranges, the molecular basis of these characteristics remains underexplored[7,10,13,14].

Despite significant advances in sequencing technologies and the increasing availability of genomic data for non-model organisms, blister beetles (Meloidae) remain markedly underrepresented in genomic databases. Currently, only one of the ~3,000 described blister beetle species has a chromosome-level genome assembly available[15]. This lack of genomic resources on blister beetles hampers the study of these destructive pests. There is a critical need for high-quality genomic data to advance studies on blister beetle evolution, adaptation, and pest biology.

[1]College of Life Sciences, Shaanxi Normal University, Xi'an, 710119, China. [2]Guangxi Key Laboratory of Sericulture Ecology and Applied Intelligent Technology, 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: xushengquan@snnu.edu.cn

Libraries	Number of Sequences	Total length (bp)	Average length (bp)	N50 length (bp)
Pacbio	283,720	4,596,014,246	16,199	16,172
Hi-C	141,007,338	21,151,100,700	150	—
RNA	4,811,560	6,721,734,000	150	—

Table 1. Sequencing raw data of the *H. marcipoli* assembly.

Genome assembly	Value
Estimated genome size (bp)	128,524,906
Total scaffold length (bp)	111,448,416
Number of chromosome-level scaffolds	11
Longest scaffold (bp)	14,184,471
N50 of scaffold (bp)	10,223,839
Total contig length (bp)	111,428,416
Longest contig (bp)	9,755,183
N50 of contig (bp)	4,646,650
GC content (%)	31.98
BUSCO completeness (genome mode) (%)	99.8
Genome annotation	
Number of protein-coding genes	13,357
Repeats in genome (%)	26.43
TEs in genome (%)	7.62
Mean coding sequence length (bp)	651
BUSCO completeness (protein mode) (%)	98.8
Gene functional annotation	
Non-redundant (NR) (%)	96.64
eggnog-mapper (%)	85.88
Swiss-prot (%)	74.69
Interproscan (%)	9.23
Percentage of overall annotated genes (%)	96.91

Table 2. Genome assembly and annotation statistics of *H. marcipoli*.

Here we present the first chromosome-level genome assembly and annotation of *H. marcipoli*. This high-resolution genome serves not only as a valuable resource for comparative genomic analyses but also as a foundational reference for future investigations into the evolutionary dynamics, functional genomics, and ecological roles of blister beetles.

Methods

Sample collection. This study collected a total of 10 female and male *H. marcipoli* specimens from Huanjiang County (24°83'N, 108°21'E), Hechi City, Guangxi Province, China in August 2019. Three female adult specimens were used for PacBio, Hi-C sequencing and transcriptome sequencing. The abdomens of all specimens were removed before DNA extraction to avoid contamination from intestinal contents, and the remaining body tissues were used for genomic DNA extraction. Genomic DNA extraction and sequencing, as well as RNA sequencing, were carried out by Biomarker (Biomarker Technologies Co., LTD in Beijing, China).

DNA extraction and genome sequencing. High-quality DNA was extracted using DNeasy Blood and Tissue Kits from QIAGEN Inc. DNA quantity and quality were then measured using a 2100 Bioanalyzer (Agilent) and a Qubit 3.0 Fluorometer (Invitrogen), with integrity confirmed via 1% agarose gel electrophoresis. For PacBio long sequencing, the DNA was purified using AMPure PB beads, and the final high-quality gDNA was used for subsequent library construction. The PacBio SMRTbell library was constructed using SMRTbell® Express Template Prep Kit 3.0. Qualified libraries were evenly loaded on SMRT Cell and sequenced using Sequel II system. The Hi-C library was constructed according to the standard protocols described previously[16]. It was then constructed and sequenced using the Illumina NovaSeq 6000 sequencing platform with 183x depth. Finally, 4.60 Gb raw PacBio continuous long reads and 21.15 Gb Hi-C data was generated (Table 1).

RNA extraction and transcriptome sequencing. Total RNA was extracted from a single adult female specimen without biological replication. RNA was extracted from tissues using standard CTAB-LiCI extraction methods[17] followed by rigorous quality control of the RNA samples by means of an Agilent 2100 bioanalyzer (Agilent Technologies, Santa Clara, CA, USA): precise detection of RNA integrity. The cDNA library was built using TruSeq RNA Sample Prep Kit v2 and sequenced on the Illumina NovaSeq 6000 platform. A total 6.72 Gb

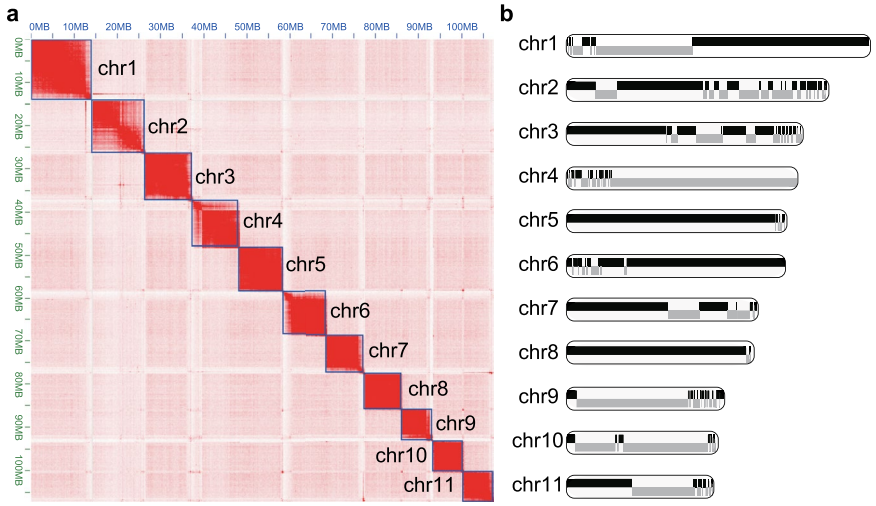

Fig. 1 (**a**) Hi-C contact map showing chromosome-level assembly validation of the *H. marcipoli* genome. The heatmap displays interaction frequencies between genomic regions, with darker colors indicating higher contact probabilities; (**b**) Distribution of contigs along the 11 chromosomes of *H. marcipoli*.

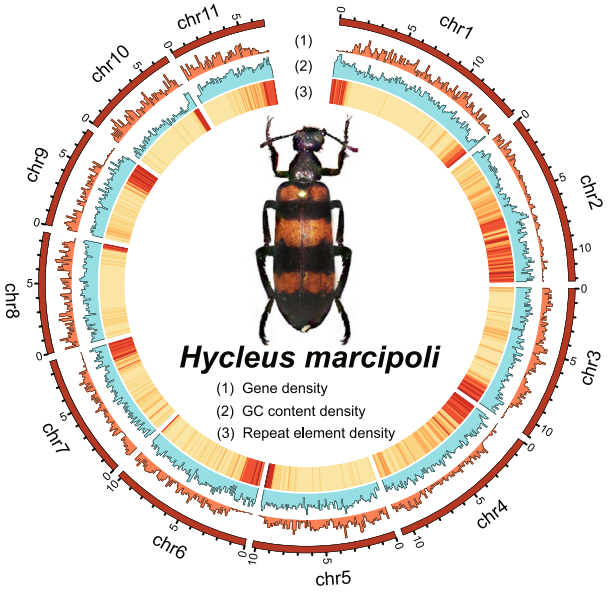

Fig. 2 Schematic representation of the genomic characteristics of *H. marcipoli*. (1) Gene density; (2) GC content density; (3) Repeat element density.

RNA data was generated (Table 1). Low quality sequences and adapter contamination in whole genome sequence data from the above steps were filtered using Trimmomatic v.0.39[18].

Genome assembly. Quality control on raw Illumina data performed using fastp v0.23.2[19] using default parameters. To estimate the genome size of *H. marcipoli*, we used PacBio reads as input data and applied KmerGenie[20]. We estimated the genome size to be approximately 128.52 Mb (Table 2). We assembled the PacBio reads using Flye v2.3.5b[21] with default parameters and used Purge Haplotigs[22] to identify and remove redundant contigs. The initial contig genome size was 126.63 Mb. After removing redundancy and identifying potential contaminants, we obtained an optimized genome of 111 Mb distributed across 168 contigs, with a contig N50 of 4.65 Mb and a scaffold N50 of 10.22 Mb (Table 2). Prior to scaffolding, the high-quality Hi-C library data were aligned to the genome draft using BWA v0.7.17[23] and Samtools v1.14[24]. The draft genome of *H. marcipoli* was further scaffolded using high-quality data from the Hi-C library with HapHic[25]. After scaffolding, manual adjustments were made using Juicebox v2.15[26]. Finally, 92.97% of the contigs (107.29 Mb) were anchored to 11 chromosomes, with chromosome lengths ranging from 6,843,577 bp to 141,844,471 bp (Fig. 1).

Class	Cov	Count	Proportion (%)	Number_of_Distinct_Classifications
DNA	6,245,850	10,932	5.41	98
LTR	1,934,482	1,631	1.68	31
LINE	599,171	1,459	0.52	20
SINE	17,649	93	0.02	2
Penelope	693,053	1,066	0.6	15
Rolling Circle	765,761	1,360	0.66	9
Other (Simple Repeat, Microsatellite, RNA)	3,685,273	54,318	3.19	2848
Unclassified	16,555,295	58,780	14.35	346

Table 3. Annotation and percentage of repeat sequences in genome from Earl Grey.

Type	Counts	Length	Min length	Ave length	Max length	N50
Gene	13,357	58,784,336	105	4,401	113,232	8,830
cDNA	13,357	26,108,082	105	1,955	44,880	2,541
CDS	13,357	20,554,587	105	1,539	44,727	1,953
Protein	13,357	6,851,529	35	513	14,909	651

Table 4. Statistics of protein-coding gene annotations of *H. marcipoli* genome.

Item	Median	Mean
Gene_length	2,162.00	4,101
Isoform_number	1	1.12
Intergenic_length > 0	929	4,711.65
cDNA_length	1,490	1,954.64
exon_number	4	4.9
CDS_length	1,170	1,538.86
CDS_num	4	4.72
Intron_length	354	2,420
Single_exon_length	222	398.87
Single_CDS_length	198	325.76
Single_intron_length	58	620.48

Table 5. Statistics of protein-coding gene structures of *H. marcipoli*.

Repeat annotation. The Earl Grey pipeline (v4.1.0)[27] was used to identify repetitive elements. Approximately 30.50 Mb of the genome was identified as repetitive sequences, constituting 26.43% of the entire genome (Fig. 2). Transposable elements (TEs) occupy 7.62% of the genome, with DNA elements being the dominant TE type at 5.41%, followed by long terminal repeats (LTRs) at 1.68%, long interspersed nuclear elements (LINEs) at 0.52%, and short interspersed nuclear elements (SINEs) at 0.02%. Notably, 14.35% of the repeat sequences were unclassified (Table 3).

Protein-coding gene annotation. For gene annotation, we combining three strategies: *ab initio* prediction, homologous gene comparison, and transcriptome-based annotation. *Ab initio* prediction was performed using BRAKER v2.1.5[28], which automatically trained Augustus v3.3.4[29] and utilized both transcriptome data and protein homology information. The RNA-seq data in BAM format were generated via HISAT2 v2.2.0[30], while protein sequences were retrieved from the OrthoDB10 v1[31] database. For transcript assembly, the mapped transcriptome data were further processed with StringTie v2.1.4[32]. For homology-based annotation, gene sets from five annotated species in Tenebrionoidea—*Tribolium madens*[33], *Tenebrio molitor*[34], *Zophobas morio*[35], *Tribolium castaneum*[36], and *Asbolus verrucosus*[37]—were downloaded. Of these, three are the closest related species published to date. Downloaded protein sequences were then aligned against *H. marcipoli* genome assembly using BLASTP[38] and were identified using GeneWise. Finally, we used the EVidenceModeler (EVM) pipeline v1.1.1[39] to integrate the results from the three strategies. We identified a total of 13,357 protein-coding genes, with an average gene length of 4,401 bp (Fig. 2). Further analysis of gene structure revealed a total cDNA length of 20.55 Mb, with the longest cDNA being 44,727 bp and the average cDNA length being 1,538 bp. The total protein length was 6.85 million amino acids, with the longest protein being 14,909 amino acids and the average protein length being 513 amino acids (Table 4). Among the 13,357 protein-coding genes in *H. marcipoli*, the average number of exons per gene was about 5, with the average length of a single exon being 398.87 bp, and the average length of an intron being approximately 2,420 bp, with single intron averaging 620 bp (Tables 5, 6). The NR (Non-redundant) database, the SwissProt database, the Interproscan database and the EggNOG-mapper database were used for

Chr	Gene counts	Length	Exon length	Exon percentage	Intron length	Intron percentage	Intergenic region length	Intergenic region percentage
chr1	1,690	14,184,471	3,604,378	25.41	4,096,848	28.88	6,483,245	45.71
chr2	973	12,192,676	1,953,476	16.02	4,358,189	35.74	5,881,011	48.23
chr3	1,307	11,002,656	2,664,459	24.22	3,273,214	29.75	5,064,983	46.03
chr4	1,188	10,780,237	2,223,213	20.62	2,994,452	27.78	5,562,572	51.6
chr5	1,419	10,223,839	2,936,349	28.72	3,034,125	29.68	4,253,365	41.6
chr6	1,221	10,208,050	2,384,340	23.36	2,764,160	27.08	5,059,550	49.56
chr7	897	8,899,765	1,730,855	19.45	2,480,060	27.87	4,688,850	52.69
chr8	1,022	8,632,337	2,045,228	23.69	2,730,976	31.64	3,856,133	44.67
chr9	695	7,335,323	1,297,339	17.69	2,000,305	27.27	4,037,679	55.04
chr10	751	6,987,866	1,714,558	24.54	2,059,018	29.47	3,214,290	46
chr11	963	6,843,577	1,867,864	27.29	2,162,669	31.6	2,813,044	41.1

Table 6. Number of protein-coding genes in different sequences of *H. marcipoli* genome and percentage of gene characteristics.

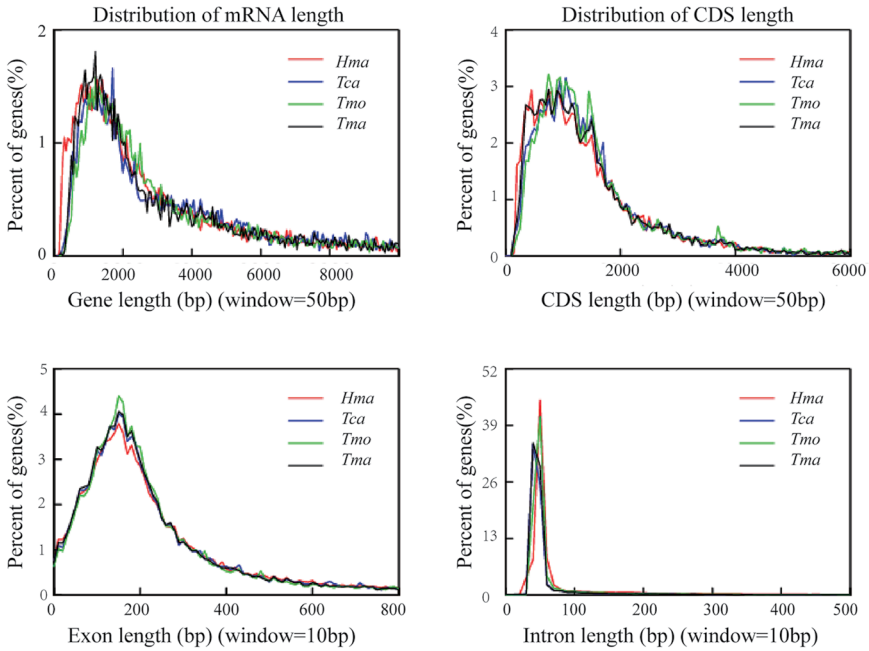

Fig. 3 Annotated genes comparison of the distribution of gene length, CDS length, exon length, and intron length in *H. marcipoli* with other species with annotation. The x-axis represents the length and the y-axis represents the density of genes. Hma, *H. marcipoli*; Tca, *Tribolium castaneum*; Tmo: *Tenebrio molitor*; Tma: *Tribolium maden*.

alignment and to functionally annotate the predicted gene structures. Based on gene functional annotation, 12,944 genes were annotated in at least one database, accounting for 96.91% of the total predicted genes (Table 1). BUSCO analysis (Insecta_odb10)[40] identified 98.80% of the genes, further confirming the accuracy and completeness of the gene prediction (Fig. 4).

Data Records

We have uploaded the raw sequencing data (including Pacbio data, Hi-C data and transcriptome data) to the NCBI database. The BioProject accession number is PRJNA1225931, BioSample accession number is SAMN46911070. The RNA-Seq are available under accession number SRR32479383[41]. The genomic PacBio sequencing data can be found in the NCBI Sequence Read Archive (SRA) database under the accession numbers SRR32489734[42]. Hi-C sequencing data refers to accession numbers SRR32479382[43] in the SRA database. The final genome assembly was deposited in the GenBank under the accession number: GCA_051167335.1[44]. Genome annotation information of repeated sequences, gene structure is available in the Figshare database[45].

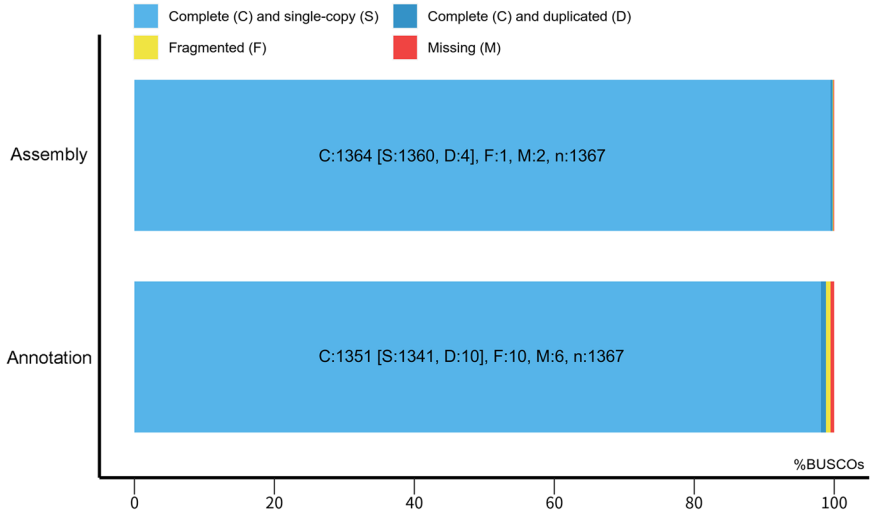

Fig. 4 BUSCO assessments of assembly and annotation.

Technical Validation

To validate the accuracy of *H. marcipoli*'s genome, we mapped our transcriptomic data to the genome, achieving a 99.47% mapping rate and thus confirming the high quality of the *H. marcipoli* genome. The BUSCO v5.2.2 assessment (Insecta_odb10) indicated a high completeness of 99.8% (Fig. 4). We further validated the accuracy and reliability of *H. marcipoli*'s gene structure by comparing its gene distribution with those of other annotated species, and found that consistent patterns across all species supported the accuracy of our gene annotation data. (Fig. 3). Overall, the evaluation results indicate that our *H. marcipoli* genome assembly is complete, accurate, and of high quality.

Code availability

No specific script was used in this work. The codes and pipelines used in data processing were all executed according to the manual and protocols of the corresponding bioinformatics software.

Received: 14 May 2025; Accepted: 30 July 2025;
Published online: 09 August 2025

References

1. Dossey, A. T. Insects and their chemical weaponry: new potential for drug discovery. *Natural product reports* **27**, 1737–1757 (2010).
2. Selander, R. B. Bionomics, systematics, and phylogeny of Lytta, a genus of blister beetles (Coleoptera, Meloidae) 28. *Illinois biological monographs; v. 28* (1960).
3. Bologna, M. & Pinto, J. The Old World genera of Meloidae (Coleoptera): a key and synopsis. *Journal of Natural History* **36**, 2013–2102 (2002).
4. Wu, Y. M., Li, J. & Chen, X. S. Draft genomes of two blister beetles *Hycleus cichorii* and *Hycleus phaleratus*. *Gigascience* **7**, giy006 (2018).
5. Riccieri, A., Mancini, E., Pitzalis, M., Salvi, D. & Bologna, M. A. Multigene phylogeny of blister beetles (Coleoptera, Meloidae) reveals extensive polyphyly of the tribe Lyttini and allows redefining its boundaries. *Systematic Entomology* **47**, 569–580 (2022).
6. Riccieri, A., Capogna, E., Pinto, J. D. & Bologna, M. A. Molecular phylogeny, systematics and biogeography of the subfamily Nemognathinae (Coleoptera, Meloidae). *Invertebrate Systematics* **37**, 101–116 (2023).
7. Riccieri, A., Mancini, E., Salvi, D. & Bologna, M. A. Phylogeny, biogeography and systematics of the hyper-diverse blister beetle genus *Hycleus* (Coleoptera: Meloidae). *Molecular Phylogenetics and Evolution* **144**, 106706 (2020).
8. Bologna, M. A. & Di Giulio, A. Biological and morphological adaptations in the pre-imaginal phases of the beetle family Meloidae. *Atti Accademia Nazionale Italiana di Entomologia* **59**, 141–152 (2011).
9. Bologna, M. A., D'Inzillo, B., Cervelli, M., Oliverio, M. & Mariottini, P. Molecular phylogenetic studies of the Mylabrini blister beetles (Coleoptera, Meloidae). *Molecular Phylogenetics and Evolution* **37**, 306–311 (2005).
10. Saul-Gershenz, L., Millar, J. G., McElfresh, J. S. & Williams, N. M. Deceptive signals and behaviors of a cleptoparasitic beetle show local adaptation to different host bee species. *Proceedings of the National Academy of Sciences* **115**, 9756–9760 (2018).
11. Lebesa, L. N. *Visual and olfactory cues used in host location by the blister beetle Hycleus apicicornis (Coleoptera: Meloidae), a pest of Desmodium (Fabaceae) species*, University of Pretoria (South Africa), (2012).
12. Wagner, L. S., Campos-Soldini, M. P. & Guerenstein, P. G. Olfactory responses of the blister beetle Epicauta atomaria, a polyphagous crop pest, to host, non-host, and conspecific odors. *Entomologia Experimentalis et Applicata* **172**, 806–817 (2024).
13. Bologna, M. A., Amore, V. & Pitzalis, M. Meloidae of Namibia (Coleoptera): taxonomy and faunistics with biogeographic and ecological notes. *Zootaxa* **4373**, 1–141 (2018).
14. Bologna, M. A. & Turco, F. The Meloidae (Coleoptera) of the United Arab Emirates with an updated Arabian checklist. *Zootaxa* **1625**, 1–33 (2007).
15. Shen, C. *et al*. A chromosome-level genome assembly of *Mylabris sibirica* Fischer von Waldheim, 1823 (Coleoptera, Meloidae). *Scientific Data* **12**, 269 (2025).
16. Belton, J.-M. *et al*. Hi–C: a comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268–276 (2012).
17. Pahlich, E. & Gerlitz, C. Deviations from michaelis-menten behaviour of plant glutamate dehydrogenase with ammonium as variable substrate. *Phytochemistry* **19**, 11–13 (1980).

18. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
19. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
20. Chikhi, R. & Medvedev, P. Informed and automated k-mer size selection for genome assembly. *Bioinformatics* **30**, 31–37 (2014).
21. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nature biotechnology* **37**, 540–546 (2019).
22. Roach, M. J., Schmidt, S. A. & Borneman, A. R. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC bioinformatics* **19**, 1–10 (2018).
23. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics* **25**, 1754–1760 (2009).
24. Danecek, P. *et al*. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
25. Zeng, X. *et al*. Chromosome-level scaffolding of haplotype-resolved assemblies using Hi-C data without reference genomes. *Nature Plants* **10**, 1184–1200 (2024).
26. Durand, N. C. *et al*. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell systems* **3**, 99–101 (2016).
27. Baril, T., Galbraith, J. & Hayward, A. Earl Grey: A Fully Automated User-Friendly Transposable Element Annotation and Analysis Pipeline. *Molecular Biology and Evolution* **41**. https://doi.org/10.1093/molbev/msae068 (2024).
28. Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR genomics and bioinformatics* **3**, lqaa108 (2021).
29. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008).
30. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature biotechnology* **37**, 907–915 (2019).
31. Kriventseva, E. V. *et al*. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic acids research* **47**, D807–D811 (2019).
32. Kovaka, S. *et al*. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome biology* **20**, 1–13 (2019).
33. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_015345945.1 (2020).
34. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_963966145.1 (2024).
35. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_036711695.1 (2024).
36. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_031307605.1 (2023).
37. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_004193795.1 (2019).
38. Altschul, S. F. *et al*. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* **25**, 3389–3402 (1997).
39. Haas, B. J. *et al*. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome biology* **9**, 1–22 (2008).
40. Wick, R. R. & Holt, K. E. Benchmarking of long-read assemblers for prokaryote whole genome sequencing. *F1000Research* **8**, 2138 (2021).
41. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR32479383 (2025).
42. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR32489734 (2025).
43. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR32479382 (2025).
44. Zhu, W. *GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_051167335.1 (2025).
45. Yao, B. Genome annotations of *Hycleus marcipoli*. *figshare* https://doi.org/10.6084/m9.figshare.28504385.v2 (2025).

Acknowledgements

This research was funded by the National Natural Science Foundation of China, grant number 31872273.

Author contributions

S.Q.X. initiated, designed, and coordinated the project. W.H.Z. and D.L.G. completed the sequencing and data processing. W.H.Z. and B.B.Y. conducted genome assembly and annotation. J.W.W. and L.L.N. participated in language editing. All authors have 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 S.-Q.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.