



OPEN

DATA DESCRIPTOR

Chromosome-level genome of the brown lacewing *Micromus angulatus* (Stephens, 1836) (Neuroptera: Hemerobiidae)

Yang Zhao¹✉, Qingbin Zhan^{2,3}, Yongjie Wang⁴, Rongxiang Cao¹, Liben Jiang¹ & Qiuqing Xu¹

The brown lacewing *Micromus angulatus* (Stephens), a member within the subfamily Microminae of the family Hemerobiidae, is a globally distributed species and an important predatory natural enemy of various agricultural and forestry crop pests. Despite its global distribution and agricultural significance, genomic resources for the development of novel pest management strategies for *M. angulatus* and the family Hemerobiidae remain limited. Here, we present the first high-quality chromosome-level reference genome for *M. angulatus*, achieved through PacBio HiFi and Hi-C technologies. The assembled genome spans 1.29 Gb, with scaffold N50 of 63.78 Mb, and complete BUSCO score of 97.00%, containing eight pseudochromosomes, demonstrates a high degree of continuity. Functional annotation identified 13,250 protein-coding genes, and repetitive sequences, which account for 83.65% of the genome, were also characterized. This comprehensive assembly offers a robust reference for in-depth research on the genetic basis of predation and adaptability in *M. angulatus*, supports further research into the understanding the genetic diversity within Hemerobiidae, contributing to the broader knowledge within the group and its related species.

Background & Summary

The Hemerobiidae (brown lacewing) is the third largest family of the order Neuropter, which plays a crucial role in natural ecosystems and agricultural settings due to its significant predatory nature, targeting various agricultural and forestry crop pests¹. Globally, approximately 600 species across 28 genera and 11 subfamilies have been identified, exhibiting a wide distribution across major biogeographical regions^{2,3}. Members of this family, including both adults and larvae of brown lacewings are predatory and effectively predator on aphids, scale insects, whiteflies, and other small soft-bodied pests and mites. Their predatory efficiency, widespread distribution, and ability to thrive in diverse environmental conditions make them essential biological control agents in integrated pest management (IPM) systems. By naturally suppressing populations of major agricultural pests, brown lacewing species contribute to reducing the need for chemical insecticides, promoting sustainable farming practices and safeguarding biodiversity⁴. Notably, brown lacewing possesses a long lifespan, high reproductive capacity, and substantial predatory efficiency, rendering them highly valuable and promising for biological control⁵. Currently, various brown lacewing species, which are regarded as reliable control agents, are extensively utilized for the biological control of *Pseudococcus citri* Risso⁶, *Aphis gossypii* Glover⁷, and *Ceratovacuna lanigera* Zehntner^{8–10}. The brown lacewing *Micromus angulatus* (Stephens), which belongs to the subfamily Microminae of the family Hemerobiidae, is an important predatory natural enemy of various pests and mites (Fig. 1). This species is widely distributed across world, exhibiting a strong capacity to adapt to various climatic conditions and habitats^{2,3}. Both the adults and larvae of *M. angulatus* can prey on *Chaetosiphon fragaefolii* (Cockerell) on strawberries and *Myzus persicae* (Sulzer) on bell peppers, exhibiting excellent control efficacy¹¹. By releasing

¹Research Institute of Pomology, Nanjing Institute of Agricultural Sciences in Jiangsu Hilly Area, Nanjing, 210000, Jiangsu, China. ²Department of Criminal Science and Technology, Nanjing Police University, Nanjing, 210000, Jiangsu, China. ³Key Laboratory of State Forestry and Grassland Administration on Wildlife Evidence Technology, Nanjing, 210000, Jiangsu, China. ⁴Guangdong Key Laboratory of Animal Conservation and Resource Utilization, Guangdong Public Laboratory of Wild Animal Conservation and Utilization, Institute of Zoology, Guangdong Academy of Sciences, Guangzhou, 510260, Guangdong, China. ✉e-mail: zhaoyangzz1222@126.com



Fig. 1 Predation at different developmental stages of *Micromus angulatus*. (a) Adult preying on *Aphis craccivora*. (b) 3rd-instar larva preying on *Aphis glycines*. (Photos were taken by Qingbin Zhan and Yang Zhao).

first-instar larvae or female adults, green peach aphids on peppers can be effectively controlled within two weeks¹². This makes *M. angulatus* an important predatory resource for the augmentative release in agricultural and forestry ecosystems.

Despite its agricultural significance, genomic research on the Hemerobiidae family remains limited, particularly when compared to other Neuropteran families such as Chrysopidae (green lacewings). To date, seven neuroptera genomes, including three families (Chrysopoidea^{13–16}, Sisyridae^{17,18}, and Myrmeleontidae¹⁹), have been reported (as of September 2024 from NCBI), with only five species sequenced via PacBio and at the chromosome level. However, studies of Hemerobiidae have focused mainly on morphology, biology and phylogenetics in recent decades, with few studies on genomic information. This has made it difficult to conduct in-depth investigations of this family. Here, we assembled a high-quality chromosome-level genome of *M. angulatus* via high-frequency (HiFi) and high-throughput chromosome conformation capture (Hi-C) technology. The high-quality chromosome-level genome assembly of *M. angulatus* provides a genetic basis for further research on this predatory natural enemy. By providing comprehensive genomic data, this study will not only enrich the genetic resources available for *M. angulatus* but also inform the development of targeted biological control methods, ultimately contributing to more sustainable pest management practices.

Methods

Sample information. The adults of *M. angulatus* used in this study were collected from Xiangerhetun, Gaojiadian Town, Xifeng County, Liaoning Province (42°67'13"N, 124°45'92"E) on May 12, 2023. The rearing conditions were set to a temperature of $25 \pm 1^\circ\text{C}$, a relative humidity of $65 \pm 5\%$, and a light: dark cycle of 16 hours: 8 hours. Six-generation male adults were selected for sequencing under these laboratory conditions. Detailed observations of molting frequency and morphological characteristics allowed for the identification and differentiation of the 1st-, 2nd-, and 3rd- instar larvae, as well as the adult stage¹.

Library construction, quality control and sequencing. Genomic DNA was extracted from a single male individual using the CTAB method, and libraries PacBio HiFi sequencing were constructed using the SMRTbell® Express Template Prep Kit 2.0 (Pacific Biosciences #PN 101-853-100, CA, USA) based on the platform PacBio Sequel IIe. After quality control, the samples were fragmented to 15 kb using a Megaruptor instrument (Diagenode B06010001, Liege, Belgium) and concentrated using AMPure® PB Beads (Pacific Biosciences 100-265-900), and SMRTbell libraries were constructed using a 2.0 kit. This process included the removal of single-stranded overhangs, DNA damage repair, end repair, A-tailing, adapter ligation, and enzyme digestion. Fragment selection was performed using the SageELF (Sage Science ELF000) or BluePippin system (Sage Science BLU0001). The construction of the PacBio HiFi 15 K libraries was conducted by Berry Genomics Corporation, Beijing, China (hereafter referred to as Berry). For second-generation whole-genome sequencing, libraries were constructed from a single male individual using the Agencourt AMPure XP-Medium Kit and sequenced on the DNBSEQ-T7 platform, with insert sizes ranging from 200 to 400 bp. The library construction was completed by Berry. For second-generation standard transcriptome sequencing, RNA was extracted through TRIzolTM Reagent, and libraries were constructed with the VAHTS mRNA-seq v2 Library Prep Kit. The library construction was completed by Berry. Third-generation full-length (ONT) transcriptome sequencing was performed using RNA extracted with the DP441 RNA Prep Pure Plant Plus Kit from a single male individual, and libraries were constructed using the SQK-PCS109 + SQK-PBK004 Kit on the ONT PromethION platform. Library construction was conducted by BenaGen, Wuhan, China. Hi-C library construction using a single male individual, including formaldehyde cross-linking, restriction enzyme digestion (using MboI), end repair, DNA cyclization, and DNA purification, was also performed by Berry. The yield and sequencing depth of the raw sequences are detailed in Table 1.

Genome assembly. The survey inferred genome characteristics based on k-mer frequency distribution, utilizing the khist.sh tool (part of BBTools v38.90²⁰) to generate k-mer frequency histograms and simultaneously assess genome features. The k-mer size was set to 21. The predicted genome size is approximately 1,229,938,700 bp, with a heterozygosity rate of 1.24% and a repeat content of 89.062%, indicating a highly heterozygous and ultra-highly repetitive genome. High-quality PacBio HiFi reads were generated by Pacific Biosciences' (PacBio)

Library type	sequence count	Total output (bp)	Average length (bp)	N50 Length (bp)	Coverage (X)
PacBio HiFi	5,726,271	62,021,847,021	10,831.11	12,240	48.13
Hi-C	912,633,886	136,895,082,900	150	—	106.24
RNA-Seq	63,823,728	9,573,559,200	150	—	—
RNA-ONT	8,453,500	11,291,288,521	1,335.69	1,612	—

Table 1. Statistics of the DNA sequence data used for the *M. angulatus* genome assembly.

SMRT (pbccs v6.4.0) sequencing technology (<https://github.com/PacificBiosciences/ccs>), and sequences with a base quality score above Q30 were retained. The initial assembly was performed by Hifiasm v0.19.6²¹ with the parameters ('-l 3'). Low-depth sequences, which are likely to be contaminants or erroneous, were excluded, and only contig sequences with a sequencing depth exceeding 6X (approximately 1/10 of the sequencing depth) were retained in the Hifiasm assembly. Given the high base quality of the assembly (QV value exceeding 60, as shown in subsequent genome assessment with Merqury), conventional polishing steps were unnecessary. BUSCO analyses indicated no significant heterozygosity in the assembly, so conventional heterozygosity removal steps were also omitted. Chromosome anchoring and assembly of the contigs were performed using Hi-C data and the YAHS v1.2.4²² pipeline. Firstly, Hi-C data were quality controlled using chromap v0.2.5²³ (including sequence read alignment, duplicate removal, and Hi-C contact extraction). Next, two rounds of anchoring were performed using YAHS v1.2 with the default parameters. Manual correction of the initial round of assembly was performed via Juicebox v1.11.08²⁴ to address issues such as misjoins, translocations, and inversions, followed by the final round of anchoring. The Hi-C interaction heatmap for *Micromys angulatus* (Fig. S1) illustrates the chromosomal interaction patterns across the assembled genome. The heatmap shows a clear distinction of eight primary scaffolds, representing the pseudochromosomes, which account for a substantial portion of the genome. Our Hi-C anchoring heatmap analysis indicated that chromosome anchoring quality was moderate, producing only 8 relatively long scaffolds (746,006,264 bp, 57.89%), with approximately 40% of the genome unable to be anchored, likely representing highly repetitive regions (later confirmed by repeat annotation). Hi-C sequencing is less effective in highly repetitive regions and fails to produce strong interaction signals²⁵. Subsequent protein-coding gene annotation revealed that the first 8 longest sequences contained 12,124 genes, accounting for 91.50% (12124/13250) of the total, indicating that the majority of the coding regions were within these 8 sequences, which can be considered pseudochromosomes.

Genome quality assessment. Genome integrity was assessed via Compleasm v0.2.47²⁶, which evaluated BUSCO to perform the completeness assessment with the Insecta_odb10 dataset (n = 1,367 single-copy orthologues), which evaluates the completeness of the genome by predicting the presence of nearly universal single-copy orthologous (USCO) genes. Additionally, second-generation genome and transcriptome (both second- and third-generation) raw sequences were aligned back to the genome assembly to verify the utilization rate of the raw data and the completeness of the assembly. The alignment was performed using Minimap2 v2.28²⁷, and the alignment rate was calculated using SAMtools v1.20²⁸. Potential contamination in the assembly was identified using a BLAST-like search with MMseq 2 v1.38²⁹ against the NCBI nt and UniVec databases. Single-base quality scores (QV values) of the genome were assessed via Merqury v1.39³⁰.

The detailed metrics of the final genome assembly are presented in Table 2. The genome size was 1.29 Gb, with 909 scaffolds and 994 contigs. The maximum scaffold and contig lengths were 169.16 Mb and 51.53 Mb, respectively. The scaffold N50 and contig N50 lengths were 63.78 Mb and 13.65 Mb, respectively, and the GC content was 30.07%. The 8 pseudochromosomes account for 746.01 Mb, accounting for 57.89% of the genome. The BUSCO evaluation of the genome assembly indicated a completeness of 97.00%, with a duplicate BUSCO proportion of 4.75%, indicating no significant redundancy. The alignment rates for the second-generation survey, second-generation RNA, third-generation RNA, and third-generation HiFi data were 96.47%, 90.82%, 97.58%, and 99.90%, respectively, all exceeding 90%, reflecting high sequencing quality. The genome assembly QV was 65.242, corresponding to an error rate of only 2.99086e-07, indicating excellent base quality.

Annotation of repetitive sequences. The species-specific de novo repeat library was constructed by RepeatModeller v2.0.510³¹ with an additional LTR search pipeline (-LTRStruct). This library is based on the principles of repetitive sequence structure-specific and de novo prediction, and it was merged with the Dfam 3.7³² and RepBase-20181026³³ databases to form the final repeat sequence reference database. Repetitive sequence identification was performed by RepeatMasker v4.1.5³⁴, and the results were compared against the constructed repeat sequence database. The results revealed 3,913,133 repetitive sequences (1,077,857,554 bp), accounting for 83.65% of the genome, which is a typical ultrahigh proportion of a repetitive genome. The six most abundant types of repetitive sequences were Unknown (58.79%), LTRs (9.30%), LINEs (7.65%), DNA (4.71%), rolling circles (0.79%), and SINEs (0.59%). The detailed statistics are provided in Table S1.

Compared with the genome of the green lacewing species *Chrysopa pallens*¹³, the *M. angulatus* genome is nearly 700 Mb larger. This genome expansion is due primarily to the proliferation of transposable elements, particularly the Unclassified, LTR, LINE, and DNA transposon categories. In the comparison of repeat composition between *M. angulatus* and *C. pallens* (Fig. 2a), notable differences in the abundance and diversity of transposable elements were observed. We found that *M. angulatus* has a significantly higher proportion of repetitive elements, particularly DNA transposons, LINEs (Long Interspersed Nuclear Elements), and LTRs (Long Terminal Repeats), across a wide range of evolutionary ages as indicated by the Kimura substitution levels. DNA

Content	Values
Assembly size (bp)	1,288,575,812
Number of pseudo-chromosomes (sizes)	8 (746,006,264 bp)
Number of scaffolds/contigs	909/994
Longest scaffold/contig (Mb)	169.16/51.53
N50 scaffold/contig length (Mb)	63.78/13.65
GC content (%)	30.07
BUSCO completeness (%)	97.00
Complete and single-copy BUSCOs (S)	92.25
Complete and duplicated BUSCOs (D)	4.75
Fragmented BUSCOs (F)	0.29
Missing BUSCOs (M)	2.71
Mapping ratio of reads (%)	
HIFI	99.28
RNA-Seq	90.82
RNA-ONT	97.58

Table 2. Statistics for *M. angulatus* genome assembly and gene annotation.

transposons in *M. angulatus* contribute a greater percentage of the genome compared to *C. pallens*, indicating some historical activities. Similarly, LINEs and LTRs show higher representation in *M. angulatus*, suggesting multiple waves of transposable element expansion. These expansions are likely key factors in the larger genome size of *M. angulatus*, reflecting the evolutionary processes that have shaped its genome architecture. We further compared the significantly expanded transposon families within the LTR, LINE, and DNA categories (defined here as families differing by more than 1 Mb). In *M. angulatus*, we observed a significant expansion of various transposable elements (TEs) compared to *C. pallens* (Fig. 2b). Specifically, DNA transposons, LINEs (Long Interspersed Nuclear Elements), and LTRs (Long Terminal Repeats) were prominently expanded. Among DNA transposons, elements such as TcMar-Tc1, hAT-Ac, and PIF-Spy showed the greatest expansion in *M. angulatus*, with total sizes reaching several megabases (Mb). For LINEs, elements like R1-LOA and I-Jockey demonstrated notable expansions, indicating an active or historically active role in genome evolution. We found that in LTR elements, Gypsy, Pao, and Copia showed extensive amplification in *M. angulatus*, with sizes exceeding tens of megabases. This substantial expansion of transposable elements, particularly in the Gypsy family of LTRs, suggests that transposable elements played a crucial role in the genome evolution of *M. angulatus*, contributing to its larger genome size relative to *C. pallens*. These findings highlight the evolutionary divergence between the two species and the impact of TE proliferation on genome architecture.

To further investigate the origins of the transposon expansions, we utilized the script calcDivergenceFromAlign.pl from RepeatMasker-4.1.5 to evaluate the TE landscape by calculating the Kimura 2-parameter divergence for each TE family, which essentially measures the degree of differentiation or genetic distance. As shown in Fig. 2c, there were three distinct waves of transposon expansion in the *M. angulatus* genome (three peaks). Assuming that the first expansion (the rightmost peak, approximately position 32) occurred around 115 million years ago, which coincides with the appearance of fossil records of family Hemerobiidae³⁵, the second expansion (approximately position 18) could be calculated as $18/32 \times 115 = 64.5$ million years ago, roughly corresponding to the era of dinosaur extinction. The third expansion (around positions 6/7) began approximately 20 million years ago and is still ongoing. Excluding Unclassified repeats, as illustrated in Fig. 2d, only the leftmost peak (expansion) remains prominent.

Annotation of noncoding RNAs. The annotation of ncRNAs was performed by two strategies: (1) annotating rRNA, snRNA, and miRNA via alignment with known noncoding RNA libraries from the Rfam database using Infernal v1.1.5³⁶ software and (2) predicting tRNA sequences within the genome using tRNAscan-SE v2.0.12³⁷ software and filtering low-confidence tRNAs via the built-in ‘EukHighConfidenceFilter’ script. The genome annotation of noncoding RNAs revealed 5,481 ncRNAs, including 4,609 rRNAs, 72 miRNAs, 334 snRNAs, 1 sRNA, 786 tRNAs, 3 ribozymes, 3 lncRNAs, and 1 leader RNA. The snRNAs included 291 spliceosomal RNAs (U1, U2, U4, U5, U6, and U11), 11 minor spliceosomal RNAs (U4atac, U6atac, and U12), 26 C/D box snoRNAs, and 5 HACA-box snoRNAs. The detailed statistics are presented in Table S2.

Identification of protein-coding gene structures. The identification of protein-coding gene structures was performed via the MAKER v3.01.04³⁸ pipeline, which integrates three types of evidence to predict the structures of protein-coding genes.

- (1) Ab initio gene prediction. Methods: BRAKER v3.0.6³⁹ and GeMoMa v1.9⁴⁰ were used to integrate transcriptome and protein evidence. The results from both methods were merged as input files (ab.gff3) for MAKER ab initio to broaden the range of potential coding gene candidates. The transcriptome data were aligned to the genome via HISAT2 v2.2.1⁴¹ to generate BAM alignment files. BRAKER automatically trained Augustus v3.4.0⁴² and GeneMark-ETP⁴³, integrating arthropod protein sequences (extracted from the OrthoDB11 database⁴⁴) and transcriptome data to increase prediction accuracy. GeMoMa predicted

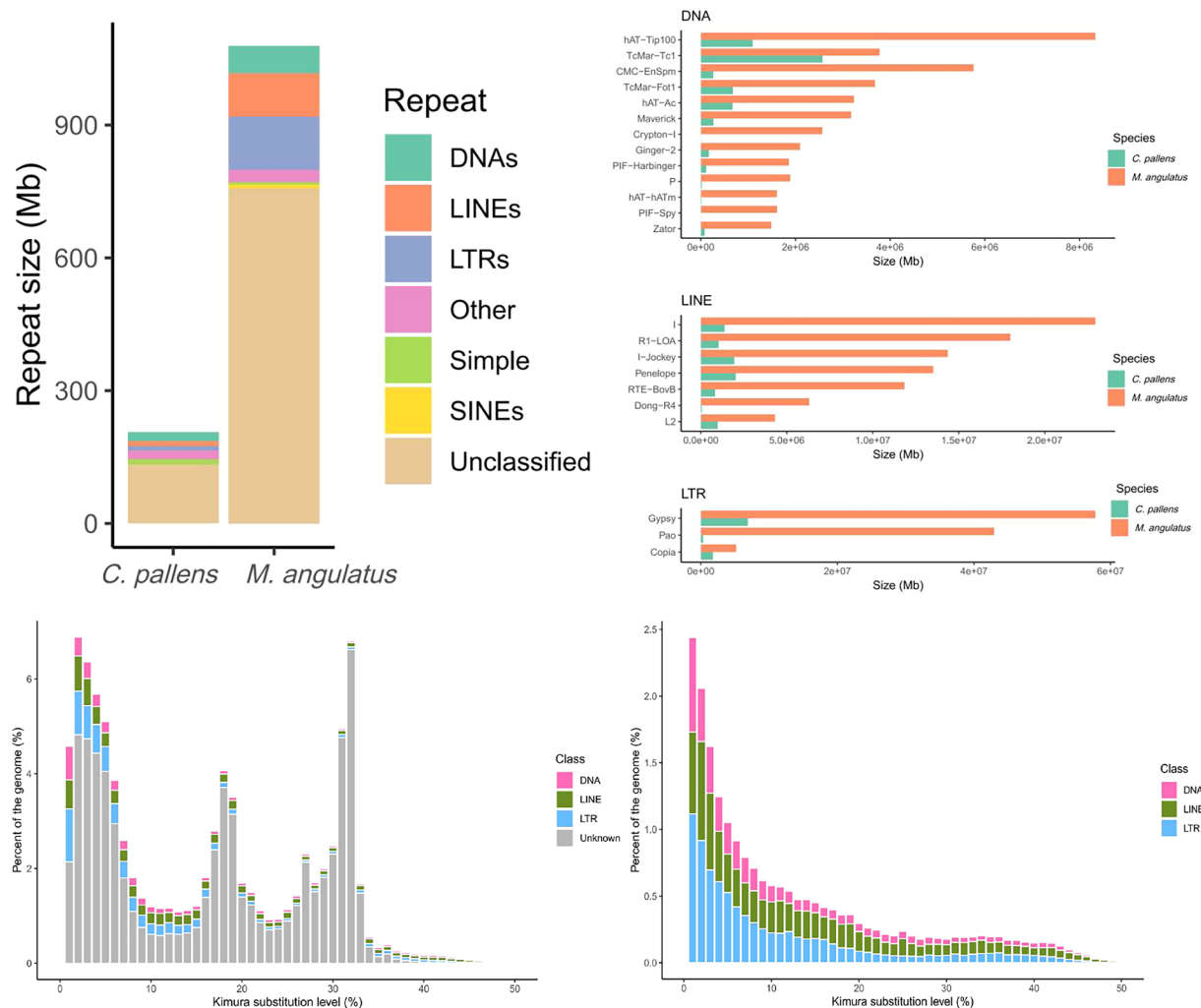


Fig. 2 Repetitive sequences of *Micromys angulatus*. **(a)** Comparison of repeat composition between *M. angulatus* and *C. pallens*. **(b)** Transposon families significantly expanded in *M. angulatus* compared to *C. pallens*, including DNA transposons, LINE (Long Interspersed Nuclear Elements), and LTR (Long Terminal Repeats). **(c)** Interspersed repeat landscape based on K2P divergence (CpG adjusted). **(d)** Interspersed repeat landscape based on K2P divergence (CpG adjusted) with unclassified repeats excluded.

genes on the basis of protein homology and intron position information via high-quality genome assemblies, protein sequences, and annotated GFF files from closely related species: *Drosophila melanogaster* (Diptera) (GCA_000001215.4⁴⁵), *Bombyx mori* (Lepidoptera) (GCA_030269925.2⁴⁶), *Apis mellifera* (Hymenoptera) (GCA_019321825.1⁴⁷), *Chrysoperla carnea* (Neuroptera) (GCA_905475395.1¹⁴), and *Tribolium castaneum* (Coleoptera) (GCA_000002335.3⁴⁸). GeMoMa uses the parameters GeMoMa.c = 0.4 and GeMoMa.m = 100000 for gene prediction on the basis of homologous information from these species.

(2) Transcript alignment for gene structure prediction. Methods: StringTie v2.2.1⁴⁹ was used for reference-guided assembly in mixed mode (-mix) on the basis of second- and third-generation transcriptome data. The BAM alignment files for the second-generation sequences were generated by HISAT2 from step (1), and the third-generation sequences were aligned to the genome via Minimap2 (-ax splice).

(3) Homology-based prediction. Methods: Homology prediction was conducted by aligning protein sequences (as used in GeMoMa analysis) to the genome and the Wilcoxon tests was performed by rstatix (<https://github.com/kassambara/rstatix>). The MAKER pipeline predicted 13,250 protein-coding genes (Table 3), with an average gene length of 18,226.6 bp. Each gene contained an average of 5.3 exons with an average exon length of 432.7 bp and 4.3 introns with an average intron length of 3,933.5 bp. Each gene had an average of 5.0 CDSs, with an average CDS length of 330.6 bp. The BUSCO completeness assessment of the predicted protein-coding genes yielded 97.6% C [S:77.7%, D:19.9%], 0.2% F, 2.2% M, and 1367% n, surpassing the genome completeness of 97.00%, indicating high-quality protein-coding gene annotation. We further compared the metrics of *M. angulatus* protein-coding genes with those of *C. pallens*.

annotation	Number
Structure annotation	
Number of protein-coding genes	13,250
Number of predicted protein sequences	16,869
Mean protein length (aa)	609
Mean gene length (bp)	18,226.6
Gene ratio	18.74%
Number of exons per gene	5.3
Mean exon length (bp)	432.7
Exon ratio	2.40%
Number of CDSs per gene	5.0
Mean CDS length (bp)	330.6
CDS ratio	1.73%
Number of introns per gene	4.3
Mean intron length (bp)	3,933.5
Intron ratio	16.34%
Function annotation	
Number of genes matching UniProt records	12,110
Number of genes labelled as “Uncharacterized protein”	780
Number of genes labelled as “unknown function”	1,217
Number of genes with InterProScan annotations	11,349
Number of genes with eggNOG annotations	12,084
Number of genes with Enzyme Codes (EC) from eggNOG annotations	2,693
Number of genes with KEGG ko terms from eggNOG annotations	7,848
Number of genes with KEGG pathway terms from eggNOG annotations	4,706
Number of genes with COG Functional Categories from eggNOG annotations	11,316
Number of genes with GO items	10,151
Number of genes with KEGG pathways items	4,706

Table 3. Summary of protein-coding genes of *Micromys angulatus*.

	<i>M. angulatus</i>	<i>C. pallens</i>
No. genes	13,250	12,840
Mean gene length (bp)	18,226.6	7,555.9
Gene ratio	18.74%	18.02%
No. exons per gene	5.3	5.7
Mean exon length (bp)	432.7	410.7
Exon ratio	2.40%	5.62%
No. CDSs per gene	5.0	5.4
Mean CDS length (bp)	330.6	293.7
CDS ratio	1.73%	3.83%
No. introns per gene	4.3	4.7
Mean intron length (bp)	3,933.5	1,120.1
Intron ratio	16.34%	12.40%

Table 4. Comparison of protein-coding genes between *M. angulatus* and *C. pallens*.

As shown in Table 4 and Fig. 3, the gene and intron lengths in *M. angulatus* are significantly longer than those in *C. pallens*, contributing to overall genome expansion. In the comparison of protein-coding genes between *M. angulatus* and *C. pallens* (Fig. 3), significant differences were observed across various genomic features. The gene length in *M. angulatus* was significantly longer than in *C. pallens*, with average gene lengths being approximately twice as large. Similarly, exon lengths also exhibited a significant difference, with *M. angulatus* having longer exons on average. We also found that coding sequences (CDS) in *M. angulatus* were longer than those in *C. pallens*. A more striking contrast was seen in intron lengths, where *M. angulatus* displayed significantly longer introns compared to *C. pallens*. Therefore, our results suggest that the genome of *M. angulatus* has undergone considerable expansion at the level of both gene structure and intron-exon architecture, which may contribute to its larger genome size and unique functional adaptations compared to *C. pallens*.

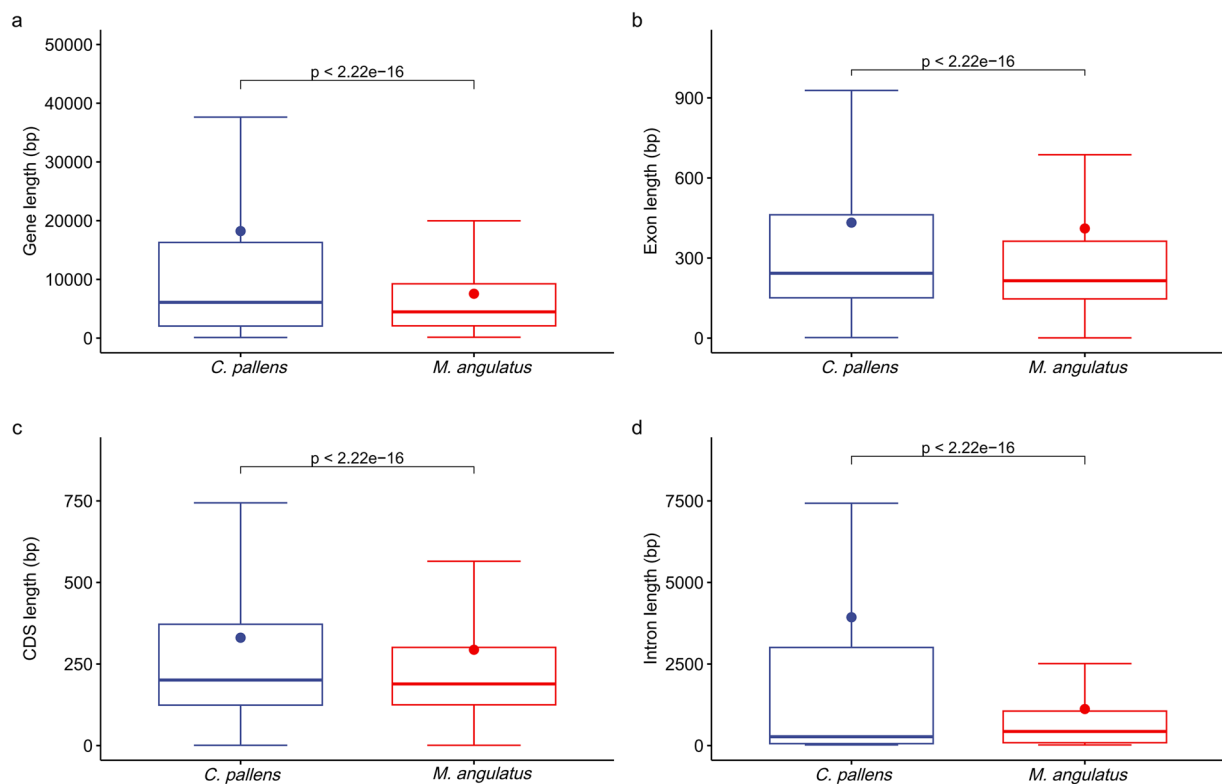


Fig. 3 Comparison of protein-coding genes between *M. angulatus* and *C. pallens*. (a) Gene length (bp). (b) Exon length (bp). (c) CDS length (bp). (d) Intron length (bp).

Functional annotation of protein-coding genes. The functional annotation of genes was conducted via three strategies: (1) Comparison with existing databases. The UniProtKB v202305 (including SwissProt and TrEMBL) database was used. Method: Using the high-sensitivity mode of Diamond v2.1.7.161⁵⁰ ('-very-sensitive -e 1e-5'), genes were searched against the UniProtKB database to predict their functions. (2) Comparison with Comprehensive Databases The predicted conserved protein sequences and domains, Gene Ontology (GO) terms, and pathways (KEGG, Reactome) were analysed. Methods: InterPro 5.65-97.0⁵¹ was used to search four databases: Pfam⁵², SMART⁵³, Superfamily⁵⁴, and CDD⁵⁵. Additionally, eggNOG-mapper v2.1.1⁵⁶ was used to search the eggNOG v5.0.2⁵⁷ database.

The results from these two methods were integrated to produce the final functional gene predictions, which are typically presented in a GFF file format. A total of 12,110 genes (91.40%) matched entries in the UniProtKB database. InterPro identified protein domains in 11,349 protein-coding genes. Combined results from InterPro and eggNOG-mapper-annotated GO terms for 10,151 genes and KEGG pathways for 4,706 genes.

The comparison of gene structure between *M. angulatus* and *C. pallens* reveals distinct differences in genomic organization. *C. pallens* has a slightly lower total number of genes (12,840) compared to *M. angulatus* (13,250), but the mean gene length in *C. pallens* is significantly shorter at 7,555.9 bp compared to 18,226.6 bp in *M. angulatus*. Both species show similar numbers of exons per gene, with *M. angulatus* averaging 5.3 exons per gene and *C. pallens* averaging 5.7. However, the mean exon length is slightly shorter in *C. pallens* (410.7 bp) compared to *M. angulatus* (432.7 bp). Notably, the coding sequence (CDS) length is longer in *M. angulatus*, with an average of 330.6 bp compared to 293.7 bp in *C. pallens*. Although *M. angulatus* exhibits a lower number of introns per gene (4.3) than *C. pallens* (4.7), the mean intron length in *M. angulatus* is significantly longer at 3,933.5 bp compared to 1,120.1 bp in *C. pallens*. These results indicate that *M. angulatus* possesses genes with longer mean lengths and introns, which contrasts with the more compact gene structure observed in *C. pallens*, characterized by shorter genes and introns. This difference might reflect variations in genome evolution and functional genomic organization between the two species. The detailed results are provided in Table 4.

Chromosomal collinearity. To investigate the chromosomal evolution of genomes, we conducted a collinearity analysis between *M. angulatus* and two closely related species within the same order, *C. pallens* and *Chrysoperla carnea*, to obtain relevant information. Given the moderate quality of chromosome assembly in *M. angulatus*, the eight longest sequences were selected for comparison. Protein sequences between the species were aligned using the Diamond software with blastp-like functionality, applying ultra-sensitive, iterative settings ('-ultra-sensitive-iterate -e 1e-5 -k 5'). Subsequently, collinearity analysis was performed using MCScanX⁵⁸, with parameters set to '-s 4 -e 1e-5', meaning that at least four genes were required to be considered a collinear block, with the e-value threshold set to 1e-5. As shown in Fig. 4, a high degree of collinearity is observed between the two species *C. pallens* and *C. carnea* from the same family, with chromosomes aligning clearly on a one-to-one

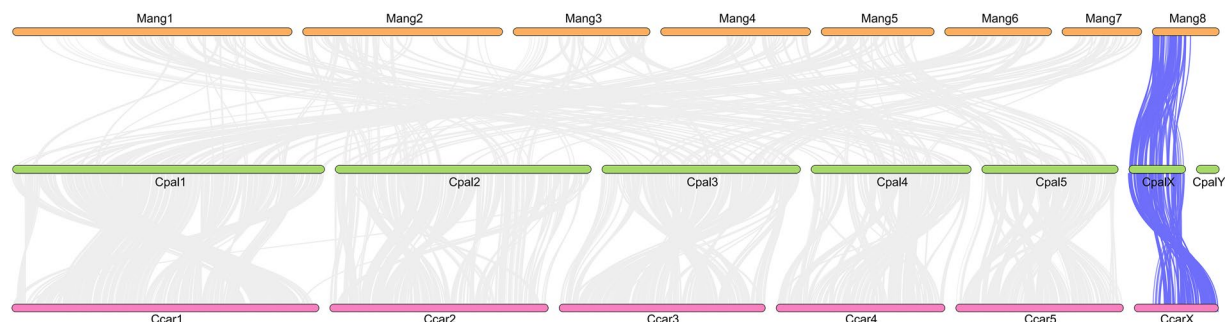


Fig. 4 Chromosomal synteny among *M. angulatus*, *C. pallens* and *C. carnea*.

basis. However, internal rearrangements and inversions within chromosomes were also evident. Key and major chromosomal variations were highlighted in the figure, notably with the X chromosome being highly conserved across all three species, suggesting that *M. angulatus* 8 may represent the X chromosome. Due to the more distant phylogenetic relationship between *M. angulatus* and the other two species from a different family, chromosomal homology was relatively low, and no clear one-to-one correspondence was observed.

Data Records

The raw sequencing data and genome assembly of *Micromus angulatus* have been deposited at the Sequence Read Archive (SRA) and GenBank databases of the National Center for Biotechnology Information (NCBI), under BioProject number PRJNA1101095. The genome assembly and annotation data have been released under GCA_047833675.1⁵⁹ (master record: JBCARC000000000). The Illumina, Pacbio, Hi-C genome sequencing data, RNA-ONT and Illumina RNASeq sequencing data were submitted to the Sequence Read Archive (SRA) database at under accession numbers SRR31897142⁶⁰, SRR29924137⁶¹, SRR29908479⁶², SRR31895759⁶³ and SRR29841511⁶⁴.

Technical Validation

The predicted protein-encoding gene sequences were subjected to BUSCO integrity assessment, yielding the following results: C: 97.6%, n: 1,367. The integrity surpassed 97.00% of the genome assessments. This indicates the high quality of the genome assembly produced in this study. The assembled genome spans 1.29 Gb with 8 pseudochromosomes and scaffold N50 of 63.78 Mb which indicating long, continuous sequences. These metrics ensure that the genome assembly is of high quality with minimal gaps and interruptions. MMseq2 was used to check for contamination by aligning the genome against the NCBI nt and UniVec databases. Merqury software assessed single-base accuracy, yielding a QV score of 65.242 (very low error rate of 2.99086e-07). All these results suggest that we got a high-quality assembly of *M. angulatus* with high integrity, continuity and accuracy.

Code availability

No specific script was used in this work. All commands and pipelines used in data processing were executed according to the manual and protocols of the corresponding bioinformatic software.

Received: 21 November 2024; Accepted: 28 February 2025;

Published online: 06 March 2025

References

1. Zhao, Y. Systematics of family Hemerobiidae from China (Insecta: Neuroptera, Hemerobiidae). *China Agricultural University*, (2016).
2. Oswald, J. D. Bibliography of the Neuropterida. Lacewing Digital Library, Research Publication No. 2. <http://lacewing.tamu.edu/Biblio/Main>. [Accessed 23th October 2024] (2024).
3. Yang, D., Liu, X.Y. & Yang, X. K. *The Color Atlas of Neuropterida From China*. Henan Science and Technology Press. 449–565 (2023).
4. Zhao, Y. *et al.* Predation Efficiency and Biological Control Potential of *Micromus angulatus* Against *Aphis craccivora*. *Agronomy* **14**, 2242, <https://doi.org/10.3390/agronomy14102242> (2024).
5. Zhao, Y. & Liu, Z. Q. Advances in the biological control application of Hemerobiidae. *China Plant Protection* **40**, 31–37 (2020).
6. Zimmerman, E. C. *Insects Of Hawaii: A Manual Of The Insects Of The Hawaiian Islands, Including An Enumeration Of The Species And Notes On Their Origin, Distribution, Hosts, Parasites, Etc.* University of Hawaii Press. **6** (1957).
7. Miller, G. L. & Cave, R. D. Bionomics of *Micromus posticus* (Walker) (Neuroptera: Hemerobiidae) with descriptions of the immature stages. *Proc Entomol Soc Wash* **89**, 776–789 (1987).
8. Lingappa, S., Patil, R. K., Mulimani, V. & Ramegowda, G. K. Brown lacewing, *Micromus igorotus* Banks - a potential predator of sugarcane woolly aphid. *Curr Sci* **87**, 1056–1057 (2004).
9. Mulimani, V., Lingappa, S., Patil, R. K. & Ramegowda, G. K. Biology and feeding potential of *Micromus timidus* Hagen (Neuroptera: Hemerobiidae) on sugarcane woolly aphid, *Ceratovacuna lanigera* Zehntner. *Karnataka J Agric Sci* **23**, 246–248 (2010).
10. Mulimani, V., Lingappa, S., Patil, R. K. & Ramegowda, G. K. Field evaluation of *Micromus igorotus* Banks (Neuroptera: Hemerobiidae) for the management of sugarcane woolly aphid, *Ceratovacuna lanigera* Zehntner. *Int J Plant Prot* **3**, 228–237 (2010).
11. Pekas, A., Smedt, L. D., Verachttert, N. & Boonen, S. The brown lacewing *Micromus angulatus*: A new predator for the augmentative biological control of aphids. *Biol Control* **186**, 105324 (2023).
12. Ntalia, P., Broufas, G. D., Wäckers, F., Pekas, A. & Pappas, M. L. Overlooked lacewings in biological control: The brown lacewing *Micromus angulatus* and the green lacewing *Chrysopa formosa* suppress aphid populations in pepper. *J Appl Entomol* **146**, 796–800 (2022).

13. Wang, Y. Y. *et al.* The first chromosome-level genome assembly of a green lacewing *Chrysopa pallens* and its implication for biological control. *Mol Ecol Resour* **22**, 755–767 (2022).
14. Crowley, L. The genome sequence of the common green lacewing, *Chrysoperla carnea* (Stephens, 1836). *Wellcome Open Res.* **6** (2021).
15. NCBI BioProject: PRJNA905226 <https://www.ncbi.nlm.nih.gov/datasets/taxonomy/1504853/> (2023).
16. NCBI BioProject: PRJNA905226 <https://www.ncbi.nlm.nih.gov/datasets/taxonomy/482137/> (2024).
17. NCBI BioProject: PRJNA273140 <https://www.ncbi.nlm.nih.gov/datasets/taxonomy/1560625/> (2023).
18. NCBI BioProject: PRJNA273139 <https://www.ncbi.nlm.nih.gov/datasets/taxonomy/279440/> (2023).
19. NCBI BioProject: PRJNA849191 <https://www.ncbi.nlm.nih.gov/datasets/taxonomy/2028601/> (2004).
20. Bushnell, B. BMap. <https://sourceforge.net/projects/bbmap/>.
21. 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).
22. Zhou, C., McCarthy, S. A. & Durbin, R. YaHS: Yet another Hi-C scaffolding tool. *Bioinformatics* **39**, btac808 (2023).
23. Zhang, H. *et al.* Fast alignment and preprocessing of chromatin profiles with Chromap. *Nat Commun* **12**, 6566 (2021).
24. Durand, N. C. *et al.* Juicer Provides a One-Click System for Analysing Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95–98 (2016).
25. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
26. Huang, N. & Li, H. compleasm: A faster and more accurate reimplement of BUSCO. *Bioinformatics* **39**, btad595 (2023).
27. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
28. Li, H. *et al.* 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
29. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive datasets. *Nat Biotechnol* **35**, 1026–1028 (2017).
30. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol* **21**, 245 (2020).
31. Flynn, J. M. *et al.* RepeatModeller2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci* **117**, 9451–9457 (2020).
32. Storer, J., Hubley, R., Rosen, J., Wheeler, T. J. & Smit, A. F. The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mobile DNA* **12**, 2 (2021).
33. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* **6**, 11 (2015).
34. Smit, A. *et al.* RepeatMasker Open-4.0. <http://www.repeatmasker.org> (2013/2015).
35. Ponomarenko, A. G. Novye setchatokrylye (Insecta, Neuroptera) iz mezozoya Mongolii (ed. Grunt, T. A.). Novye taksony iskopaemykh bespozonochnykh Mongolii. *Akademiia Nauk SSSR, Sovmestnaya Sovetskoye-Mongol'skaya Paleontologicheskaya Ekspeditsiya* **41**, 101–111 (1992).
36. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).
37. Chan, P. P. & Lowe, T. M. tRNAscan-SE: Searching for tRNA genes in genomic sequences. *Methods Mol Biol* **1962**, 1–14 (2019).
38. Holt, C. & Yandell, M. MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinf* **12**, 491 (2011).
39. Bruna, 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 Bioinf* **3**, lqaa108 (2021).
40. Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S. O. & Grau, J. Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinf* **19** (2018).
41. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907–915 (2019).
42. 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).
43. Bruna, T., Lomsadze, A. & Borodovsky, M. A new gene finding tool GeneMark-ETP significantly improves the accuracy of automatic annotation of large eukaryotic genomes. *bioRxiv* **34**, 757–768 (2024).
44. Kuznetsov, D. *et al.* OrthoDB V11: Annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Res* **51**, D445–D451 (2023).
45. Hoskins, R. A. *et al.* Sequence finishing and mapping of *Drosophila melanogaster* heterochromatin. *Science* **316**(5831), 1625–8 (2007).
46. Zhao, Q. *et al.* An Orphan Gene Enhances Male Reproductive Success in *Plutella xylostella*. *Mol Biol Evol* **41**(7), msae142 (2024).
47. Cao, L., Zhao, X., Chen, Y. & Sun, C. Chromosome-scale genome assembly of the high royal jelly-producing honeybees. *Sci Data* **8**(1), 302 (2021).
48. Tribolium Genome Sequencing Consortium. The genome of the model beetle and pest *Tribolium castaneum*. *Nature* **452**, 949–955 (2008).
49. Kovaka, S. *et al.* Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol* **20**, 278 (2019).
50. Buchfink, B., Reuter, K. & Drost, H. G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* **18**, 366–368 (2021).
51. Paysan-Lafosse, T. *et al.* InterPro in 2022. *Nucleic Acids Res* **51**, D418–D427 (2023).
52. El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res* **47**, D427–D432 (2019).
53. Letunic, I., Khedkar, S. & Bork, P. SMART: Recent updates, new developments and status in 2020. *Nucleic Acids Res* **49**, D458–D460 (2021).
54. Wilson, D. *et al.* SUPERFAMILY—sophisticated comparative genomics, data mining, visualization and phylogeny. *Nucleic Acids Res.* **37**, D380–D386 (2009).
55. Wang, J. *et al.* The conserved domain database in 2023. *Nucleic Acids Res.* **51**, D384–D388 (2023).
56. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol Biol Evol.* **38**, 5825–5829 (2021).
57. Huerta-Cepas, J. *et al.* eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).
58. Wang, Y. *et al.* MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49–e49 (2012).
59. NCBI GeneBank https://identifiers.org/ncbi/insdc.gca:GCA_047833675.1 (2024).
60. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR31897142> (2024).
61. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR29924137> (2024).
62. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR29908479> (2024).
63. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR31895759> (2024).
64. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR29841511> (2024).

Acknowledgements

This research was supported by the National Natural Science Foundation of China (No. 32100366) and Fundamental Research Funds for the Central Universities under Grant LGZD202405.

Author contributions

Y.Z. wrote the draft manuscript and performed the data analysis. Q.Z. assembled and annotated the genome. R.C., Y.W. and Q.X. analysed the data and generated the images. Y.W. and Q.Z. collected the first-generation samples from Liaoning Province. Q.X. and L.J. reared *M. angulatus* to the sixth generation under laboratory conditions. Q.Z. and L.J. improved the manuscript and provided critical suggestions. Y.Z. and Y.W. designed and coordinated the study. All the authors have read and approved the 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/s41597-025-04739-w>.

Correspondence and requests for materials should be addressed to Y.Z.

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