scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Chromosome-level genome assembly of the wasp *Vespa mandarinia* (Hymenoptera: Vespidae)

Xiao-Yun Niu[1], Yuan-Chong Shi[1], Hai-Feng Mo[1] & Pu Yang[1,2] ✉

The wasp *Vespa mandarinia* raised in Yunnan Province of China exhibits a relatively large body size, potent toxicity, invasive potential, and ecological impacts. These characteristics are closely linked to its acute olfactory sensitivity, venom-aided predation, and flight capacity. However, the lack of high-quality genomic resources limits understanding of its adaptive evolution and invasive potential. In this study, DNBSEQ, Oxford Nanopore sequencing, and Hi-C sequencing were employed to sequence and assemble the genome of *V. mandarinia*. The assembled genome consisted of 25 chromosomes. The genome size was 191.80 Mb. The scaffold N50 was 9.20 Mb. The complete BUSCOs of genome was 99.5%. A total of 12,023 genes in the *V. mandarinia* genome were predicted and 11,541 genes (99.84%) were functionally annotated. *V. mandarinia* wasps showed high level of collinearity with *V. velutina*. The genome will provide valuable resources for understanding the evolution of wasp, also provide information for venom research on *V. mandarinia*.

Background & Summary

The wasp *Vespa mandarinia*, the largest hornet in the world, exhibits remarkable morphological adaptations, with adults reaching 4 cm in body length andmore than 7 cm in wingspan[1–3]. Notably, a record-breaking individual measuring 6 cm body length with a 9.35 cm wingspan was documented in Yunnan, China, inhabiting a nest exceeding 1.5 m in diameter[4]. These traits strengthen its ecological dominance, enhancing predatory efficiency, defensive capabilities, and sustained flight performance[5].

As a eusocial predator, its cooperative brood care and reproductive division of labor ensure colony stability, facilitating resource provisioning during critical developmental phases[5]. These characteristics lay foundation to the aggressive ability of the wasp. While its role in regulating pest populations highlights potential biocontrol applications[6,7], rapid proliferation risks ecological imbalance, especially pollination disruptions caused by some invasive wasps[8]. Although *V. mandarinia* mainly live in temperate Asian regions with high rainfall and dense vegetation[9] and thrives in habitats supporting abundant bee populations[3,10], the super adaptation ability and captive breeding of some wasps have increased the population reproduction and expansion. Recently, global invasions, including detections in Canada[11], the United States[12], and Europe[13] have raised urgent ecological and agricultural concerns.

The wasp *V. mandarinia* raised in Yunnan Province of China (Fig. 1A) exhibits exceptional adaptations, including enlarged body size, heightened venom potency, and a restricted geographic range. Its pupae serve as delicious and nutritious food resource, and their venom has important medicinal value[14–16]. Due to the expensive value, its cultivation has expanded rapidly across Yunnan and neighboring regions over the past decade. The invasive potential and ecological impacts are closely related to acute olfactory sensitivity[17], eusocial cooperation[18], potent venom-aided predation[19], and superior flight capacity[20]. However, the lack of high-quality genomic resources has limited understanding of its adaptive evolution and invasive potential.

This study presents chromosome-level genome assembly of *V. mandarinia*. The genome comprised 25 chromosomes and the genome size was 191.80 Mb. The complete BUSCOs of genome was 99.5%. A total of 12,023

[1]Institute of Highland Forest Science, Chinese Academy of Forestry, Kunming, 650224, China. [2]Key Laboratory of Breeding and Utilization of Resource Insects of National Forestry and Grassland Administration, Kunming, 650224, China. ✉e-mail: zjuyangpu@aliyun.com

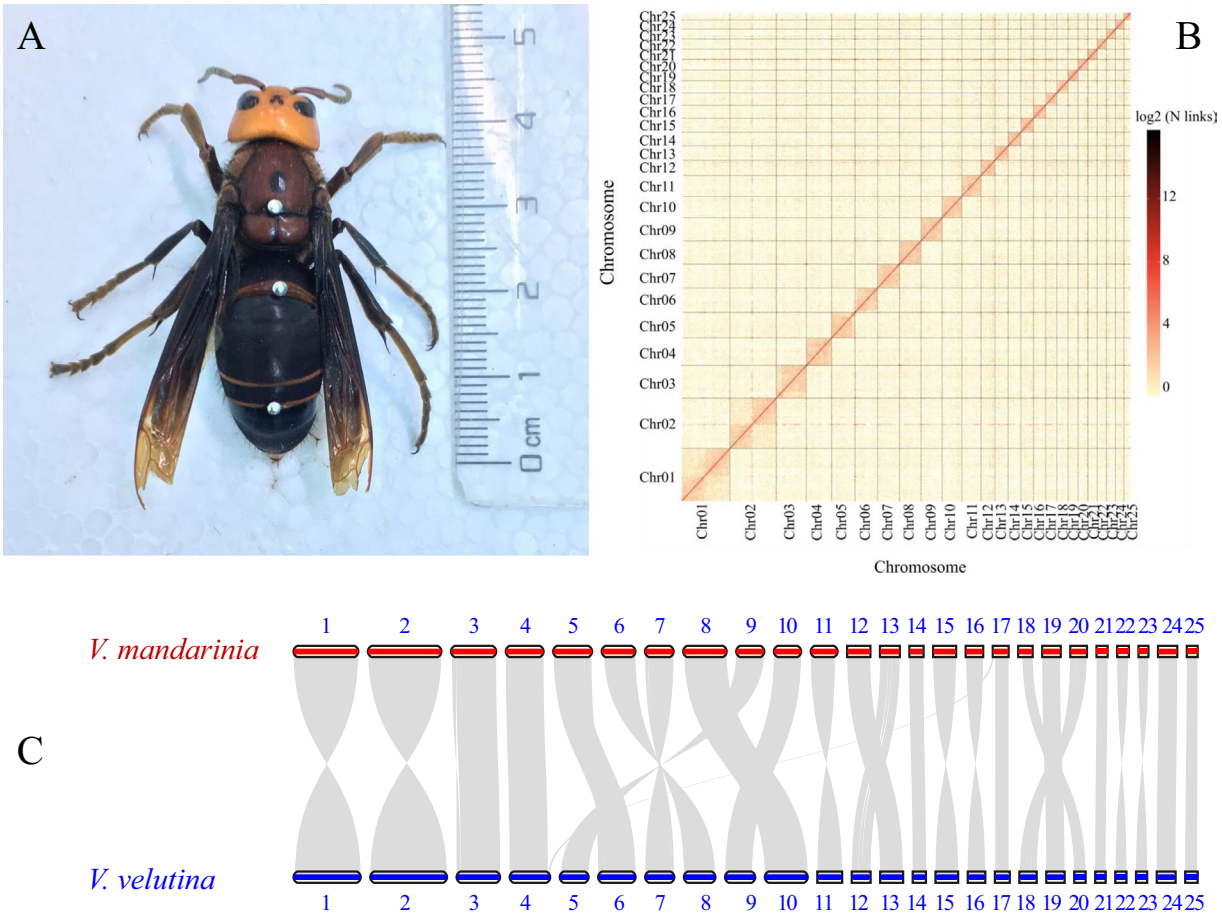

Fig. 1 The wasp *Vespa mandarinia* and its genome characteristics. (**A**) The picture of *V. mandarinia* with a ruler showed the body size. (**B**) HIC chromosome interaction map showed the 25 chromosomes pseudo-chromosomes. (**C**) Collinearity analysis result between *V. Insecta* and *V. velutina*.

genes were predicted. The study provided information for research on the sociality origin of Vespinae subfamily and serves as a foundational resource for managing *V. mandarinia*'s ecological risks, while facilitating further research on venom.

Methods

Sample collection and sequencing. A male individual of *V. mandarinia* from Yunnan Province, China, was selected for genomic sequencing. Tissue samples comprising the head, thorax, and legs were dissected used for second-generation sequencing (DNBSEQ), third-generation sequencing (Oxford Nanopore), and Hi-C sequencing. Genomic DNA was fragmented (300–400 bp) using Covaris (Covaris LE220, Covaris, Woburn, USA), size-selected with Agencourt AMPure XP-Medium Kit (Beckman Coulter, California, USA), and processed through end repair, adenylation, and adapter ligation. PCR-amplified libraries were purified and converted into DNA nanoballs (DNBs) for sequencing on the DNBSEQ platform, yielding about 24.03 Gb clean data (80.09 million reads) with 98.41% Q20 and 30.93% GC content. 20 kb DNA fragments were size-selected via BluePippin, end-repaired, and ligated to sequencing adapters. Libraries were sequenced by Nanopore platform. 54.06 Gb of filtered data (3.19 million reads) was obtained with reads N50 of 19,761 bp. Chromatin was crosslinked with formaldehyde, digested with MboI, and ligated after biotin labeling. DNA was sheared to 400 bp for library construction and sequencing. The DNBSEQ-T7 platform (BGI) sequencing produced about 24.27 Gb of paired-end reads (80.90 million clean reads) with 96.86% Q20.

Genome assembly. The Nanopore reads was initially assembled using NextDenovo (v2.5). The assembled genome was subsequently rectified by the second-generation sequencing data using Pilon (v1.24). The raw genome data was about 200.4 Mb and the longest sequence was 16.58 Mb. For chromosomal scaffolding, Hi-C reads were used for auxiliary genome assembly. The valid pair read was mapped to the genome to construct the interaction map[21]. Contigs were clustered, ordered, and oriented into pseudo-chromosomes[22,23]. Final assembly genome was polished by DNBSEQ datasets using Pilon (v1.24)[24]. The chromosome-level assembly comprising 25 pseudochromosomes totaling 191.80 M, representing 95.60% of the total genomic sequences (Tables 1,2). The contig N50 was 9.57 M (Fig. 1B; Tables 1,2). BUSCO assessment result showed complete BUSCOs was 99.5% (Fig. 2A).

Statistic	Contig Length (bp)	Number of Contigs
Max length	16595814	
N90	3687391	23
N80	4653447	18
N70	5950160	14
N60	9078813	11
N50	9568788	9
N40	9851539	7
N30	10341143	5
N20	10465094	3
N10	15056922	2
Total length	200616667	
number > = 100 bp		64
number > = 2000bp		64

Table 1. Statistics result of the assembled *Vespa mandarinia* genome.

ChrName	Length (bp)
Chr01	20,673,010
Chr02	19,847,621
Chr03	12,922,423
Chr04	10,642,621
Chr05	10,112,509
Chr06	9,580,673
Chr07	9,331,938
Chr08	9,224,667
Chr09	9,131,832
Chr10	8,383,776
Chr11	8,316,283
Chr12	5,959,961
Chr13	5,497,135
Chr14	5,487,616
Chr15	5,359,751
Chr16	5,281,391
Chr17	4,835,482
Chr18	4,657,750
Chr19	4,277,920
Chr20	4,160,515
Chr21	4,079,524
Chr22	4,048,532
Chr23	3,783,699
Chr24	3,654,485
Chr25	2,805,912

Table 2. The length of 25 chromosomes of *V. mandarinia* genome.

Transcriptome sequencing. To establish a comprehensive transcriptomic profile of *V. mandarinia*, tissues and developmental stages were sampled: one queen wasp, one male pupa, adult, one female pupa, adult, one female head, ten antennae harvested from five female adults, and alimentary canal. The DNBSEQ short-reads enabled high-depth gene expression quantification. The eight libraries from the eight samples were constructed separately and paired-end sequenced (150 bp) by DNBSEQ, yielding more than 43 M raw reads per sample.

The PacBio long-reads resolved full-length isoforms from pooled RNA. A mixture of total RNA from the eight samples mentioned above, larvae at different stages, and pupae at early and late stages were used for SMRTbell library construction. The library was sequenced via the Sequel System (PacBio Sequel, PacBio, USA, MenloPark). A total of 734,980 reads (63.17 G) were generated, and 665,005 CCS were obtained.

Gene prediction. RepBase database (v23.06) and RepeatMasker (v4.1.1) were used to detect conserved transposable elements (TEs). In addition, ltr_finder and RepeatModeler were used for *de novo* prediction.

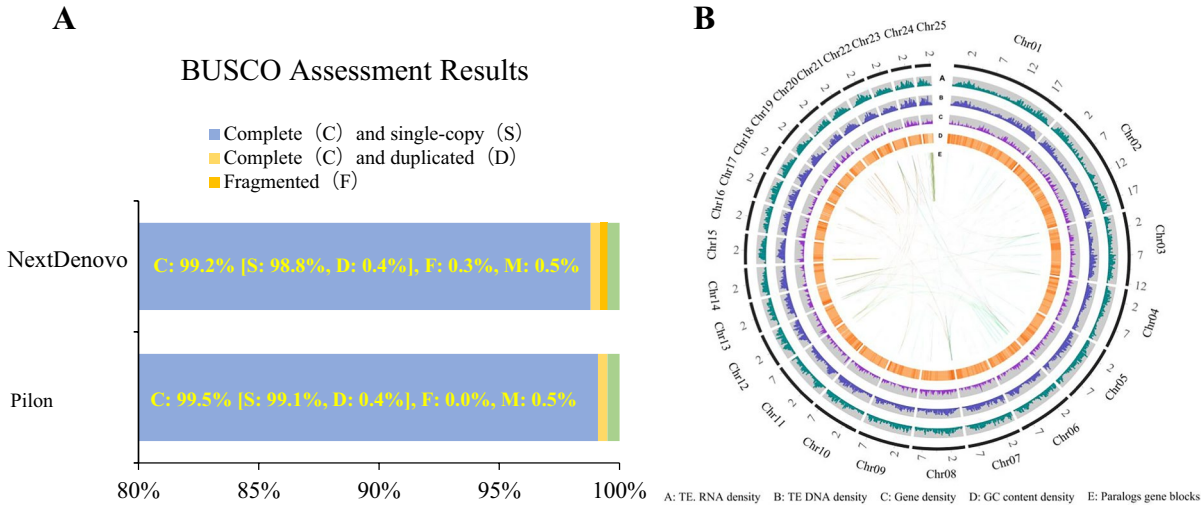

Fig. 2 The BUSCO assessments and genome circus map of *V. mandarinia* genome. (**A**) BUSCO assessment results of preliminarily assembled genome and polished genome of *V. mandarinia*. (**B**) The circus map showed TE RNA density, TEs DNA density, gene density, GC content density, and paralog blocks.

Comprehensive annotation revealed that TEs account for 28.10% of the *V. mandarinia* genome (Supplemental Table S1). Retrotransposons accounted for 3.40% of the genome and 4.30% were DNA transposons.

The transcriptome sequencing reads of *V. mandarinia* were aligned to the genome using STAR. The annotated genome of seven insects (*Drosophila melanogaster*, *V. mandarinia*, *Apis mellifera*, *Vespula pensylvanica*, *Vespula vulgaris*, *Vespa crabro*, and *Vespa velutina*) were selected for homologous gene prediction using Exonerate within the MAKER3 (v3.01.03). Initial MAKER predictions yielded 1,000 high-confidence genes, which were used for SNAP (v2013-11-29) and Augustus (v3.4.0) training. The training datasets were used for the second around annotation by MAKER. Final integration of transcriptomic and homology prediction identified 12,023 protein-coding genes.

Noncoding RNA (ncRNA) elements were identified using a multi- approach. Ribosomal RNAs (rRNAs), small nuclear RNAs (snRNAs), and microRNAs (miRNAs) were predicted by aligning the genome to the Rfam database using Infernal (v1.1.4). Transfer RNAs (tRNAs) were annotated using tRNAscan-SE (v2.0.12) with default parameters. Ribosomal RNA subunits were further identified using RNAmmer. The final annotation identified 189 tRNAs, 135 rRNAs, 58 miRNAs and 10 snRNAs.

Functional annotations of predicted genes were achieved through systematic alignment against eight major biological databases: NR (Nonredundant Protein Database), KEGG (Kyoto Encyclopedia of Genes and Genomes), COG (Clusters of Orthologous Groups), GO (Gene Ontology), Pfam (Pfam Protein Families Database), and Swiss-Prot databases. To obtain toxin-specific annotations, iVenomDB (Integrated Venom Database) andT3DB (The Toxin and Toxin-Target Database) databases were also used. A total of 11,541 genes (99.84%) received functional annotations. Classification coverage across key database as following: NR (11,485; 99.36%), Swiss-Prot (10,331; 89.38%), iVenomDB (100; 0.87%), T3DB (2,476; 21.42%).

Synteny analysis. Synteny analysis using JCVI (v1.1.19) revealed extensive genomic conservation between *V. mandarinia* and *V. velutina* (Fig. 2B). Chr04 of *V. velutina* was syntenic with Chr04 and Chr17 of *V. mandarinia*, which indicated that translocation took place at Chr04 and Chr17 of *V. mandarinia* during evolution (Fig. 1C).

Data Record

The genomics data was submitted to an INSDC database (NLM-NCBI) with accession number GCA_050613455.1[25], and BioProject number PRJNA1165714. The transcriptome data was submitted to the NCBI BioProject database with SRA accession numbers of SRX29056093- SRX29056100 (https://identifiers.org/ncbi/insdc.sra:SRP589988)[26].

Technical Validation

In order to evaluate the integrity of assembly, BUSCO assessment was carried out to assess the genome assembly results[27]. BUSCO assessment result showed complete BUSCOs was 99.5%. The complete and single-copy BUSCOs was 95.1%. The missing BUSCOs was about 0.5% (Fig. 2A). BUSCO assessment of the functional annotation within the final gene set found that complete BUSCOs was 98.5% (Supplemental Table S2). These results confirmed the high quality and integrity of the genome assembly. The mapping rate of DNBSEQ reads to the *V. mandarinia* genome was 99.92%. *V. mandarinia* and *V. velutina* showed a high level of collinearity (Fig. 1C). The information revealed that this sequencing was of high quality.

Code availability

The bioinformatics analysis was conducted by software developers using the documentation and blueprint. The software version and method used are described in the method.

Received: 4 June 2025; Accepted: 31 July 2025;
Published online: 11 August 2025

References

1. Dong, S., Sun, A., Tan, K. & Nieh, J. C. Identification of giant hornet Vespa mandarinia queen sex pheromone components. *Curr Biol.* **32**(5), R211–R212, https://doi.org/10.1016/j.cub.2022.01.065 (2022).
2. Wen, P. *et al*. The sex pheromone of a globally invasive honey bee predator, the Asian eusocial hornet, Vespa velutina. *Sci Rep.* **7**(1), 12956, https://doi.org/10.1038/s41598-017-13509-7 (2017).
3. Nuñez-Penichet, C. *et al*. Geographic potential of the world's largest hornet, *Vespa mandarinia* Smith (Hymenoptera: Vespidae), worldwide and particularly in North America. *PeerJ.* **9**, e10690, https://doi.org/10.7717/peerj.10690 (2021).
4. Li, Z. The discovery of the world's largest individual wasp. *Discov Nat.* **000**(006), 10–20 (2019).
5. Kapheim, K. M. Nutritional, endocrine, and social influences on reproductive physiology at the origins of social behavior. *Curr Opin Insect Sci.* **22**, 62–70, https://doi.org/10.1016/j.cois.2017.05.018 (2017).
6. McGruddy, R. A. *et al*. Invasive paper wasps have strong cascading effects on the host plant of monarch butterflies. *Ecol Entomol.* **46**(2), 459–469, https://doi.org/10.1111/een.12992 (2021).
7. Moreyra, S., Lozada, M., Fernández-Arhex, V. & Pietrantuono, A. L. *Polistes dominula* (Hymenoptera: Vespidae) paper wasps detect native grasshoppers using olfactory cues. *Austral Ecol.* **48**(8), 2041–2052, https://doi.org/10.1111/aec.13442 (2023).
8. Rojas-Nossa, S. V. *et al*. Predator and pollinator? An invasive hornet alters the pollination dynamics of a native plant. *Basic Appl Ecol.* **71**, 119–128, https://doi.org/10.1016/j.baae.2023.07.005 (2023).
9. Zhu, G., Gutierrez Illan, J., Looney, C. & Crowder, D. W. Assessing the ecological niche and invasion potential of the Asian giant hornet. *Proc Natl Acad Sci.* **117**(40), 24646–24648, https://doi.org/10.1073/pnas.2011441117 (2020).
10. Kim, W. M., Kim, S. Y. & Song, W. Microhabitat characteristics affecting the occurrence and diversity of queen hornets (genus Vespa) in an urban green area. *Landsc Ecol Eng.* **16**(2), 173–186, https://doi.org/10.1007/s11355-020-00415-w (2020).
11. Freeman, A. & Xia, X. Phylogeographic reconstruction to trace the source population of Asian giant hornet caught in Nanaimo in Canada and Blaine in the USA. *Life.* **14**(3), 283, https://doi.org/10.3390/life14030283 (2024).
12. Wilson, T. *et al*. Insights into the prey of Vespa mandarinia (Hymenoptera: Vespidae) in Washington state, obtained from metabarcoding of larval feces. *Front Insect Sci.* **3**, 1134781, https://doi.org/10.3389/finsc.2023.1134781 (2023).
13. Otis, G. W., Taylor, B. A. & Mattila, H. R. Invasion potential of hornets (Hymenoptera: Vespidae: Vespa spp.). *Front Insect Sci.* **3**, 1145158, https://doi.org/10.3389/finsc.2023.1145158 (2023).
14. Li, H. *et al*. Antennal sensilla of two subspecies of Vespa mandarinia (Smith) (Hymenoptera:Vespidae) observed with scanning electron microscopy. *J Fujian Agric Univ Nat Sci Ed.* **50**(04), 466–471 (2021).
15. Tan, J. L., van Achterberg, C., Chen, X. X. Potentially lethal social wasps. Fauna of the Chinese Vesplnae (Hymenoptera: Vespidae). Science Press: Beijing, China (2015).
16. Feng, Y., Chen, X. M. C. & Ye, S. D. The common edible species of wasps in Yunnan and their value as food. *Rorest research* **145**, 578–581 (2001).
17. Couto, A., Arnold, G., Ai, H. & Sandoz, J. C. Interspecific variation of antennal lobe composition among four hornet species. *Sci Rep.* **11**(1), 20883, https://doi.org/10.1038/s41598-021-00280-z (2021).
18. Suenami, S., Konishi Nobu, M. & Miyazaki, R. Community analysis of gut microbiota in hornets, the largest eusocial wasps, Vespa mandarinia and V. simillima. *Sci Rep.* **9**(1), 9830, https://doi.org/10.1038/s41598-019-46388-1 (2019).
19. Carriazo, S. & Ortiz, A. Wasp stings and plasma exchange. *Clin Kidney J.* **15**(8), 1455–1458, https://doi.org/10.1093/ckj/sfac055 (2022).
20. Mattila, H. R., Nguyen, L. T. P., Perrard, A., Bain, M. & Otis, G. W. Biology of the southern giant hornet, Vespa soror: nest architecture, morphological differences among castes, and the genetic structure of colonies. *Front Insect Sci.* **3**, 1136297, https://doi.org/10.3389/finsc.2023.1136297 (2023).
21. Durand, N. C. *et al*. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**(1), 95–98, https://doi.org/10.1016/j.cels.2016.07.002 (2016).
22. Dudchenko, O. *et al*. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science.* **356**(6333), 92–95, https://doi.org/10.1126/science.aal3327 (2017).
23. Servant, N. *et al*. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**(1), 259, https://doi.org/10.1186/s13059-015-0831-x (2015).
24. Walker, B. J. *et al*. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. Wang J., ed. *PLoS ONE.* **9**(11),e112963. https://doi.org/10.1371/journal.pone.0112963 (2014).
25. Yang, P. *Genbank* https://identifiers.org/ncbi/insdc.gca:GCA_050613455.1 (2025).
26. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRP589988 (2025).
27. 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. Kelley J., ed. *Mol Biol Evol.* **38**(10),4647–4654. https://doi.org/10.1093/molbev/msab199 (2021).

Acknowledgements

The work was supported by the Chinese Academy of Forestry Outstanding Youth Support Program [grant number CAFYBB2023QB007]; the Young- and Middle-Aged Academic and Technical Leaders Reserve Talent Project of Yunnan Province [grant number 202105AC160031].

Author contributions

Xiao-Yun Niu: drafted the manuscript (lead), analyzed the data (supporting), made some figures (lead). Yuan-Chong Shi: drafted part of the manuscript (supporting), analyzed the data (supporting). Hai-Feng Mo: drafted part of the manuscript (supporting). Pu Yang: conceived and designed the experiments, analyzed the data (lead), edited and reviewed 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/s41597-025-05749-4.

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

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.