



OPEN

DATA DESCRIPTOR

Chromosome-scale genome assembly of the dune cricket *Schizodactylus jimo*

Zhengzhong Huang^{1,5}, Le Zong^{1,5}, Zequn Dong^{1,2}, Zhiwei Dong³, Jinwu He³, Congqiao Li¹, Xueyan Li^{3,4} & Siqin Ge^{1,2}

The dune cricket *Schizodactylus jimo* (Orthoptera: Schizodactylidae) is a unique insect because of its folded forward wings. Despite its captivating wing morphology and potential bionic significance, the genomic basis underlying the evolutionary and ecological significance of its wing folding is still lacking. Here, we construct a high-quality chromosome-level genome assembly of *S. jimo*, utilizing PacBio HiFi long reads, Illumina short reads, and Hi-C sequencing data. The assembled genome is 1.33 Gb in length, with a scaffold N50 of 216.2 Mb and nine pseudochromosomes. The assembly demonstrates 98.1% completeness based on insect universal single-copy orthologs (BUSCO). Comprehensive annotation identified 12,612 protein-coding genes, extensive repeat elements (45.7%), and diverse non-coding RNA genes. This genomic dataset provides a valuable resource for understanding the evolutionary adaptations and ecological specialization of the dune cricket with folded forward wings, and facilitates future comparative genomics and functional studies within Orthoptera and other sand-dwelling insects.

Background & Summary

The family Schizodactylidae Brullé, 1835 represents one of the most morphologically distinctive lineages within the suborder Ensifera (e.g. crickets, katydids, and their allies). Members of this family are easily recognized by their conspicuously expanded, lobiform tarsi and are uniquely adapted to psammophilous (sand-dwelling) environments¹. They are nocturnal burrowers, remaining concealed in self-excavated tunnels during the day and emerging at night for feeding and activity^{2,3}. Extant diversity within Schizodactylidae is extremely limited, comprising only 20 described species⁴ in two genera: *Comicus* (11 species) and *Schizodactylus* (9 species). Both genera are restricted to dynamic dune environments characterized by poor vegetation cover and high wind activity, which continuously reshapes their habitats^{5,6}. Even so, the two genera exhibit a disjunct biogeographic pattern. *Comicus* is confined to the arid sand-dune systems of southern Africa, including the Namib and Kalahari deserts, and all species are wingless throughout all developmental stages⁷. *Schizodactylus*, on the other hand, is mainly distributed across semi-stable or shifting dunes in South and Southeast Asia countries such as China, India, Pakistan, Sri Lanka, and Myanmar, with only one brachypterous species, *Schizodactylus inexpectatus*, endemic to Turkey^{8,9}. Very interestingly, all South and Southeast Asian species are characterized by fully developed or reduced wings. The family has a deep evolutionary history, with fossil representatives dating back to the Lower Cretaceous, although phylogenetic evidence suggests an even more ancient origin¹⁰.

Among the two extant genera, detailed morphological investigations of the wing-bearing thoracic structures have only been conducted for the species of *Comicus*¹¹, but no related published data are currently available for *Schizodactylus*. Most species within *Schizodactylus* exhibit a particularly distinctive wing-folding morphology that sets them apart from other orthopterans (Fig. 1). Although only limited molecular data, such as partial mitochondrial genomes or short gene fragments, have been reported for certain species of Schizodactylidae and related orthopterans¹², these resources remain too sparse and fragmented to elucidate the genetic mechanisms underlying their unique morphological traits or to establish a robust phylogenomic framework for the

¹State Key Laboratory of Animal Biodiversity Conservation and Integrated Pest Management, Institute of Zoology, Chinese Academy of Sciences, Beijing, 100101, China. ²University of Chinese Academy of Sciences, Beijing, 100049, China. ³State Key Laboratory of Genetic Evolution & Animal Models, Chinese Academy of Sciences, Kunming, 650223, China. ⁴Yunnan Key Laboratory of Biodiversity Information, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, 650223, China. ⁵These authors contributed equally: Zhengzhong Huang, Le Zong. ✉e-mail: lixy@mail.kiz.ac.cn; gesq@ioz.ac.cn



Fig. 1 *Schizodactylus jimo* He & Liu, 2021. (a) Dorsal view of specimen showing the right wing fully extended and the left wing in a folded position. The scale bar is 2 cm. (b) Field photograph of a living individual in its natural dune habitat, with both wings fully folded over the abdomen.

group. In China, the family Schizodactylidae was undocumented until the recent discovery of a new species, *Schizodactylus jimo*, in Yunnan Province in 2021⁹. To investigate its evolutionary significance and ecological adaptations, we assembled a high-quality chromosome-level genome for *S. jimo*. Comprehensive annotation was performed, including the identification of repetitive elements, non-coding RNA genes, and protein-coding genes. This genomic assembly provides resources for investigating the evolutionary trajectory, habitat specialization, and functional genomics of Schizodactylidae, and serves as a foundation for future comparative and ecological studies of dune-adapted orthopterans.

Methods

Sample information. *Schizodactylus jimo* adult individuals were collected from sandy dune habitats at the riverside of the Nujiang River, Mangkuan Township, Baoshan City, Yunnan Province, China, during 2022–2024 (98.8891°E, 25.4412°N, alt. 735 m) by Zhiwei Dong and Zhengzhong Huang. Two male adults were dissected to remove guts and mouthparts to reduce microbial contamination and then their tissue samples were flash-frozen in liquid nitrogen and stored at -80°C until use. The tissue of one male adult was used for Illumina, PacBio HiFi, and Hi-C sequencing. The tissue of another individual was used for RNA sequencing. Voucher specimens were deposited in both Insect Collection of Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China (KIZ 0139654) and Institute of Zoology, Chinese Academy of Sciences, China (IOZ 20250710).

Illumina, PacBio HiFi, Hi-C, and RNA sequencing. Genomic DNA was extracted from one adult male individual using a CTAB-based protocol optimized for insect tissues to obtain high-molecular-weight DNA. DNA degradation and contamination were monitored on 1% agarose gels. DNA purity was checked using the NanoPhotometer[®] spectrophotometer (IMPLEN, CA, USA). DNA concentration was measured using Qubit[®] DNA Assay Kit in Qubit[®] 2.0 Fluorometer (Life Technologies, CA, USA).

Short-read libraries were generated using the TruSeq Nano DNA HT Sample Preparation Kit (Illumina, USA) following the manufacturer's protocol. The libraries were then sequenced on an Illumina NovaSeq 6000 platform, generating 150 bp paired-end reads, with post-sequencing validation confirming an average insert size of ~350 bp.

PacBio HiFi read libraries were prepared using the SMRTbell[®] prep kit 3.0 according to the manufacturer's instructions. After DNA quality control, DNA shearing and cleanup, DNA repair and A-tailing reactions, adapter ligation and cleanup, nuclease treatment, size selection, and sequencing was performed using the PacBio Revio system according to the manufacturer's operating manual.

Hi-C libraries were also generated using the TruSeq Nano DNA HT Sample Preparation Kit (Illumina, USA) following the manufacturer's protocol. After cross-linking, cells were lysed and digested with the restriction enzyme DpnII, followed by end repair, DNA cyclization and DNA purification. Finally, the Hi-C libraries were sequenced on the Illumina NovaSeq 6000 platform.

RNA sequencing libraries were prepared using the TruSeq RNA Library Preparation Kit (Illumina, USA), and then sequenced on an Illumina NovaSeq 6000 with 150 bp paired-end reads.

All sequencing was performed at NextOmics (Wuhan, China). In total, the sequencing effort produced 83.51 Gb of PacBio HiFi reads, 110.19 Gb of Hi-C reads, 327.65 Gb of Illumina reads and 11.71 Gb of RNA-seq reads (Table 1).

Genome survey and assembly. Before assembling the *S. jimo* genome, raw sequencing reads were processed with fastp v0.23.4 under stringent parameters (-q 20 -D -g -x -u 10 -5 -r -c) to remove low-quality

Library type	Platform	Average read length (bp)	Raw data (Gb)	Depth (×)
Short-reads	Illumina	150	327.65	245.53
HiFi-reads	PacBio Revio	17,944	83.51	62.58
Hi-C-reads	Illumina	150	110.19	82.57
RNA-seq	Illumina	150	11.71	—

Table 1. Summary of sequence reads.

Content	Values
Assembly size (bp)	1,334,471,050
Number of pseudo-chromosomes (sizes)	9 (1,229,291,788 bp)
Number of scaffolds/contigs	111 / 164
Longest scaffold/contig (Mb)	275.066 / 139.074
N50 scaffold/contig length (Mb)	216.223 / 45.883
GC content (%)	39.82
BUSCO Completeness (%)	98.1
S	96.4
D	1.7
F	0.4
M	1.5

Table 2. Genome assembly summary.

bases, duplicates, and artifacts, thereby reducing sequencing-level contamination, genome survey analysis was performed using a 21-mer frequency distribution of short-reads generated with the script “kit.sh” in BBTools (v38.90)¹³. The genome size was estimated to be 1,288,053,473 bp (i.e., 1.29 Gb) with 1.04% heterozygosity and 63.41% repetitive sequences.

High-quality HiFi reads (Q20 or higher) were assembled for the draft genome by Hifiasm (v0.24.0)¹⁴. In the depth-based filtration during assembly, with the ‘-l 3’ parameter to explicitly retain contigs having a sequencing depth above 3× and discard very low-depth contigs (<3×), which are highly likely to represent foreign contamination or assembly errors. Haplotypic duplication sequences were removed using Purge_dups (v1.2.5)¹⁵ pipeline with the default parameters (‘-2 -a 70’). These two steps were applied sequentially to address distinct problems. The first step performs depth-based elimination of contamination, and the second step removes haplotigs in order to obtain a non-redundant primary haplotype assembly. To further construct the chromosome-level genome, Hi-C data were subsequently mapped to these contigs using Chromap (v0.2.6)¹⁶. The uniquely mapped reads were then used to anchor the contigs onto the pseudochromosomes using YAHS (v1.2) software¹⁷. Manual correction of misjoins was conducted in Juicebox (v1.11.08)¹⁸. Explicit contamination screening of the final assembly was systematically performed using MMseqs2 v13 with a blastn-like search strategy against two comprehensive databases—the NCBI nt database and the UniVec database for common vectors and contaminants. This process includes an integrated step wherein NCBI’s Foreign Contamination Screen (FCS) tool is employed to identify and filter out contaminant sequences.

The final chromosome-level genome is 1.33 Gb, with scaffold and contig numbers of 111 and 164, closely matches our 1.29 Gb preliminary estimate, supporting the robustness of both estimation and assembly methods. The longest scaffold/contig length is 275.07/139.07 Mb, and the scaffold/contig N50 length is 216.22/45.88 Mb. The assembly exhibits high continuity, as supported by the metrics in Table 2. The GC content was 39.82%. A total of 9 pseudochromosomes (1.23 Gb) were assembled, with an anchoring rate of 92.12% (Fig. 2).

The sequencing depth of each chromosome was assessed using the Minimap2 (v2.29)¹⁹ and SAMtools (v1.10)²⁰. With referring to the sex determination mechanism of male in the suborder Ensifera being XO^{21,22}, in which the sequencing depth of male X chromosome is theoretically only half of that of autosomes, chromosome 2 is identified as X chromosome (Table 3). The single-base QV was assessed using Merqury v1.3¹⁴, with values around 60 for each chromosome, corresponding to an estimated error rate of approximately 1×10^{-7} . These results indicate that the assembly achieves near single-base accuracy while retaining the structural completeness provided by long-read sequencing.

Genome completeness was assessed using BUSCO v5.7.1²³ against the Insecta_odb10 database (1,367 genes), revealing 98.1% completeness with only 1.7% multi-copy BUSCOs, indicating minimal assembly redundancy. The mapping ratio of short-reads, HiFi-reads, and RNA-seq reads was 99.01%, 98.99% and 97.04%, respectively, indicating extremely high sequencing quality. All the above data indicate that the assembled genome has reached an extremely high level of continuity and integrity.

Repeat annotation. The repeat sequence library was constructed based on the principle of repeat sequence-specific structure and *de novo* prediction by using RepeatModeler v2.0.5²⁴ with LTR search process (‘-LTRStruct’). This library was then merged with the Dfam 3.816²⁵ and RepBase-2018102617²⁶ databases to form the final repeat sequence reference database.

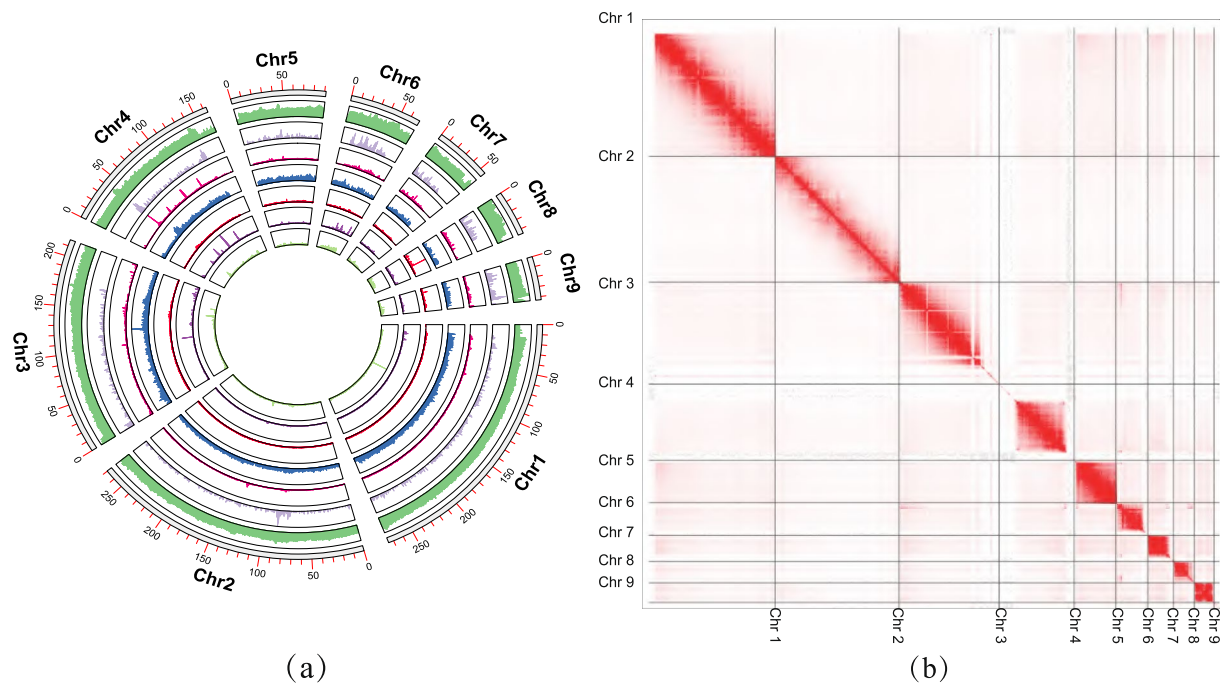


Fig. 2 Assembly of chromosome-level genome of *Schizodactylus jimo*. **(a)** Circos plot with a window size of 100k bp. **(b)** The heatmap displays all interactions among 9 chromosomes. The outer to inner circle of **(a)** were chromosome length, GC content, gene density, DNA transposon density, SINE transposon density, LINE transposon density, LTR transposon density, and Simple repeats density, respectively.

Chromosome	Length (bp)	HiFi (×)	Illumina (×)	QV	Error rate
Chr1	275,066,161	67.1946	262.186	61.8767	6.49121e-07
Chr2	270,227,376	35.8213	134.227	58.559	1.39348e-06
Chr3	216,222,903	76.3587	285.07	62.3625	5.80426e-07
Chr4	163,219,514	84.0188	297.799	62.3849	5.7745e-07
Chr5	91,347,396	66.4627	257.042	60.8941	8.13934e-07
Chr6	69,093,254	113.98	443.72	60.14	9.68276e-07
Chr7	55,978,253	69.1737	409.634	62.0831	6.18996e-07
Chr8	44,905,194	64.1839	288.39	59.0565	1.24265e-06
Chr9	43,231,737	46.6943	249.445	56.3232	2.33176e-06

Table 3. Genome assembly summary of length, sequencing coverage and QV value for each chromosome.

The repeat sequence content of *S. jimo* was annotated using RepeatMasker v4.1.5²⁷ and the custom repeat library described above. This analysis identified 609.87 Mb of repetitive sequences, accounting for 45.74% of the total genome, which is a typical medium to high repetitive genome (Table S1).

The top six types of repetitive sequences in terms of proportion are: LINE (23.30%), Unknown (11.85%), SINEs (3.10%), Simple repeats (2.66%), DNA (2.97%), and LTRs (0.59%). As can be seen from Fig. 3, the transposons in the genome are mainly caused by two expansions. The distant expansion (the peak on the right) is mainly contributed by Unknown repeats, while the recent expansion is mainly formed by LINE repeats (the peak on the left) (Fig. 3).

Non-coding RNA genes. Non-coding RNA annotation was performed by comparing sequences with the Rfam database using Infernal v1.1.5²⁸ to identify various ncRNAs including rRNA, snRNA and miRNA, while tRNA prediction was conducted using tRNAscan-SE v2.0.12²⁹ with low-confidence tRNAs filtered by the built-in EukHighConfidenceFilter script.

The genomic non-coding RNA annotation identified a total of 5,395 ncRNAs, consisting of 1,699 rRNAs, 74 miRNAs, 1,306 snRNAs (including 1,289 spliceosomal RNAs [U1, U2, U4, U5, U6, U11], 2 minor spliceosomal RNAs [U4atac, U6atac], 12 C/D box snoRNAs, and 1 HACA-box snoRNA), 1,475 tRNAs, 4 ribozymes, and 2 lncRNAs (Table S2).

Gene structure prediction. The protein-coding genes (PCGs) of the repeat-masked genome were predicted based on *de novo* prediction, homology prediction, and RNA-seq prediction. For *de novo* prediction, BRAKER³⁰

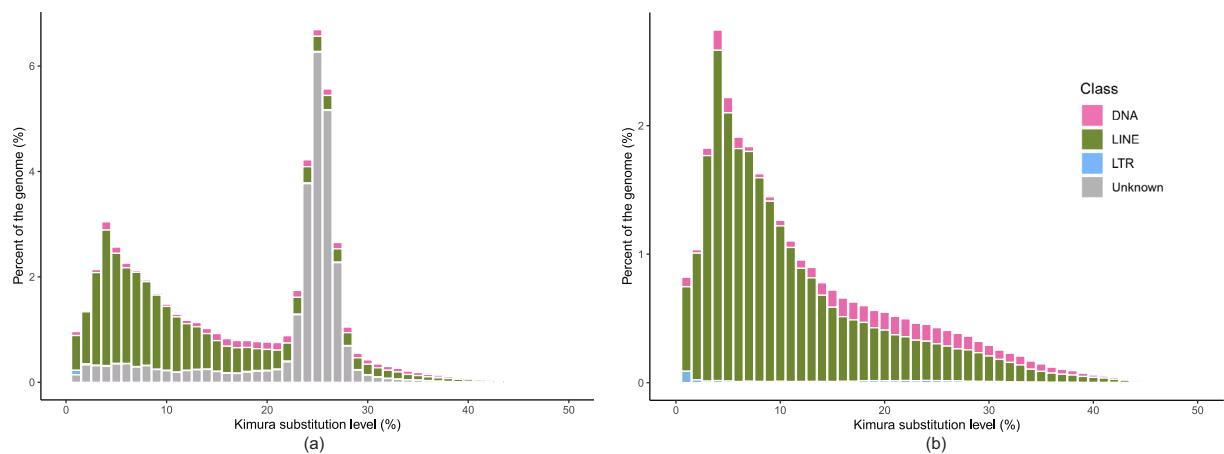


Fig. 3 Interspersed repeat landscape of the *S. jimo* genome based on Kimura 2-parameter (K2P) divergence (CpG-adjusted). **(a)** Repeat landscape of all annotated repeat classes, including DNA transposons, LINEs, LTRs, and unclassified/unknown repeat elements. The x-axis represents K2P divergence, while the y-axis shows the proportion of the genome contributed by each repeat class. Peaks indicate periods of transposable element expansion. **(b)** Same K2P landscape with unclassified/unknown repeat elements removed to highlight the dynamics of well-annotated TE families.

Structure annotation	Number
Number of protein-coding genes	12,612
Number of predicted protein sequences	17,609
Mean protein length (aa)	627
Mean gene length (bp)	27,906.60
Gene ratio	26.40%
Number of exons per gene	8.7
Mean exon length (bp)	355.6
Exon ratio	2.94%
Number of CDSs per gene	8.3
Mean CDS length (bp)	209.8
CDS ratio	1.66%
Number of introns per gene	7.7
Mean intron length (bp)	3,463.40
Intron ratio	23.46%

Table 4. Summary of gene structure prediction.

automatically trained two *ab initio* prediction tools, Augustus3.4.0³¹ and GeneMark-ETP³², and integrated arthropod protein sequences (extracted from the OrthoDB12v1 database³³) along with transcriptome data to enhance prediction accuracy. (The transcriptome alignment was performed by mapping the RNA-seq short-read transcriptome data to the genome using minimap2 to generate BAM alignment files with the parameter ‘-x splice:sr’). For homology-based prediction, protein datasets from the holometabolous insect *Drosophila melanogaster* (Diptera, GCA_000001215.4) and four neopteran insects: *Anabrus simplex* (Orthoptera, GCA_040414725.1), *Schistocerca nitens* (Orthoptera, GCA_023898315.2), *Bacillus rossius redtenbacheri* (Phasmatodea, GCA_032445375.1), and *Periplaneta americana* (Blattodea, GCA_040183065.1) were used in GeMoMa (v1.9; parameters: GeMoMa.c = 0.4, GeMoMa.m = 130000)³⁴ to perform gene prediction based on sequence homology. For RNA-seq prediction, we aligned RNA-seq using minimap2^{19,35} (‘-x splice:sr’) and assembled unigenes using StringTie v2.2.1³⁶, and the coding sequences were identified using GeneMark-ETP.

We predicted a total of 12,612 PCGs (Table 4 and Table S3) with an average gene length of 27,906.6 bp in the genome by MAKER (v3.01.04)³⁷. On average, each gene contained 8.7 exons (355.6 bp each), 7.7 introns (3,463.4 bp each), and 8.3 coding sequences (209.8 bp each). The integrity of the gPCGs evaluated by BUSCO software (v5.7.1)²³ was 97.9%, which is close to 98.1% of the genomic evaluation. It can be seen that the annotation quality of protein coding genes is excellent.

Gene functional prediction. Functional annotation of the predicted proteins was performed by aligning against UniProtKB v202503 (SwissProt + TrEMBL) using DIAMOND v2.1.7.161³⁸ (--very-sensitive, --evalue 1e-5), searching the Pfam³⁹ database via InterProScan 5.70-102.0⁴⁰, and annotating with eggNOG-mapper

Function annotation	Number of genes
matching Uniprot records	11,761
labelled as “Uncharacterized protein”	820
labelled as “unknown function”	929
InterProScan annotations	10,438
GO items from InterProScan annotations	6,289
eggNOG annotations	11,610
GO items from eggNOG annotations	8,895
Enzyme Codes (EC) from eggNOG annotations	2,610
KEGG ko terms from eggNOG annotations	7,811
KEGG pathway terms from eggNOG annotations	4,738
COG Functional Categories from eggNOG annotations	10,887
GO items (combining InterProScan and eggNOG results)	9,921

Table 5. Summary of gene functional prediction.

v2.1.12⁴¹ against eggNOG v5.0.2⁴² to predict conserved domains, Gene Ontology (GO) terms, and pathways (KEGG, Reactome).

The result showed that 11761 (93.25%) genes in the genome matched records from the UniProtKB database. InterPro identified the protein domains of 10438 protein coding genes. InterPro and eggNOG mapper identified 9921 GO genes and 4738 KEGG pathway entries (Table 5).

Data Records

The dataset is available at the NCBI under the BioProject PRJNA1279776⁴³. RNA-seq data were deposited in the Sequence Read Archive (SRA) at NCBI under accession number SRR34590780⁴⁴. PacBio HiFi long-read sequencing data were deposited in the SRA under accession numbers SRR34590781⁴⁵. Hi-C sequencing data were deposited in the SRA under accession numbers SRR34590782⁴⁶. Illumina sequencing data were deposited in the SRA under accession numbers SRR34590783⁴⁷. The assembled genome is available at Genbank⁴⁸. The genome assembly and genome annotation results have additionally been deposited in the Figshare repository under the <https://doi.org/10.6084/m9.figshare.29880239>⁴⁹.

Technical Validation

The completeness and integrity of the chromosome-level genome assembly for *S. jimo* were assessed using two complementary approaches. First, all original short-read and long-read sequencing data were aligned back to the assembly using Minimap2 (v2.29)¹⁹. The resulting alignment files were processed with SAMtools (v1.10)²⁰ to calculate the coverage depth across the assembled chromosomes, verifying the effective incorporation of raw data. Second, the completeness of the gene space was benchmarked with BUSCO (v5.7.1)²³ against the Insecta_odb10 dataset, which contains 1,367 conserved single-copy orthologs. This analysis identified 98.1% of the benchmark genes as complete and 0.4% as fragmented, collectively confirming a high level of assembly completeness.

Data availability

All sequencing and assembly data generated in this study have been deposited in public repositories. Raw sequencing data, including RNA-seq reads, PacBio HiFi long reads, Hi-C libraries, and Illumina short reads, have been deposited in the NCBI Sequence Read Archive (SRA) under the BioProject accession PRJNA1279776⁴³, with the following individual accession numbers: SRR34590780⁴⁴, SRR34590781⁴⁵, SRR34590782⁴⁶, and SRR34590783⁴⁷. The assembled genome is available at Genbank⁴⁸. The assembled genome and its corresponding annotation files are archived in the Figshare repository under the <https://doi.org/10.6084/m9.figshare.29880239>⁴⁹.

Code availability

No custom code was generated or used in this study. All analyses were conducted with publicly available software as cited in the Methods section.

Received: 28 August 2025; Accepted: 10 December 2025;

Published online: 22 December 2025

References

1. Kevan, D. K. M. Orthoptera. In *Synopsis and Classification of Living Organisms*, 2 (ed. Parker, S. P.) 352–383 (McGraw-Hill, New York, 1982).
2. Channa, S. A., Sultana, R. & Wagan, M. S. Morphology and burrowing behaviour of *Schizodactylus minor* (Ander, 1938) (Grylloptera: Schizodactylidae: Orthoptera) of Pakistan. *Pakistan Journal of Zoology* **45**, 1191–1196 (2013).
3. Ramme, W. Systematisches, Verbreitung und Morpho-Biologisches aus der Gryllacriden-Unterfamilie Schizodactylinae (Orth.). *Zeitschrift für Morphologie und Ökologie* **22**, 163–172, <https://doi.org/10.1007/BF00417967> (1931).
4. Cigliano, M. M., Braun, H., Eades, D. C. & Otte, D. Orthoptera Species File. *Orthoptera Species File Online* <http://orthoptera.speciesfile.org/> (accessed 2025).
5. Aydın, G. & Khomutov, A. The biology, nymphal stages, and life habits of the endemic sand dune cricket *Schizodactylus inexpectatus* (Werner, 1901) (Orthoptera: Schizodactylidae). *Turkish Journal of Zoology* **32**, 427–432 (2008).

6. Khattar, N. Description of adult and nymphal stages of *Schizodactylus monstrosus* (Drury) (Orthoptera). *Journal of Natural History* **6**, 589–600, <https://doi.org/10.1080/00222937200770521> (1972).
7. Irish, J. The species of *Comicus* Brunner v. Wattenwyl (Orthoptera: Schizodactylidae) with a discussion of their origin: systematics. *Navorsinge van die Nasionale Museum* **5**, 254–263, https://doi.org/10.10520/AJA00679208_404 (1986).
8. Aydin, G. Distribution of the dune cricket, *Schizodactylus inexpectatus* (Orthoptera: Schizodactylidae), in the Cukurova Delta, southern Turkey. *Zoology in the Middle East* **36**, 111–113 (2005).
9. He, Z. Q. & Liu, Y. J. A new species of dune cricket from China (Orthoptera: Ensifera: Schizodactylidae). *Zootaxa* **4999**, 356–362, <https://doi.org/10.11646/zootaxa.4999.4.5> (2021).
10. Heads, S. W. & Leuzinger, L. On the placement of the Cretaceous orthopteran *Brauckmannia groeningae* from Brazil, with notes on the relationships of Schizodactylidae (Orthoptera, Ensifera). *ZooKeys* **77**, 17–30, <https://doi.org/10.3897/zookeys.77.769> (2011).
11. Leubner, F., Bradler, S. & Wipfler, B. The thoracic morphology of the wingless dune cricket *Comicus calcaris* (Orthoptera: Schizodactylidae): Novel apomorphic characters for the group and adaptations to sand desert environments. *Arthropod Structure and Development* **46**, 449–461, <https://doi.org/10.1016/j.asd.2017.03.006> (2017).
12. Jost, M. C. & Shaw, K. L. Phylogeny of Ensifera (Hexapoda: Orthoptera) using three ribosomal loci, with implications for the evolution of acoustic communication. *Molecular Phylogenetics and Evolution* **38**, 510–530, <https://doi.org/10.1016/j.ympev.2005.09.004> (2006).
13. Bushnell, B. *BBMap* <https://sourceforge.net/projects/bbmap/> (2014).
14. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods* **18**, 170–175, <https://doi.org/10.1038/s41592-020-01056-5> (2021).
15. Guan, D. *et al.* Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896–2898, <https://doi.org/10.1093/bioinformatics/btaa025> (2020).
16. Zhang, H. *et al.* Fast alignment and preprocessing of chromatin profiles with Chromap. *Nature Communications* **12**, 6566, <https://doi.org/10.1038/s41467-021-26865-w> (2021).
17. Zhou, C., McCarthy, S. A. & Durbin, R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* **39**(1), 1–3, <https://doi.org/10.1093/bioinformatics/btac808> (2023).
18. Durand, N. C. *et al.* Juicer provides a One-Click system for analyzing Loop-Resolution Hi-C experiments. *Cell Systems* **3**, 95–98, <https://doi.org/10.1016/j.cels.2016.07.002> (2016).
19. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100, <https://doi.org/10.1093/bioinformatics/bty191> (2018).
20. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, 1–4, <https://doi.org/10.1093/gigascience/giab008> (2021).
21. Bai, J. R. & Shi, F. M. Karyotypes of two species of Meconematinae (Orthoptera, Tettigoniidae) from China. *Acta Zootaxonomica Sinica* **38**, 483–487 (2013).
22. Chang, Y. L. & Lian, Z. M. Chromosomal studies of three species of Tettigoniidae. *Yi Chuan = Hereditas* **24**, 155–158 (2002).
23. 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, <https://doi.org/10.1093/molbev/msab199> (2021).
24. Flynn, J. M. *et al.* RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 9451–9457, <https://doi.org/10.1073/pnas.1921046117> (2020).
25. 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, <https://doi.org/10.1186/s13100-020-00230-y> (2021).
26. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* **6**, 11, <https://doi.org/10.1186/s13100-015-0041-9> (2015).
27. Smit, A., Hubley, R. & Green, P. RepeatMasker Open-4.0 <http://www.repeatmasker.org> (2013–2015).
28. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**(22), 2933–2935, <https://doi.org/10.1093/bioinformatics/btt509> (2013).
29. Chan, P. P. & Lowe, T. M. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. *Methods in Molecular Biology* **1962**, 1–14, https://doi.org/10.1007/978-1-4939-9173-0_1 (2019).
30. Brůna, T. *et al.* BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics* **3**, lqaa108 (2021).
31. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**(5), 637–644, <https://doi.org/10.1093/bioinformatics/btn013> (2008).
32. Brůna, T., Lomsadze, A. & Borodovsky, M. GeneMark-ETP significantly improves the accuracy of automatic annotation of large eukaryotic genomes. *Genome Research* **34**(5), 757–768, <https://doi.org/10.1101/gr.278373.123> (2024).
33. Tegenfeldt, F. *et al.* OrthoDB and BUSCO update: Annotation of orthologs with wider sampling of genomes. *Nucleic Acids Research* **53**, D516–D522, <https://doi.org/10.1093/nar/gkac987> (2025).
34. 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. *Bioinformatics* **19**, 189, <https://doi.org/10.1186/s12859-018-2203-5> (2018).
35. Li, H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* **37**, 4572–4574, <https://doi.org/10.1093/bioinformatics/btab705> (2021).
36. Kovaka, S. *et al.* Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biology* **20**, 278, <https://doi.org/10.1186/s13059-019-1910-1> (2019).
37. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *Bioinformatics* **12**, 491, <https://doi.org/10.1186/1471-2105-12-491> (2011).
38. Buchfink, B., Reuter, K. & Drost, H. G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods* **18**, 366–368, <https://doi.org/10.1038/s41592-021-01101-x> (2021).
39. El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Research* **47**, D427–D432, <https://doi.org/10.1093/nar/gky995> (2018).
40. Paysan-Lafosse, T. *et al.* InterPro in 2022. *Nucleic Acids Research* **51**, D418–D427, <https://doi.org/10.1093/nar/gkac993> (2022).
41. 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. *Molecular Biology and Evolution* **38**, 5825–5829, <https://doi.org/10.1093/molbev/msab293> (2021).
42. 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 Research* **47**, D309–D314, <https://doi.org/10.1093/nar/gky1085> (2018).
43. NCBI BioProject <https://identifiers.org/ncbi/bioproject:PRJNA127976> (2025).
44. NCBI Sequence Read Archive <https://idxers.org/ncbi/insdc.sra:SRR34590780> (2025).
45. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR34590781> (2025).
46. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR34590782> (2025).
47. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR34590783> (2025).
48. NCBI Genbank https://identifiers.org/ncbi/insdc.gca:GCA_053542775.1 (2025).
49. Huang, Z. *et al.* Chromosome-level genome assembly of *Schizodactylus jimo* using PacBio HiFi and Hi-C sequencing. *Figshare* <https://doi.org/10.6084/m9.figshare.29880239> (2025).

Acknowledgements

We appreciate Zheng Zhou, Yuhao Wu, Yuhao Yao for their helping in breeding insects. We thank Jiazhi Zhang for providing the pictures of *S. jimo*. This work was supported by grants from Yunnan Provincial Science and Technology Department (Grant No. 202401BC070017), from Institute of Zoology, Chinese Academy of Sciences (Grant No. 2023IOZ0310, 2023IOZ0104), from National Natural Science Foundation of China (Grant No. 32270460), from State Key Laboratory of Animal Biodiversity Conservation and Integrated Pest Management (Grant No. SKLA2502).

Author contributions

S.Q.G. and X.Y.L. contributed to the research design. Z.Q.D. and Z.W.D. collected the samples. Z.Z.H. identified the species. L.Z. and J.W.H. performed the genome assembly and annotation analyses. Z.Z.H., L.Z. and C.Q.L. analyzed the data. Z.Z.H., L.Z., Z.Q.D. and X.Y.L. wrote the draft manuscript and revised the manuscript. All co-authors contributed to this manuscript and approved it.

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/s41597-025-06454-y>.

Correspondence and requests for materials should be addressed to X.L. or S.G.

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