scientific data

OPEN

DATA DESCRIPTOR

The chromosome-level genome assembly of the early monocot species *Tofieldia thibetica*

Hong Chen[1,2], Xiang-Yu Wang[1,2], Jia-Le Wang[1], Ning-Hua Tan[1]✉ & Ling-Yun Chen[1]✉

Tofieldia thibetica **is a species native to Southwest China. It holds an important phylogenetic position as one of the earliest-diverging lineages within monocots. Here we report a chromosome-level genome assembly of *T. thibetica*, assembled using PacBio HiFi and Hi-C sequencing data. The 1,641 Mb assembly has a contig N50 of 55.66 Mb, with 1,362 Mb of sequences anchored to 15 pseudochromosomes. Genome annotation predicted 53,411 protein-coding genes, of which 53,075 were functionally annotated. Repetitive sequences account for 79.53% of the genome. This assembly fills a critical genomic gap in Tofieldiaceae and provides an essential resource for phylogenomic and functional studies. This assembly also supports efforts to explore the family's medicinal potential and to advance our understanding of monocot evolution.**

Background & Summary

Tofieldia thibetica, a species in the family Tofieldiaceae[1], is endemic to Southwest China. It typically inhabits scrublands, grassy slopes, and rock crevices at elevations ranging from 700 to 2,300 meters[2]. The whole plant of *T. thibetica* has traditionally been used in medicine for promoting digestion, alleviating pain, detoxifying and reducing swelling, relieving gastric discomfort, treating wind-induced skin rashes, and mitigating injuries from falls or trauma[3].

Tofieldia thibetica was treated as a member of Liliaceae in *Flora of China*[2], but it is now classified within the family Tofieldiaceae[1,4–8]. Tofieldiaceae was recognized as an early-branching lineage within Alismatales, which is sister to core monocots. *T. thibetica*, characterized by distinctive morphological traits such as free carpels developing into obovoid capsules and seeds featuring a prominent white raphe (Fig. 1a), serves as an ideal model for studying monocot evolution and speciation. However, previous studies have reported conflicting results regarding the phylogenetic relationships among Tofieldiaceae, core Alismatales, and Araceae[9,10]. The paucity of genomic data of Tofieldiaceae has limited our understanding of the evolution of this family.

Advances in high-throughput sequencing have revolutionized phylogenetic and comparative genomic studies. We generated a chromosome-level genome assembly of *T. thibetica* by integrating PacBio HiFi long-read sequencing with Hi-C data. This assembly fills a key genomic gap by clarifying the evolutionary placement of this species, supporting functional gene discovery and conservation biology applications, and facilitating broader phylogenetic and adaptive evolution studies within Alismatales and monocots.

Methods

Plant materials and sequencing. Fresh roots, stems, leaves, and inflorescences of *T. thibetica* were collected from one individual plant in Chuxiong, Yunnan (101.3947°E, 25.0962°N; 1,877 m) on July 17, 2023. After cleaning, tissues were flash-frozen in liquid nitrogen, preserved on dry ice, and subsequently used for genomic DNA extraction using the cetyltrimethylammonium bromide (CTAB) method[11] and total RNA extraction using the TRIzol Reagent kit (Invitrogen, USA). A SMRTbell library (insert size ~15 kb) was prepared according to the manufacturer's protocol (Pacific Biosciences, CA, USA), and long-read sequencing was conducted using the PacBio Revio platform. A Hi-C library was also constructed, followed by sequencing using the BGI DNBSEQ-T7 platform (Beijing Genomic Institute, Shenzhen, China). PacBio Revio HiFi sequencing yielded 88.26 Gb of long reads (Table 1), while Hi-C sequencing generated 264 Gb of raw data, with 262.38 Gb retained after quality filtering with Fastp (v0.23.2)[12]. Transcriptome sequencing on the DNBSEQ-T7 platform produced 26.60 Gb of raw data, from which 25.46 Gb of high-quality reads were obtained for downstream analyses.

[1]Department of Resources Science of Traditional Chinese Medicines, School of Traditional Chinese Pharmacy, China Pharmaceutical University, 211198, Nanjing, China. [2]These authors contributed equally: Hong Chen, Xiang-Yu Wang. ✉e-mail: nhtan@cpu.edu.cn; lychen@cpu.edu.cn

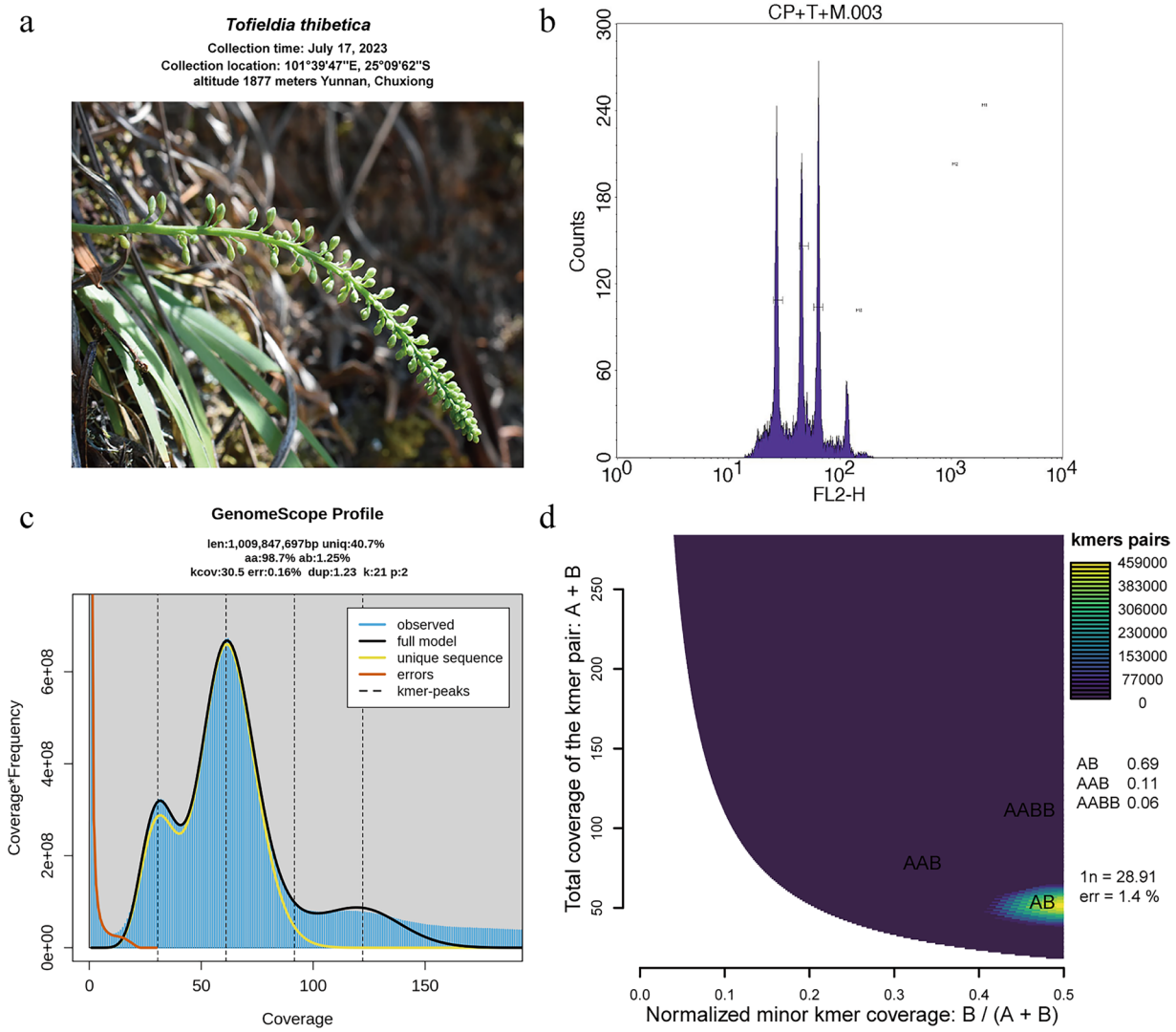

Fig. 1 Plant photo (**a**), genome size (**b,c**), and ploidy estimation (**d**) of *T. thibetica* using flow cytometry, k-mer analysis, and Smudgeplot. In the flow cytometry histograms (**b**), FL2-H represents the propidium iodide (PI) fluorescence intensity, reflecting relative nuclear DNA content. CP indicates the co-processed sample. From left to right, the first G1 peak represents tomato (*Solanum lycopersicum*, 1 C = 900 Mbp), the second corresponds to the tested sample, and the third represents maize (*Zea mays* B73, 1 C = 2.3 Gbp). Tomato and maize were used as internal reference standards. The value 0.03 denotes the coefficient of variation (CV) of the G1 peak, indicating high measurement precision.

Sequence	Platform	Total reads	Total bases (bp)	Sequence depth (×)
HiFi reads	PacBio Revio	5,527,556	88,262,875,770	55.44
Hi-C reads	DNBSEQ-T7	1,759,106,000	262,376,374,000	198.46
RNA reads	DNBSEQ-T7	178,140,764	25,460,198,000	—

Table 1. Summary of sequencing data generated for *T. thibetica*.

Estimate of genome size. Using flow cytometry with tomato (900 Mb) and maize (2.3 Gb) as internal standards, we estimated a genome size of ~1.61 Gb for *T. thibetica*, based on the sample-to-standard fluorescence intensity ratio of 0.50 (Fig. 1b). *K*-mer frequency distribution (21-mer) of PacBio HiFi reads was calculated using the count and histogram sub-commands of Jellyfish[13]. The resulting profile was modeled with GenomeScope2 (v1.0.0)[14], which indicated a diploid genome (p = 2) with a haploid size of 1.01 Gb, 59.30% repetitive content, and 1.25% heterozygosity (Fig. 1c). Smudgeplot (v0.1.6)[15] analysis (k = 31) also supported diploidy (Fig. 1d). Collectively, these results indicated that *T. thibetica* has a large, highly repetitive, and moderately heterozygous genome.

Assembly statistic	Contig-level assembly	Chromosome-level assembly
Number of contigs/chromosomes	730	15
Largest contig/scaffold (bp)	107,953,961	132,319,398
Total length (bp)	1,641,809,050	1,362,655,837
GC content (%)	43.87	43.98
Contig/scaffold N50 (bp)	55,660,170	93,597,260
Contig/scaffold N90 (bp)	21,285,891	73,790,276
Contig/scaffold L50	12	7
Contig/scaffold L90	29	13
BUSCO completeness (%)	98.40	95.60
Complete single-copy BUSCOs	1,531	1,515

Table 2. Statistics for contig-level and chromosome-level genome assemblies of *T. thibetica*.

***De Novo* genome assembly and quality assessment.** The chromosome-scale genome assembly of *T. thibetica* was generated using a hierarchical strategy that combined PacBio HiFi long reads and Hi-C data. PacBio HiFi reads were assembled with hifiasm (v0.19.6)[16], and redundant sequences were removed using Purge_dups (v1.2.5)[17], resulting in 730 contigs with a total length of 1,641 Mb. The contig N50 was 55.66 Mb with a GC content of 43.87%. BUSCO (v5.1.2, embryophyta_odb10; 1,614 single-copy orthologs)[18] analysis identified 1,531 complete single-copy genes, corresponding to 98.40% completeness (Table 2). Integration of Hi-C data with Juicer (v1.6)[19] and 3D-DNA (v180922)[20] anchored 1,362 Mb of sequences (83.0% of the genome) to 15 pseudochromosomes andincreased the scaffold N50 to 93.60 Mb. The final assembly exhibited a GC content of 43.98% and a BUSCO completeness of 95.60% with 1,515 complete single-copy genes (Table 2). Quality was further supported by an LTR Assembly Index (LAI) of 17.68 estimated with LTR_retriever (v2.9.7)[21], a base-level quality value (QV) of 63.99, and a *k*-mer completeness of 83.96% calculated with Merqury (v1.3)[22]. A genome-wide Hi-C interaction heatmap was generated with HapHiC (v1.0.6)[23] (Fig. 2a), and a circos plot was constructed using Circos (v0.69)[24] (Fig. 2b).

Repetitive sequence annotation. Repetitive sequences were identified through a combination of homology-based prediction with RepeatMasker (v4.1.2)[25] and *de novo* identification with RepeatModeler (v2.0.5)[26]. Transposable elements were further classified using EDTA (v2.1.1)[27], and elements that remained unclassified were subjected to deep annotation with DeepTE[28]. In total, 1,084.71 Mb of repetitive sequences were identified, representing 79.53% of the genome (Table 3). Long terminal repeats (LTRs) were the predominant class, with Copia and Gypsy accounting for 16.04% and 53.71% of the genome, respectively.

Gene structure annotation. Protein-coding genes in the *T. thibetica* genome were annotated by integrating multiple approaches: homology-based prediction with GeneWise (v2.4.1)[29], transcriptome-based evidence from RNA-seq read alignment with HISAT2 (v2.2.1)[30], transcript assembly with StringTie (v2.2.1)[31], and ORF prediction with TransDecoder (v5.7.1, https://github.com/TransDecoder/TransDecoder), and *ab initio* prediction with Augustus (v3.5.0)[32]. The results were merged using GETA (v2.7.1, https://github.com/chenlianfu/geta), yielding 53,411 protein-coding genes. Gene structural features included an average gene length of 6,847 bp, mean exon length 1,663 bp, mean intron length 5,184 bp, and mean coding sequence length 1,290 bp. BUSCO analysis (v5.1.2, embryophyta_odb10) indicated a completeness of 88.60%.

Non-coding RNA annotation. We employed a suite of computational tools to identify and annotate non-coding RNAs within the *T. thibetica* genome. tRNAs were predicted with tRNAscan-SE (v2.0.11)[33] in eukaryotic mode (default). rRNAs, snoRNAs, and miRNAs were predicted with Infernal (v1.1.4)[34] using homology searches against the Rfam database (e-value: 1e-5). The annotation identified 1,840 rRNA genes, 296 tRNA genes, 52 snoRNA genes, and 22 miRNA genes.

Gene functional annotation. Functional annotation of the 53,411 protein-coding genes in *T. thibetica* was performed using seven databases: EggNOG[35], Non-Redundant Protein Sequence Database (Nr), Swiss-Prot[36], GO[37,38], KEGG[39,40], InterPro[41], and Pfam[42]. 53,075 genes (99.37%) were successfully functionally annotated (Table 4). The Nr database achieved the highest annotation coverage at 98.61%, followed by eggNOG at 96.71%, with Swiss-Prot covering 57.31% of genes. For conserved domain prediction, InterPro and Pfam annotated 86.23% and 85.52% of genes, respectively. KEGG pathway annotation was assigned to 57.98% of genes, while 37.64% received GO terms.

Data Record

The raw sequencing data of *T. thibetica*, including PacBio HiFi reads, Hi-C reads, and transcriptome reads, have been deposited in NCBI Sequence Read Archive (SRA) with the following accession codes: PacBio HiFi, SRR35263715[43], SRR35263717[44], SRR35263718[45]; Hi-C, SRR35263713[46], SRR35263714[47]; transcriptome, SRR35263711[48], SRR35263712[49], SRR35263716[50]. The corresponding BioProject and BioSample accessions are PRJNA1319348 and SAMN51150105, respectively.

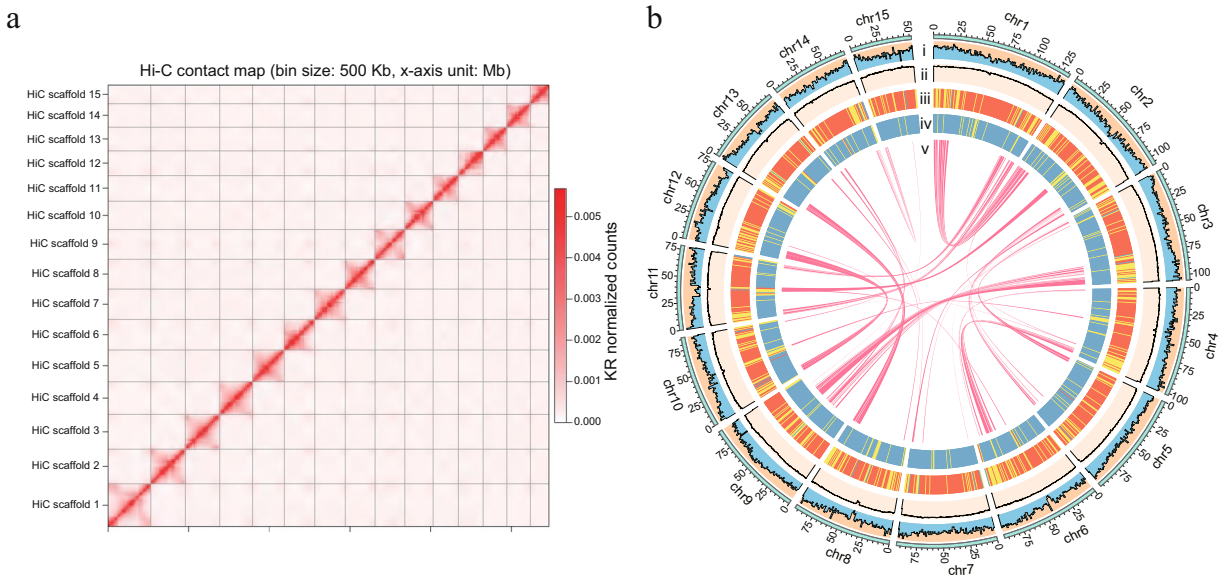

Fig. 2 Chromosome-level assembly and features of *T. thibetica*. (**a**) Hi-C contact heatmap displaying chromosomal interactions. (**b**) Circos plot illustrating genomic features, including (i) pseudochromosomes, (ii) gene density, (iii) GC content, (iv) LTR retrotransposon density, and (v) syntenic links.

Class		Count	Masked (bp)	Masked (%)
LTR	Gypsy	750,392	731,900,128	53.71
	Copia	171,450	218,616,531	16.04
	unknown	22,660	19,667,807	1.44
LINE		10,343	8,486,789	0.62
TIR		229,264	83,844,818	6.15
Non-LTR		406	275,390	0.02
Non-TIR		37,090	9,454,686	0.69
Repeat fragment		42,840	11,461,359	0.84

Table 3. Classification and genomic composition of repetitive elements in *T. thibetica*.

Database	Annotated genes	Annotation ratio (%)
Nr	52,667	98.61
EggNOG	51,654	96.71
Swiss-Prot	30,609	57.31
InterPro	46,056	86.23
Pfam	45,677	85.52
KEGG	30,967	57.98
GO	20,101	37.64
Total annotated genes	53,075	99.37

Table 4. Functional annotation statistics of protein-coding genes in *T. thibetica*.

Technical Validation

The integrity and accuracy of *T. thibetica* genome assembly were evaluated through multiple approaches. The Hi-C contact map revealed strong intra-chromosomal interaction signals along the diagonal (Fig. 2), confirming the integrity of the genome structure. The GC content distribution showed no evidence of sequence contamination (Fig. 2b). For genomic completeness assessment, BUSCO analysis demonstrated that 95.6% of complete single-copy genes were assembled from a set of 1,614 conserved orthologs (C: 95.6% [S: 93.9%, D: 1.7%], F: 0.9%, M: 3.5%, n: 1,614). To validate accuracy, Merqury produced a base-level quality value (QV) of 63.99, and LTR_retriever generated an LTR Assembly Index (LAI) of 17.68. These results indicated high accuracy in both base calling and repetitive region assembly. Additionally, 53,075 (99.37%) gene models were successfully annotated against databases including NR, KEGG, EggNOG, GO, Pfam, Swiss-Prot, and InterPro. Taken together, these results provided compelling evidence that the *de novo* assembled *T. thibetica* genome is of high quality.

Data availability

The chromosome-level genome assembly of *T. thibetica* has been deposited in NCBI GenBank under the accession number GCA_054954025.1[51], associated with BioProject PRJNA1404966 and BioSample SAMN54730446. The corresponding genome annotation files are available on Figshare (https://doi.org/10.6084/m9.figshare.30058588).

Code availability

No custom code was developed in this study. All code and pipelines employed in data processing were rigorously executed according to the operation manuals and standardized protocols of publicly available bioinformatics tools. Detailed descriptions of software versions, code implementations, and parameter configurations are provided in the Methods section.

Received: 18 September 2025; Accepted: 9 March 2026;
Published online: 21 March 2026

References

1. Lin, C. T., Liu, B., Zhao, M. J., Ma, K. P. & Ji, L. Q. Catalogue of life China: Towards an index of known species present in China. *The Innovation life* **3**, 100140, https://doi.org/10.59717/j.xinn-life.2025.100141 (2025).
2. Editorial Committee of Flora of China. In *Flora Reipublicae Popularis Sinicae* Vol. 14 P. 10 (Science Press, 1980).
3. Sichuan Academy of Natural Resources Sciences. *Tofieldia thibetica*. Sichuan Plant Resources Information Network http://www.scpri.ac.cn/mplant/#!/list/specinfo/2357/1 (1978).
4. Li, X. X. & Zhou, Z. K. The higher-level phylogeny of monocots based on *mat*K, *rbc*L and 18S rDNA sequences. *J. Syst. Evol.* **45**, 113–133 (2007).
5. Davis, J. I. *et al*. A phylogeny of the monocots, as inferred from *rbc*L and *atp*A sequence variation, and a comparison of methods for calculating jackknife and bootstrap values. *Syst. Bot.* **29**, 467–510, https://doi.org/10.1600/0363644041744365 (2004).
6. Davis, J. I. *et al*. Are mitochondrial genes useful for the analysis of monocot relationships? *Taxon* **55**, 857–870, https://doi.org/10.2307/25065681 (2006).
7. Holloway, S. J. & Friedman, W. E. Embryological features of *Tofieldia glutinosa* and their bearing on the early diversification of monocotyledonous plants. *Ann. Bot.* **102**, 167–182, https://doi.org/10.1093/aob/mcn084 (2008).
8. Azuma, H. & Tobe, H. Molecular phylogenetic analyses of Tofieldiaceae (Alismatales): family circumscription and intergeneric relationships. *J. Plant Res.* **124**, 349–357, https://doi.org/10.1007/s10265-010-0387-5 (2011).
9. Chen, L. Y. *et al*. Phylogenomic analyses of Alismatales shed light into adaptations to aquatic environments. *Mol. Biol. Evol.* **39**, msac079, https://doi.org/10.1093/molbev/msac079 (2022).
10. Luo, Y. *et al*. Plastid phylogenomic analyses resolve Tofieldiaceae as the root of the early diverging monocot order Alismatales. *Genome Biol. Evol.* **8**, 932–945, https://doi.org/10.1093/gbe/evv260 (2016).
11. Porebski, S., Bailey, L. G. & Baum, B. R. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol. Biol. Rep.* **15**, 8–15, https://doi.org/10.1007/BF02772108 (1997).
12. Chen, S. F., Zhou, Y. Q., Chen, Y. R. & Gu, J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890, https://doi.org/10.1093/bioinformatics/bty560 (2018).
13. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770, https://doi.org/10.1093/bioinformatics/btr011 (2011).
14. Vurture, G. W. *et al*. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204, https://doi.org/10.1093/bioinformatics/btx153 (2017).
15. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **11**, 1432, https://doi.org/10.1038/s41467-020-14998-3 (2020).
16. Cheng, H. Y., Concepcion, G. T., Feng, X. W., Zhang, H. W. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175, https://doi.org/10.1038/s41592-020-01056-5 (2021).
17. Guan, D. F. *et al*. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896–2898, https://doi.org/10.1093/bioinformatics/btaa025 (2020).
18. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212, https://doi.org/10.1093/bioinformatics/btv351 (2015).
19. Durand, N. C. *et al*. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98, https://doi.org/10.1016/j.cels.2016.07.002 (2016).
20. Dudchenko, O. *et al*. *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95, https://doi.org/10.1126/science.aal3327 (2017).
21. Ou, S. & Jiang, N. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* **176**, 1410–1422, https://doi.org/10.1104/pp.17.01310 (2018).
22. 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, https://doi.org/10.1186/s13059-020-02134-9 (2020).
23. Zeng, X. F. *et al*. Chromosome-level scaffolding of haplotype-resolved assemblies using Hi-C data without reference genomes. *Nat. Plants* **10**, 1184–1200, https://doi.org/10.1038/s41477-024-01755-3 (2024).
24. Krzywinski, M. *et al*. Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645, https://doi.org/10.1101/gr.092759.109 (2009).
25. Tarailo-Graovac, M. & Chen, N. S. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **4**, 4.10.1–4.10.14, https://doi.org/10.1002/0471250953.bi0410s25 (2004).
26. Flynn, J. M. *et al*. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci.* **117**, 9451–9457, https://doi.org/10.1073/pnas.1921046117 (2020).
27. Ou, S. J. *et al*. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* **20**, 275, https://doi.org/10.1186/s13059-019-1905-y (2019).
28. Yan, H. D., Bombarely, A. & Li, S. DeepTE: a computational method for de novo classification of transposons with convolutional neural network. *Bioinformatics* **36**, 4269–4275, https://doi.org/10.1093/bioinformatics/btaa519 (2020).
29. Birney, E., Clamp, M. & Durbin, R. GeneWise and Genomewise. *Genome Res.* **14**, 988–995, https://doi.org/10.1101/gr.1865504 (2004).
30. 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, https://doi.org/10.1038/s41587-019-0201-4 (2019).
31. Pertea, M. *et al*. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295, https://doi.org/10.1038/nbt.3122 (2015).

32. Stanke, M. *et al*. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439, https://doi.org/10.1093/nar/gkl200 (2006).
33. Chan, P. P., Lin, B. Y., Mak, A. J. & Lowe, T. M. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* **49**, 9077–9096, https://doi.org/10.1093/nar/gkab688 (2021).
34. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935, https://doi.org/10.1093/bioinformatics/btt509 (2013).
35. 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, https://doi.org/10.1093/nar/gky1085 (2019).
36. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**, 45–48, https://doi.org/10.1093/nar/28.1.45 (2000).
37. The Gene Ontology Consortium. The gene ontology resource: 20 years and still GOing strong. *NucleicAcids Res.* **47**, D330–D338, https://doi.org/10.1093/nar/gky1055 (2019).
38. Ashburner, M. *et al*. Gene ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29, https://doi.org/10.1038/75556 (2000).
39. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30, https://doi.org/10.1093/nar/28.1.27 (2000).
40. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–D462, https://doi.org/10.1093/nar/gkv1070 (2016).
41. Mitchell, A. L. *et al*. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* **47**, D351–D360, https://doi.org/10.1093/nar/gky1100 (2019).
42. Mistry, J. *et al*. Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419, https://doi.org/10.1093/nar/gkaa913 (2021).
43. *NCBI Sequence Read Archive* https://identifiers.org/insdc.sra:SRR35263715.
44. *NCBI Sequence Read Archive* https://identifiers.org/insdc.sra:SRR35263717.
45. *NCBI Sequence Read Archive* https://identifiers.org/insdc.sra:SRR35263718.
46. *NCBI Sequence Read Archive* https://identifiers.org/insdc.sra:SRR35263713.
47. *NCBI Sequence Read Archive* https://identifiers.org/insdc.sra:SRR35263714.
48. *NCBI Sequence Read Archive* https://identifiers.org/insdc.sra:SRR35263711.
49. *NCBI Sequence Read Archive* https://identifiers.org/insdc.sra:SRR35263712.
50. *NCBI Sequence Read Archive* https://identifiers.org/insdc.sra:SRR35263716.
51. NCBI Genome Assembly for *Tofieldia thibetica* isolate wild_individual_1, GenBank accession GCA_054954025.1, https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_054954025.1/ (2026).

Acknowledgements

We thank Wen-Da Zhang and Pu-Guang Zhao for assistance in data analyses. This work was supported by the National Natural Science Foundation of China (No. 32370242) and the Fundamental Research Funds for the Central Universities (No. 2632024TD04).

Author contributions

L.Y.C. and N.H.T. designed this study; H.C., X.Y.W. and J.L.W. conducted data analyses; H.C. drafted this manuscript. All authors read and revised this manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to N.-H.T. or L.-Y.C.

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.