Scientific Data

https://doi.org/10.1038/s41597-026-07516-5

Article in Press

A chromosome-level genome assembly of the endangered rose *Rosa anemoniflora*

Received: 30 November 2025

Accepted: 26 May 2026

Cite this article as: Zhou, X., Yu, Q., Fu, C. *et al.* A chromosome-level genome assembly of the endangered rose *Rosa anemoniflora*. *Sci Data* (2026). https://doi.org/10.1038/s41597-026-07516-5

Xianzhen Zhou, Qun Yu, Chengjie Fu, Guojiang Lin, Xiaochun Huang, Tianqi Kang & Wei Gao

A chromosome-level genome assembly of the endangered rose *Rosa anemoniflora*

Xianzhen Zhou[1,#], Qun Yu[1,#], Chengjie Fu[1,#], Guojiang Lin[1], *Xiaochun Huang[2], Tianqi Kang[3]*, Wei Gao[1,*]

[1] Institute of Forest Ecology and Carbon Sink Measurement, Department of Forestry, Fujian Forestry Vocational and Technical College, Nanping 353000, Fujian, China

[2] Fujian Forestry Prospect and Design Institute, Fuzhou 353001, Fujian, China

[3] National Nature Reserve of Mangdang Mountain of Fujian Province, Nanping 353000, Fujian, China

[#]These authors contributed equally to this work: Xianzhen Zhou and Qun Yu

* Correspondence: Wei Gao

Email address

Xianzhen Zhou: zxz92322@nwafu.edu.cn

Qun Yu: yuqun@fjlzy.com

Chengjie Fu: fuchengjie@fjlzy.com

Guojiang Lin: linguojiang@fjlzy.com

Xiaochun Huang: hxc13559326373@163.com

Tianqi Kang: 1059165545@qq.com

Wei Gao: gao01271@163.com

Abstract

Rosa anemoniflora, a critically endangered climbing rose endemic to China, holds significant ornamental and conservation value but lacks genomic resources. Here, we present a chromosome-level genome assembly generated using PacBio HiFi, Illumina, Iso-Seq, and Hi-C sequencing. The final assembly spans 547.52 Mb, with 95.72% of sequences anchored to seven pseudochromosomes and a contig N50 of 73.49 Mb. BUSCO assessment reveals 97.36% completeness (eudicotyledons_odb12). Repetitive elements constitute 55.31% of the genome, and 33,623 protein-coding genes were annotated, 97.68% of which are supported by BUSCO. We also identified 1,333 rRNA, 838 tRNA, and 1,043 other non-coding RNA genes. All data are publicly available at National Genomics Data Center (NGDC) (PRJCA052329). This high-quality genome provides a foundational resource for conservation genomics,

evolutionary studies, and the genetic dissection of adaptive and reproductive traits in this threatened species.

Background & Summary

The genus *Rosa* comprises approximately 200 species distributed widely across temperate to subtropical regions of Asia, Europe, and North America. Rosa species are characterized by complex ploidy levels ranging from diploid to decaploid and high heterozygosity, which together have fostered remarkable morphological and ecological diversity[1-3]. Cultivated roses have been extensively studied and utilized due to their significant economic value in the cut-flower, horticultural, and essential oil industries[4]. Meanwhile, wild rose resources—such as *R. roxburghii* and *R. davurica*—show great potential in functional foods and pharmaceuticals owing to their rich content of antioxidants and vitamin C[5-8]. In recent years, advances in high-throughput sequencing have led to chromosome-level genome assemblies for several cultivated and wild roses, such as *R. chinensis*[9] and *R. gigantea*[10], laying a foundation for evolutionary and functional genomics studies in the genus.

However, genomic information remains nearly absent for many wild rose species of unique biological value that are currently threatened or endangered. One such example is *R. anemoniflora* (Section *Synstylae*), a climbing shrub endemic to Nanping City, Fujian Province, China, growing at elevations of 400–1,000 m[11]. This species holds exceptional ornamental and germplasm value, with its vibrant flower color, elegant floral morphology, concentrated blooming period, and climbing habit making it highly promising for garden landscaping and urban greening. Naturally occurring on hillsides, wastelands, roadsides, and riverbanks within its elevation range, *R. anemoniflora* may possess genetic adaptations to environmental stresses such as drought and salinity-traits of potential relevance for breeding programs. Unfortunately, field surveys indicate extremely low population sizes, far below the

minimum viable threshold for long-term persistence. In 2021, the species was listed as Near Threatened (NT) by the International Union for Conservation of Nature (IUCN) and included as a Class II protected species in China's National Key Protected Wild Plants List (https://www.mee.gov.cn/xxgk2018/xxgk/xxgk01/202305/W020230522536560832337.pdf). Its critically low reproductive success (fruit set <15%, seed germination <20%) and high sensitivity to habitat fragmentation have exacerbated its conservation status[12,13], creating an urgent need for genomic resources to inform evidence-based protection strategies.

To address this gap, we present the first chromosome-level, high-quality genome assembly of *R. anemoniflora*, generated by integrating PacBio HiFi[14] long-read sequencing (~150× coverage, based on the estimated genome size of ~547 Mb) and Hi-C[15] chromatin conformation capture (~105×coverage). The final assembly spans approximately 547.52 Mb, with a contig N50 of 73.49 Mb and a BUSCO completeness score of 97.29%. Hi-C scaffolding successfully anchored 95.72% of the assembly onto seven pseudochromosomes. All raw sequencing data, the assembled genome, and structural and functional annotation files have been deposited in public repositories at National Genomics Data Center (NGDC)[16] with the accession number of PRJCA052329 in accordance with FAIR principles. This genomic resource fills a critical void in the genomics of endangered *Rosa* species. It provides a foundational platform for future research into the genetic mechanisms underlying *R. anemoniflora*'s endangered status, as well as for comparative genomics and conservation genetics studies in closely related taxa.

Methods

Sample collection and sequencing

The *R. anemoniflora* research material was collected from the Mangdangshan National Nature Reserve in Fujian (26°36'12" N, 118°02'30" E).

Genomic DNA was extracted from fresh leaves using the CTAB method. DNA integrity was assessed by electrophoresis on a 1% agarose gel stained with ethidium bromide. As shown in Supplementary Figure 1, a single high-molecular-weight band (>15 kb) without smearing or RNA contamination was observed, indicating that the extracted DNA was of high integrity and suitable for downstream library construction. The DNA was used to construct Illumina short-read, PacBio HiFi long-read, and Hi-C chromatin conformation capture libraries.

Total RNA was extracted from the same leaf tissue using the RNeasy Plant Mini Kit (Qiagen) according to the manufacturer's instructions. RNA integrity was assessed by electrophoresis on a 1% denaturing agarose gel stained with ethidium bromide. As shown in Supplementary Figure 2, two clear bands corresponding to 28S rRNA and 18S rRNA were observed, with a 28S:18S intensity ratio of approximately 2:1 and no visible smearing, confirming high RNA integrity. cDNA was synthesized using the SMARTer PCR cDNA Synthesis Kit, size-selected (>1 kb) with BluePippin, and sequenced on a PacBio Sequel II platform, producing 4,782,271 CCS reads (N50 = 1,426 bp; GC content = 43.59%).

The Illumina library was sequenced on a NovaSeq 6000 platform using paired-end 150 bp reads, yielding 72.89 Gb of raw data (~138× genome coverage) for genome characterization. PacBio HiFi sequencing was performed on a Revio system, generating 28.72 Gb of high-fidelity long reads with an N50 read length of 20,455 bp. For the full-length transcriptome, cDNA was synthesized using the SMARTer PCR cDNA Synthesis Kit, size-selected (>1 kb) with BluePippin, and sequenced on a PacBio Sequel II platform, producing 4,782,271 CCS reads (N50 = 1,426 bp; GC content = 43.59%). The Hi-C library was sequenced on a DNBSEQ-T7 platform to support chromosome-scale scaffolding.

Genome assembly

Genome assembly followed a multi-step strategy. First, PacBio HiFi reads were assembled de novo using hifiasm[17] v0.19.8 (default parameters, 32 threads) to generate an initial contig set. Redundant haplotigs were purged using purge_dups[18] v1.2.5 based on k-mer depth distribution (main peak at 58.1×) and read coverage depth (Figure 1). The cut-off value for haplotig identification was set to 70 (as estimated by the calcuts function), and sequences with coverage depth below this threshold were removed as haplotypic duplications, resulting in a haploid-resolved assembly (purged.fa) of 638.52 Mb across 128 scaffolds. After removing lowSubsequently, gene models containing frameshifts or coding sequence (CDS) lengths not divisible by three were filtered out to ensure the integrity of protein-coding sequence-quality and short sequences, a haploid-resolved assembly (draft.fa) of 542.26 Mb across 109 contigs was retained for subsequent Hi-C scaffolding. Hi-C data were then integrated using 3D-DNA[19] v201008 with default parameters with 2 rounds of misassembly correction, -r 2) to anchor contigs into chromosome-scale pseudomolecules, followed by manual correction with Juicebox Assembly Tools. The final assembly spans 547.52 Mb with a contig N50 of 73.49 Mb, and 534.43 Mb (95.72%) was anchored to seven pseudochromosomes (Table 1 and Figure 2), corresponding to the haploid chromosome number of this diploid species (2n = 2x = 14). Assembly completeness was assessed with BUSCO[20] v6.0.0 using the eudicotyledons_odb12 (2,805 genes): 97.36% of BUSCOs were complete (Table 1), 0.82% were fragmented, and 1.82% were missing. Additionally, Illumina short reads from the genome survey were remapped to the final assembly using BWA-MEM[21] v0.7.17 (default settings), achieving a mapping rate of 98.10% and a coverage breadth of 99.27%.

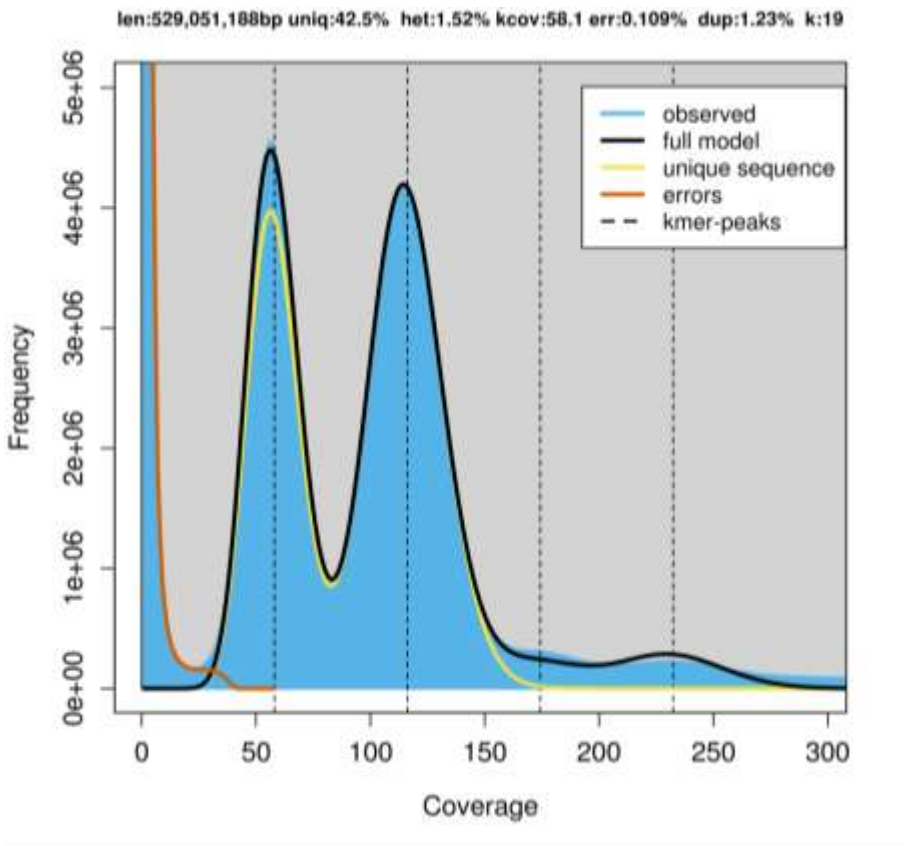

Figure 1 Distribution of k-mer frequencies employed to estimate the genome size.

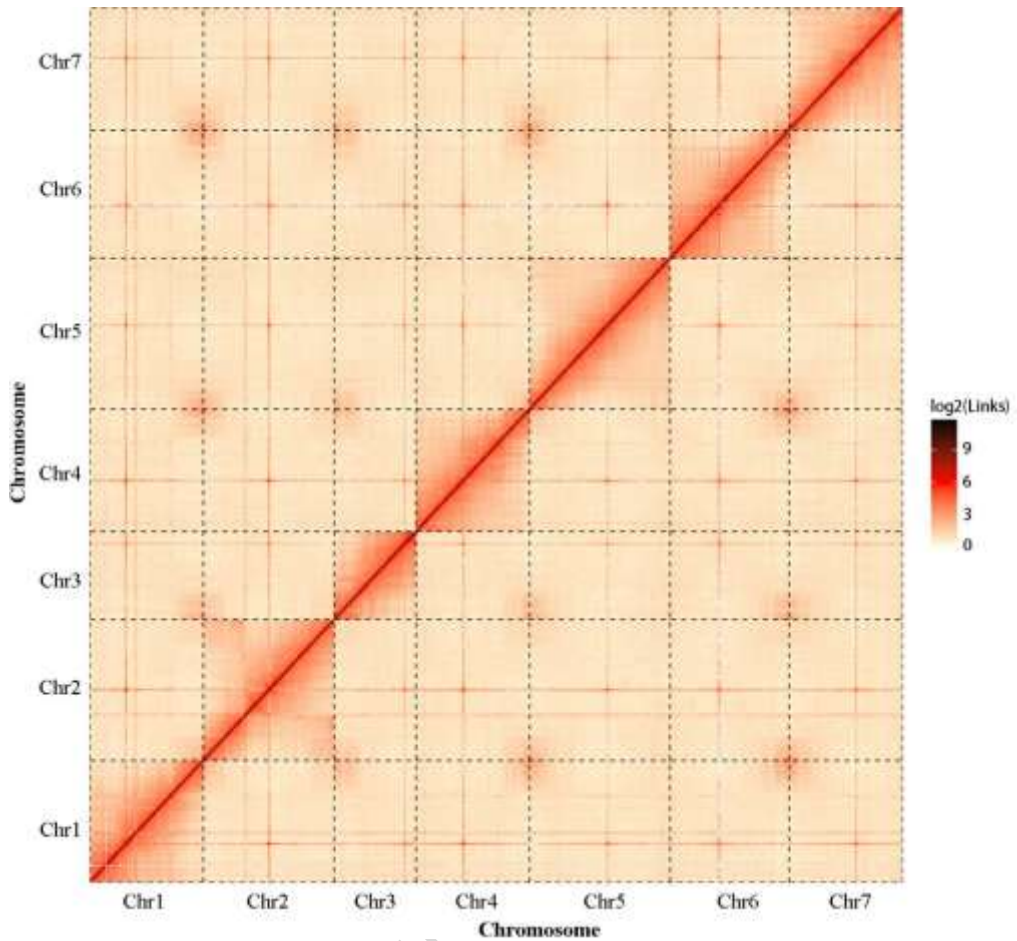

Figure 2 Hi-C intra-chromosomal contact map of the *R. anemoniflora* genome assembly

Genome annotation

Repeat elements were annotated using a combined approach: a species-specific repeat library was constructed with RepeatModeler[22] v2.0.6 (https://github.com/Dfam-consortium/TETools), supplemented with RepBase[23], and applied via RepeatMasker[24] v4.1.2. LTR retrotransposons were further refined using EDTA[25] v2.2.2 for structural prediction and classification. Repeats constitute 55.31% of the genome, dominated by LTR retrotransposons (31.48%), followed by unclassified repeats (13.92%), DNA transposons (7.10%), LINEs (2.81%), and SINEs (0.00%)（Table 2）.

Protein-coding genes were predicted by integrating *ab initio*, homology-based, and transcriptome evidence. *Ab initio* predictions were generated with Augustus[26] v3.3.2 and GeneMark-ES[27] v4.71.

Homology-based annotation leveraged protein sequences from *Arabidopsis thaliana*, *Fragaria vesca*, and *Rosa chinensis* using GeMoMa[28] v1.9 and exonerate[29] v2.2.0. Transcriptomic evidence came from PacBio Iso-Seq CCS reads processed with PASA[30] v2.5.2. These three lines of evidence were weighted (homology: 10, transcriptome: 12, ab initio: 4) and merged using EVidenceModeler[31] v1.1.1 to produce a high-confidence consensus gene set. Subsequently, gene models containing frameshifts or coding sequence (CDS) lengths not divisible by three were filtered out to ensure the integrity of protein-coding sequences. A total of 33,623 protein-coding genes were annotated, with an average of 4.8 exons per gene and a mean gene length of 3,487 bp (Table 1). BUSCO analysis of the gene set recovered 97.68% of conserved eudicot genes as complete. Functional annotation was performed by aligning proteins to NR and TrEMBL[32] databases using DIAMOND[33] v2.0.14 (e-value ≤ 1e−5) and identifying domains with InterProScan[34] v5.50-84.0, enabling mapping to KEGG[35] and Gene Ontology[36] (GO) categories. Genomic features were visualized using Circos[37] (Figure 3).

Non-coding RNAs were identified as follows: tRNAs were predicted with tRNAscan-SE[38] v1.3.1; rRNA genes (5S, 18S, 28S) were annotated using RNAmmer[39] v1.2; other ncRNAs were detected by homology search against Rfam[40] v1.0. Annotation statistics are summarized in Table 3.

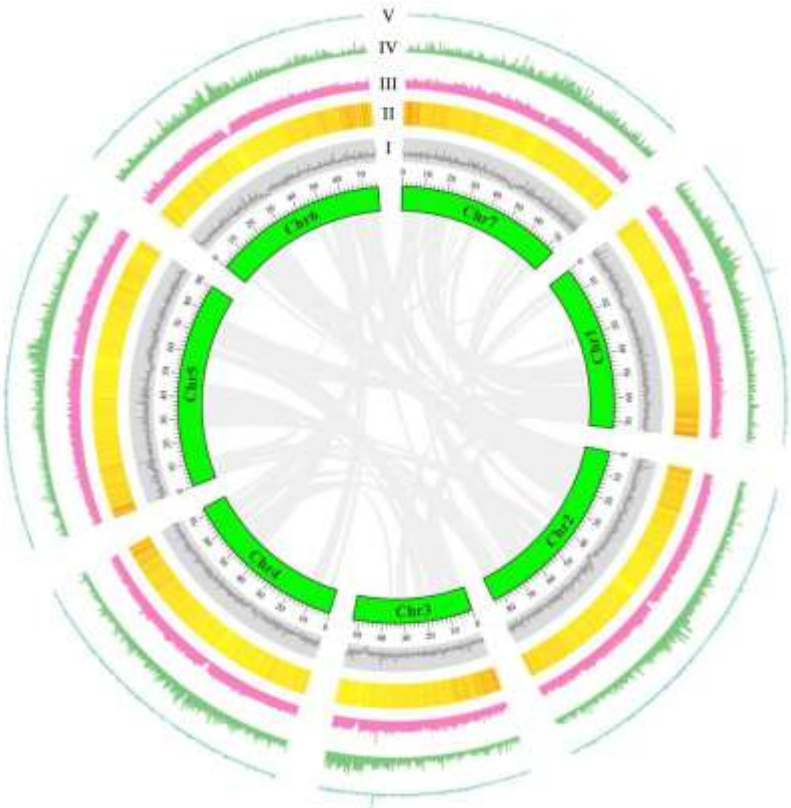

Figure 3. Genome landscape of *Rosa anemoniflora*. Tracks from outer to inner circles: (I) GC content (%) in 100-kb windows; (II) gene density (number of genes per 100-kb window); (III) repeat density (number of repetitive sequences per 100-kb window); (IV) density of LTR/Gypsy elements; (V) density of LTR/Copia elements. Central lines indicate intra-genomic collinear blocks identified by MCScanX.

Table 1 Basic statistics of the *Rosa anemoniflora* genome assembly and annotation

Item	Value
Total assembly size (bp)	547,519,537
Number of chromosomes anchored	7
Proportion of sequence anchored to chromosomes	0.9572
Contig N50 (bp)	73,487,902
GC content (%)	38.75
BUSCO completeness (%)	97.36
Total number of genes	33,623
Total length of genic regions (bp)	117,246,647
Proportion of genic regions in the genome (%)	21.41
Average gene length (bp)	3,487
Average number of exons per gene	4.8
Average exon length (bp)	238.6
Average intron length (bp)	603.5

Table 2 Repetitive element composition in the *Rosa anemoniflora* genome

Repeat Class	Repeat Size (bp)	Percentage of Genome (%)
LTR retrotransposons	172,344,069	31.48
DNA transposons	38,855,010	7.1
LINEs	15,411,239	2.81
SINEs	2,692	0
Penelope	41,541	0.01
Retroviral	233,424	0.04
Unclassified	76,235,085	13.92
Satellites	781,359	0.14
Simple repeats	134,576	0.02
Total interspersed repeats	302,848,095	55.31

Table 3 Statistics of non-coding RNA prediction results

ncRNA Type	Copy	Avg. length (bp)	Total length (bp)	% of genome
5s_rRNA	563	115.23	64,872	0.0118
18s_rRNA	383	1792.3	686,449	0.1254
28s_rRNA	387	3,864.67	1,495,629	0.2732
tRNA	838	74.29	62,254	0.0114
other ncRNA	1,043	132.49	138,190	0.0252

Data Records

The raw sequencing data has been deposited in the Genome Sequence Archive (GSA) at the China National Center for Bioinformation (CNCB; Project Accession: PRJCA052329). The Hi-C reads are available under GSA accession CRA034365 (Run ID: CRR2373096)[41]. The final genome assembly (in FASTA format) has been deposited in NCBI GenBank under accession number GCA_054131425.1[42] Gene structural annotations (in GFF3 format) and functional annotation files are publicly available on Figshare[43].

Technical Validation

The quality of the *R. anemoniflora* genome assembly and annotation was evaluated using multiple complementary approaches.

Assembly contiguity and completeness. The final assembly achieved high contiguity, with a contig

N50 of 73.49 Mb, an N90 of 52.83 Mb, and the longest sequence reaching 90.62 Mb. A total of 95.72% of the assembly (534.43 Mb) was successfully anchored to seven pseudochromosomes, corresponding to the haploid chromosome number of this diploid species (2n = 2x = 14). Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis against the eudicotyledons_odb12 (2,805 genes) recovered 97.36% complete genes (93.90% single-copy, 3.46% duplicated), with only 0.82% fragmented and 1.82% missing, indicating near-complete representation of the expected gene space. As an additional validation, Compleasm[44] analysis using the more recent eudicotyledons_odb12 lineage (2,805 genes) yielded 99.08% complete BUSCOs (95.19% single-copy, 3.89% duplicated), further confirming the high completeness of the assembly.

Annotation validation. The gene annotation pipeline integrated *ab initio*, homology-based, and transcriptome evidence, yielding 33,623 protein-coding genes. BUSCO assessment of the gene set recovered 97.68% complete BUSCOs (93.58% single-copy, 4.10% duplicated), with 1.00% fragmented genes and only 1.32% missing, demonstrating high annotation completeness. Functional annotation assigned putative functions to 95.90% of the predicted genes across major public databases (NR, Swiss-Prot, Pfam, GO, KEGG), providing a valuable resource for downstream functional studies.

Together, these multi-faceted validation results confirm that the *R. anemoniflora* genome assembly and annotation are of high quality and suitable for conservation genomics, evolutionary analyses, and comparative genomic studies within the genus *Rosa*.

Data availability

This dataset has not been previously published or presented, and no conflicts of interest are declared. The raw sequencing reads for *R. anemoniflora* have been deposited in the National Genomics Data Center (NGDC) under Project Accession PRJCA052329, with Hi-C reads available under accession CRA034365

(Run ID: CRR2373096). The chromosome-scale genome assembly has been deposited in NCBI GenBank under accession number GCA_054131425.1. Annotation data (including structural and functional annotations) are publicly available in the figshare database (https://doi.org/10.6084/m9.figshare.30742499.v1).

Code availability

All analyses were performed strictly in accordance with the manuals and protocols of the publicly available tools, and no custom code was utilized in this study.

References

1 Gahlaut, V., Kumari, P., Jaiswal, V. & Kumar, S. Genetics, genomics and breeding in *Rosa* species. *J Hortic Sci Biotech* **96**, 545-559, doi:10.1080/14620316.2021.1894078 (2021).

2 Jian, H. *et al.* Decaploidy in *Rosa praelucens* Byhouwer (Rosaceae) Endemic to Zhongdian Plateau, Yunnan, China. *Caryologia* **63**, 162-167, doi:10.1080/00087114.2010.10589722 (2010).

3 Chtourou, K. *et al.* Genetic diversity and relationships among tunisian wild and cultivated *Rosa* L. Species. *Plants* **13**, 3563, doi:10.3390/plants13243563 (2024).

4 Raymond, O. *et al.* The *Rosa* genome provides new insights into the domestication of modern roses. *Nat. Genet.* **50**, 772-777, doi:10.1038/s41588-018-0110-3 (2018).

5 Xu, J., Vidyarthi, S. K., Bai, W. & Pan, Z. Nutritional constituents, health benefits and processing of *Rosa Roxburghii*: A review. *J. Funct. Foods* **60**, 103456, doi:10.1016/j.jff.2019.103456 (2019).

6 Ni, M. *et al.* UPLC-ESI-MS/MS-Based Analysis of Various Edible Rosa Fruits Concerning Secondary Metabolites and Evaluation of Their Antioxidant Activities. *Foods* **13**, 796, doi:10.3390/foods13050796 (2024).

7 Liu, M.-H. *et al.* Chemical analysis of dietary constituents in *Rosa roxburghii* and *Rosa sterilis* fruits. *Molecules* **21**, 1204, doi:10.3390/molecules21091204 (2016).

8 Kalisz, A. *et al.* Petals of different ornamental rose cultivars as a rich source of bioactive compounds for functional foods. *Sci. Hortic.* **321**, 112240, doi:10.1016/j.scienta.2023.112240 (2023).

9 Hibrand Saint-Oyant, L. *et al.* A high-quality genome sequence of *Rosa chinensis* to elucidate ornamental traits. *Nat. Plants* **4**, 473-484, doi:10.1038/s41477-018-0166-1 (2018).

10 Zhou, L. *et al.* Multi-omics analyzes of *Rosa gigantea* illuminate tea scent biosynthesis and release mechanisms. *Nat. Commun.* **15**, 8469, doi:10.1038/s41467-024-52782-9 (2024).

11 Chinese Academy of Sciences, E. C. o. F. o. C. *Flora of China*. Vol. 37 (Science Press, 1985).

12 Qun, Y. *et al.* Flowering characteristics and breeding system of the rare plant *Rosa anemoniflora*. *Acta Bot. Boreali-Occidential Sinica* **44**, 1937-1945, doi:10.7606/j.issn.1000-4025.20240072 (2024).

13 Qun, Y. *et al.* Response of leaf functional traits of rare plant *Rosa anemoniflora* to environmental changes. *Journal of Tropical and Subtropical Botany* **32**, 705-714, doi:10.11926/jtsb.4787.

14 Wenger, A. M. *et al.* Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* **37**, 1155-1162, doi:10.1038/s41587-019-0217-9 (2019).

15 Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293, doi:doi:10.1126/science.1181369 (2009).

16 CNCB-NGDC Members and Partners. a. Database resources of the National Genomics Data Center, China National Center for Bioinformation in 2026. *Nucleic Acids Res.* **54**, D28-d47, doi:10.1093/nar/gkaf1172 (2026).

17 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, doi:10.1038/s41592-020-01056-5 (2021).

18 Guan, D. *et al.* Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896-2898, doi:10.1093/bioinformatics/btaa025 (2020).

19 Dudchenko, O. *et al.* De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92-95, doi:10.1126/science.aal3327 (2017).

20 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. *Mol. Biol. Evol.* **38**, 4647-4654, doi:10.1093/molbev/msab199 (2021).

21 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589-595, doi:10.1093/bioinformatics/btp698 (2010).

22 Flynn, J. M. *et al.* RepeatModeler2 for automated genomic discovery of transposable element families. *P Natl Acad Sci Usa* **117**, 9451-9457, doi:10.1073/pnas.1921046117 (2020).

23 Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* **6**, 11, doi:10.1186/s13100-015-0041-9 (2015).

24 Chen, N. Using Repeat Masker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinform.* **5**, 4-10, doi:10.1002/0471250953.bi0410s05Digital (2004).

25 Ou, S. *et al.* Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* **20**, 275, doi:10.1186/s13059-019-1905-y (2019).

26 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, doi:10.1093/bioinformatics/btn013 (2008).

27 Lomsadze, A., Burns, P. D. & Borodovsky, M. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res.* **42**, e119-e119, doi:10.1093/nar/gku557 (2014).

28 Keilwagen, J., Hartung, F. & Grau, J. in *Gene Prediction: Methods and Protocols* (ed Martin Kollmar) 161-177 (Springer New York, 2019).

29 Slater, G. S. C. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinf.* **6**, 31, doi:10.1186/1471-2105-6-31 (2005).

30 Haas, B. J. *et al.* Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654-5666, doi:10.1093/nar/gkg770 (2003).

31 Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol.* **9**, R7, doi:10.1186/gb-2008-9-1-r7 (2008).

32 Kriventseva, E. V., Servant, F. & Apweiler, R. Improvements to CluSTr: the database of SWISS-PROT+TrEMBL protein clusters. *Nucleic Acids Res.* **31**, 388-389, doi:10.1093/nar/gkg035 (2003).

33 Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **18**, 366-368, doi:10.1038/s41592-021-01101-x (2021).

34 Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236-1240, doi:10.1093/bioinformatics/btu031 (2014).

35 Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27-30, doi:10.1093/nar/28.1.27 (2000).

36 Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25-29, doi:10.1038/75556 (2000).

37 Chen, C., Wu, Y. & Xia, R. A painless way to customize Circos plot: from data preparation to visualization using TBtools. *iMeta* **1**, e35, doi:https://doi.org/10.1002/imt2.35 (2022).

38 Chan, Patricia P., Lin, Brian Y., Mak, Allysia J. & Lowe, Todd M. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* **49**, 9077-9096, doi:10.1093/nar/gkab688 (2021).

39 Lagesen, K. *et al.* RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**, 3100-3108, doi:10.1093/nar/gkm160 (2007).

40 Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A. & Eddy, S. R. Rfam: an RNA family database. *Nucleic Acids Res.* **31**, 439-441, doi:10.1093/nar/gkg006 (2003).

41 *National Genomics Data Center* https://ngdc.cncb.ac.cn/gsa/browse/CRA034365/CRR2373096 (2025).

42 *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_054131425.1 (2025).

43 Zhou, X. Gene structural annotations (in GFF3 format) and functional annotation files of *Rosa anemoniflora*. *Figshare*. https://doi.org/10.6084/m9.figshare.30742499.v1 (2025).

44 Huang, N. & Li, H. compleasm: a faster and more accurate reimplementation of BUSCO. *Bioinformatics* **39**, btad595, doi:10.1093/bioinformatics/btad595 (2023).

Funding

This research was funded by the Fujian Province Forestry Science and Technology Research Project (2024FKJ30, 2025FKJ24)

Author information

Author notes

These authors contributed equally: Xianzhen Zhou, Qun Yu.

Authors and Affiliations

Institute of Forest Ecology and Carbon Sink Measurement, Department of Forestry, Fujian Forestry Vocational and Technical College, Nanping 353000, Fujian, China

Xianzhen Zhou, Qun Yu, Chenjie Fu, Guojiang Lin & Wei Gao

Fujian Forestry Prospect and Design Institute, Fuzhou 353001, Fujian, China

Xiaochun Huang

National Nature Reserve of Mangdang Mountain of Fujian Province, Nanping 353000, Fujian, China

Tianqi Kang

Contributions

Q.Y. and W.G. planned the project. C.F., G.L., X.H., and T.K. performed the experiments. X.Z. performed the data analyses. Q.Y., T.K. and X.H. assisted with sampling and experimentation. X.Z. and W.G. wrote and revised the manuscript. Also, all authors read, edited and approved the final manuscript.

Corresponding author
Correspondence to Wei Gao.

Ethics declarations

Competing interests
The authors declare no competing interests.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.