scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Chromosome-level genome assembly of *Phoebe chekiangensis*, an endemic and endangered tree in east Asia

Ying-Feng Hu[1] ✉, Ju Tang[1], Xiao-Hong Li[1], Zhi-Zhong Li[2] & Jian-Wen Shao[1,3] ✉

Phoebe chekiangensis **is a rare nanmu species endemic to southeastern China that provides high-quality timber and has ecological and economic importance. Genomic resources for this genus are limited, constraining research on its genetic diversity, adaptive evolution, and conservation. Here we report a chromosome-level reference genome of *P. chekiangensis* generated using PacBio HiFi long reads, MGI short reads, and Hi-C sequencing. The assembled genome spans 918.59 Mb with a scaffold N50 of 68.99 Mb, and 98.88% of sequences were anchored to 12 pseudochromosomes. Repetitive elements account for 49.45% of the genome. A total of 34,381 protein-coding genes were predicted, of which 91.91% were functionally annotated. Genome quality was supported by a QV of 44.96 (99.99% base accuracy), a BUSCO completeness of 98.4%, and an LAI score of 14.32, meeting reference-quality standards. This high-quality genome provides a valuable resource for conservation genomics, evolutionary biology, and molecular breeding within *Phoebe* and the Lauraceae family.**

Background & Summary

The genus *Phoebe* (Lauraceae) comprises more than 55 recognized species, mainly distributed across tropical and subtropical regions of Asia[1]. Members of this genus produce hard, fine-textured, and aromatic timber that has long been valued for construction, furniture making, and artisanal carving[2,3]. In addition, volatile aromatic compounds and diverse secondary metabolites derived from *Phoebe* species have been reported to exhibit pharmacological activities, underscoring their economic and medicinal value[4,5]. Despite this importance, genomic studies of *Phoebe* remain scarce, with only the genome of *P. bournei* currently available[2,6]. The lack of genomic resources has limited both basic research and applied studies within this genus.

Phoebe chekiangensis C. B. Shang is a rare species endemic to southeastern China and a key source of high-value nanmu timber. Historically, it was widely employed in the construction and furnishing of imperial palaces, while its distinctive morphology also makes it suitable for ornamental planting[7,8]. However, natural populations have declined sharply due to environmental change, human disturbance, and poor natural regeneration[9], leading to its designation as a Grade II species in the List of National Key Protected Wild Plants in China and Vulnerable in the IUCN Red list of threatened species[10,11]. Previous studies have primarily focused on its physiology, ecology, and population dynamics, whereas conservation genetics efforts have largely relied on conventional molecular markers, including simple sequence repeats (SSRs), inter-simple sequence repeats (ISSRs), and chloroplast DNA fragments[10,12,13]. Although plastome[14] and transcriptome[15] data are available, the absence of a reference genome has hindered systematic investigations of its genetic diversity, adaptive evolution, and conservation management.

Here, we present a chromosome-level reference genome of *P. chekiangensis* constructed using PacBio HiFi long reads, MGI short reads, and Hi-C sequencing. The assembled genome spans 918.58 Mb with a contig N50 of 10.38 Mb (Table 1). Hi-C scaffolding anchored 98.88% of sequences to 12 pseudochromosomes, yielding a scaffold N50 of 68.99 Mb (Figs. 1A,2). This assembly shows markedly improved contiguity compared with the

[1]College of Life Sciences, Anhui Normal University, Wuhu, 241000, Anhui, China. [2]Collaborative Innovation Center of Recovery and Reconstruction of Degraded Ecosystem in Wanjiang Basin, Anhui Normal University, Wuhu, 241002, China. [3]The Anhui Provincial Key Laboratory of Biodiversity Conservation and Ecological Security in the Yangtze River Basin, Anhui Normal University, Wuhu, 241002, China. ✉e-mail: huyf@ahnu.edu.cn; shaojw@ahnu.edu.cn

Genome assembly	
Total scaffold length (bp)	918,593,100
Number of chromosome-level scaffolds	12
Longest scaffold (bp)	141,852,477
N50 of scaffold (bp)	68,986,285
Total contig length (bp)	918,578,000
Number of contigs	166
N50 of contig (bp)	10,378,288
GC content (%)	40.05
BUSCO completeness (genome mode) (%)	98.40%
LAI vaule	14.32
QV vaule	44.96
mapping rate(%)	98.07%
Genome annotation	
Number of protein-coding genes	34,381
Repeats in genome (%)	49.45%
TEs in genome (%)	47.81%

Table 1. Summary of the *Phoebe chekiangensis* genome assembly and annotation.

published *P. bournei* genome[2] (contig N50 = 5.64 Mb). Repetitive sequences account for 50.88% of the genome, with transposable elements comprising ~95.11% of the repeats (Table 1). A total of 34,381 protein-coding genes (PCGs) were predicted, of which 31,598 (91.91%) received functional annotations. This reference genome provides an essential resource for the conservation and sustainable utilization of *P. chekiangensis* and serves as a valuable foundation for evolutionary studies and molecular improvement within the genus *Phoebe*.

Methods

Sample collection. In 2024, a healthy and wild individual of *P. chekiangensis* was collected from the type locality at West Lake, Hangzhou, Zhejiang, China (30.19°N, 120.09°E) for genome sequencing (Fig. 1B; Figure S1). Young leaves, mature leaves, stems, and fruits were sampled, immediately flash-frozen in liquid nitrogen, and transported on dry ice to Novogene Corporation (Tianjin, China) for sequencing. A voucher specimen was deposited in the Herbarium of Anhui Normal University (ANUB) under accession number ANUB100117 (Figure S2).

Library construction and genome sequencing. High-quality genomic DNA was extracted from frozen young leaves of *P. chekiangensis* using a modified CTAB method[16], and DNA concentration and purity were assessed with a NanoDrop spectrophotometer. Three sequencing libraries were prepared. First, a short-read library with an insert size of 300–500 bp was constructed using the MGIEasy FS DNA Library Prep Set and sequenced on the DNBSEQ-T7 platform, yielding ~118.37 Gb of clean data (Table S1). Second, a PacBio HiFi library was generated by randomly shearing high-molecular-weight DNA into 15–18 kb fragments, followed by SMRTbell library preparation and Circular Consensus Sequencing (CCS) on the PacBio Revio platform, producing ~45.53 Gb of HiFi reads with an N50 of 18.36 kb (Table S1). Third, a Hi-C library was constructed using the restriction enzyme DpnIII (GATC). After crosslink reversal, DNA was fragmented into 300–700 bp, and interacting fragments were enriched and sequenced on the DNBSEQ-T7 platform (PE150), generating ~96.74 Gb of data for chromosome-scale scaffolding (Table S1).

In addition, total RNA was extracted from flash-frozen tissues of young leaves, mature leaves, stems, and fruits using the Plant Total RNA Isolation Kit (Sangon Biotech Co., China) following the manufacturer's protocol. RNA-seq libraries were prepared with the NEBNext Ultra™ RNA Library Prep Kit for Illumina (NEB) and sequenced on the Illumina NovaSeq. 6000 platform, generating ~48.96 Gb of paired-end (PE150) clean data in total (Table S2).

Genome survey. The genome size of *P. chekiangensis* was estimated using flow cytometry, with *Zea mays* (2.3 Gb) serving as the internal reference. Relative fluorescence intensity suggested that the genome size of *P. chekiangensis* was ~0.39 × that of maize (~0.90 Gb, Figure S3). In addition, ~118.37 Gb of short-read sequencing data were analyzed using Jellyfish v2.3.0[17] with *k-mer* = 19 to generate the *k-mer* frequency distribution, and genome characteristics were inferred with GCE v1.0.2[18], yielding an estimated genome size of 907 Mb, a heterozygosity rate of 2.33%, and a repeat content of 52.19% (Figure S4). Ploidy analysis performed with Smudgeplot v0.4.0[19] indicated that *P. chekiangensis* is diploid, with the most frequent haplotype structure being ab (frequency = 0.82, Figure S5).

Genome assembly. A total of 45.53 Gb (~50.59 × coverage) of PacBio HiFi reads were assembled using Hifiasm v0.25.0-r726[20] with default parameters. Redundant and heterozygous regions were removed with Purge_Dups v1.2.6[21], in which short reads were aligned to the assembly using BWA v0.7.17-r1188[22] to identify and eliminate duplicated sequences. Hi-C data were subsequently used to anchor the primary contigs onto 12 pseudochromosomes with HapHiC v1.0.7[23] under default settings, and potential assembly or scaffolding errors were

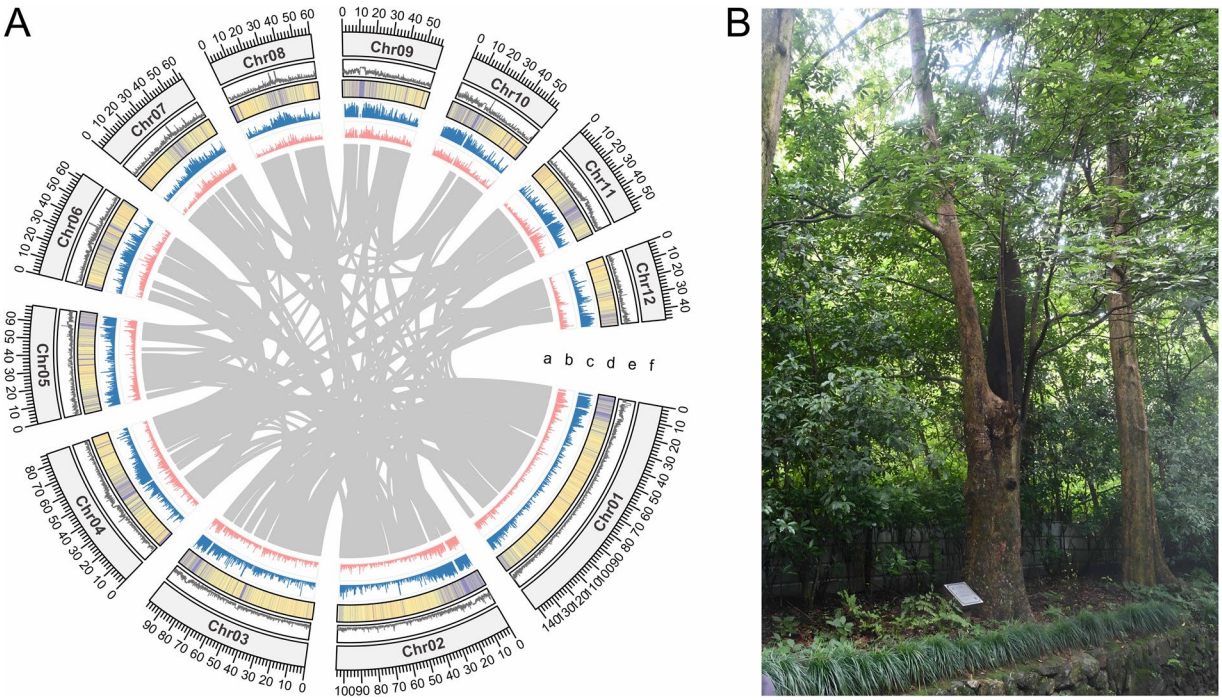

Fig. 1 The circus map (**A**) and habitat (**B**) of *Phoebe chekiangensis*. The circles, from innermost to outermost, show track the syntenic regions within each pseudochromosomes (a), track the LTR-*Copia* density (b), track the LTR-*Gypsy* density (c), track the gene density (d), track the GC content (e), and track the length of pseudochromosomes (f), respectively.

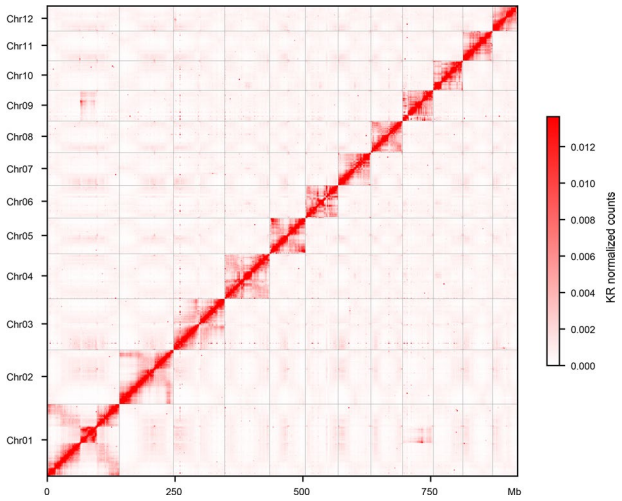

Fig. 2 Hi-C heat map of the finally assembled genome of *Phoebe chekiangensis*. There were strong interactive signals within the 12 chromosomes around the diagonal based on the spatial positional relationship of chromatin DNA. The strength of the interaction was shown by different colors from white (low) to dark red (high).

manually corrected with Juicebox v2.20.0[24] (Fig. 2). Assembly quality was assessed with Merqury v1.3[25] to estimate the QV value, while BUSCO v6.0.0[26] and the LTR Assembly Index (LAI) were employed to evaluate genome completeness[27]. The final chromosome-level assembly of *P. chekiangensis* was 918.59 Mb in size with a GC content of 40.05%, comprising 12 pseudochromosome scaffolds and 20 unanchored scaffolds, and achieved a scaffold N50 of 68.99 Mb (Table 1).

Gene prediction and functional annotation. For genome annotation, repetitive elements were first identified using both homology-based and *de novo* approaches. In the homology-based analysis, RepeatMasker v4.1.9 (http://www.repeatmasker.org) with the Repbase database[28] was used to detect known repeats under default

Type	Number	Length (bp)	Percentage of the genome (%)
DNA transposons	360,296	22,315,883	2.42%
LTR	126,177	221,572,083	24.12%
LINE	20,607	17,933,450	1.95%
Rolling-circles	474	1,645,810	0.17%
Unclassifed	107,019	39,203,031	4.26%
Non-interspersed Repeats(Simple repeat, Low complexity and rRNA)	521,228	22,874,499	2.49%

Table 2. Repeat annotation of the *Phoebe chekiangensis* genome.

Type		Count	% of all protein coding genes
Annotation	NR	30,234	87.94
	Swiss-Prot	20,701	60.21
	KOG	10,791	31.39
	InterPro	29,197	84.92
	eggNOG	28,624	83.26
	KEGG	8711	25.34
Total	Annotated	31,598	91.91
	Protein coding genes	34,381	—

Table 3. Functional annotation of protein coding genes in *Phoebe chekiangensis*.

parameters. For the *de novo* prediction, RepeatModeler v2.0.1[29] was employed to generate a species-specific repeat library, which was subsequently used by RepeatMasker for genome-wide repeat annotation. Unclassified repeat elements were further classified using DeepTE with the plant model[30]. In total, ~454.33 Mb of repetitive sequences were identified (Table 2), representing 49.55% of the genome, including DNA transposons (17.82%), LTR retrotransposons (24.12%), LINEs (1.95%), rolling-circle elements (0.17%), unclassified repeats (4.26%), and other elements (2.49%).

We next annotated non-coding RNAs, including miRNAs, tRNAs, rRNAs, and snRNAs, across the genome. tRNAs were identified using tRNAscan-SE v1.3.1[31], while rRNAs were predicted with RNAmmer v1.2[32] under default parameters. Other ncRNAs, including miRNAs and snRNAs, were annotated using Infernal v1.1.3[33] with recommended settings. In total, 201 miRNAs, 551 tRNAs, 3,229 rRNAs, and 750 snRNAs were identified in the *P. chekiangensis* genome (Table S3).

For PCG annotation, we employed an integrative strategy combining ab initio prediction, homology-based annotation, and RNA-seq–assisted annotation. Ab initio and homology-based predictions were conducted using BRAKER v2.1.5[34] and GeMoMa v1.9[35], respectively, with PCGs from *Oryza sativa*, *Arabidopsis thaliana*, *Phoebe bournei*, *Cinnamomum camphora*, *Lindera megaphylla*, and *Chimonanthus salicifolius* serving as homologous references[2,36–39]. For the RNA-seq–based approach, transcriptomic reads were aligned to the genome with HISAT v2.1.0[40], and transcripts were assembled with TRINITY v2.4.0[41] in both genome-guided and *de novo* modes. The assembled transcripts were further processed with the PASA v2.4.1 pipeline[42] to predict gene structures. Results from the three approaches were integrated using EVM v1.1.1[43], with weighting ratios of PASA:GeMoMa:BRAKER set at 10:5:1, and PASA was subsequently applied to refine the consensus models by adjusting exon–intron boundaries. The final annotation yielded 34,381 PCGs (Table 1) in the genome of *P. chekiangensis*, with mean gene models spanning 7,953.43 bp and containing an average of 4.68 coding sequences (CDSs). Functional annotation was performed by aligning the predicted protein sequences against six public databases (NR, KEGG, Swiss-Prot, InterPro, KOG, and eggNOG) using DIAMOND v4.0.515[44] with an E-value cutoff of 1E-05. In total, 31,598 PCGs (91.91%) were annotated in at least one database (Table 3).

Comparative genomic analysis. To further compare the assembly quality of *P. chekiangensis* with that of the previously reported *P. bournei*, we performed interspecific synteny analysis using JCVI v1.5.3[45] with default parameters. The results revealed a very high degree of collinearity between the two genomes, with 20,898 syntenic gene pairs identified (Fig. 3). Moreover, the syntenic depth exhibited a clear 1:1 relationship, indicating high assembly accuracy and strong genomic consistency between the two species (Figure S6). In addition, assembly quality of *P. bournei* was further assessed using BUSCO v6.0.0[46] in genome mode, which provided a reference for subsequent comparisons with the *P. chekiangensis* assembly.

Data Records

The sequencing datasets generated in this study are available from the National Center for Biotechnology Information (NCBI) under project accession PRJNA1440944. These datasets comprise genome survey (SRX32595579)[47], PacBio HiFi (SRX32595574)[48], Hi-C (SRX32595573)[49], and RNA-seq (SRX32595575–SRX32595578)[50–53] reads. The assembled genome sequence has been deposited in the National Center for Biotechnology Information (NCBI) under accession JBQMFG000000000[54], and the associated genome annotation files are accessible via Figshare (https://doi.org/10.6084/m9.figshare.29644925.v1)[55].

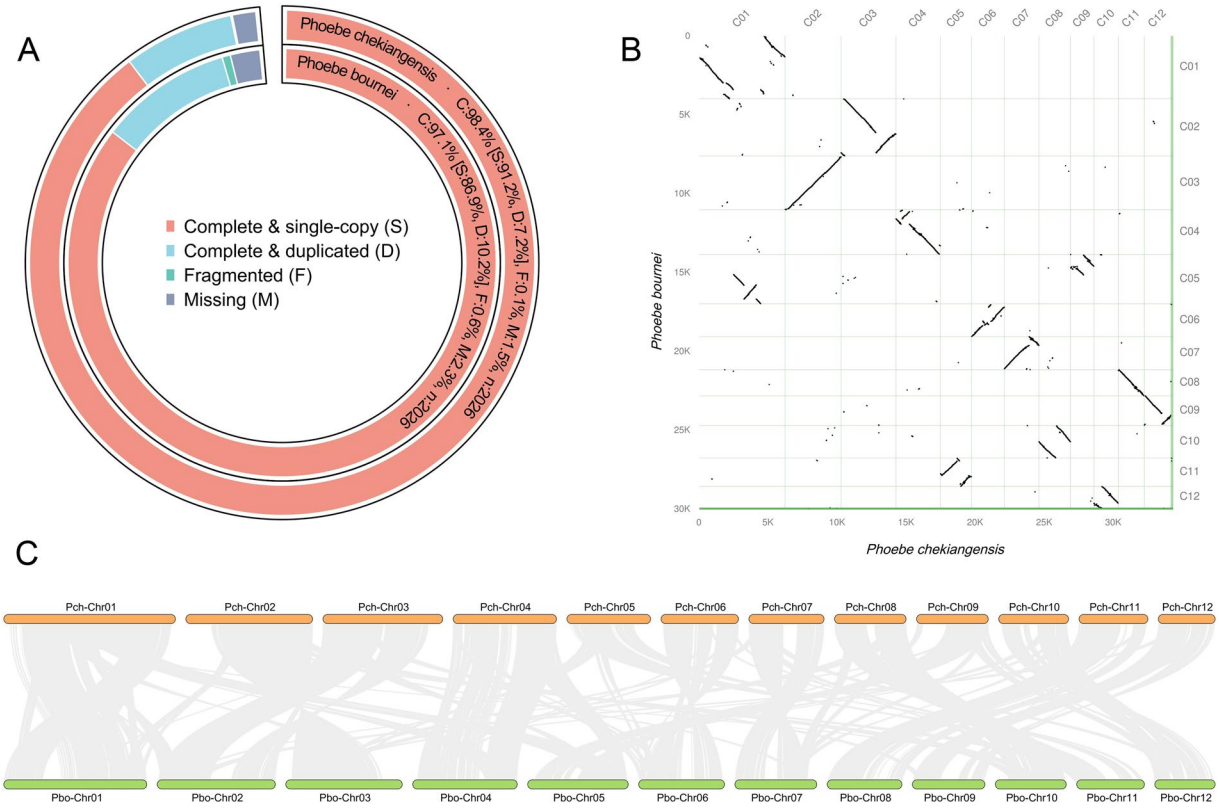

Fig. 3 The BUSCO assessments of *Phoebe chekiangensis* (**A**) and synteny analysis result between *P. chekiangensis* and *P. bournei* (**B,C**).

Technical Validation

The quality of the *P. chekiangensis* genome assembly was evaluated using BUSCO v6.0.0[46] with the embryophyta_odb12 dataset (2,026 core embryophyte genes), which identified 98.4% complete BUSCOs (91.2% single-copy and 7.2% duplicated, Fig. 3). This value exceeds the 97.1% completeness reported for the closely related species *P. bournei* (86.9% single-copy and 10.2% duplicated), reflecting an improvement in assembly quality. The QV estimated with Merqury v1.3[56] was 44.96, corresponding to a base-level accuracy of 99.99% (Table 1). Additionally, resequencing reads mapped to the assembly with a rate of 98.40% and an average coverage of >99.88%, further supporting its completeness and accuracy (Table 1). The LTR Assembly Index (LAI) score was 14.32, confirming that the *P. chekiangensis* assembly meets the reference-quality standard (Figure S7).

Data availability

The data that support the findings of this study are available on the Genbank under accession JBQMFG000000000 (assembled genome sequence, https://identifiers.org/ncbi/insdc:JBQMFG000000000), and the Figshare (associated genome annotation files, https://doi.org/10.6084/m9.figshare.29644925.v1), and the NCBI under project accession PRJNA1440944 (sequencing datasets, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1440944).

Code availability

No custom scripts were used in this study. All analyses were performed with standard bioinformatics tools following the recommended parameters and protocols.

Received: 1 September 2025; Accepted: 1 April 2026;
Published online: 07 April 2026

References

1. Xiao, T. W. *et al.* Plastid phylogenomics of tribe Perseeae (Lauraceae) yields insights into the evolution of East Asian subtropical evergreen broad-leaved forests. *BMC Plant Biol.* **22**(1), 32, https://doi.org/10.1186/s12870-021-03413-8 (2022).
2. Han, X. *et al.* The chromosome-scale genome of *Phoebe bournei* reveals contrasting fates of terpene synthase (TPS)-a and TPS-b subfamilies. *Plant Commun.* **3**(6), 100410, https://doi.org/10.1016/j.xplc.2022.100410 (2022).
3. Wu, J. *et al.* Study on the structural characteristics and physical and mechanical properties of *Phoebe bournei* thinning wood. *J. Renew. Mater.* **10**(11), 3025, https://doi.org/10.32604/jrm.2022.019989 (2022).
4. Ding, W. *et al.* Essential oils extracted from *Phoebe* hui Cheng ex Yang: Chemical constituents, antitumor and antibacterial activities, and potential use as a species identifier. *J. Wood Chem. Technol.* **37**(3), 201–210, https://doi.org/10.1080/02773813.2016.1271435 (2017).

5. Ding, W. *et al*. Essential oil extracted from leaf of *Phoebe bournei* (Hemsl.) yang: chemical constituents, antitumor, antibacterial, hypoglycemic activities. *Nat. Prod. Res.* **34**(17), 2524–2527, https://doi.org/10.1080/14786419.2018.1542393 (2020).
6. Chen, S. P. *et al*. The Phoebe genome sheds light on the evolution of magnoliids. *Hortic. Res.* **7**(1), 146, https://doi.org/10.1038/s41438-020-00368-z (2020).
7. Xie, C. P. *et al*. Analysis on relationship between suitable habitat of *Phoebe chekiangensis* and climatic environmental factors. *J. Sichuan Agric. Univ.* **38**(3), 264–271, https://doi.org/10.16036/j.issn.1000-2650.2020.03.003 (2020).
8. Xu, M. *et al*. Dissecting organ-specific aroma-active volatile profiles in two endemic *Phoebe* species by integrated GC-MS metabolomics. *Metabolites* **15**(8), 526, https://doi.org/10.3390/metabo15080526 (2025).
9. Wu, X. *et al*. Chloroplast spacer DNA analysis revealed insights into phylogeographical structure of *Phoebe chekiangensis*. *Forests* **15**(7), 1073, https://doi.org/10.3390/f15071073 (2024).
10. Lu, Z. L. *et al*. On the necessity, principle, and process of updating the List of National Key Protected Wild Plants. *Biodivers. Sci.* **29**, 1577–1582, https://doi.org/10.17520/biods.2021394 (2021).
11. World Conservation Monitoring Centre. *Phoebe chekiangensis*. The IUCN Red List of Threatened Species **1998**, e.T32438A9707020, https://doi.org/10.2305/IUCN.UK.1998.RLTS.T32438A9707020.en (1998).
12. Zhang, R. *et al*. Genetic diversity and differentiation within three species of the family Lauraceae in southeast China. *Biochem. Syst. Ecol.* **44**, 317–324, https://doi.org/10.1016/j.bse.2012.06.007 (2012).
13. Ding, Y. *et al*. Development of EST-SSR markers and analysis of genetic diversity in natural populations of endemic and endangered plant *Phoebe chekiangensis*. *Biochem. Syst. Ecol.* **63**, 183–189, https://doi.org/10.1016/j.bse.2015.10.008 (2015).
14. Li, Y. *et al*. Complete chloroplast genome sequences of two endangered *Phoebe* (Lauraceae) species. *Bot. Stud.* **58**(1), 37, https://doi.org/10.1186/s40529-017-0192-8 (2017).
15. He, B. *et al*. Transcriptome sequencing and SNP detection in *Phoebe chekiangensis*. *PeerJ* **5**, e3193, https://doi.org/10.7717/peerj.3193 (2017).
16. Doyle, J. J. *et al*. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**, 11–15 (1987).
17. Marcais, G. *et al*. 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).
18. Liu, B. *et al*. Estimation of genomic characteristics by analyzing *k*-mer frequency in *de novo* genome projects. *arXiv* 1308, https://doi.org/10.48550/arXiv.1308.2012 (2013).
19. Ranallo-Benavidez, T. R. *et al*. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **11**, 1–10, https://doi.org/10.1038/s41467-020-14998-3 (2020).
20. Cheng, H. *et al*. Haplotype-resolved assembly of diploid genomes without parental data. *Nat. Biotechnol.* **40**, 1332–1335, https://doi.org/10.1038/s41587-022-01261-x (2022).
21. 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).
22. Li, H. *et al*. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760, https://doi.org/10.1093/bioinformatics/btp324 (2009).
23. Zeng, X. *et al*. Chromosome-level scaffolding of haplotype-resolved assemblies using Hi-C data without reference genomes. *Nat. Plants* **10**(8), 1184–1200, https://doi.org/10.1038/s41477-024-01755-3 (2024).
24. Chakraborty, M. *et al*. Contiguous and accurate *de novo* assembly of metazoan genomes with modest long read coverage. *Nucleic Acids Res.* **44**, e147, https://doi.org/10.1093/nar/gkw654 (2016).
25. Rhie, A. *et al*. Mercury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 245, https://doi.org/10.1186/s13059-020-02134-9 (2020).
26. Manni, M. *et al*. 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**(10), 4647–4654, https://doi.org/10.1093/molbev/msab199 (2021).
27. Ou, S. *et al*. Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Res.* **46**(21), e126, https://doi.org/10.1093/nar/gky730 (2018).
28. Bao, W. *et al*. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* **6**, 11, https://doi.org/10.1186/s13100-015-0041-9 (2015).
29. Flynn, J. M. *et al*. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. USA* **117**, 9451–9457, https://doi.org/10.1073/pnas.1921046117 (2020).
30. Yan, H. D. *et al*. DeepTE: a computational method for de novo classification of transposons with convolutional neural network. *Bioinformatics* **36**(15), 4269–4275, https://doi.org/10.1093/bioinformatics/btaa519 (2020).
31. Schattner, P. *et al*. The tRNAscan-SE, snoscan and snoGPSweb servers for the detection oftRNAs and snoRNAs. *Nucleic Acids Res.* **33**, 686–689, https://doi.org/10.1093/nar/gki366 (2005).
32. Lagesen, K. *et al*. RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**(9), 3100–3108, https://doi.org/10.1093/nar/gkm160 (2007).
33. Nawrocki, E. P. *et al*. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935, https://doi.org/10.1093/bioinformatics/btt509 (2013).
34. Hoff, K. J. *et al*. BRAKER1: Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32**(5), 767–769, https://doi.org/10.1093/bioinformatics/btv661 (2015).
35. Keilwagen, J. *et al*. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res.* **44**, e89, https://doi.org/10.1093/nar/gkw092 (2016).
36. Van Bel, M. *et al*. PLAZA 5.0: extending the scope and power of comparative and functional genomics in plants. *Nucleic Acids Res.* **50**(1), 1468–1474, https://doi.org/10.1093/nar/gkab1024 (2022).
37. Tian, X. C. *et al*. Unique gene duplications and conserved microsynteny potentially associated with resistance to wood decay in the Lauraceae. *Front. Plant Sci.* **14**, 1122549, https://doi.org/10.3389/fpls.2023.1122549 (2023).
38. Jiang, R. *et al*. A Chromosome-Level Genome of the Camphor Tree and the Underlying Genetic and Climatic Factors for Its Top-Geoherbalism. *Front. Plant Sci.* **13**, 827890, https://doi.org/10.3389/fpls.2022.827890 (2022).
39. Lv, Q. *et al*. The *Chimonanthus salicifolius* genome provides insight into magnoliid evolution and flavonoid biosynthesis. *Plant J.* **103**(5), 1910–1923, https://doi.org/10.1111/tpj.14874 (2020).
40. Kim, D. *et al*. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**(4), 357–360, https://doi.org/10.1038/nmeth.3317 (2015).
41. Grabherr, M. G. *et al*. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652, https://doi.org/10.1038/nbt.1883 (2011).
42. Haas, B. J. *et al*. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**(19), 5654–5666, https://doi.org/10.1093/nar/gkg770 (2003).
43. Haas, B. J. *et al*. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**(1), R7, https://doi.org/10.1186/gb-2008-9-1-r7 (2008).
44. Buchfink, B. *et al*. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60, https://doi.org/10.1038/nmeth.3176 (2015).
45. Tang, H. *et al*. JCVI: A versatile toolkit for comparative genomics analysis. *iMeta* **3**(4), e211, https://doi.org/10.1002/imt2.211 (2024).

46. Simao, F. A. *et al*. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**(19), 3210–3212, https://doi.org/10.1093/bioinformatics/btv351 (2015).
47. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR37718282 (2026).
48. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR37718287 (2026).
49. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR37718288 (2026).
50. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR37718283 (2026).
51. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR37718284 (2026).
52. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR37718285 (2026).
53. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR37718286 (2026).
54. Hu, Y. F. The chromosome-level genome sequence of the *Phoebe chekiangensis*. *GenBank* https://identifiers.org/ncbi/insdc:JBQMFG000000000 (2025).
55. Li, Z. Z. Chromosome-level genome assembly of the endemic and endangered plant *Phoebe chekiangensis*. *figshare* https://doi.org/10.6084/m9.figshare.29644925.v1 (2025).
56. Rhie, A. *et al*. 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).

Acknowledgements

This study was supported by grants from the National Natural Science Foundation of China (No. 32470380) and the Peak Discipline Construction Project of Anhui Province (No. 2025GFXK057).

Author contributions

Ying-Feng Hu: Methodology, Formal analysis, Writing-original draft. Ju Tang, Xiao-Hong Li: Formal analysis. Zhi-Zhong Li: Investigation, Funding acquisition, Writing-review & editing. Jian-Wen Shao: Conceptualization, Writing-review & editing, Supervision.

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/s41597-026-07211-5.

Correspondence and requests for materials should be addressed to Y.-F.H. or J.-W.S.