Plant Diversity 47 (2025) 991–994

Contents lists available at ScienceDirect

Plant Diversity

journal homepage: http://www.keaipublishing.com/en/journals/plant-diversity/
http://journal.kib.ac.cn

Case Report

The first haplotype-resolved genome assembly of *Prunus* s.l. subgenus *Laurocerasus* (*Prunus spinulosa*)

Sining Zhang (张飚宁)[a], Jun Chen (陈军)[a,*], Pan Li (李攀)[a,b,**]

[a] Laboratory of Systematic & Evolutionary Botany and Biodiversity, College of Life Sciences, Zhejiang University, Hangzhou 310058, China
[b] Key Laboratory of Biodiversity and Environment on the Qinghai-Tibetan Plateau, Ministry of Education, School of Ecology and Environment, Xizang University, Lhasa 850000, China

ARTICLE INFO

Article history:
Received 26 May 2025
Received in revised form
10 September 2025
Accepted 11 September 2025
Available online 22 September 2025

Keywords:
Haplotype-resolved genome
Phylogenomics
Prunus
Subgenus *Laurocerasus*
Rosaceae
Tetraploidy

ABSTRACT

Prunus spinulosa (2n = 4x = 32) is an evergreen species of significant medicinal usage and ecological value. However, the lacking of a high-quality genome of *P. spinulosa* has obstructed further exploration of its ecological study and phylogenetic relationship of *Prunus*. In this study, we present the first haplotype-resolved genome assembly of *Prunus* s.l. subgenus *Laurocerasus*, the tetraploid genome of *P. spinulosa* was phased into 32 pseudochromosomes with 4 haplotypes, the genome size of each haplotype ranged from 249.82 Mb to 259.69 Mb, and N50 fluctuated from 31.35 Mb to 33.25 Mb, the protein-coding genes vary from 21,272 to 22,668. Different evaluation methods showed that the *P. spinulosa* genome assembly has high quality of completeness, continuity and accuracy. Being the first complete genome of *P. spinulosa*, it provides a valuable genetic resource for the *Prunus* tetraploid species database and supports further functional genomic study of this species.

Prunus sensu lato (s.l.), a genus in the Rosaceae family, contains ten subgenera with approximately 250–400 flowering tree or shrub species (Chin et al., 2014). This genus comprises numerous species of significant economic value, including peach, plum, and cherry, along with oilseed species like almond (Govaerts et al., 2021). Members of *Prunus* share several conserved morphological traits (Rehder, 1927), such as simple leaves bearing basal or petiolar glands, a superior single ovary, and drupaceous fruits. Furthermore, all species in this genus possess a basic chromosome number of eight (Lin et al., 1991). Historically, subgenera within *Prunus* have been categorized into three major groups based on inflorescence structure: (1) racemose inflorescence in subgenera *Padus* Mill., *Laurocerasus* Tourn. ex Duh., *Pygeum* Gaertn., *Prinsepia* Royle., and *Maddenia* Hook. f. et Thoms.; (2) corymbose inflorescence in subgenus *Cerasus* Mill.; (3) solitary flowers in subgenera *Amygdalus* L., *Prunus* L. sensu stricto (s.s.), *Armeniaca* Mill., and *Emplectocladus* Torr. (Potter et al., 2007; Hodel et al., 2021).

The *Prunus* subgenus *Laurocerasus* represents a basal lineage among these subgenera (Hodel et al., 2021), characterized by a high frequency of tetraploidy (Meurman, 1929; Zhao et al., 2016). It comprises approximately 80 species, including *Prunus spinulosa* Siebold & Zucc., an evergreen species distributed in southeastern China, the Philippines, and parts of Japan (Deng et al., 2024). *P. spinulosa* typically inhabits moist, forested environment adjacent to large rivers, at elevations from 300 to 1500 m (Govaerts et al., 2021). Its glossy leaves superficially resemble those of holly (*Ilex bioritsensis* Hayata). In traditional Chinese medicine, seeds of *P. spinulosa* have been used to treat dysentery (Zhang et al., 1994).

At the time this study was conducted, 18 genome assemblies of *Prunus* species were publicly available (Table S1). However, only two of them, *Prunus padus* L. and *P. fruticosa* Pall., are polyploid and belong to subgenera *Padus* and *Cerasus*, respectively (Table S2). This imbalance in available genomic resources has limited our understanding of *Prunus*, particularly among polyploid species and underrepresented subgenera. Here, we report the first chromosome-level haplotype-resolved genome assembly of *P. spinulosa* (the first reference genome for subgenus *Laurocerasus*) with PacBio HiFi and Hi-C sequencing, which contains 87,838 genes. This high-quality tetraploid genome provides a valuable genetic resource for resolving the complex phylogenetic relationships and genome polyploidization evolution of *Prunus*.

Fresh leaves of *Prunus spinulosa* were collected from an individual in Hangzhou Botany Garden, Zhejiang Province, China

* Corresponding author.
** Corresponding author.
E-mail addresses: siningzhang@zju.edu.cn (S. Zhang), cjevol@zju.edu.cn (J. Chen), panli_zju@126.com (P. Li).
Peer review under the responsibility of Editorial Office of Plant Diversity.

https://doi.org/10.1016/j.pld.2025.09.007

(30°24′96″N, 120°12′20″E) for genome sequencing. A voucher specimen (No. ZSN240501-PSpin-001) has been deposited at the laboratory of systematic and evolutionary botany and biodiversity, Zhejiang University. Fresh young leaves, roots, and twigs were also collected from a seedling in Wuyun Mountain, Hangzhou (30°19′11″N, 120°09′41″E) for gene annotation with transcriptome sequencing. All samples were immediately frozen in liquid nitrogen and then stored at −80 °C before further use. Plant materials were delivered to Novogene (Beijing, China) for whole genome sequencing, while transcriptome sequencing was performed by Anoroad (Beijing, China).

We generated 98.32 Gb of Hi-C data, 21.28 Gb of HiFi long-read data, and 235.49 Gb of Illumina short-read resequencing data with the Illumina NovaSeq X Plus and Illumina HiSeq 2500 platforms, corresponding to an overall sequencing depth of approximately 318×. Transcriptome sequencing produced 2.44 Gb, 2.33 Gb, and 2.55 Gb from leaf, twig, and root, respectively (Table S3). Sequencing adapters were trimmed and clean reads were quality-filtered using fastp v.0.23.12 (Chen et al., 2018). Reads with low-quality bases ($Q \leq 5$) or with more than 10% ambiguous bases (N) were removed.

Flow cytometry analysis was conducted with the maize (*Zea mays* L.) inbred line B73 as the internal standard, which gave an estimate of the genome size of *Prunus spinulosa* approximately equal to 1.14 Gb. This estimate is substantially larger than those of known diploid *Prunus* species, which typically range from 227.6 Mb to 344.3 Mb (Table S1). Additionally, to assess the heterozygosity and ploidy of *P. spinulosa*, we conducted a genome survey using Illumina short reads following a two-step *k*-mer based protocol. First, Jellyfish v.2.3.0 (Marçais and Kingsford, 2011) was employed to compute the 21-mer frequency, which was further analyzed in GenomeScope v.2.0 (Vurture et al., 2017). The estimated haploid genome size was 219 Mb, with 59% unique *k*-mer (Fig. S1A). Results of GenomeScope also illustrated that the *k*-mer distribution of *P. spinulosa* was highly consistent with typical patterns of heterozygous polyploid genomes (Ranallo-Benavidez et al., 2020). Next, Smudgeplot v.0.4.0 (James et al., 2020; Ranallo-Benavidez et al., 2020) was used to estimate the heterozygosity and ploidy of *P. spinulosa*, given the same 21-mer histogram with the lower and upper thresholds of *k*-mer set to 15 bp and 285 bp, respectively. The results suggested the genomic composition of *P. spinulosa* most likely match that of either an AB-type diploid (48% probability) or an AABB-type allotetraploid species (36% probability). The probabilities of being other ploidies were negligible, including AAB (7%), AAAB (6%), and AAAABB (3%) (Fig. S1B). Collectively, all above analyses support an allotetraploid origin for *P. spinulosa*.

All Pacbio HiFi and Hi-C reads were first *de novo* assembled into contigs using Hi-C integration mode implemented in hifiasm v.0.19.8 (Cheng et al., 2021) with default parameters. The initial assembly gave a full genome size of 1131.54 Mb, which was consistent with the estimate of flow cytometry. Hi-C reads were aligned to contigs by bwa v.0.7.17 (Li, 2013) with parameters '-mem -5SP -S -h -b -F 3340'. Reads with mapping quality lower than 30 were discarded using SAMtools v.1.6 (Li et al., 2009). The contigs were then scaffolded into 32 pseudochromosomes with HapHiC (Zeng et al., 2024), which were further ordered and oriented based on Hi–C interaction signals using Juicebox (Durand et al., 2016). Ultimately, 90.23% of the contigs was anchored to 32 pseudochromosomes with a final size of 1021.06 Mb and a contig N50 of 31.9 Mb (Table S2). The reference genome assembly comprised two representative subgenome pairs, with sizes of 501.9 Mb and 519.16 Mb, respectively. Hi-C contact maps showed strong intra-subgenomic interactions with clear diagonal patterns (Fig. 1B). The completeness of the *P. spinulosa* genome assembly was evaluated with BUSCO v.5.4.7 (Manni et al., 2021) based on the eudicots_odb10 dataset, which contained 1614 conserved single-copy genes. The result showed an overall completeness of 99.4% (Fig. S1C and Table S2).

To resolve the subgenome structure, synteny analysis was used to compare the genome of *Prunus spinulosa* to that of *P. campanulata*, which had the highest assembly quality among available closely related species. A total of 16 pseudochromosomes showing higher mapping percentages (46.46–81.18%) than *P. campanulata* were designated as subgenome A using RagTag (Alonge et al., 2022). The other 16 pseudochromosomes with lower mapping percentages (15.75–37.52%) were designated as subgenome B (Fig. 1C). The final genome assembly comprised four haplotypes with each size ranging from 249.82 Mb to 259.69 Mb (Fig. 1A). Scaffold N50 values were calculated by QUAST v.5.2.0 (Gurevich et al., 2013) and varied from 31.35 Mb to 33.25 Mb depending on haplotypes. The LTR Assembly Index (LAI) scores (Ou et al., 2018) varies from 17.25 to 17.83, suggesting a high level of repeat sequence continuity across all haplotypes (Table 1).

Given the biological significance of tandem repeats, we constructed a repeat sequence library using RepeatModeler v.2.0.5 (Flynn et al., 2020) for *Prunus spinulosa* genome, which was combined with Repbase v.21.12 (Jurka, 2000), and then used to identify repetitive elements by RepeatMasker v.4.1.5 (Tarailo-Graovac and Chen, 2009). In total, 129.89 Mb–145.99 Mb of repeat contents were identified for *P. spinulosa* four haplotypes, taking up to more than 55% of the whole genome (Table 1).

Subsequently, a comprehensive protocol integrating transcriptome-assisted prediction, homology-based prediction, and ab initio prediction, was performed to annotate the protein coding genes for *Prunus spinulosa* genome. For transcriptome-assisted prediction, RNA-seq reads were aligned to the genome using HISAT2 v.2.2.1 (Kim et al., 2015) as well as *de novo* assembled into transcripts using Trinity v.2.15.1 (Grabherr et al., 2011). Gene models were then predicted using PASA v.2.3.3 (Haas et al., 2003) based on both read alignments and assembled transcripts. Next, homology-based gene prediction was conducted using GeMoMa v.1.9 (Keilwagen et al., 2018) with protein sequences from *P. campanulata* as the reference. Later, ab initio gene prediction was performed by GeneMark-ET v.4.68_lic (Lomsadze et al., 2005) and AUGUSTUS v.3.3.3 (Stanke et al., 2004), respectively. The results were combined by BRAKER v.3.0.3 (Hoff et al., 2016). Finally, gene models predicted by all three approaches were integrated using EvidenceModeler v.1.1.1 (Haas et al., 2008) with weighted contributions set to 'PASA: BRAKER: GeMoMa = 10: 5: 1'. The longest transcript containing both start and stop codons was retained for each gene model. In total, 87,838 protein-coding genes were annotated across the four haplotypes, with gene counts equal to 22,599 in A1, 21,272 in A2, 22,668 in B1, and 21,299 in B2. Correspondingly, 115,922, 105,134, 114,848, and 105,811 exons were identified in A1, A2, B1, B2 haplotype, respectively (Table 1). The completeness of protein coding gene annotation was assessed with BUSCO, which suggested 99.3% of conserved genes were complete, only 0.2% were fragmented and 0.5% were missing.

To assess the collinearity of genome structure within *Prunus*, we performed synteny analyses using NGenomeSyn v.1.41 (He et al., 2023). Analyses were performed both among the four haplotypes of *P. spinulosa* (Fig. S1E) and across *Prunus* genomes from different subgenera (Fig. 1D). Although *P. kanzakura* and *P. yedoensis* were included in the initial analyses, they were not shown in the final visualization due to their relatively fragmented assemblies. Overall, most syntenic blocks were well preserved across subgenera, suggesting genome structure is highly conserved in *Prunus* despite differences in ploidy. In particular, *P. spinulosa* exhibited high synteny with closely related species

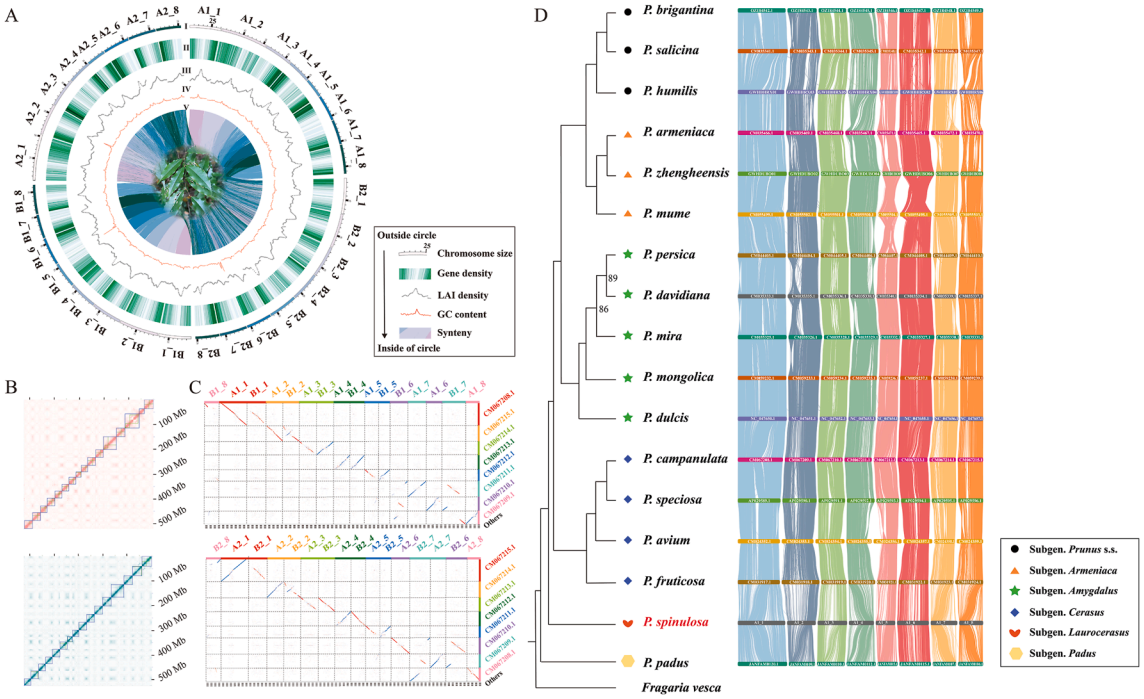

Fig. 1. Genomic features and comparative analysis of *Prunus spinulosa*. (A) The genomic circle plot of *P. spinulosa*. From outside to inside tracks: (I) chromosomes size, (II) gene density, (III) full LAI density, (IV) GC content, and (V) synteny. (B) The Hi-C contact maps of the genome of *P. spinulosa* (scale in Mb). The map shows the calculated interaction frequency distribution of Hi-C links both between and within pseudochromosomes. Individual pseudochromosomes are represented by blue boxes, and contigs are outlined in green. (C) Synteny relationships between the two representative subgenome pairs of *P. spinulosa* and the genome of *P. campanulata*. Different colors on the x-axis and y-axis represent individual pseudochromosomes of *P. spinulosa* and their corresponding chromosomes in *P. campanulata*, respectively. The results support haplotype-specific differentiation within each subgenome. (D) Genome-wide synteny relationships and phylogenetic tree across 17 *Prunus* species, with *P. spinulosa* highlighted in red font. Each line represents a collinearity region, and different shapes of species name labels denote their subgenus origin. The conserved syntenic blocks reveal shared chromosomal architecture and evolutionary rearrangements among the *Prunus* species. Nodes without explicit bootstrap values have 100% support.

Table 1
Summary of *Prunus spinulosa* genome assembly and annotations.

Haplotype	A1	A2	B1	B2
Genome size (bp)	264,324,096	261,958,656	272,305,152	272,074,752
Number of contigs	14	27	24	22
Contig N50 (bp)	31,941,975	31,351,131	32,964,957	33,254,002
Contig N75 (bp)	26,604,073	27,913,756	31,353,280	31,067,524
Number of scaffolds	8	8	8	8
Scaffold N50 (bp)	32,509,856	32,075,939	33,372,689	33,606,860
GC (%)	37.33	37.31	37.32	37.32
LAI	17.71	17.41	17.25	17.83
Repeat elements				
LINE (bp)	14,995,246	16,712,474	16,024,931	15,832,556
SINE (bp)	5,004,735	4,586,067	4,789,382	4,898,584
LTR (bp)	87,422,975	87,139,547	97,684,422	90,745,371
LTR-Gypsy (bp)	7,803,560	7,314,402	8,436,203	8,670,484
LTR-Copia (bp)	2,547,308	2,690,541	2,510,245	2,650,919
LTR-Unknown (bp)	912,021	779,513	920,698	969,489
Unclassified (bp)	22,207,987	16,972,837	22,360,862	29,312,468
Total (bp)	138,600,832	136,195,371	152,726,743	153,079,871
Total (%)	52.43	50.02	58.30	56.26
Protein-coding gene				
Gene number	22,599	22,668	21,272	21,299
mRNA	22,599	22,668	21,272	21,299
CDS	115,922	114,848	105,134	105,811
Exon	115,922	114,848	105,134	105,811

from subgenera *Cerasus* and *Padus*, further supporting the overall quality and structural accuracy of the *P. spinulosa* genome assembly.

Phylogenetic relationships among the *Prunus* species were reconstructed to provide an evolutionary framework for interpreting synteny patterns (Fig. 1D). Orthologous genes were identified with OrthoFinder v.2.5.5 (Emms and Kelly, 2019), focusing on single-copy orthologs (SCOs) and low-copy orthologs (LCOs), the latter defined as genes present in more than 80% of species with less than 5 copies per species. LCO sequences were aligned using MAFFT v.7.505 (Katoh et al., 2002) with '-gt 0.8 -st 0.001 -cons 60', trimmed with trimAL (Capella-Gutiérrez et al., 2009), and concatenated via AMAS.py (Borowiec, 2016) under 'concat -f phylip -d dna'. The maximum likelihood tree was inferred in IQ-TREE v.2.4.0 (Nguyen et al., 2015) with *Fragaria vesca* (Shulaev, 2011) as outgroup. ModelFinder Plus ('-m MFP') tested 1223 models, and node support was assessed with 1000 ultrafast bootstrap replicates and SH-like aLRT tests ('-bb 1000 –alrt 1000'). The resulting topology was consistent with previous classifications of *Prunus* (Hodel et al., 2021; Su et al., 2023), confirming the close relationship of *P. spinulosa* with subgenus *Padus* (Fig. 1D).

In summary, the chromosome-level, haplotype-resolved reference genome of *Prunus spinulosa* presented in this study is of high quality and shall serve as an important genetic resource for future phylogenomic analysis and evolutionary research within *Prunus*.

Data availability

The genome assembly, annotation, and raw reads of HiFi, Hi-C, Illumina resequencing data, and RNA-seq of *Prunus spinulosa* that support this study have been deposited into the China National Center for Bioinformation with bioproject accession number: PRJCA040363.

CRediT authorship contribution statement

Sining Zhang: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal

analysis, Data curation, Investigation. **Jun Chen**: Methodology, Writing – review & editing, Conceptualization, Supervision, Resources, Project administration. **Pan Li:** Methodology, Writing – review & editing, Funding acquisition, Conceptualization, Supervision, Resources, Investigation, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 32570239) and Key Technology Research and Development Program of Zhejiang Province (Grant No. 2023C03138). We thank Dr. Junjie Wu (Zhejiang University) for his help during the genome annotation of this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.pld.2025.09.007.

References

Alonge, M., Lebeigle, L., Kirsche, M., et al., 2022. Automated assembly scaffolding elevates a new tomato system for high-throughput genome editing. Genome Biol. 23, 258. https://doi.org/10.1186/s13059-022-02823-7.

Borowiec, M.L., 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ 4, e1660. https://doi.org/10.7717/peerj.1660.

Capella-Gutiérrez, S., Silla-Martínez, J., Gabaldón, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

Chen, S., Zhou, Y., Chen, Y., et al., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

Cheng, H., Concepcion, G.T., Feng, X., et al., 2021. 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.

Chin, S.W., Shaw, J., Haberle, R., et al., 2014. Diversification of almonds, peaches, plums and cherries - molecular systematics and biogeographic history of *Prunus* (Rosaceae). Mol. Phylogenet. Evol. 76, 34–38. https://doi.org/10.1016/j.ympev.2014.02.024.

Deng, J., Liu, C., Yang, H., et al., 2024. Floristic geographical components and distribution characteristic of species in subgen. *Laurocerasus* (*Prunus* Linn.) in China. J. Plant Resour. Environ. Times 33, 11–21. https://doi.org/10.3969/j.issn.1674-7895.2024.06.02.

Durand, N.C., Robinson, J.T., Shamin, M.S., et al., 2016. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. 3, 99–101. https://doi.org/10.1016/j.cels.2015.07.012.

Emms, D.M., Kelly, S., 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20, 238. https://doi.org/10.1186/s13059-019-1832-y.

Flynn, J.M., Robert, H., Clement, G., et al., 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. U.S. A. 117, 9451–9457. https://doi.org/10.1073/pnas.1921046117.

Govaerts, R., Nic-Lughadha, E., Black, N., et al., 2021. The world checklist of vascular plants, a continuously updated resource for exploring global plant diversity. Sci. Data 8, 215. https://doi.org/10.1038/s41597-021-00997-6.

Grabherr, M., Brian, H., Moran, Y., et al., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652. https://doi.org/10.1038/nbt.1883.

Gurevich, A., Saveliev, V., Vyahhi, N., et al., 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075. https://doi.org/10.1093/bioinformatics/btt086.

Haas, B.J., Delcher, A., Mount, S., et al., 2003. Improving the arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 31, 5654–5666. https://doi.org/10.1093/nar/gkg770.

Haas, B.J., Salzberg, S.L., Zhu, W., et al., 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol. 9, 1–22. https://doi.org/10.1186/gb-2008-9-1-r7.

He, M., Yang, J., Jing, Y., et al., 2023. NGenomeSyn: an easy-to-use and flexible tool for publication-ready visualization of syntenic relationships across multiple genomes. Bioinformatics 39, btad121. https://doi.org/10.1093/bioinformatics/btad121.

Hodel, R., Zimmer, E., Wen, J., 2021. A phylogenomic approach resolves the backbone of *Prunus* (Rosaceae) and identifies signals of hybridization and allopolyploidy. Mol. Phylogenet. Evol. 160, 107118. https://doi.org/10.1016/j.ympev.2021.107118.

Hoff, K., Lange, S., Lomsadze, A., 2016. BRAKER1: unsupervised RNA-seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics 32, 767–769. https://doi.org/10.1093/bioinformatics/btv661.

James, M.P., Valerie, R.H., Crystal, B., et al., 2020. Measuring genome sizes using read-depth, *k*-mers, and flow cytometry: methodological comparisons in beetles (Coleoptera). G3: Genes Genom. Genet. 10, 3047–3060. https://doi.org/10.1534/g3.120.401028.

Jurka, J., 2000. Repbase update: a database and an electronic journal of repetitive elements. Trends Genet. 9, 418–420. https://doi.org/10.1016/s0168-9525(00)02093-x.

Katoh, K., Misawa, K., Kuma, K., et al., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066. https://doi.org/10.1093/nar/gkf436.

Keilwagen, J., Hartung, F., Paulini, M., et al., 2018. Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. BMC Bioinformatics 19, 189. https://doi.org/10.1186/s12859-018-2203-5.

Kim, D., Langmead, B., Salzberg, S., 2015. HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12, 357–360. https://doi.org/10.1038/nmeth.3317.

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint 1303, 3997. https://doi.org/10.48550/arXiv.1303.3997.

Li, H., Handsaker, B., Wysoker, A., et al., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Lin, S., Pu, F., Zhang, J., et al., 1991. Observations on the chromosome number of *Prunus*. China Fruits 2, 8–10. https://doi.org/10.16626/j.cnki.issn1000-8047.1991.02.005.

Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y.O., et al., 2005. Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res. 33, 6494–6506. https://doi.org/10.1093/nar/gki937.

Manni, M., Berkeley, M.R., Seppey, M., et al., 2021. 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. https://doi.org/10.1093/molbev/msab199.

Marçais, G., Kingsford, C., 2011. 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.

Meurman, O., 1929. *Prunus laurocerasus* L., a species showing high polyploidy. J. Genet. 21, 85–94. https://doi.org/10.1007/BF02983360.

Nguyen, L.T., Schmidt, H., Haeseler, A., et al., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274. https://doi.org/10.1093/molbev/msu300.

Ou, S., Chen, J., Jiang, N., 2018. Assessing genome assembly quality using the LTR assembly index (LAI). Nucleic Acids Res. 46, e126. https://doi.org/10.1093/nar/gky730.

Potter, D., Eriksson, T., Evans, R., et al., 2007. Phylogeny and classification of Rosaceae. Plant Syst. Evol. 266, 5–43. https://doi.org/10.1007/s00606-007-0539-9.

Ranallo-Benavidez, T.R., Jaron, K.S., Schatz, M.C., 2020. 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.

Rehder, A., 1927. Manual of Cultivated Trees and Shrubs Hardy in North America, Second ed. Macmillan, New York.

Shulaev, V., Sargent, D., Crowhurst, R., et al., 2011. The genome of woodland strawberry (*Fragaria vesca*). Nat. Genet. 43, 109–116. https://doi.org/10.1038/ng.740.

Stanke, M., Steinkamp, R., Waack, S., et al., 2004. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res. 32, W309–W312. https://doi.org/10.1093/nar/gkh379.

Su, N., Hodel, R., Wang, X., et al., 2023. Molecular phylogeny and inflorescence evolution of *Prunus* (Rosaceae) based on RAD-seq and genome skimming analyses. Plant Divers. 45, 397–408. https://doi.org/10.1016/j.pld.2023.03.013.

Tarailo-Graovac, M., Chen, N., 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Protoc. Bioinformat. 4, 4.10.11–14.10.14. https://doi.org/10.1002/0471250953.bi0410s25.

Vurture, G.W., Sedlazeck, F.J., Nattestad, M., et al., 2017. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics 33, 2202–2204. https://doi.org/10.1093/bioinformatics/btx153.

Zeng, X., Yi, Z., Zhang, X., et al., 2024. 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.

Zhang, H., Zhang, Z., Yue, J., 1994. Summary of Chinese Materia Medical Resources, first ed. Science Press.

Zhao, L., Jiang, X., Zuo, Y., et al., 2016. Multiple events of allopolyploidy in the evolution of the racemose lineages in *Prunus* (Rosaceae) based on integrated evidence from nuclear and plastid data. PLoS One 11, e0157123. https://doi.org/10.1371/journal.pone.0157123.