



OPEN

DATA DESCRIPTOR

Haplotype-resolved T2T genome assembly of the *Populus nigra* NL-1976

Fenfen Liu^{1,6}, Chenggong Liu^{1,6}, An Vanden Broeck², Petra Štochlová³, Xiaolong Jiang⁴, Chengcheng Gao¹, Xueli Zhang¹, Ning Liu^{1,5} & Qinjun Huang^{1✉}

Poplar as a diploid and model plant, it possesses abundant genetic diversity and rapid growth characteristics, making it an important species for artificial timber forests. *Populus nigra* is an important parent in poplar hybrid breeding. With the rapid development of bioinformatics, higher standards of genome analysis are now required. Therefore, haplotype genome assembly of *P. nigra* is essential for accurately distinguishing homologous chromosomes and for identifying genes associated with significant traits. In this study, we successfully generated a haplotype-resolved near telomere-to-telomere (T2T) chromosome-scale genome and achieved gap-free level of *P. nigra*. The genome sizes of the two haploid assemblies were 385,184,975 bp and 390,479,648 bp with contig N50 of 22,312,907 bp and 22,054,730 bp, and both had 19 chromosomes. A total of 49,077 and 50,129 genes were annotated for nigraHap1 and nigraHap2. The high-quality assembly in this study not only provides a reference genome for poplar but also could be a foundational basis for research in tree systems biology.

Background & Summary

Poplar is one of the most widely distributed and cultivated tree species in the world¹. They occupy highly diverse ecological zones, ranging from arid deserts to humid tropical regions, and are classified into 5 to 8 intrageneric sections within the genus². *Populus nigra*, from the sect. *Aigeiros*, is a dioecious, deciduous tree native to Europe, Western Asia, and North Africa³. It is known for its fast growth, stress resistance, low wood density, moderate strength, and attractive tree form⁴. *P. nigra* is one of the preferred species for shelterbelts, timber plantations, and pulpwood, playing a key-role in the sustainable development of softwood forests⁵. It is also considered a promising candidate and important feedstock for second-generation biofuels⁶.

In addition, *P. nigra* is a vital parental species in hybrid poplar breeding programs^{7,8}. Its hybrid offspring with *P. deltoides*, known as *P. × euramericana*, has become an important cultivated poplar species worldwide^{1,9}. However, traditional hybrid breeding and introduction efforts are labor-intensive, time-consuming, and inefficient, limiting the speed of cultivar upgrading¹⁰. Unfortunately, although the national conservation programmes of *P. nigra* exist in most European countries and include the protection of *in situ* populations in conservation units¹¹. However, as climate change intensifies, the *P. nigra* has become one of the most threatened tree species in Europe, mainly because of the loss of its natural alluvial habitats especially sand and gravel banks that allow for successful reproduction¹². Therefore, it is crucial to implement necessary measures to accurately reveal the genetic characteristics of *P. nigra*. This will enable the exploration of its genetic potential, shorten the breeding cycle, and accelerate the renewal of varieties. Such efforts are essential for the improvement of poplar species and for ensuring timber security and ecological safety in the face of future global climate change.

The development of molecular breeding technology has enabled the achievement of these objectives, with high-quality genome assembly and annotation serving as essential tools in conservation strategies^{13,14}. Genome assemblies and annotations aid to identify adaptive traits crucial for survival such as drought tolerance in trees¹⁵. Genome data can therefore help conservationists to identify distinct conservation units. By this way, functional

¹State Key Laboratory of Tree Genetics and Breeding, Research Institute of Forestry, Chinese Academy of Forestry, Beijing, 100091, China. ²Research Institute for Nature and Forest (INBO), Gaverstraat 4, 9500, Geraardsbergen, Belgium. ³Silva Tarouca Research Institute for Landscape Research Institute CZ-252 43 Průhonice, Průhonice, Czech Republic. ⁴College of Forestry, Central South University of Forestry and Technology, Changsha, 410004, China. ⁵UGent-Woodlab (Laboratory of Wood Technology), Department of Environment, Ghent University, Ghent, 9000, Belgium. ⁶These authors contributed equally: Fenfen Liu, Chenggong Liu. ✉e-mail: huangqj@caf.ac.cn

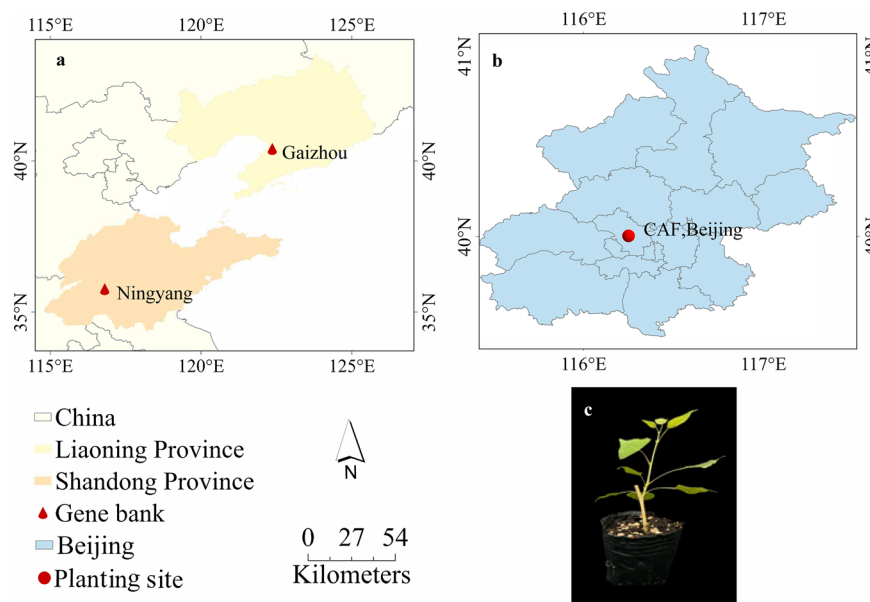


Fig. 1 Information on the *P. nigra* 'NL-1976' germplasm in China. (a) Different provenance of *P. nigra* 'NL-1976' introduced to China. (b) Planting site of experimental sample *P. nigra* 'NL-1976'. (c) One-year-old plant of *P. nigra* 'NL-1976' from cutting.

genomics contributes to the understanding of a species' adaptive potential and resilience to environmental changes¹⁶. For instance, with the rapid development of sequencing techniques, an increasing number of complex genomes from both plants and animals have been successfully phased and constructed to the T2T level. The release of these high-quality genomes has laid a basis for species evolution, genetic variation, hybrid vigor, and more researches^{2,17–19}. Extending the T2T method to a broader array of germplasm resources will help bridge the gap between genomic data and phenotypic outcomes, providing researchers with a unique opportunity to implement genome-wide haplotype-based plant improvement initiatives¹⁴.

An examination of the history of poplar breeding and genetic improvement indicates that following the release of the poplar genome through shotgun sequencing technology in 2006, genomes of various poplar species have been published. Notably, both poplar '84K' (*P. alba* × *P. tomentosa*) and *P. trichocarpa* have achieved the T2T level^{2,20–22}. For the *P. nigra*, while its chromosome-scale genome has been reported, it has not yet reached T2T level, with 277 gaps remaining²³. The genome still exhibits relatively low continuity and quality, and many challenges remain unresolved. Furthermore, plant genome assembly encounters distinct challenges. Polyploidy, both ancient and recent, is prevalent in plants, and their genomes are characterized by an abundance of highly similar long repetitive sequences¹⁴. The high heterozygosity and the presence of these repetitive sequences in the *Populus* genome result in genome assemblies that are often not highly contiguous, and the assemblies in repetitive regions, centromeres, and telomeres tend to be incomplete²⁴. Thanks to advances in sequencing methods, Oxford Nanopore sequencing combined with the increasingly mature PacBio HiFi technology-offering has become the primary data type for high-quality genome assembly. Moreover, the HiFi + ONT + Hi-C methods, which leverages longer reads to resolve complex chromosome structures and effectively assemble repetitive areas, is currently the best assembly strategy available²⁵.

In this study, the clonal variety *P. nigra* 'NL-1976', introduced from the Netherlands, was the material. Using PacBio HiFi sequencing, Oxford Nanopore ultra-long sequencing, and Hi-C techniques, we first constructed a high-fidelity T2T haplotype genome for *P. nigra*, filling the gap in the T2T haplotype genome of this species. Through the analysis of Illumina RNA and ONT-RNA sequencing data, we conducted haplotype gene annotation and functional prediction, resulting in the generation of two gapless haplotype genomes that annotated 49,077 and 50,129 genes, respectively. Our results can provide some experience and basis for the systemic genomics research of poplars and other forest trees.

Methods

Sample collection and DNA extraction. The *P. nigra* 'NL-1976' (male) used in this study was originally from the Netherlands and introduced to China in 2000 by the Chinese Academy of Forestry (CAF), where it was planted in Ningyang, Shandong and Gaizhou, Liaoning (Fig. 1a). In March 2024, we collected one-year-old cuttings and expanding planted them in automatic artificial controlled greenhouse of CAF, Beijing (Fig. 1b). In July 2024, fresh young healthy leaves, stems, and roots were collected (Fig. 1c), rapidly frozen in liquid nitrogen, and then stored at -80°C for subsequent genome sequencing. Genomic DNA was extracted from young leaves using the modified CTAB method for genome sequencing.

Genome sequencing. *Long-read library construction and sequencing.* For HiFi sequencing, PacBio SMRT sequencing was performed on the PacBio revio platform, with single-molecule real-time circular consensus

Sequencing	Reads number	Reads base (bp)	Average reads length (bp)	Reads N50 (bp)	Depth (×)
ONT	464,052	47,822,043,164	103,053.2	100,000	111.97
Hi-C	372,371,116	111,711,334,800	2 × 150	2 × 150	—
HiFi	2,852,700	53,139,588,030	18,641.5	18,836	124.42
RNA-seq	39,519,254	11,855,776,200	2 × 150	2 × 150	—
ONT-RNA	9,120,715	9,478,694,728	1,039.25	1,144	222.93

Table 1. Summary of DNA sequencing data of *P. nigra* 'NL-1976' genome.

sequencing (CCS) library preparation. High-quality genomic DNA was extracted from the leaves, sheared, and purified. Sequencing libraries were prepared and subjected to fragment size selection prior to sequencing on the PacBio Sequel II platform. The DNA libraries were sequenced on 2 SMRT cells. A total of 53.14 Gb HiFi reads were generated, with approximately $124.42 \times$ coverage of the haploid genome, an N50 of 18.84 kb, and 2,852,700 reads, the smallest fragment length being 17.62 kb (Table 1).

High-quality DNA were used for ONT PromethION library preparation and sequencing, following the manufacturer's guidelines (Oxford Nanopore Technologies). The final sequencing produced 47.82 Gb of data, with 464,052 reads, an average length of 103,053.2 bp, an N50 of 100 kb, and an N90 of 78,120 bp (Table 1). For ONT full-length RNA sequencing, after quality control, total RNA was reverse transcribed and labeled. The resulting transcript was purified using magnetic beads. Sequencing adapters were then ligated to the purified product using the SQK-PCS109 kit, and the cDNA library was precisely quantified with Qubit. The sequencing was performed following the Nanopore library construction protocol described in Jain²⁶. The results generated 9.48 Gb of data, with a total of 9,120,715 reads. The longest sequence was 54,286 bp, with the average length of 1,039.25 bp, an N50 of 1,144 bp, and an N90 of 626 bp (Table 1).

Hi-C sequencing. The Hi-C technique, developed by Lieberman *et al.* in 2009²⁷, was used in this study. The sample cells were fixed and biotinylated with the DPNII restriction enzyme. DNA fragments with interactions were then ligated, purified, and fragmented. The 5' ends were phosphorylated, and a dA tail was added to the 3' ends before ligating adapters. The captured Hi-C DNA was PCR amplified, and the library concentration and insert size were determined using a Qubit 3.0 and Agilent 2100. High-throughput sequencing was performed on the MGI platform, with PE150 read length. Generated 372,371,116 reads and 111.71 Gb of data, with a GC content of 38.98% and a Q30 value of 94.85% (Table 1).

Second generation transcriptome sequencing. RNA was extracted from roots, stems, and leaves. mRNA was enriched using oligo dT magnetic beads, followed by fragmentation, cDNA synthesis, end repair, and dA-tailing. The fragments were ligated, PCR amplified, denatured, and circularized. Finally, high-throughput sequencing was performed on the DNBSEQ platform. Generated 11.86 Gb of data with 39,519,254 reads, a GC content of 44.65%, and a Q30 value of 94.42% (Table 1). All the work of sequencing above relies on Wuhan Benagen Technology Co., Ltd. (Wuhan, China).

Genome survey analysis. Before genome assembly, a genome survey was conducted to understand the genome size, GC content, and heterozygosity to develop an appropriate sequencing strategy. After DNA extraction and library preparation, sequencing was performed on the BGI platform, generating a large amount of short-read sequence data. Next, we used Fastp v0.21.0²⁸ software to remove low-quality sequences and contaminants. Then, Jellyfish v2.3.026²⁹ was used to calculate the frequency distribution of the depth of clean data with 19-mer. with a K-mer depth of $129.6 \times$. Genome size was estimated to be 427.1 Mb with a heterozygosity of 1.05% using GenomeScope v2.0³⁰ (Fig. 2).

Raw Data Filtering and Genome Assembly. We performed *de novo* genome assembly to obtain haploid T2T genomes (Fig. 3). Before assembly, we first using Fastp and SeqKit v2.10.0³¹ software to retain HiFi raw sequences longer than 10 kb as clean reads. For ONT data, we retained sequences longer than 100 kb and filtered adapter sequences based on Porechop v0.2.4²⁷ to. Fastp was used to filter Hi-C data. Then, we performed the primary assembly using Hifiasm v0.25.0^{32–35} to assemble three genome sets: pure HiFi, pure ONT, and HiFi + ONT + Hi-C. The pure HiFi and pure ONT genomes were later used to fill gaps and resolve telomeres. The genome generated by the parameters "hifiasm-h1 hic_R1.fq.gz-h2 hic_R2.fq.gz -ul ont.fastq.gz hifi.fa" serves as the backbone of the T2T genome, producing two haploid assemblies, representing the diploid genome's parental haplotypes. We used Purge_dups v1.2.5³⁶ to remove redundancy from all three genome sets.

Next, the Hi-C reads were mapped to the genome assembly using Juicer v1.7.6³⁷. We applied an automated process to correct and order orientation errors through the 3D-DNA v1.03.73³⁸ scaffolding pipeline. Juicebox v1.10.8³⁹ was used for manual adjustments of the assembled scaffolds in a graphical and interactive manner. To further improve accuracy, each chromosome was individually re-scaffolded with 3D-DNA and manually adjusted using Juicebox. Gaps were filled with ONT and HiFi contigs, and ONT ultra-long reads were further utilized for gap filling using TGS-GapCloser v1.2.1⁴⁰. Telomere sequences were identified using the quarTeT v1.1.6⁴¹ software with the "-c plant" parameter, and telomeres were filled using minimap2 v2.28⁴². Genome polishing was performed using NextPolish v1.4.1⁴³ with the task set as "rewrite = 1212".

We used Mummer v4.0.1⁴⁴ for synteny analysis of two haploid genomes with the *P. trichocarpa*¹⁵ reference genome to validate the genome's correctness. Genome continuity was evaluated using BUSCO v1.0.0 software

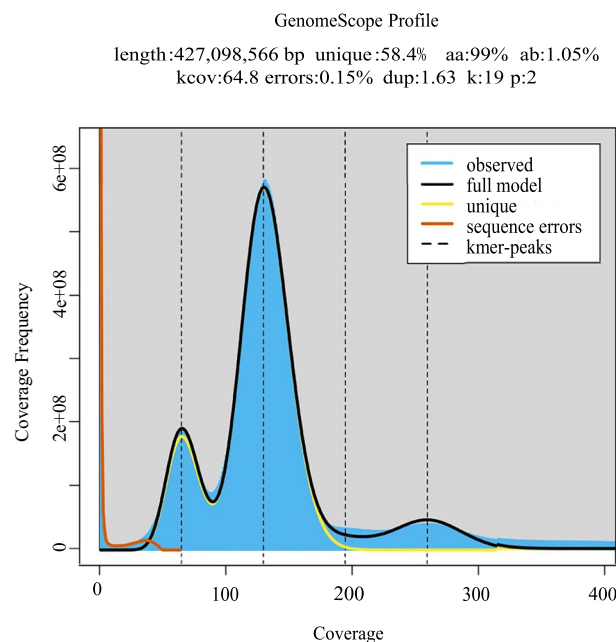


Fig. 2 K-mer depth and K-mer individual frequency distribution plot at K-mer = 19.

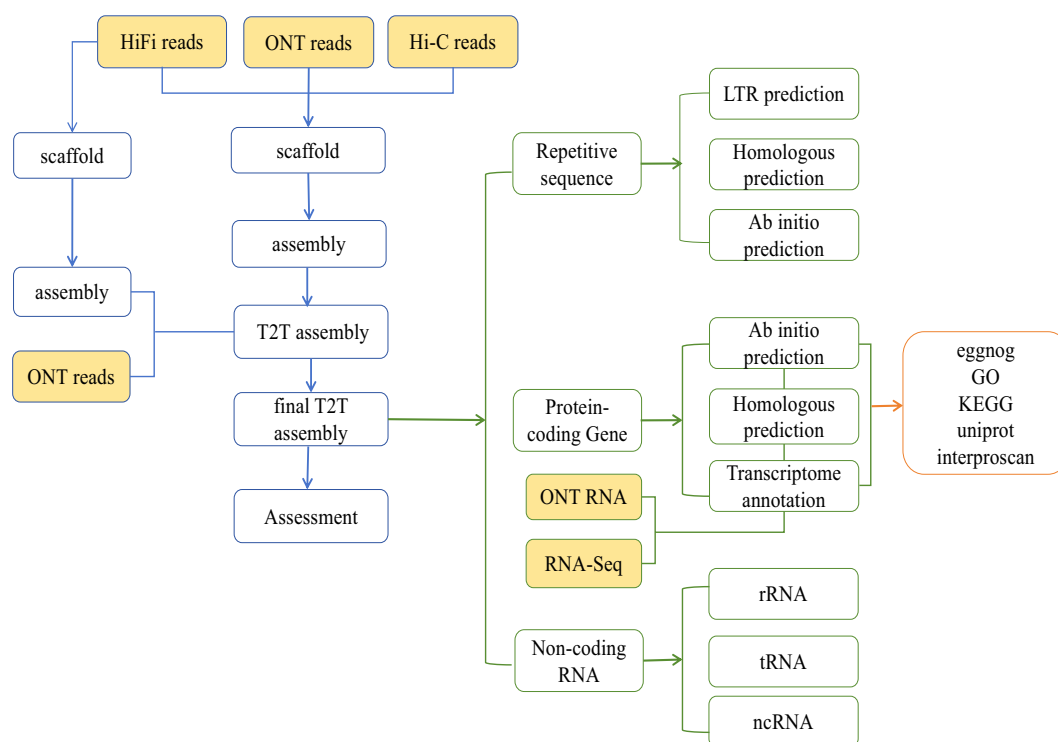


Fig. 3 The workflow of the genome assembly (in blue), the genome annotation (in green) and the functional annotation (in orange). The yellow filled box represents sequencing data. The software used are indicated in red.

with the specified “embryophyta_odb10” database⁴⁵. Both PacBio long reads and Illumina reads were mapped to the genome assembly using minimap2 and bwa v0.7.17⁴⁶, respectively.

The final assembled genome contained two fully separated haplotypes, named nigraHap1 and nigraHap2, each with 19 chromosomes ($2n = 38$). Compared to other published poplar genomes (*P. deltoides* ‘I-69’⁴⁷, *P. nigra* subsp. *betulifolia* L., 1753²³, *P. trichocarpa* Nisqually-122) as shown in Table 2, for the same *Populus nigra*, our genome assembly has reached the near-T2T level and achieved a 0-gap status, with BUSCO assessment results also being 0.3% and 0.4% higher. The genome sizes were 390.48 Mb and 385.14 Mb, respectively. The contig N50

Species	<i>P. nigra</i> 'NL-1976' nigraHap1	<i>P. nigra</i> 'NL-1976' nigraHap2	<i>P. nigra</i> 'NL-1976'	<i>P. deltoides</i> 'I-69'	<i>P. nigra</i> betulifolia L., 1753	<i>P. trichocarpa</i> Nisqually-1 Trahap1	<i>P. trichocarpa</i> Nisqually-1 Trahap2
Genome size (Mb)	385.18	390.48	389.43	424.59	414.18	391.76	397.43
GC content (%)	33.6	33.72	33.64	33.34	35	33.86	33.91
Number of Gap	0	0	0	0	277	6	2
N50 (Mb)	22.31	22.05	22.04	21.51	22.49	21.8	20.9
Maximum scaffold sequence length (bp)	48,831,123	49,313,916	48,827,307	53,044,917	50,570,738	52,164,823	52,084,038
Minimum scaffold sequence length (bp)	12,830,606	13,255,773	6,224,961	15,239	55,380	15,726	10,965
Average length (bp)	20,272,893.4	20,551,560.4	18,544,479	1,608,294.8	16,567,390	3,211,530.1	1,252,100
Complete BUSCOs (%)	98.7	98.8	98.7	98.6	98.4	98.5	98.2

Table 2. Four species of poplar statistics of the genome quality for the final assembly.

Chromosome	nigraHap1	nigraHap2
chr01	48,831,123	49,313,916
chr02	24,687,210	24,637,615
chr03	20,438,832	21,377,002
chr04	21,719,839	22,054,730
chr05	23,107,642	22,762,548
chr06	25,552,461	26,084,403
chr07	14,470,518	14,648,781
chr08	24,560,674	25,781,307
chr09	12,830,606	13,255,773
chr10	22,312,907	22,126,957
chr11	17,371,465	17,510,466
chr12	15,794,386	15,961,851
chr13	15,120,485	15,196,773
chr14	23,581,014	23,839,270
chr15	14,006,661	14,330,654
chr16	14,065,549	13,996,680
chr17	15,317,265	15,587,876
chr18	14,270,132	14,569,084
chr19	17,146,206	17,443,962

Table 3. Chromosome length statistics of the *P. nigra* 'NL-1976' haploid genomes.

lengths for the two haplotypes were 22.31 Mb and 22.05 Mb, with no gaps in either genome. The chromosome lengths are listed in Table 3.

Genome annotation. The annotation process for the final T2T haploid genome follows the workflow (Fig. 3), which includes repetitive sequence annotation, gene structure annotation, and non-coding RNA annotation. LTRs were identified using LTR_HARVEST_parallel⁴⁸ and LTR_FINDER_parallel⁴⁹ with the parameter settings: -size 5000000 -threads 60 -finder_para -w 2 -C -D 15000 -d 1000 -L 7000 -l 100 -p 20 -M 0.85. For repetitive sequence annotation, we used LTR_retriever v3.0.1⁵⁰ to annotate long terminal repeats (LTRs). Homology-based prediction was carried out using RepeatMasker v4.18⁵¹ and RepeatModeler v2.0.6⁵² for de novo prediction of repetitive sequences (Fig. 4). The lengths of the repetitive sequences in nigraHap1 and nigraHap2 were 178,528,825 bp and 184,871,473 bp, respectively, accounting for 46.35% and 47.35% of the total genome size. Among the interspersed repeats, five types of transposable elements (TEs) were identified, including long terminal repeats (LTRs), Long interspersed nuclear elements (LINEs), DNA elements (DNA), short interspersed nuclear elements (SINEs), and unclassified elements. The quantities and sizes of these elements are shown in Table 4. Among these, the unclassified regions were the most abundant, followed by LTR repeats, and the least abundant were SINE repeats.

The structural annotation consists of three parts: homology prediction, de novo prediction, and transcript prediction, with the results of the three methods combined at the end. First is homology prediction, where data including *Arabidopsis*⁵³, *P. trichocarpa*²², and *P. alba* × *P. glandulosa* (84 K)²¹. Homology prediction was performed using Miniport v0.14⁵⁴. For transcript prediction, data from both third-generation full-length transcriptomes and second-generation transcriptomes were combined. Hisat2 v2.2.1⁵⁵ and Stringtie v3.0.0^{56,57} were used for second-generation transcript prediction. Third-generation ONT data were filtered using NanoComp v2.0⁵⁸ and Chopper v2.7.10⁵⁹ to remove fragments with a quality lower than 7 and a length shorter than 50 bp. The command used was “chopper -q 7 -l 50”, resulting in a sequence of 14,138,072,330 bp. These sequences were

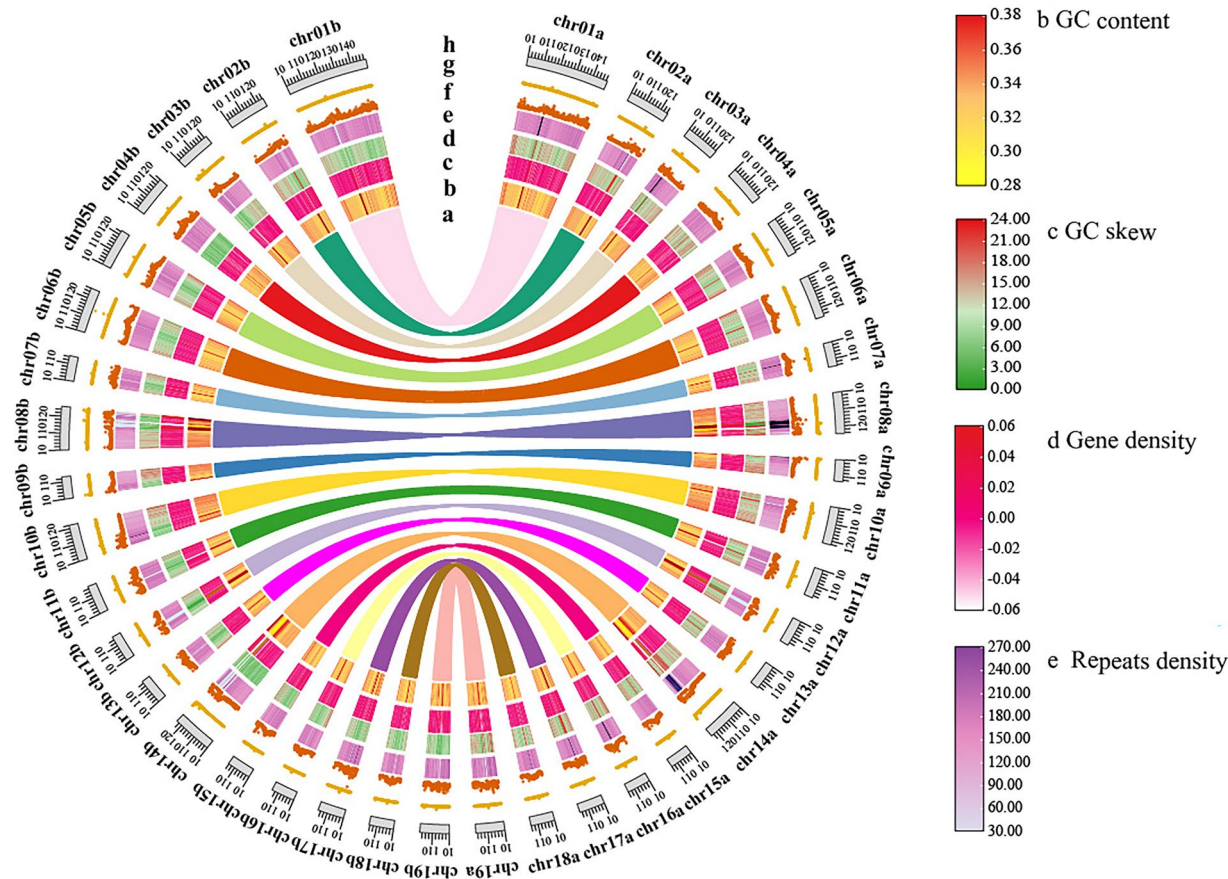


Fig. 4 Circos plot of *P. nigra* ‘NL-1976’ haploid. The genomic overview is presented from outer to inner circles as follows. (a) Synteny between nigraHap1 and nigraHap2. (b) GC content in non-overlapping 1 Mb windows. (c) GC skew in non-overlapping 1 Mb windows. (d) Gene density in non-overlapping 1 Mb windows. (e) Percentage of interspersed repeats in non-overlapping 1 Mb windows. (f) LTR content in non-overlapping 1 Mb windows. (g) LINE content in non-overlapping 1 Mb windows. (h) Chromosome length (Mb).

Type	nigraHap1			nigraHap2		
	Number	Length (bp)	Percentage	Number	Length (bp)	Percentage (%)
LTR	68,118	57,036,655	14.81	71,134	60,326,817	15.45
LINEs	5,802	4,131,930	1.07	6,112	4,312,031	1.10
DNA transposons	19,768	13,999,522	3.63	20,422	13,467,978	3.45
SINEs	4,402	802,120	0.21	3,457	575,447	0.15
Unclassified	371,412	91,171,807	23.67	302,784	94,326,075	24.16
Total	469,502	178,528,825	46.35	403,909	184,871,473	47.35

Table 4. Statistics of repeat sequence annotation.

then mapped to the genome using minimap, with stringtie predicting the transcripts, and gffcompare v0.12.6⁶⁰ merging the results of the second and third-generation predictions.

Finally, TransDecoder v5.7.1⁶¹ was used for coding prediction, and the results were concatenated. For de novo prediction, the second-generation RNA was first assembled using Trinity v2.15.2⁶² and PASA v2.53⁶³, followed by redundancy removal with cd-hit v4.8.1⁶⁴. Third-generation ONT full-length RNA was used to train the model “etraining–species = populus-nigra”. Augustus v3.5.0⁶⁵ and GeneMark-ES v2.0⁶⁶ were used for gene prediction. Finally, we used EvidenceModeler v2.1.0⁶⁷ to combine the three types of evidence. Predictions encoding fewer than 50 amino acids were discarded. A total of 49,077 and 50,129 protein-coding genes were predicted for both nigraHap1 and nigraHap2. The total lengths of the protein-coding genes were 145.54 Mb and 146.79 Mb, with a total of 285,417 and 206,628 exons for nigraHap1 and nigraHap2, respectively. The structural prediction GFF files were evaluated using BUSCO, yielding a result of 97.6% (Table 5).

Non-coding RNA prediction included rRNA, tRNA, and ncRNA predictions. Rnammer v1.2⁶⁸ was used for rRNA prediction, tRNAscan-SE v2.0.12⁶⁹ was used for tRNA prediction, and Infernal v1.1.5⁷⁰ was used for ncRNA prediction. The identification of transfer RNAs (tRNAs) was performed using tRNAscan-SE. Other

Assembly	nigraHap1	nigraHap2
Total number of gene	49,077	50,129
Total length of gene (bp)	145,536,768	146,785,024
Average length of mRNA (bp)	2,965.5	2,928.15
Total number of exons	285,417	279,775
mRNAs per gene	1	1
Average length of exon (bp)	216.1	223.47
Average exon number per gene	5.8	5.4

Table 5. Statistics of predicted protein-coding genes.

Type		nigraHap1			nigraHap2		
		Cope number	Average length (bp)	Total length (bp)	Cope number	Average length (bp)	Total length (bp)
miRNA		630	140.20	88,324	231	125.02	28,879
tRNA		114	74.11	8,448	487	77	70,827
rRNA	total	10	509.2	5,092	14	470.64	6,589
	LSU rRNA	5	671.6	3,358	6	768.5	4,611
	SSU rRNA	4	406	1,624	7	266	1,862
	5S	1	110	110	1	116	116
snRNA	total	80	101.19	8,095	364	104.71	38,114
	CD-box	69	95.36	6,580	254	91.67	23,285
	HACA-box	3	134.33	403	72	124.54	9,111
	Splicing	8	139	1,112	38	150.47	5,718

Table 6. Classification of repetitive sequences and ncRNAs of the *P. nigra* 'NL-1976' genome.

non-coding RNAs (ncRNAs), such as microRNAs (miRNAs), ribosomal RNAs (rRNAs), and small nuclear RNAs (snRNAs), were identified using Infernal by searching against the Rfam v.14.147⁷¹ database. Finally, the number of miRNAs, tRNAs, rRNAs, and snRNAs predicted from the nigraHap1 and nigraHap2 genomes were 630 and 231, 114 and 487, 10 and 14, and 80 and 364, respectively (Table 6).

Protein function prediction. For functional annotation of protein-coding genes, the gene function annotation is based on the previously predicted gene structure results. The `gff3_file_to_proteins.pl` script was used to extract protein sequences from the gene structure `gff3` results. The functional predictions for GO, KEGG, and Pfam were performed using the online version of egg-nog-mapper (<http://egg-nog-mapper.embl.de/>), while UniProt (EBI)⁷² was used for SWISS-PROT database functional predictions. The InterProScan v5.60⁷³ software was used for domain prediction. InterProScan database predicted the highest number of genes, with 42,965 (87.55%) and 43,683 (87.14%) for the two groups, respectively (Fig. 5).

Identification of centromeres and telomeres. Using the `quarTeT` software with the “-c plant” program to identify telomeric sequences. The `T2Tvalidator` plugin in `TBtools` v2.154⁷⁴ automatically identifies and visualizes the centromere and telomere regions of the T2T genome based on the sequences. No telomeric sequences were found at the right ends of chr08 and chr14 (Table 7 and Fig. 6).

Data Records

The raw data of PacBio HiFi sequencing reads, Hi-C sequencing reads, ONT sequencing reads and RNA-seq sequencing data described in this study have been deposited at the National Genomics Data Center (NGDC)⁷⁵ in GSA database under BioProject accession number PRJCA040372, title is “*Populus nigra* Raw sequence reads”, all are “.gz” compressed files, and all raw sequencing data are publicly accessible. The accession numbers of PacBio HiFi reads, Hi-C sequencing reads, ONT sequencing data are CRX1730796⁷⁶, CRX1730797⁷⁷ and CRX1730798⁷⁸, they are publicly available. The Illumina RNA-seq and full-length RNA-seq data are available in the GSA database with the accession number CRX1730799⁷⁹ and CRX1730800⁸⁰.

The genome assembly, annotation sequences and protein sequence are available at Figshare⁸¹. The genome assembly has also been deposited in NCBI GenBank with accession number JQZWWZ000000000⁸² for nigraHap1 and JQZWY000000000⁸³ for nigraHap2, respectively, all data are publicly available.

Technical Validation

To ensure the completeness and accuracy of the two haplotype-resolved genomes of *P. nigra* 'NL-1976', The assembly and annotation's completeness of the set of highly conserved single-copy orthologous genes in the genome was assessed using BUSCO (within the `embryophyta_odb10` database). The genome assembly completeness results showed that the two haplotypes covered 98.7% (1,593) and 98.8% (1,595) of complete BUSCOs, with single-copy genes accounting for 82.3% and 82.0%, respectively, and duplicated genes numbering 265 and

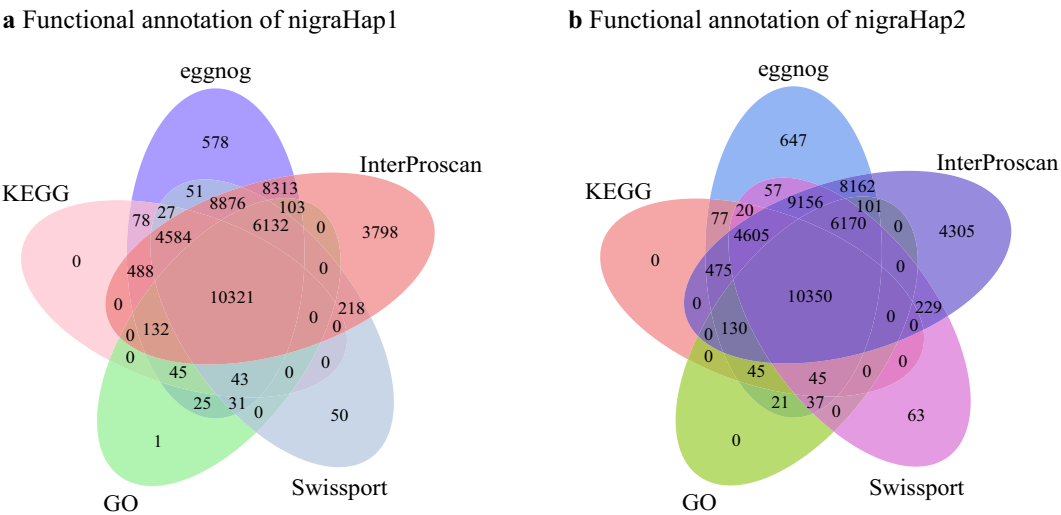


Fig. 5 Venn diagram of the functionally annotated protein-coding genes based on different databases. **(a)** Venn diagrams for functional annotation of nigraHap1. **(b)** Venn diagrams for functional annotation of nigraHap2.

Chromosome	nigraHap1			nigraHap2		
	Status	Left number	Right number	Status	Left number	Right number
chr01	both	419	837	both	163	661
chr02	both	1,008	396	both	921	170
chr03	both	373	506	both	602	172
chr04	both	284	500	both	123	1,573
chr05	both	627	655	both	1257	671
chr06	both	476	606	both	479	654
chr07	both	689	174	both	947	479
chr08	left	170	0	left	546	0
chr09	both	253	519	both	315	506
chr10	both	761	1041	both	191	525
chr11	both	293	692	both	654	294
chr12	both	486	1100	both	376	319
chr13	both	285	1134	both	736	662
chr14	left	352	0	left	568	0
chr15	both	678	286	both	734	527
chr16	both	1,495	403	both	754	1,387
chr17	both	1,834	147	both	391	326
chr18	both	300	407	both	401	471
chr19	both	862	121	both	255	710

Table 7. The number of telomere sequences.

270. Additionally, 0.7% and 0.6% of genes were fragmented, while 0.6% were missing (Table 8). When compared to other poplar species (Fig. 7), the assembly quality is comparable to that of the *P. trichocarpa* haplotype genomes. The BUSCOs completeness assessment using annotated sequences showed that nigraHap1 and nigraHap2 genome annotation covered 97.6% and 98.0% of complete BUSCOs, with single-copy genes accounting for 82.3%, duplicated genes for 15.4% and 15.6%, fragmented genes numbering 14 and 11, and missing genes numbering 24 and 21, respectively, it indicates that almost all core functional genes have been completely captured, with very few key genes missing or fragmented, demonstrating high coverage and structural integrity of the genome assembly (Table 8).

To assess the reliability of the two haplotype-resolved genomes of *P. nigra* ‘NL-1976’, we used Merqury v1.368 with meryl v1.384 (under 19-mer) to evaluate the consensus quality value (QV). The QVs for nigraHap1 and nigraHap2, based on comparisons with second-generation sequencing data, were 41.57% and 40.70%, with completeness of 84.86% and 85.18%, respectively. When compared with HiFi data, the QVs were 41.41% and 40.53%, with completeness of 84.85% and 85.16% it indicates that no more than 1 error per 10,000 bases.

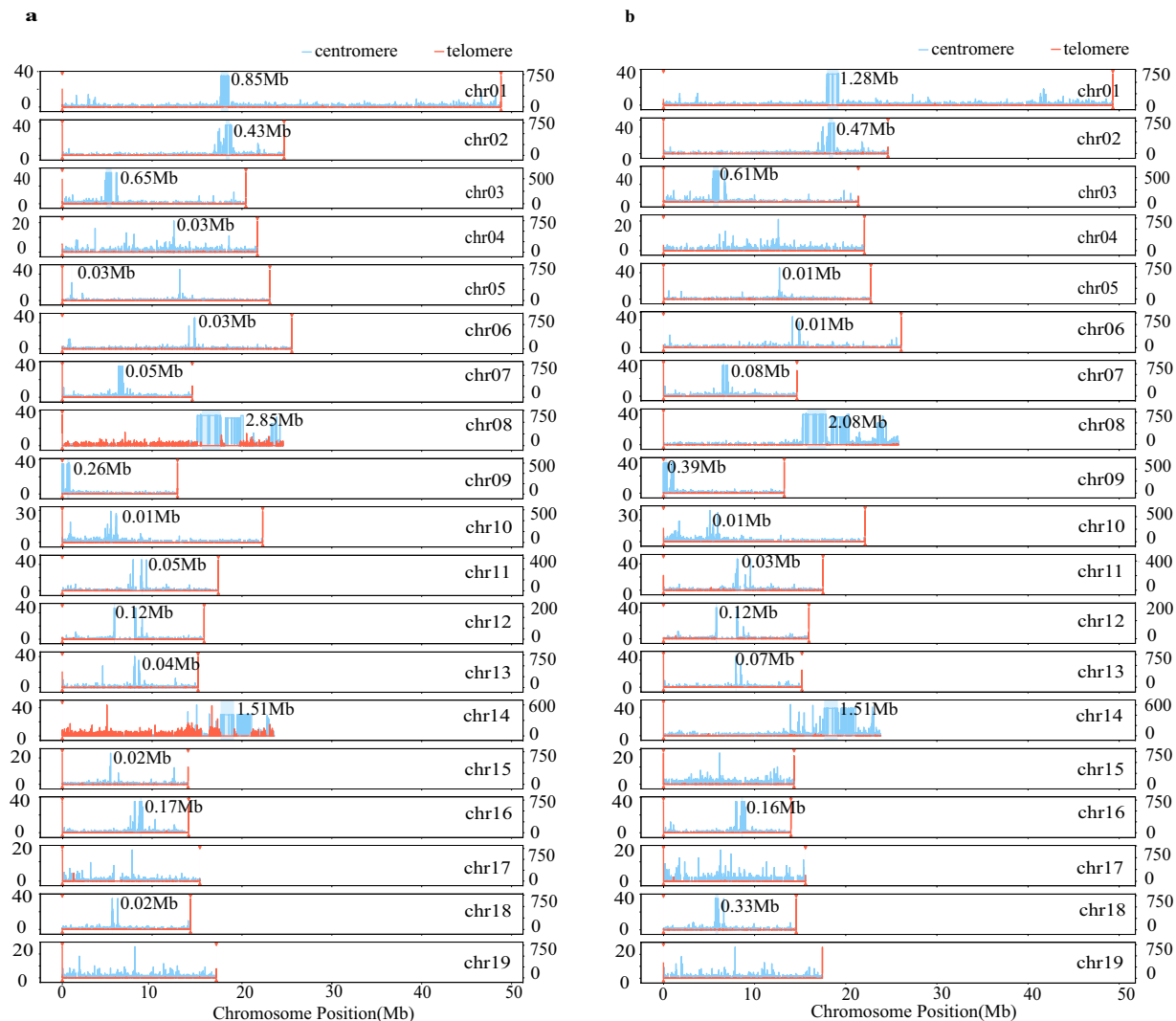


Fig. 6 Distribution of telomeres and centromeres in two haplotypes. **(a)** Telomeres and centromeres nigraHap1. **(b)** Telomeres and centromeres nigraHap2.

Statistic	Haploid genomes		Protein-coding genes	
	nigraHap1	nigraHap2	nigraHap1	nigraHap2
Total BUSCO groups searched	1,614		1,614	
Complete BUSCOs (%)	1,593 (98.7%)	1,595 (98.8%)	1,576 (97.6%)	1,575 (98%)
Complete and single-copy BUSCOs (%)	1,328 (82.3%)	1,323 (82%)	1,328 (82.3%)	1,329 (82.3%)
Complete and duplicated BUSCOs (%)	265 (16.4%)	270 (16.7%)	248 (15.4%)	252 (15.6%)
Fragmented BUSCOs (%)	11 (0.7%)	10 (0.6%)	14 (0.9%)	11 (0.7%)
Missing BUSCOs (%)	10 (0.6%)	10 (0.6%)	24 (1.5%)	21 (1.4%)

Table 8. BUSCO results of the haploid genome and protein-coding genes.

Data availability

All raw sequencing data (PacBio HiFi, Hi-C, ONT, RNA-seq) about this study are available in the NGDC GSA database under BioProject PRJCA040372⁸⁵. The accession numbers of PacBio HiFi reads, Hi-C sequencing reads, ONT sequencing data are [CRX1730796](#)⁷⁶, [CRX1730797](#)⁷⁷ and [CRX1730798](#)⁷⁸, the Illumina RNA-seq and full-length RNA-seq data are available in the GSA database with the accession number [CRX1730799](#)⁷⁹ and [CRX1730800](#)⁸⁰.

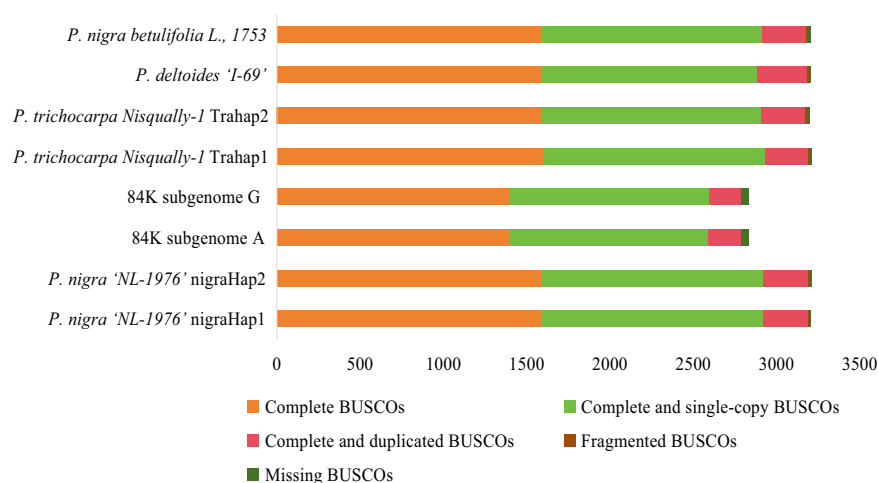


Fig. 7 Detailed BUSCO results of five poplar species.

Code availability

No specific code was developed in this study. The software and their versions used in this study are described in the Methods section. Any modified parameters are also specified therein, and unless otherwise stated, the default parameters of the software were used.

Received: 25 April 2025; Accepted: 27 November 2025;

Published online: 09 December 2025

References

- Liu, N. *et al.* Enhancing large-diameter timber production: Evaluating poplars by genotype and spacing. *Ind. Crop. Prod.* **223**, 120148, <https://doi.org/10.1016/j.indcrop.2024.120148> (2025).
- Shi, T. *et al.* The super-pangenome of *Populus* unveils genomic facets for its adaptation and diversification in widespread forest trees. *Mol. Plant* **17**, 725–746, <https://doi.org/10.1016/j.molp.2024.03.009> (2024).
- Allwright, M. R. *et al.* Biomass traits and candidate genes for bioenergy revealed through association genetics in coppiced European *Populus nigra* (L.). *Biotechnol. Biofuels* **9**, 1–22, <https://doi.org/10.1186/s13068-016-0603-1> (2016).
- Gupta, A. *et al.* Bioethanol production from hemicellulose rich *Populus nigra* involving recombinant hemicellulases from *Clostridium thermocellum*. *Bioresour. Technol.* **165**, 205–213, <https://doi.org/10.1016/j.biortech.2014.03.132> (2014).
- Vanden Broeck, A. *et al.* Reintroduced native *Populus nigra* in restored floodplain reduces spread of exotic poplar species. *Front. Plant Sci.* **11**, 580653, <https://doi.org/10.3389/fpls.2020.580653> (2021).
- Guerra, F. P. *et al.* Association genetics of chemical wood properties in black poplar (*Populus nigra*). *New Phytol.* **197**, 162–176, <https://doi.org/10.1111/nph.12003> (2013).
- Benetka, V., Novotná, K. & Štochllová, P. Wild populations as a source of germplasm for black poplar (*Populus nigra* L.) breeding programmes. *Tree Genet. Genomes* **8**, 1073–1084, <https://doi.org/10.1007/s11295-012-0487-6> (2012).
- Vanden Broeck, A. *et al.* Paternity analysis of *Populus nigra* L. offspring in a Belgian plantation of native and exotic poplars. *Ann. Forest Sci.* **63**, 783–790, <https://doi.org/10.1051/forest:2006060> (2006).
- Liu, C. *et al.* Growth of *Populus × euramericana* plantlet under different light durations. *Forests* **14**, 579, <https://doi.org/10.3390/f14030579> (2023).
- Han, F. *et al.* One-step creation of CMS lines using a *BoCENH3*-based haploid induction system in *Brassica* crop. *Nat. Plants* **10**, 581–586, <https://doi.org/10.1038/s41477-024-01643-w> (2024).
- Alimpić, F. *et al.* The status and role of genetic diversity of trees for the conservation and management of riparian ecosystems: A European experts' perspective. *J. Appl. Ecol.* **59**, 2476–2485, <https://doi.org/10.1111/1365-2664.14247> (2022).
- Michalak, M. *et al.* Desiccation tolerance and cryopreservation of seeds of black poplar (*Populus nigra* L.), a disappearing tree species in Europe. *Eur. J. Forest Res.* **134**, 53–60, <https://doi.org/10.1007/s10342-014-0832-4> (2015).
- Nevers, Y. *et al.* Quality assessment of gene repertoire annotations with OMARK. *Nat. Biotechnol.* **43**, 124–133, <https://doi.org/10.1038/s41587-024-02147-w> (2025).
- Garg, V. *et al.* Unlocking plant genetics with telomere-to-telomere genome assemblies. *Nat. Genet.* **56**, 1788–1799, <https://doi.org/10.1038/s41588-024-01830-7> (2024).
- Li, Q. *et al.* The *Cissus quadrangularis* genome reveals its adaptive features in an arid habitat. *Hortic. Res.* **11**, uhae038, <https://doi.org/10.1093/hr/uhae038> (2024).
- Liang, Y. Y. *et al.* Pan-genome analysis reveals local adaptation to climate driven by introgression in oak species. *Mol. Biol. Evol.* **42**, msaf088, <https://doi.org/10.1093/molbev/msaf088> (2025).
- Hu, G. *et al.* Two divergent haplotypes from a highly heterozygous lychee genome suggest independent domestication events for early and late-maturing cultivars. *Nat. Genet.* **54**, 73–83, <https://doi.org/10.1038/s41588-021-00971-3> (2022).
- Bredemeyer, K. R. *et al.* Single-haplotype comparative genomics provides insights into lineage-specific structural variation during cat evolution. *Nat. Genet.* **55**, 1953–1963, <https://doi.org/10.1038/s41588-023-01548-y> (2023).
- Shi, D. *et al.* Single-pollen-cell sequencing for gamete-based phased diploid genome assembly in plants. *Genome Res.* **29**, 1889–1899, <http://www.genome.org/cgi/doi/10.1101/gr.251033.119> (2019).
- Tuskan, G. A. *et al.* The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**, 1596–1604, <https://doi.org/10.1126/science.1128691> (2006).
- NCBI GenBank https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_033621325.1/ (2024).
- NGDC GWH <https://ngdc.cncb.ac.cn/gwh/Assembly/83710/show> (2025).
- ENA European Nucleotide Archive <https://identifiers.org/ena.embl/PRJEB62046> (2023).

24. Liu, W. *et al.* A nearly gapless, highly contiguous reference genome for a doubled haploid line of *Populus ussuriensis*, enabling advanced genomic studies. *For. Res.* **4**, e019, <https://doi.org/10.48130/forres-0024-0016> (2024).
25. Altemose, N. *et al.* Complete genomic and epigenetic maps of human centromeres. *Science* **376**, eabl4178, <https://doi.org/10.1126/science.abl4178> (2022).
26. Jain, M. *et al.* The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* **17**, 1–11, <https://doi.org/10.1186/s13059-016-1103-0> (2016).
27. Lieberman, A. E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293, <https://doi.org/10.1126/science.1181369> (2009).
28. Chen, S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *Imeta* **2**, e107, <https://doi.org/10.1002/imt2.107> (2023).
29. Marcais, 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).
30. 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).
31. Shen, W., Sipos, B. & Zhao, L. SeqKit2: A Swiss army knife for sequence and alignment processing. *Imeta* **3**, e191, <https://doi.org/10.1002/imt2.191> (2024).
32. De Coster, W. *et al.* NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* **34**, 2666–2669, <https://doi.org/10.1093/bioinformatics/bty149> (2018).
33. Cheng, H. *et al.* Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**(170–175), 5, <https://doi.org/10.1038/s41592-020-01056-6> (2021).
34. 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).
35. Cheng, H. *et al.* Scalable telomere-to-telomere assembly for diploid and polyploid genomes with double graph. *Nat. Methods* **21**, 967–970, <https://doi.org/10.1038/s41592-024-02269-8> (2024).
36. Deng, F. *et al.* Purge_dups: efficient removal of haplotigs and false duplications in genome assemblies. *Bioinformatics* **37**, 4234–4236, <https://doi.org/10.1093/bioinformatics/btaa025> (2021).
37. 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).
38. 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).
39. Durand, N. C. *et al.* 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> (2016).
40. Xu, M. *et al.* TGS-GapCloser: A fast and accurate gap closer for large genomes with low coverage of error-prone long reads. *Gigascience* **1**, g1aa094, <https://doi.org/10.1093/gigascience/g1aa094> (2020).
41. Lin, Y. *et al.* QuarTeT: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. *Hortic. Res.* **10**, uhad127, <https://doi.org/10.1093/hr/uhad127> (2023).
42. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100, <https://doi.org/10.1093/bioinformatics/bty191> (2018).
43. Hu, J. *et al.* NextPolish: a fast and efficient genome polishing tool for long read assembly. *Bioinformatics* **36**, 2253–2255, <https://doi.org/10.1093/bioinformatics/btz891> (2019).
44. Marçais, G. *et al.* MUMmer4: A fast and versatile genome alignment system. *PLoS Comput. Biol.* **14**, e1005944, <https://doi.org/10.1371/journal.pcbi.1005944> (2018).
45. 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).
46. Li, H. & Durbin, R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* **25**, 1754–1760, <https://doi.org/10.1093/bioinformatics/btp324> (2009).
47. NCBI GenBank https://www.ncbi.nlm.nih.gov/search/all/?term=GCA_015852605.1 (2021).
48. Frank, M. Y., Sylvie, C., Shan, Y. F. & Raja, R. LTR annotator: Automated identification and annotation of LTR retrotransposons in plant genomes. *Bioinformatics* **5**, 165–174, <https://doi.org/10.17706/ijbbb.2015.5.3.165-174> (2015).
49. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, 265–268, <https://doi.org/10.1186/s13100-019-0193-0> (2007).
50. Ou, S. J. & 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).
51. Chen, N. Using Repeat Masker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics* **5**, 4.10.11–4.10.14, <https://doi.org/10.1002/0471250953.bi0410s05> (2004).
52. Flynn, J. M. *et al.* RepeatModeler2 for automated genomic discovery of transposable element families. *P. Natl. A. Sci.* **117**, 9451–9457, <https://doi.org/10.1073/pnas.1921046117> (2020).
53. NCBI GenBank https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_902651935.1/ (2025).
54. Li, H. Protein-to-genome alignment with minimap2. *Bioinformatics* **39**, btad014, <https://doi.org/10.1093/bioinformatics/btad014> (2023).
55. Kim, D. *et al.* 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).
56. Shumate, A. *et al.* Improved transcriptome assembly using a hybrid of long and short reads with StringTie. *PLoS Comput. Biol.* **18**, e1009730, <https://doi.org/10.1371/journal.pcbi.1009730> (2022).
57. Kovaka, S. *et al.* Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* **20**, 1–13, <https://doi.org/10.1186/s13059-019-1910-1> (2019).
58. De Coster, W. & Rademakers, R. NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics* **39**, btad311, <https://doi.org/10.1093/bioinformatics/btad311> (2023).
59. Logsdon, G. A., Vollger, M. R. & Eichler, E. E. Long-read human genome sequencing and its applications. *Nat. Rev. Genet.* **21**, 597–614, <https://doi.org/10.1038/s41576-020-0236-x> (2020).
60. Pertea, M. *et al.* Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, 1650–1667, <https://doi.org/10.1038/nprot.2016.095> (2016).
61. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512, <https://doi.org/10.1038/nprot.2013.084> (2013).
62. 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).
63. Haas, B. J. *et al.* Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666, <https://doi.org/10.1093/nar/gkg770> (2003).

64. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659, <https://doi.org/10.1093/bioinformatics/btl158> (2006).
65. Stanke, M., Steinkamp, R. & Waack, S. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res.* **2**, 309–312, <https://doi.org/10.1093/nar/gkh379> (2004).
66. Brůna, T., Lomsadze, A. & Borodovsky, M. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genomics and Bioinformatics* **2**, lqaa026, <https://doi.org/10.1093/nargab/lqaa026> (2020).
67. Haas, B. J. *et al.* 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> (2008).
68. Lagesen, K. *et al.* RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**, 3100–3108, <https://doi.org/10.1093/nar/gkm160> (2007).
69. Martin K. Gene prediction: methods and protocols (Humana Press, 2019).
70. 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).
71. Griffiths-Jones, S. *et al.* Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33**, D121–D124, <https://doi.org/10.1093/nar/gki081> (2005).
72. Cao, K. *et al.* A unified computational framework for single-cell data integration with optimal transport. *Nat. Commun.* **13**, 7419, <https://doi.org/10.1038/s41467-022-35094-8> (2022).
73. Zdobnov, E. M. & Apweiler, R. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847–848, <https://doi.org/10.1093/bioinformatics/17.9.847> (2001).
74. Chen, C. *et al.* TBtools-II: A “One for all, all for one” bioinformatics platform for biological big-data mining. *Mol. Plant.* **16**, 1733–1742, <https://doi.org/10.1016/j.molp.2023.09.010> (2023).
75. Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2025.
76. NGDC GSA <https://ngdc.cncb.ac.cn/gsa/browse/CRA025879/CRX1730796> (2025).
77. NGDC GSA <https://ngdc.cncb.ac.cn/gsa/browse/CRA025879/CRX1730797> (2025).
78. NGDC GSA <https://ngdc.cncb.ac.cn/gsa/browse/CRA025879/CRX1730798> (2025).
79. NGDC GSA <https://ngdc.cncb.ac.cn/gsa/browse/CRA025879/CRX1730799> (2025).
80. NGDC GSA <https://ngdc.cncb.ac.cn/gsa/browse/CRA025879/CRX1730800> (2025).
81. Liu, F. F. The genome assembly and annotation results of the haplotype of *Populus nigra* NL-1976. *figshare. Dataset.* <https://doi.org/10.6084/m9.figshare.29850356.v1> (2025).
82. NCBI GenBank https://identifiers.org/ncbi/insdc.gca:GCA_052753605.1 (2025).
83. NCBI GenBank https://identifiers.org/ncbi/insdc.gca:GCA_052724285.1 (2025).
84. Rhie, A. *et al.* Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 1–27, <https://doi.org/10.1186/s13059-020-02134-9> (2020).
85. *Nucleic Acids Research* **53**, D30–D44. <https://academic.oup.com/nar/article/53/D1/D30/7893335?login=true> (2025).

Acknowledgements

This work was supported by the Fourteenth Five-Year National Key Research and Development Program of China (2022YFD2200301), and the Basic Research Fund of Research Institute of Forestry, Chinese Academy of Forestry (CAFYBB2024QF004).

Author contributions

F.L. and Q.H. conceived this research. C.L. and Q.H. acquired the funding and designed the methodology. F.L., C.G. and X.Z. collected and prepared the tissue samples for sequencing. F.L., C.L., N.L. and X.J. analyzed the data and developed the figures. F.L. and C.L. wrote the original draft manuscript. A.V.B., Q.H. and P.Š. supervised the data analysis. C.L., A.V.B., Q.H. and P.Š. revised the manuscript. All authors have reviewed and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Q.H.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025