Horticulture Research

Article

Populus cathayana genome and population resequencing provide insights into its evolution and adaptation

Xiaodong Xiang[1,2,‡], Xinglu Zhou[1,2,‡], Hailing Zi [3,‡], Hantian Wei[1], Demei Cao[1], Yahong Zhang[1], Lei Zhang[1,2,*] and Jianjun Hu[1,2,*]

[1]State Key Laboratory of Tree Genetics and Breeding, Key Laboratory of Tree Breeding and Cultivation of National Forestry and Grassland Administration, Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, China
[2]Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing, Jiangsu 210037, China
[3]Novogene Bioinformatics Institute, Beijing 100083, China
*Corresponding authors. E-mails: leizhang1142@caf.ac.cn; hujj@caf.ac.cn
‡Co-first authors contributing equally to this work.

Abstract

Populus cathayana Rehder, an indigenous poplar species of ecological and economic importance, is widely distributed in a high-elevation range from southwest to northeast China. Further development of this species as a sustainable poplar resource has been hindered by a lack of genome information the at the population level. Here, we produced a chromosome-level genome assembly of *P. cathayana*, covering 406.55 Mb (scaffold N50 = 20.86 Mb) and consisting of 19 chromosomes, with 35 977 protein-coding genes. Subsequently, we made a genomic variation atlas of 438 wild individuals covering 36 representative geographic areas of *P. cathayana*, which were divided into four geographic groups. It was inferred that the Northwest China regions served as the genetic diversity centers and a population bottleneck happened during the history of *P. cathayana*. By genotype–environment association analysis, 947 environment-association loci were significantly associated with temperature, solar radiation, precipitation, and altitude variables. We identified local adaptation genes involved in DNA repair and UV radiation response, among which *UVR8*, *HY5*, and *CUL4* had key roles in high-altitude adaptation of *P. cathayana*. Predictions of adaptive potential under future climate conditions showed that *P. cathayana* populations in areas with drastic climate change were anticipated to have greater maladaptation risk. These results provide comprehensive insights for understanding wild poplar evolution and optimizing adaptive potential in molecular breeding.

Introduction

The genus *Populus* (Salicaceae) is widely distributed in the northern hemisphere throughout the subtropical to boreal forests, and is classified into six sections (*Abaso*, *Aigeiros*, *Leucoides*, *Populus*, *Tacamahaca*, and *Turanga*) [1]. China is a unique region concerning the genus *Populus*, with about half of *Tacamahaca* poplars having their natural range completely or partlyin China [2]. *Populus cathayana* Rehder (Salicaceae, *Tacamahaca*) is a native tree in China, widely distributed in northern, central, and southwestern areas [3, 4]. This species predominantly grows at altitudes of 1000–3000 m, with some occurring at 3900 m on the eastern margin of the Qinghai–Tibet Plateau [5]. Despite the challenging climates of high altitudes and short growing seasons, *P. cathayana* demonstrates exceptional growth potential and high adaptability [6]. Wild *P. cathayana* is commonly used as a hybrid parent to culture elite poplar varieties that adapt to harsh environments and have beneficial traits such as fast growth and cold tolerance.

Disentangling the genetic mechanisms driving local adaptation and genetic differentiation is essential for conservation efforts and the development of breeding strategies [7, 8]. Extensive experimental research on tree populations has shown that local populations consistently exhibit higher adaptability in their native habitats [9, 10]. Local adaptation related to spatial variations in natural selection pressures can be explained by factors such as latitude, longitude, altitude, and slope [11, 12]. The distribution of populations is significantly influenced by dramatic climatic oscillations and historical geology, leaving diagnostic signatures in the genomes [13]. The use of population genetics methods and genomic data can be a crucial strategy to examine local adaptation across wild species [14]. High-quality genomes of poplars such as *Populus trichocarpa*, *P. alba*, *P. euphratica*, and *P. alba* × *P. glandulosa* provide a foundation for understanding adaptive evolution and genetic differentiation [15–18]. Wild populations of *P. cathayana* were considered to be important genetic resources in meeting ecological and forest product needs [4]. However, previous studies have used limited markers, and have not provided understanding of the genomic genetic variation of *P. cathayana*.

Future global environmental changes are predicted to negatively affect ecosystem homeostasis, and climate change increases our interest in the adaptation of species and populations to new environments [12]. Trees are long-lived organisms with slow adaptive evolution, making them vulnerable to rapid environmental changes. In recent years, researchers have increasingly used genomic methods to predict the impact of climate change on local adaptation and species vulnerability [19]. Genotype–environmental association (GEA) is used as an approach to elucidate the genetic bases of adaptation of *Betula nana* [20], *Thlaspi arvense* [21], *Quercus lobata* [22], *Picea sitchensis* [23],

and *Populus trichocarpa* [24]. Due to the polygenic nature of these adaptive variants, the genetic architecture of local adaptation to climate can be very diverse among even closely related species. For the identified key adaptive mutation sites, genomic shifts can be measured to assess the amount of population genetic composition change required to track future environmental conditions [25, 26]. Therefore, the use of genomic tools can not only simulate changes in species ranges over time but also provide novel understanding to assess evolutionary adaptation potential and the ability to withstand future climate risks.

Populus cathayana has been recognized for its high-altitude growth and wide adaptation to harsh environmental conditions, such as low temperatures and intense solar radiation. However, the genetic mechanisms of local adaptation of *P. cathayana* at broad spatial scales are still unknown. As a wild tree species population, it has rarely been subjected to artificial selection, making it an ideal model for understanding species adaptability and local adaptation. Here, we assembled a *de novo* chromosome-scale *P. cathayana* genome and resequenced the genomes of 438 individuals from 36 geographic regions in China. The population structure, genetic diversity, and demographic processes of the *P. cathayana* population were revealed by genome-wide variants and phylogenetic analysis. Based on these genomic datasets, we uncovered genetic evidence of the environmental adaptability of *P. cathayana*. This study offers new insights into genome evolution and adaptation in a major forest species, and serve as a valuable resource for genome-based poplar improvement.

Results

Populus cathayana genome assembly and annotation

The genome of *P. cathayana* was sequenced, and 60.71 Gb of PacBio long-read sequencing data ($\sim$149$\times$), 47.35 Gb Illumina data ($\sim$116$\times$), and 50.63 Gb of Hi-C (chromosome conformation capture) paired-end reads data ($\sim$124$\times$) were obtained for *de novo* genome assembly (Supplementary Data Tables S1 and S2). The estimated size of the nuclear genome of *P. cathayana* by flow cytometry and *K*-mer analysis was $\sim$412.4 and $\sim$423.7 Mb, respectively (Supplementary Data Fig. S1, Supplementary Data Table S3). Based on PacBio and Illumina data, we obtained an initial genome assembly of 411 Mb (Supplementary Data Table S4). The scaffold extension and chromosomal mapping were performed using Hi-C data, resulting in a final assembly of the *P. cathayana* genome with a size of 406.5 Mb (Table 1). This assembly had 21 scaffolds, with a scaffold N50 of 20.86 Mb (Table 1, Supplementary Data Table S5). Specifically, 405.9 Mb (99.84%) of the assembly was successfully ordered and oriented to 19 pseudo-chromosomes (Supplementary Data Fig. S2, Supplementary Data Table S6). The 97.96% Benchmarking Universal Single-Copy Orthologs (BUSCO) and 98.25% Core Eukaryotic Genes Mapping Approach (CEGMA) could be completely detected in the assembly. In addition, 95.99% of Illumina reads were also mapped to the assembled genome, which indicated that the genome assembly was high-quality (Supplementary Data Tables S7 and S8).

In total, 35 977 protein-coding genes were predicted in the *P. cathayana* genome, with an average length of 3444 bp per gene and an average of 5.4 exons per transcript (Table 1, Supplementary Data Table S9). In addition, 35 366 (98.3%) genes could be annotated using functional databases (Supplementary Data Table S10). A total of 173.85 Mb sequences in the

Table 1. Statistics for genome assembly and annotation for *P. cathayana*.

Assembly feature	Statistic
Assembly	
Genome size (bp)	406 549 807
Number of scaffolds	21
N50 of scaffolds (bp)	20 860 933
Chromosome-scale scaffolds (bp)	405 899 219 (99.84%)
Number of contigs	77
N50 of contigs (bp)	10 281 541
Number of gaps	56
Complete BUSCOs	1581 (97.96%)
GC content of genome (%)	33.84
Annotation	
Number of predicted protein-coding genes	35 977
Average gene length (bp)	3444
Average coding sequence length (bp)	1317
Average exons per transcript	5.4
Repeat sequences (bp)	173 824 474 (42.76%)

P. cathayana genome were annotated as repetitive sequences, comprising 150.9 Mb (37.11%) transposable repeats and 22.95 Mb (5.64%) tandem repeats (Supplementary Data Tables S11 and S12). Long-terminal repeats (LTRs) constituted the highest proportion of retrotransposons in the genome, accounting for 22.15%, with 9.69% of Gypsy superfamilies and 4.17% of Copia superfamilies dominating (Fig. 1A). In addition, a set of non-coding RNAs (rRNA, tRNA, miRNA, snRNA, and snoRNA) was identified (Supplementary Data Table S13).

Genome evolution

Utilizing 805 single-copy orthologs shared by nine species, we performed phylogenetic reconstruction and estimated species divergence time (Supplementary Data Table S14). *Populus cathayana* shared a common ancestor with four Salicaceae species, and the estimated divergence time was between 3.46 and 17.68 million years ago (Mya) (Fig. 1B). In four-fold synonymous third-codon transversion (4DTV) analysis, the appearance of sharp peaks of 4DTV in *P. cathayana*, *P. alba*, *P. trichocarpa*, and *Salix purpurea* represents an outbreak of gene duplication, corresponding to recent salicoid duplication ($\sim$65 Mya) and core eudicot triplication ($\sim$117 Mya). The distribution of 4DTV from orthologs indicated that divergence between *P. cathayana* and *P. trichocarpa* occurred $\sim$4.1 Mya (4DTV $\sim$0.01), which was consistent with the phylogenetic result (Fig. 1C, Supplementary Data Fig. S3). The 29 710 duplicated genes discovered in *P. cathayana* were categorized into five types: 20 120 whole-genome duplicates (WGDs, 67.7%), 2426 tandem duplicates (TDs, 8.2%), 1528 proximal duplicates (PDs, 5.1%), 829 transposed duplicates (TRDs, 2.8%), and 4807 dispersed duplicates (DSDs, 16.2%). The higher K_a/K_s ratios were found in TD and PD types (Fig 2A), indicated that TDs and PDs underwent more rapid sequence divergence than genes originated from other duplication types.

All of the predicted gene models for the nine species were clustered into 29 525 orthogroups (Supplementary Data Fig. S4), of which 508 expanded and 610 contracted in *P. cathayana*. The expanded genes were significantly enriched in defense response, hormone signaling pathway, response to abiotic stimulus, and DNA repair (Fig. 2B, Supplementary Data Table S15). Specifically, the *P. cathayana* genomes exhibited an increase in genes related to defense, such as plant disease resistance (R) genes, containing

A

B

C
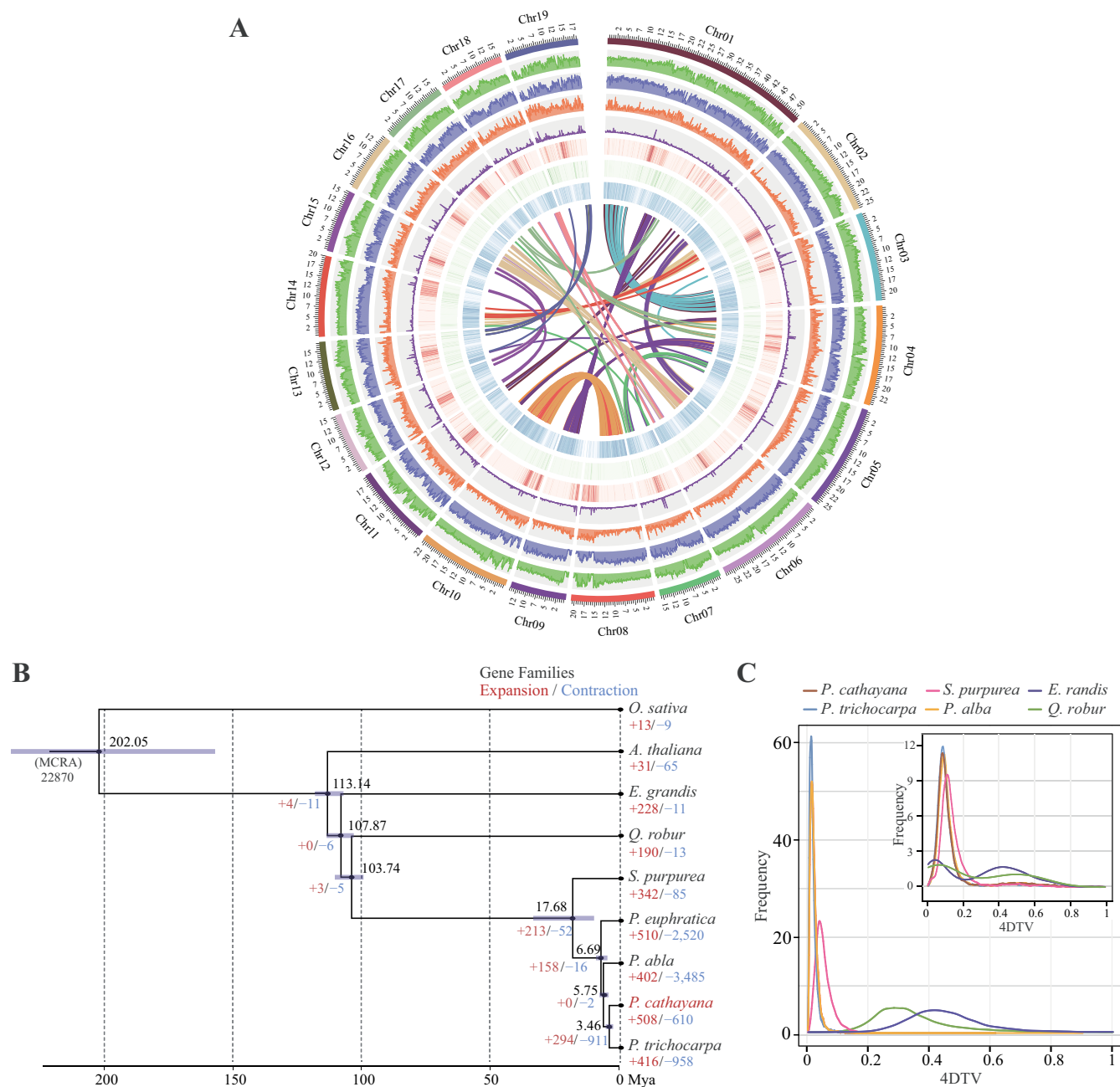
Figure 1. Evolutionary and collinearity analyses of the *P. cathayana* genome. **A** Tracks from outside to inside represent chromosomes, density of SNPs, indels, CNVs, SVs, Gypsy LTR-RT density, Copia LTR-RT density, gene density, and genomic collinearity. Variant density was calculated in non-overlapping 100-kb windows. **B** Phylogenetic tree, divergence time, and expansion and contraction of gene families for nine species. Evolutionary relationships and divergence times were calculated based on 805 single-copy orthologous genes. **C** The 4DTV distribution in the upper right (insert) shows 4DTV distribution from paralogs within *P. cathayana*, *p. trichocarpa*, *P. alba*, *Eucalyptus grandis*, *Salix purpurea*, and *Quercus robur*. The 4DTV distribution on the lower left shows the 4DTV distribution from orthologs between *P. cathayana* and the other five species.

NB-ARC and NBS domains (Fig. 2C). Among the members of expanded gene families (EGFs), 1260 (57.6%) genes originated as TD duplications and PD duplications, accounting for the largest proportion (Fig. 2D). The TD-EGF and PD-EGF gene-enriched categories were implicated in defense, response to stimulus, and stress (Supplementary Data Fig. S5). TDs and PDs have been identified as significant factors contributing to the expansion of gene families in *P. cathayana*, particularly in the case of newly formed duplications. These results demonstrated that gene family expansion plays a crucial role in the local adaption of *P. cathayana* during long-term evolution.

Population structure and genetic diversity of *P. cathayana*

We collected 438 individuals from 36 natural distribution regions of *P. cathayana* and generated whole-genome resequencing data (Fig. 3A, Supplementary Data Table S16). The cleaned reads were aligned to the *P. cathayana* genome, generating an average mapping rate of 93.36% and an average depth of 32.3 (Supplementary Data Table S17). We obtained 30 829 763 raw single-nucleotide polymorphisms (SNPs) and 12 374 210 high-quality SNPs after further filtering. A set of 569 435 insertion–deletions (indels), 35 416 copy-number variations (CNVs), and
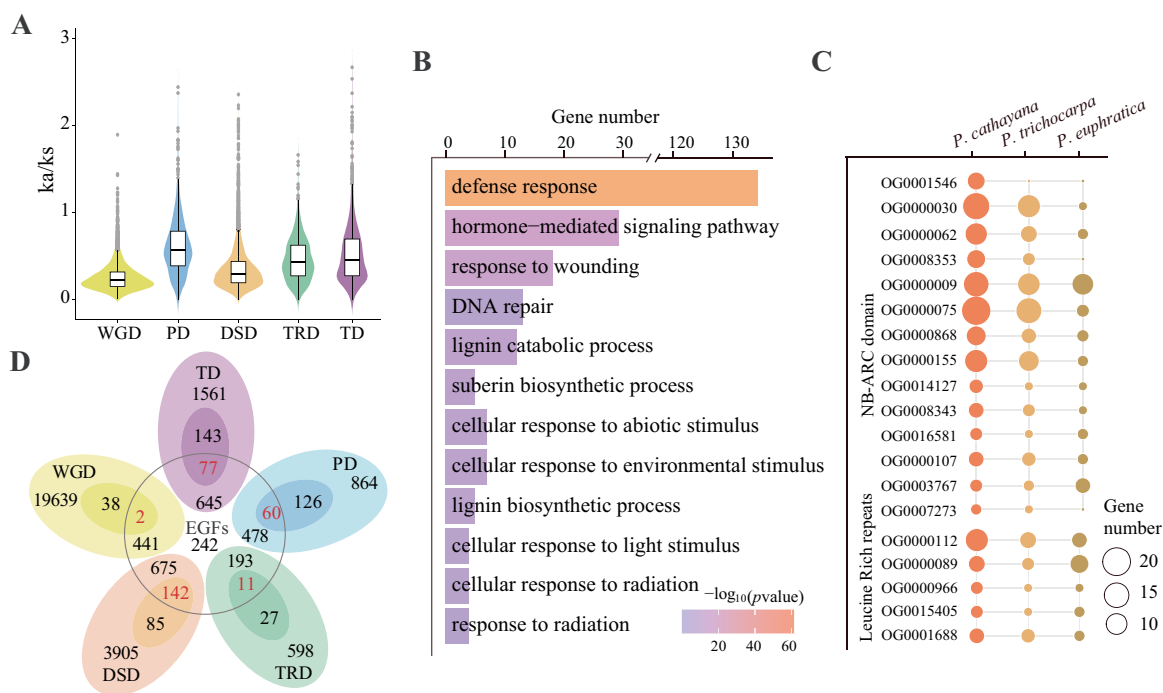
Figure 2. Genome duplication and evolution. **A** K_a/K_s ratio distributions of duplication gene pairs derived from five modes (WGD, TD, PD, TRD, and DSD). **B** Functional enrichment analysis of gene family expansion with *P. cathayana*. **C** Numbers of NBS and NR-ARC domains in the *P. trichocarpa*, *P. euphratica*, and *P. cathayana* genomes. **D** Venn diagram showing overlap between members of EGFs and five duplicated types.

9771 structure variants (SVs) was identified after filtering (Supplementary Data Table S18). Nearly half of the SNP variations were located within transposable elements (TEs), and the density of variations in TEs was higher than in mRNA regions (Supplementary Data Table S19).

The *P. cathayana* population could be divided into four groups based on the phylogenetic and population structure analyses, including NW (Northwest China), SW (Southwest China), TH (Taihang Mountains), and NC (North China) (Fig. 3B and C, Supplementary Data Fig. S6A and B). The four genetic groups that reflect the geographic distribution pattern were further supported by principal component analysis (PCA) (Fig. 3D). The Mantel test also revealed the strong significant correlation between geographic distance and genetic distance (Supplementary Data Fig. S6C). Moreover, we discovered population differentiation coefficient (F_{ST}) values between the four groups that were high or moderate (0.14–0.39), which was consistent with population admixture (Fig. 3E). In addition, numerous subclades among closely related subgroups shared historical gene flow and ancestral variation (Fig. 3F). A clear signature of genetic admixture was present in the SW group, while extensive gene flow was detected between the SW and NW subclades. Allelic admixture in some areas was evident, probably due to the pollen transmission and reproduction history.

Among the *P. cathayana* genetic groups, the highest nucleotide diversity (π) was found in the NW group ($\pi = 8.50 \times 10^{-3}$), followed by the TH group ($\pi = 7.87 \times 10^{-3}$), NC group ($\pi = 5.05 \times 10^{-3}$), and SW group ($\pi = 5.53 \times 10^{-3}$) (Fig. 3E, Supplementary Data Table S17). Overall, the genetic diversity of the *P. cathayana* natural population was 6.1×10^{-3}. The linkage disequilibrium (LD) of the whole population was 9.8 kb ($r^2 = 0.2$); the NW group presented the fastest LD decay (9 kb), followed by the NC (11 kb) and TH (15 kb) groups (Fig. 3G). The SW group exhibited the slowest LD decay (22 kb) and had low nucleotide diversity. The *P.*

cathayana population differentiation originated from the ancestor population [1649 thousand years ago (kya)], with TH and NW diverging from the common ancestor at ~1430 kya and SW and NC diverging at ~987 kya, as indicated by the demographic fluctuations model (Fig. 3H). The patterns of genomic diversity and population history strongly indicated an origin for *P. cathayana* in NW China.

Demographic history and potential distribution of *P. cathayana*

The past climate history for *P. cathayana* had a profound effect on the species' distribution and evolutionary history. We calculated the effective population size (N_e) of *P. cathayana* from millions of years ago to the last 10 000 years. Three groups of *P. cathayana* (NW, SW, TH) exhibited concordant demographic trajectories, which reached the maximum N_e at ~200 kya, while N_e decreased continuously from the penultimate glaciation (PG) to the last glaciation (LG) (200–20 kya) (Fig. 4A). The NC group maintained a lower N_e and decreased slowly during the PG period. All four groups experienced a drastic contraction in the recent 100 kya and showed relatively strong positive Tajima's D values, which indicated that a population bottleneck and balancing selection occurred.

When the present species distribution with environmental niche models (ENMs) was projected onto the Last Glacial Maximum (LGM) climate conditions, the predicted distribution showed a significant contraction for *P. cathayana* compared with its current distribution (Fig. 4B and C). The distribution of *P. cathayana* population contracted during the LG, which might have existence of glacial refuges. The demographic trajectories and species distribution support that this species experienced severe Quaternary climate shifts and associated environmental changes and our speculation about a very strong bottleneck during the history of *P. cathayana*.
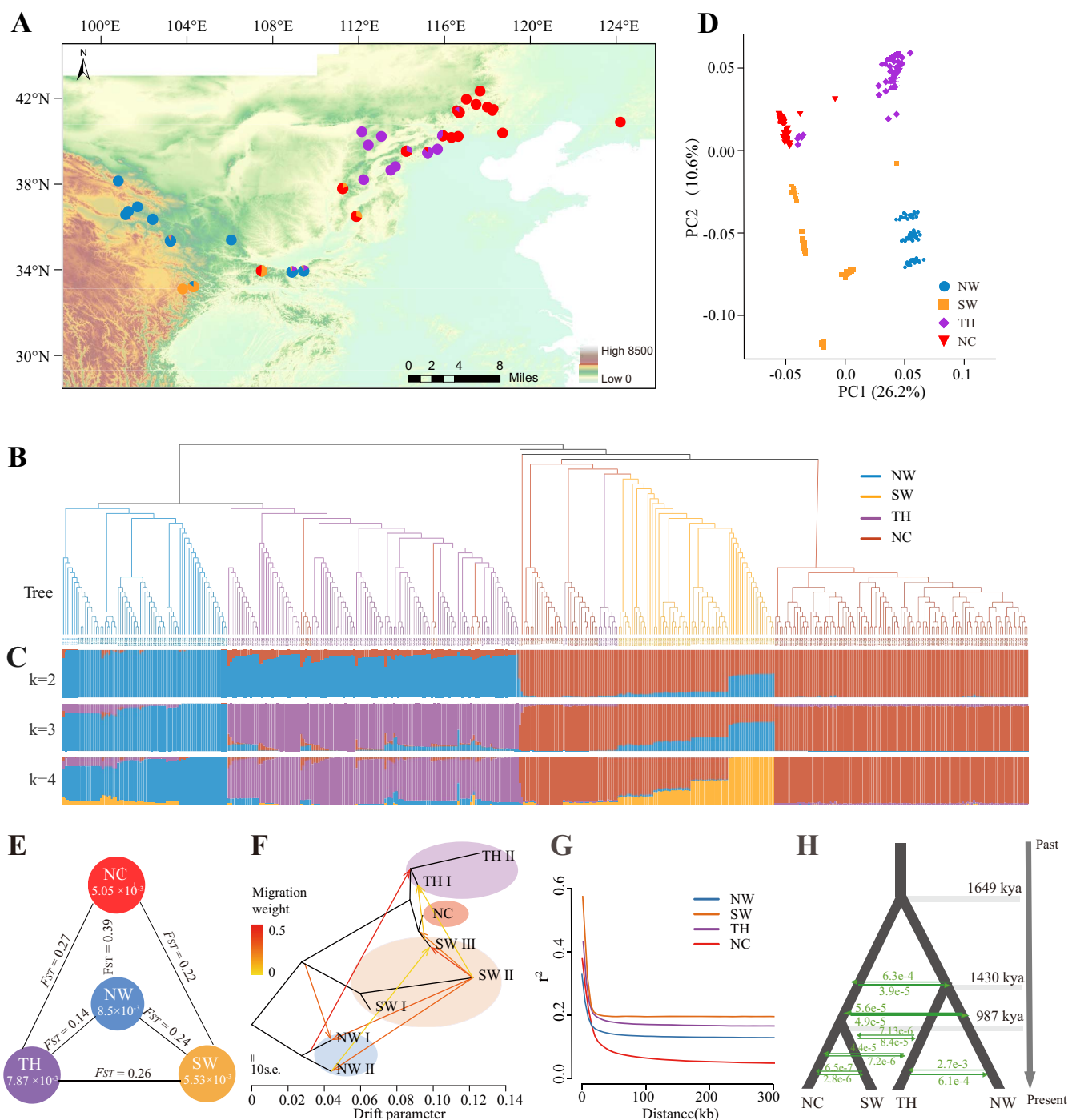
Figure 3. Geographic distribution, population structure, and genetic diversity of *P. cathayana* species. **A** Geographic distribution of sampling locations. The genetic composition at each sampling location is displayed as a pie chart, and the colors of the pies represent ancestral components according to the structure at K = 4. **B** Phylogenetic tree of all *P. cathayana* individuals. **C** Population structure of *P. cathayana* at cluster values (K) of 2–4. Each individual is denoted by a vertical bar composed of different colors corresponding to its proportion of genetic ancestry. More clusters values (K = 5–8) are shown in Supplementary Data Fig. S6. **D** PCA plot of the first two eigenvectors of the *P. cathayana* population. PC1 and PC2 divide the *P. cathayana* populations into four groups. **E** Nucleotide diversity (π) and population differentiation coefficient (F_{ST}) among the four groups. **F** Population gene flow shown by TreeMix. Arrows indicate inferred gene flow events between populations, with direction from the source population to the recipient population. Gene flow is colored according to their weight. The classification of sample groups is shown in Supplementary Data Fig. S6. **G** LD decay for the four groups of *P. cathayana*. X-axis: physical distances between two SNPs; Y-axis: R^2 used to measure linkage disequilibrium. **H** The best-fitting demographic model deduced by fastsimcoal2 is shown. Arrows represent estimated gene flow among clades.

Genome-wide selective sweep signals of four *P. cathayana* groups

Four groups of *P. cathayana* with significant altitude differences had undergone long-term natural selection (Fig. 5A).

According to the selective sweep results in NW, SW, TH, and NC, 713, 1455, 603, and 1760 candidate selection genes were identified combined with the top 5% of F_{ST} values and θπ ratios, respectively (Fig. 5B and C, Supplementary Data Fig. S7,
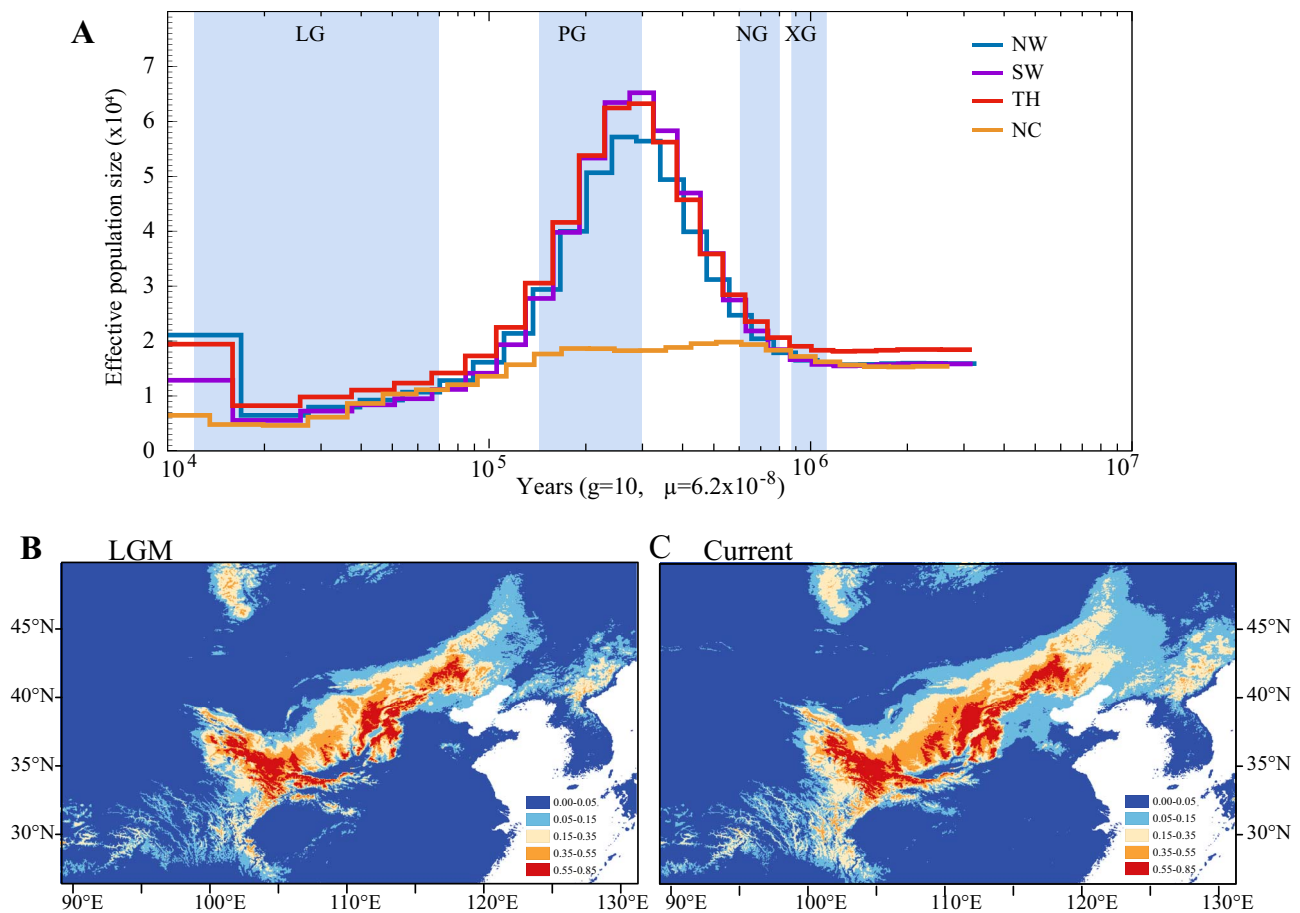
Figure 4. Demographic history and prediction of suitable habitats during the current time and the LGM. **A** PSMC for demographic history in four *P. cathayana* groups. Effective population size (N_e) was inferred using the PSMC model. Four periods, the last glacial (LG, 11–70 kya), penultimate glaciation (PG, 130–300 kya), Naynayxungla Glaciation (NG, 500–780 kya), and Xixiabangma Glaciation (XG, 800–1170 kya), are shaded in blue. **B**, **C** Environmental niche modeling (ENMs) with Maxent predicted palaeodistributions and current distributions for *P. cathayana*: habitats that were predicted to be suitable at the LGM (**B**) and the current time predicted distribution (**C**).

Supplementary Data Table S20). Although few genes were shared by different groups, significant functional enrichment in the same terms was detected. Selective sweep genes were enriched in photosynthesis, defense response, DNA repair, and hormone biosynthesis and metabolic process by Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) annotations (Supplementary Data Fig. S8, Supplementary Data Table S21). These genes with strong selection signals were viewed as the foundation for local adaptation of *P. cathayana* in heterogeneous environments.

Most notably, the selection on DNA repair and photomorphogenesis genes may represent a major strategy for providing tolerance to UV irradiation damage in *P. cathayana* at high altitude. The enriched KEGG pathways under selective sweep contained nucleotide excision repair (NER), DNA mismatch repair (MMR), and homologous recombination (HR) (Supplementary Data Table S21). REPLICATION PROTEIN A1/2 protein (RPA1/2), XPE complex (*CUL4* and *DDB2*), and homo-TFIIH complex (*TF2H1*, *TF2H2*, and *XPD*) were noted for nucleotide excision repair (Fig. 5D), with *DDB2* and *RPA1* exhibited low nucleotide diversity and high genetic differentiation between NW and other groups (Fig 5F). In addition, MRN complex (*RAD50*), *RAD51*, *NBA1*, and *BRCA2* were found to participate in homologous recombination (Fig. 5E). The HYPOCOTYL 5 (*HY5*) gene associated with photomorphogenesis was located at the center of the protein–protein

interaction (PPI) network hub resulting from selective sweep, which acted on genes related to DNA repair and response to light stimulus process (Supplementary Data Fig. S9). In addition, three genes (*DDB2*, *DCAF1*, and *CUL4*) were associated with *HY5* stability, and also exhibited a strong selection signal. Some key transcription factors (*bHLH*, *PIF3*, *APRR5*, *ARR2*) associated with *HY5* were under positive selection (Supplementary Data Fig. S9).

Genomic variants associated with environmental adaptation

Local adaptation has the potential to create correlations between genomic loci and environmental variables (EVs). Nine EVs (Biol, Bio2, Bio3, Bio13, Bio19, Srad6, Srad12, Vapr6, and altitude) related to local adaptation were used for GEA analysis. These variables represent the extremes and seasonality of temperature, precipitation, and radiation pressure (Supplementary Data Table S22, Supplementary Data Fig. S10). First, 5773 environmental association loci (EALs) were identified as significantly associated with one or more EVs by use of the latent factor mixed model (LFMM); they were found not to cluster in specific genomic regions but to be widely distributed throughout the whole (Fig. 6A, Supplementary Data Fig. S11). Also, the strong and significant isolation by environment (IBE) observed in the adaptive variants by the Mantel test indicated that the adaptive genetic variants were mainly influenced by the environment (Fig. 6B). Next, 947
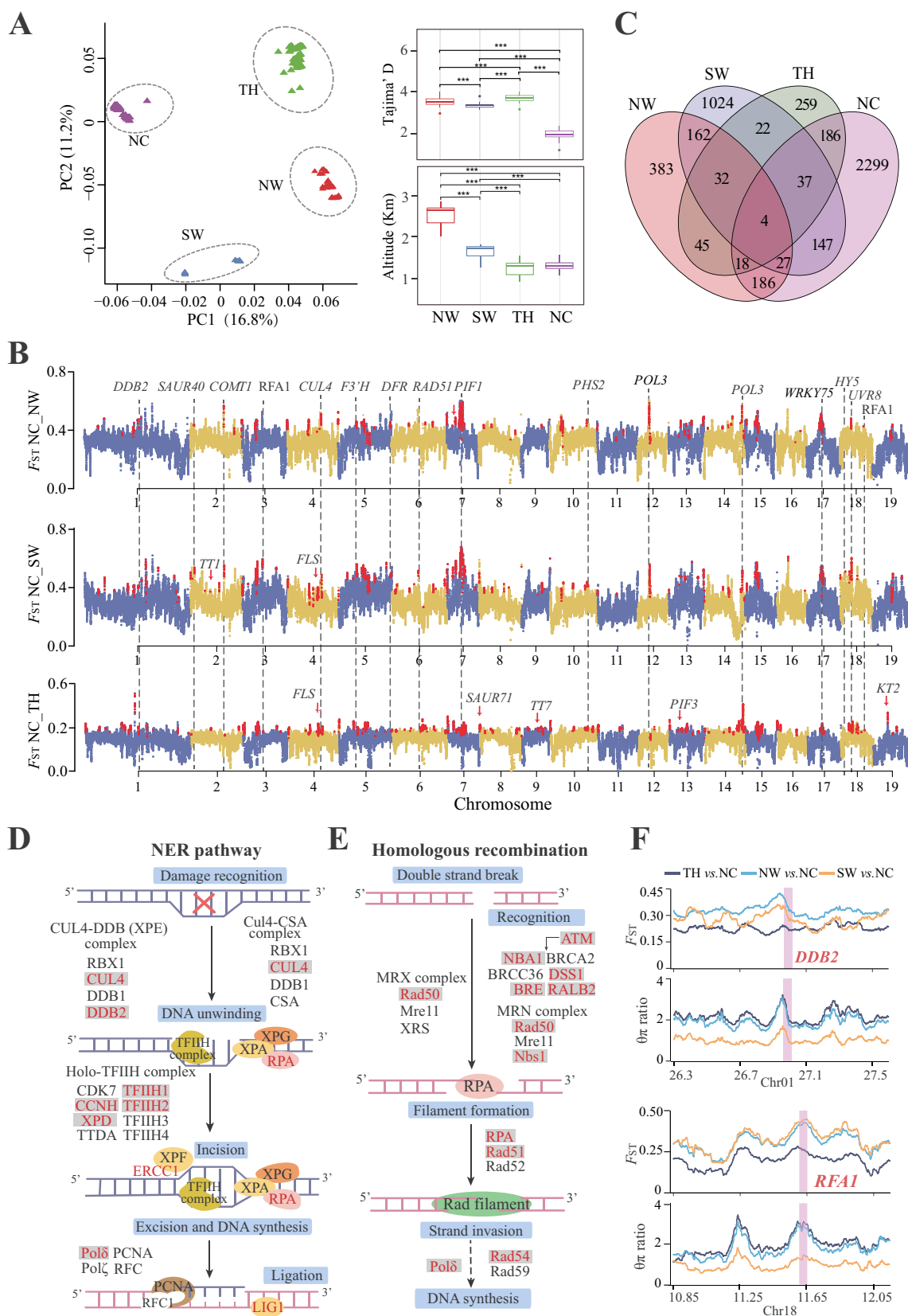
Figure 5. Population genetic differences and signatures of selective sweep. **A** PCA of selective sweep samples. Altitude and Tajima's *D* have significant differences among the four groups. **B** Pairwise F_{ST} in selective sweeps for the compared groups. Each point represents one region. Other selective sweep regions are shown in Supplementary Data Fig. S7. **C** Venn diagram of genes in selective sweeps for the four compared groups. **D**, **E** Nucleotide excision repair (NER) and homologous recombination pathway enrichment distribution based on selective sweep genes. Represents the selected gene. **F** Two genes within regions having positive selection signatures, *DDB2* and *RAF1*, are visualized with shadows in line chart plots.
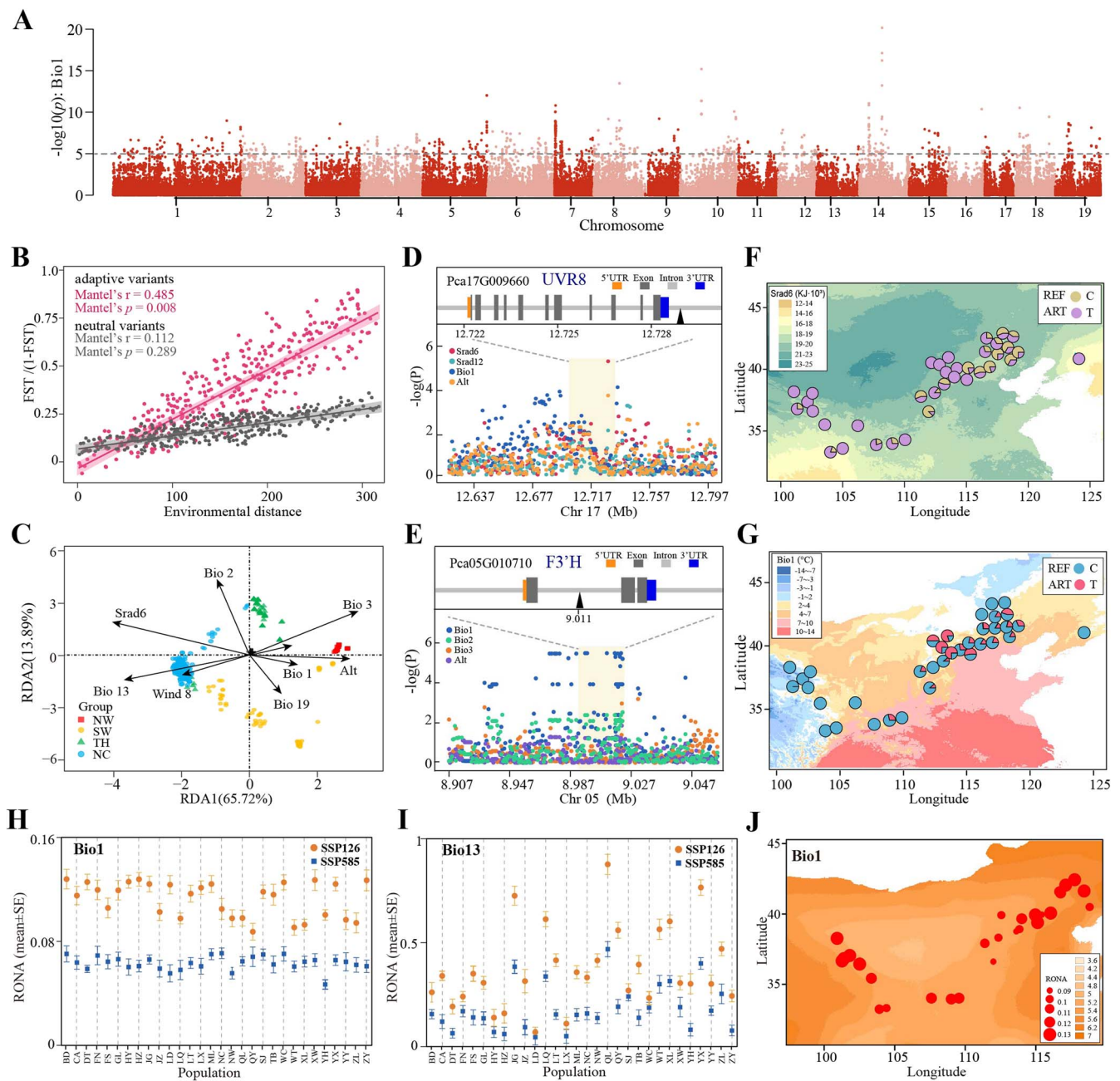
Figure 6. Environmental adaptation variables and response to future climate conditions in the *P. cathayana* population. **A** Manhattan plot for outlier loci associated with the Bio1 (annual mean temperature) variable. **B** IBE analyses for populations (*n* = 438) based on neutral variants and adaptive variants. The shadow of the linear regression represents the 95% confidence interval. **C** Partial RDA reveals the relationships between the independent environmental factors and population structure of *P. cathayana* along the RDA1 and RDA2 axes. Individuals are color-coded based on the four populations (NW, SW, TH, and NC). Vectors represent environmental variables. **D, E** Upper panels show the gene structure of *UVR8* (**D**) and *F3'H* (**E**) (triangles: candidate adaptive SNPs related to Srad6 or Bio1). Lower panels show local magnification of the Manhattan plots with environmental variables around the selected genes. **F, G** Candidate adaptive SNP allele frequencies (**F**, Chr17: 12725631; **G**, Chr05: 9005657) associated with Srad6 or Bio1 were investigated across the 36 populations. Color mapping is based on variations observed in the relevant range of environmental variables. **H, I** Comparison of average RONA values under two different climate scenarios (SSP126 and SSP585) in 2081–2100 across populations for Bio1 (*n* = 877 LFMM variants, *P* < .001) (**H**) and Bio13 (*n* = 1565 LFMM variants, *P* < .001) (**I**). Error bars represent the standard error of average RONA calculated from four different climate models. **J** Average RONA estimates of Bio1 across four climate models for the 24 populations under the SSP585 climate scenarios in 2081–2100. The raster colors on the map represent the degree of projected future climate change (absolute change). Areas with darker color are predicted to experience more substantial change in the respective climate variables. The size of the circles on the map represents the RONA values of different natural populations.

EALs associated with multiple environmental variables were identified using partial redundancy analysis (RDA) (Fig. 6C, Supplementary Data Fig. S12, Supplementary Data Table S23). Among the EALs in RDA, 11.62% were non-synonymous and 6.44% were synonymous mutations, and the non-synonymous rate showed

no significant difference from all outlier loci (two-proportion Z-test, *P* = .267) (Supplementary Data Table S24). In addition, variation ratios of the 5'UTR (4.86%) and 3'UTR (7.50%) were significantly higher than those in the outlier loci (two-proportion Z-test, *P* < .001), indicated that environmental adaptation might

mainly rely on selection through regulatory rather than protein-coding changes (Supplementary Data Table S24).

The key EALs for RDA were annotated to 391 genes (Supplementary Data Table S25, Supplementary Data Table S26). Among them, UV RESISTANCE LOCUS 8 (*UVR8*) was a photoreceptor mediating photomorphogenic responses to UV-B, and essential for photomorphogenesis and UV tolerance. To highlight the distribution pattern of allele frequencies, we focused on one essential adaptive SNP located downstream of *UVR8* (Pca17G009660) as an example (Fig. 6D). The T allele was mainly distributed in the NW China group, where solar radiation was high, whereas the C allele was more common in N and SW regions with weak solar radiation (Fig. 6E). Moreover, we found that a set of key UV irradiation genes, including cullin-4 gene (*CUL4*, Pca04G012380), DNA ligase 1 (*LIG1*, Pca04G008510), and Su(var)3–9 homologs (*SUVH4*, Pca05G019100), showed similar geographic distribution in frequencies (Supplementary Data Fig. S13). We identified multiple loci in temperature-related genes (Bio1, Bio2, and Bio3). For instance, the three tandem-duplicated genes encoding flavonoid 3′-hydroxylase (*F3′H*: Pca05G010730, Pca05G010700, and Pca05G010710) were significantly associated with Bio1 (Fig. 6F). A representative SNP variation was located in the intron of *F3′H*, which carried the T allele more frequently in high-altitude areas with lower temperature (Fig. 6G). The 75 genes identified in the GEA analysis were found within the genetic differentiation region, providing evidence of local adaptation in the *P. cathayana* population (Supplementary Data Fig. S14, Supplementary Data Table S27). The adaptive genes for selective sweep and GEA analysis were enriched in DNA repair and UV radiation response, such as *UVR8*, *HY5*, and *CUL4*. This finding emphasizes the role of these genes in high-altitude adaptation.

Using EALs by LFMM, we assessed the potential spatial pattern of maladaptation for *P. cathayana* using the risk of non-adaptedness (RONA). The values of RONA across the three future climate models were highly correlated across populations (Supplementary Data Fig. S15A, Supplementary Data Table S28). Average RONA values were calculated for each model, highlighting significant differences among populations (Fig. 6H and I). As expected, RONA increases under more severe climate change scenarios for most populations, as indicated by the comparison between SSP585 and SSP126 (Supplementary Data Fig. S15B–D). The temperature variable (Bio1) and precipitation variable (Bio13) indicated that the RONA value was stronger in areas with more drastic climate changes (Fig. 6J, Supplementary Data Fig. S15E–I). Considering climate change and the adaptive potential of variation, *P. cathayana* populations in North China will be confronted with maladaptation risks.

Discussion

Populus cathayana is an important component of the *Tacamahaca* poplars. Using comprehensive sequencing technology, we have assembled the first high-quality *P. cathayana* genome, with 99.84% of the sequences ordered and orientated onto chromosomes. The favorable results of multiple evaluation approaches further confirm the high quality and completeness of assembly [27–29]. The WGD event provides original genetic material and plays a driving role in specialization and the creation of unique traits and functions [30]. Like other species of Salicaceae, *P. cathayana* underwent two WGD events, which led to the enrichment of gene duplications in defense and stimulation responses, and partially explained the expansion of the defense response-related gene family in the genome. In particular, we have discovered the expansion of disease resistance genes in *P. cathayana*, which may provide support for disease resistance breeding in poplar.

China has abundant germplasm resources of poplar belonging to the *Tacamahaca* section, mostly distributed in areas with variable terrain and a complex climate [1, 31, 32]. We present the first report on the collection and population genomic analyses of a large natural population of *P. cathayana* that is continuously distributed in China. Based on the population genetics, the NW group exhibits a broader genetic background, suggesting it as the genetic diversity and origin center of *P. cathayana*. Population structure can reflect past divergence events and gene flow [33–35]. Previous studies on *P. trichocarpa*, *P. davidiana*, and *P. deltoides* have shown that the demographic legacy of the Pleistocene climate change had a significant effect on genome-wide patterns of diversity [14, 36–38]. The *P. cathayana* population also has experienced bottleneck effects and contraction of its potential distribution range, as suggested by our ENM model, which may lead to the existence of glacier refuges in the eastern margin of the Qinghai–Tibet Plateau, the Qinling Mountains and North China. Some northern temperate plants in East Asia have hidden refugia at much higher latitudes than previously expected [35, 39–44]. The stable demographic history and variation haplotypes in the NC group suggest that 'cryptic' refugia exist and persist in poplar. We propose that the *P. cathayana* population has a complex history of large-scale populations affected by glaciers, and migration and diffusion from refugia after the ice age to form the present-day distribution.

The wild germplasm offers rich genetic resources that can be used to improve and enhance the growth characteristics and adaptability of trees [34, 45–47]. Through genome level analysis, the adaptive features of *P. cathayana* were studied. As observed in other widely distributed species [42, 48], the differences among *P. cathayana* lineages were highly heterogeneous in the genome. Additionally, GEA was employed to highlight loci with allele frequencies closely associated with environmental gradients [49, 50]. This rich information on the genetic basis of environmental adaptation may help in the development of effective strategies to mitigate the impact of climate change on species. Indeed, *P. cathayana* is primarily adapted to high altitude and grows near streams or ponds, where the genome is continuously challenged by intensive strong ultraviolet radiation and low temperature. Many previously reported genes, such as *HY5*, *CUL4*, and *RAD51* involved in UV-B signal transduction and DNA repair pathway, were prominent in the genome region associated with local adaptation [51–54]. This study identified hundreds of genes for local adaptation, many of which were functionally involved in the response to low temperature, providing support for low-temperature tolerance breeding [55, 56].

There are significant differences in RONA values among populations, indicated that adaptive differences in species stem from the genetic background and location. When considering the impact of future climate, it is necessary to consider not only the variability of allele frequencies, but also the climate risk of geographical location. Local adaptation provides only partial solutions at the genetic level for species to cope with the challenges of climate change. In summary, we conducted high-quality reference genome and population genome analysis of *P. cathayana*, which has enhanced our understanding of the genetic diversity of local poplar species in China. Given the context of genome-scale knowledge, we have the opportunity to address the challenge of addressing population vulnerability to future climate change [48, 53, 57, 58]. Our results can be used to help screen materials suitable for various environmental conditions

from a large number of germplasms *in situ*, as they contain many unique and climate-adaptive genetic resources.

Materials and methods

Plant materials

A male *P. cathayana* individual was collected from Jiangou, Mentougou District, Beijing, China (116.05° E, 40.07° N, 922 m) for genome assembly. Tissue culture plants were produced using material from this individual. Young leaves, buds, and roots from the tissue-cultured seedlings were subsequently collected for RNA sequencing (Supplementary Data Methods S1). Next, a total of 438 wild germplasms of *P. cathayana* were collected for genome resequencing from 36 locations across its distribution range in China (Supplementary Data Table S16). Sampled individuals of each location were separated by at least 100 m. The geographic location information, including altitude, latitude, and longitude, of each individual was recorded using an Etrex GIS monitor (Garmin). All the above materials were preserved in the nursery of the Chinese Academy of Forestry in Beijing, China.

DNA isolation and genome sequencing

The DNA used for Illumina sequencing and PacBio sequencing was extracted from leaves of tissue culture seedlings using a modified CTAB method. The Illumina library, with 270-bp insert size, was constructed following the manufacturer's instructions and sequenced on the Illumina HiSeq X Ten platform. For PacBio library construction, genomic DNA was sheared to 20 kb and sequenced on the PacBio Sequel system. To satisfy the specific requirements of Hi-C sequencing, the fixed scheme was used to process and extract genomic DNA from leaf tissue, yielding DNA fragments ranging from 300 to 700 bp. The Hi-C library was sequenced on the Illumina HiSeq X Ten platform [59].

Chromosome-level genome assembly

To obtain a high-quality genome assembly, the quality-controlled PacBio reads were initially corrected using a self-align method in NextDenovo software (v2.0) (https://github.com/Nextomics/NextDenovo). The corrected reads were then used to assemble the draft genome using the correction-before-assembly strategy. Indels and SNPs were subsequently corrected using Illumina sequencing data by Pilon software over three rounds [60]. The resulting assembly underwent further processing with purge_haplotigs to remove redundant haplotypes, producing a final contig-level assembly. The Hi-C data were utilized to further remove non-significant genome regions [61]. To refine the assembly, the clean reads were aligned to the contig-level assembly using Bowtie2 (v2.2.3) with default parameters. Using HiC-Pro (v2.10), the uniquely mapped paired-end reads from Hi-C data were retained for further analysis [62]. The uniquely mapped read pairs were used to clustered, ordered and orientated scaffolds onto chromosomes by LACHESIS [63]. Before chromosome assembly, scaffolds were divided into segments of ~50 kb. BWA software (v0.7.10-r789) was used to map the Hi-C data to these segments [64]. The placement and orientation that displayed distinct chromatin interaction patterns were corrected manually. To evaluate the final genome assembly, Illumina short reads were mapped to the genome using BWA (v0.7).

Genome prediction and annotation

Firstly, three *de novo* prediction programs, RepeatModeler2 (v2.0.1) [65], RECON (v1.0.8) [66], and RepeatScout(v1.0.6) [67], were used to estimate the repeat composition in the genome (Supplementary Data Methods S1). Then, repeat sequences in this genome were classified using RepeatMasker (v19.06), REXdb (v3.0), Dfam (v3.2), and LTR_retriever (v2.8) [68–71]. The combined strategy of three approaches was used to predict protein-coding genes (Supplementary Data Methods S1). For the prediction of different types of non-coding RNA, tRNAscan-SE (v1.3.1) and miRBase (v21) were used to predict tRNA and miRNA with eukaryote parameters, respectively [72]. The rRNA genes were identified by Rfam (v12.0), and snoRNA and snRNA were predicted using INFERNAL against the Rfam (v1.1) [73, 74]. The qualities of the assembly and gene annotation were assessed using BUSCO (v5.2) and CEGMA (v2.5).

Comparative genome analysis

Genome sequences from eight eudicot genomes of *P. trichocarpa*, *P. euphratica*, *P. alba*, *S. purpurea*, *Eucalyptus grandis*, *Quercus robur*, *Arabidopsis thaliana*, and *Oryza sativa* were obtained. The download addresses of the genomes are shown in Supplementary Data Table S14. OrthoFinder2 (v2.2.7) was used to identify orthologous genes among nine species [75]. The phylogenetic relationships were determined using PhyML (v3.1) based on single-copy orthologous genes, and the resulting phylogenetic tree was visualized using FigTree (v1.4.3) [76]. Divergence times were estimated using MCMCtree in PAML (v4.9i) [77]. To determine the expansion and contraction of gene families, the CAFÉ program (v3.1) was used to compare cluster size differences between ancestors and species [78]. 4DTV was used to evaluate the selection pressure and evolution rate of genome sequences, and PAML (v4.9i) software was used for calculation [77].

We utilized the BLAST (v2.12) tool to identify orthologous genes within the *P. cathayana* genomes (E-value <1e−5) [79] and MCScanX (python) was used to identify syntenic blocks and visualize the collinearity [80]. Using DupGen_finder with default parameters, we identified five duplicate types in the *P. cathayana* genome. The K_a/K_s values of duplicate gene pairs were calculated using TBtools (v1.120) [81].

Population resequencing and variation detection

Total genomic DNA from fresh and undamaged leaves was extracted using the modified CTAB method for resequencing. Paired-end sequencing libraries were constructed with an insert size of 150 bp. Whole-genome resequencing was conducted on the Illumina HiSeq 2500 platform, aiming for a target coverage of 30× per individual. After quality assessment using FastQC, low-quality bases with a Phred score <30 were trimmed from the reads using Trimmomatic. The high-quality paired-end sequencing reads were aligned to the *P. cathayana* genome assembly using the 'mem' algorithm of BWA software (v0.7.8). Duplicate reads were subsequently removed using SAMtools (v0.1.19) [82]. SNP calling was performed using GATK (v4.1.9.0) with the joint calling method [83]. We considered indel calls produced by SAMtools mpileup within a 1- to 50-bp window. For SV identification, BreakDancerMax (v1.4.4) was used to detect insertion (INS), deletion (DEL), and inversion (INV) [84]. We only retained variation supported by more than two read pairs. To detect CNVs, CNVnator (v0.3.2) with the parameter -call 100 was used [85]. ANNOVAR was utilized to annotate the genomic regions and quantities of SNPs, indels, CNVs, and SVs based on the *P. cathayana* reference genome with GFF3 files.

Population structure and genetic diversity analysis

After variant calling SNPs were filtered with VCFTools [86] with parameters -min-meanDP 3 -minQ 30 -maf 0.05 -remove-indels -max-missing 0.2 -min-alleles 2 -max-alleles 2. A total of 12 374 210 SNPs were retained for further analysis. First, population structure was analyzed using ADMIXTURE (v1.23), and cross-validation error was tested for each cluster value (K) (2–15). Next, the TreeBeST program (v1.92) was employed to construct a neighbor-joining (NJ) tree. For PCA, GCTA (v1.24.2) was employed [87]. Homozygous individuals were retained for genetic diversity analysis, and grouping of individuals was shown in Supplementary Data Table S17. Pairwise genetic differentiation (F_{ST}) was calculated using VCFtools (v0.1.16) with a window size of 100 kb and 1-kb steps to evaluate population differentiation [86]. The population recombination rate was estimated using FastEPRR, within a 100-kb window and 1-kb steps [88]. To assess LD decay, the LD coefficient (r^2) between pairwise SNPs within a 500 kb window was calculated using PopLDdecay (v3.40) [89]. The Mantel test was executed between the $F_{ST}/(1 - F_{ST})$ matrix and geographical distance (km) matrix of the *P. cathayana* population using the R package vegan. Significance was determined based on 999 permutations. To calculate nucleotide diversity (π) and Tajima's D value, VCFtools (v0.1.16) was utilized with a 100-kb sliding window for four groups [87].

Demographic history and potential distribution prediction

PSMC (v0.6.5) was used to infer the history of effective population size (N_e) of *P. cathayana* ($\mu = 6.2 \times 10^{-8}$, $g = 10$ years) [90, 91]. Fastsimcoal2 (v2.1) was used to derive the best-fitting demographic model, estimate population differentiation time, and determine the per-generation migration rate of gene flow between clades [92]. TreeMix (v1.11) was used to model gene flow among subgroups of *P. cathayana* [93]. This inferred the maximum likelihood tree and identified potential gene flow based on the residual covariance matrix. Nine migration event models were established, and the optimal model was selected considering the AIC value and stability. Using the ENMs to predict the potential distribution areas of species, occurrence data were obtained from the sampling location of *P. cathayana*. The data for 19 bioclimatic variables during the current period (1970–2000) and the LGM were obtained from WorldClim (http://www.worldclim.org/) at a spatial resolution of 2.5 arcmin (~4.5 km). The bioclimatic data for the LGM were derived from the CCSM circulation model [94]. To calibrate and select the best ENMs, the Maxent algorithm implemented in the R package ENMeval was used [95, 96].

Identification of selection sweep signatures

To detect selection signals during adaptation in the four groups, the admixed genotypes were first removed according to the population structure. A total of 358 samples of four groups were used to detect selective sweep signals [97]. Regions with signals for selective sweeps were identified using the top 5% F_{ST} values and the top 5% $\theta\pi$ ratios [25]. For the calculation of F_{ST}, a window size of 100 kb and a step size of 1 kb were used. For the calculation of nucleotide diversity (π), a sliding window of 100 kb was employed. Genes located within selective sweep regions were considered as candidate genes. We performed GO and KEGG enrichment analysis for the candidate genes [98, 99]. The STRING database (v11.5) was also used to annotate homology genes and construct the PPI network [100].

Collection of bioclimatic variable data

To assess the impact of environmental factors, we obtained environmental variables from WorldClim. These variables included 19 bioclimatic variables (Bio1–19), wind speed (Wind6 and Wind12), water vapor pressure (Vapr6 and Vapr12), and solar radiation (Srad6 and Srad12) for the period 1970–2000, with a resolution of 2.5 arcmin (Supplementary Data Fig. S21). These variables were extracted using DIVA-GIS (v10.6). The Pearson correlation between environmental variables was calculated, and variables with a correlation >0.8 were excluded, retaining nine environmental variables for further analysis. For each population, we downloaded future (2081–2100) data for the five EVs (Bio1, Bio2, Bio3, Bio13, and Bio19) from the WorldClim CMIP6 dataset of three climate models (BCCCSM2-MR model, ACCESS-CM2 model, and CMCC-ESM2 model; resolution 2.5 arcmin), with two shared socioeconomic pathways (SSPs): SSP126 and SSP585.

Genotype–environment association analysis

Outlier loci were identified using BayeScan (v2.172) for GEA analysis, and the cutoff thresholds were a posterior probability >0.76 and a q value <0.05 [101]. The LFMM was used to test the relationship between outlier SNPs and nine environmental variables to determine the significant EALs [102, 103]. LFMMs were implemented by the R package Latent Environmental Association analysis (LEA), and natural genetic structure was incorporated as a random effect. SNP loci with $|z|$ values ≥4 and P-values ≤1.0 e−5 associated with at least one environmental variable were considered as key EALs [102]. To investigate the role of genetic variation in adaptation (the key EALs from LFMM) and neutral variation (further LD pruned of SNPs), Mantel tests were separately used to test for associations between $F_{ST}/(F_{ST}/1 - F_{ST})$ and environmental (IBE) distance, with significance determined using the R package vegan [7].

To explore the independent effects of environmental variables on environmental adaptability, the partial RDA model was conducted. The key adaptive EALs were used as response variables. Nine environmental variables were included as explanatory variables, while longitude and latitude were used as control variables to account for geographic factors. The RDA function in the R package vegan was used to perform partial RDA, and the anova.cca function was used to check the significance of environmental variables [104, 105]. The proportion of genetic variation explained by each RDA axis was calculated. SNPs that loaded on the tails of the distribution in RDA axes 1–3, with a cutoff of a 95% confidence interval, were considered as candidate loci and subjected to genomic annotation [105].

Analysis of the adaptation potential of the genome to future climate change

Using the RONA approach, we calculated the theoretical allele frequency change needed to cope with future climate for EVs (Bio1, Bio2, Bio3, Bio13 and Bio19) [26]. The linear relationship between allele frequencies at significantly associated loci by LFMM and EVs was established using linear regressions (RONA parameter: LFMM, $P < .001$) [106, 107]. To avoid insufficient statistical efficacy or bias, 30 populations with more than six individuals were retained for RONA analysis. The difference (absolute value) between the present (1970–2000) and future (2081–2100, SSP585, BCC-CSM2-MR) was displayed on a map using Arcmap (v10.6).

Acknowledgements

Author contributions

X.X., L.Z., and J.H. designed the project. D.C., Y.Z., L.Z., and J.H. collected the experimental materials. X.X. and J.H. performed the genome analyses. X.X., X.Z., H.W., and H.Z. performed resequencing and genetic analyses. X.X. and H.Z. contributed to the interpretation of results. X.X. wrote the manuscript. X.Z., L.Z. and J.H. revised the manuscript.

Data availability

The *P. cathayana* genome and the sequencing data for genome assembly (Illumina reads, PacBio long reads, Hi-C reads, and RNA-seq data) have been deposited in the National Genomics Data Center (https://ngdc.cncb.ac.cn/?lang=en) under BioProject PRJCA014016. The whole-genome resequencing data have been deposited under National Genomics Data Center BioProject PRJCA014017 with the accession SAMC1029105-SAMC1029542.

Conflict of interest

The authors declare no competing interests.

Supplementary data

Supplementary data is available at *Horticulture Research* online.

References

1. Eckenwalder JE. Systematics and evolution of *Populus*. In: Stettler RF, Bradshaw HD, Heilman PE, Hinckley TM (eds). *Biology of Populus and its Implications for Management and Conservation*. Ottawa: NRC Research Press, 1996, 7–32

2. Chen K, Peng Y, Wang Y. *et al.* Genetic relationships among poplar species in section *Tacamahaca* (*Populus* L.) from western Sichuan, China. *Plant Sci.* 2007;**172**:196–203

3. Weisgerber H, Zhang Z. *Populus cathayana* Rehder, 1931. *Encyclopedia of Woody Plants* 41. Nicol, Hamburg, 2005.

4. Li KY, Huang MR, Yang ZX. *et al.* Genetic differentiation of *Populus cathayana*. *Acta Bot Sin.* 1997;**39**:753–8

5. He Y, Zhu Z, Guo Q. *et al.* Sex-specific interactions affect foliar defense compound accumulation and resistance to herbivores in *Populus cathayana*. *Sci Total Environ.* 2021;**774**:145819

6. Lu Z, Wang Y, Peng Y. *et al.* Genetic diversity of *Populus cathayana* Rehd populations in southwestern China revealed by ISSR markers. *Plant Sci.* 2006;**170**:407–12

7. Whitlock MC, Lotterhos KE. Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of F_{ST}. *Am Nat.* 2015;**186 Suppl 1**:S24–36

8. Sang Y, Long Z, Dan X. *et al.* Genomic insights into local adaptation and future climate-induced vulnerability of a keystone forest tree in East Asia. *Nat Commun.* 2022;**13**:6541

9. Jia H, Liu G, Li J. *et al.* Genome resequencing reveals demographic history and genetic architecture of seed salinity tolerance in *Populus euphratica*. *J Exp Bot.* 2020;**71**:4308–20

10. Wang J, Street NR, Park EJ. *et al.* Evidence for widespread selection in shaping the genomic landscape during speciation of *Populus*. *Mol Ecol.* 2020;**29**:1120–36

11. Li Y, Cao K, Li N. *et al.* Genomic analyses provide insights into peach local adaptation and responses to climate change. *Genome Res.* 2021;**31**:592–606

12. Hall D, Luquez V, Garcia VM. *et al.* Adaptive population differentiation in phenology across a latitudinal gradient in European aspen (*Populus tremula*, L.): a comparison of neutral markers, candidate genes and phenotypic traits. *Evolution.* 2007;**61**:2849–60

13. Capblancq T, Fitzpatrick MC, Bay RA. *et al.* Genomic prediction of (mal)adaptation across current and future climatic landscapes. *Annu Rev Ecol Evol Syst.* 2020;**51**:245–69

14. Evans LM, Slavov GT, Rodgers-Melnick E. *et al.* Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. *Nat Genet.* 2014;**46**:1089–96

15. Liu YJ, Wang XR, Zeng QY. *De novo* assembly of white poplar genome and genetic diversity of white poplar population in Irtysh River basin in China. *Sci China Life Sci.* 2019;**62**:609–18

16. Qiu D, Bai S, Ma J. *et al.* The genome of *Populus alba* x *Populus tremula* var. *glandulosa* clone 84K. *DNA Res.* 2019;**26**:423–31

17. Hofmeister BT, Denkena J, Colome-Tatche M. *et al.* A genome assembly and the somatic genetic and epigenetic mutation rate in a wild long-lived perennial *Populus trichocarpa*. *Genome Biol.* 2020;**21**:259

18. Zhang Z, Chen Y, Zhang J. *et al.* Improved genome assembly provides new insights into genome evolution in a desert poplar (*Populus euphratica*). *Mol Ecol Resour.* 2020;**20**:781–94

19. Blanquart F, Kaltz O, Nuismer SL. *et al.* A practical guide to measuring local adaptation. *Ecol Lett.* 2013;**16**:1195–205

20. Borrell JS, Zohren J, Nichols RA. *et al.* Genomic assessment of local adaptation in dwarf birch to inform assisted gene flow. *Evol Appl.* 2020;**13**:161–75

21. Geng Y, Guan Y, Qiong L. *et al.* Genomic analysis of field pennycress (*Thlaspi arvense*) provides insights into mechanisms of adaptation to high elevation. *BMC Biol.* 2021;**19**:143

22. Gugger PF, Fitz-Gibbon ST, Albarran-Lara A. *et al.* Landscape genomics of *Quercus lobata* reveals genes involved in local climate adaptation at multiple spatial scales. *Mol Ecol.* 2021;**30**:406–23

23. Holliday JA, Ritland K, Aitken SN. Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (*Picea sitchensis*). *New Phytol.* 2010;**188**:501–14

24. Zhang M, Suren H, Holliday JA. Phenotypic and genomic local adaptation across latitude and altitude in *Populus trichocarpa*. *Genome Biol Evol.* 2019;**11**:2256–72

25. Abhilash PC, Tripathi V, Dubey RK. *et al.* Coping with changes: adaptation of trees in a changing environment. *Trends Plant Sci.* 2015;**20**:137–8

26. Rellstab C, Zoller S, Walthert L. *et al.* Signatures of local adaptation in candidate genes of oaks (*Quercus spp.*) with respect to present and future climatic conditions. *Mol Ecol.* 2016;**25**:5907–24

27. Vanneste K, Maere S, Van de Peer Y. Tangled up in two: a burst of genome duplications at the end of the cretaceous and the consequences for plant evolution. *Philos Trans R Soc Lond Ser B Biol Sci.* 2014;**369**:20130353

28. Yang FS, Nie S, Liu H. *et al.* Chromosome-level genome assembly of a parent species of widely cultivated azaleas. *Nat Commun.* 2020;**11**:5269

29. Yang S, Zhang X, Yue JX. *et al.* Recent duplications dominate NBS-encoding gene expansion in two woody species. *Mol Gen Genomics*. 2008;**280**:187–98

30. Hou J, Ye N, Dong Z. *et al.* Major chromosomal rearrangements distinguish willow and poplar after the ancestral "salicoid" genome duplication. *Genome Biol Evol*. 2016;**8**:1868–75

31. Cao DM, Zhang YH, Cheng XQ. *et al.* Genetic variation of leaf phenotypic traits in different populations of *Populus cathayana*. *Scientia Silvae Sinicae*. 2021;**57**:56–7.

32. Wang M, Zhang L, Zhang Z. *et al.* Phylogenomics of the genus *Populus* reveals extensive interspecific gene flow and balancing selection. *New Phytol*. 2020;**225**:1370–82

33. Murray KD, Janes JK, Jones A. *et al.* Landscape drivers of genomic diversity and divergence in woodland *Eucalyptus*. *Mol Ecol*. 2019;**28**:5232–47

34. Shen Y, Xia H, Tu Z. *et al.* Genetic divergence and local adaptation of *Liriodendron* driven by heterogeneous environments. *Mol Ecol*. 2022;**31**:916–33

35. Wu S, Wang Y, Wang Z. *et al.* Species divergence with gene flow and hybrid speciation on the Qinghai-Tibet plateau. *New Phytol*. 2022;**234**:392–404

36. Fahrenkrog AM, Neves LG, Resende MFR Jr. *et al.* Population genomics of the eastern cottonwood (*Populus deltoides*). *Ecol Evol*. 2017;**7**:9426–40

37. Li A, Hou Z. Phylogeographic analyses of poplar revealed potential glacial refugia and allopatric divergence in Southwest China. *Mitochondrial DNA A DNA Mapp Seq Anal*. 2021;**32**:66–72

38. Hou Z, Li A. Population genomics reveals demographic history and genomic differentiation of *Populus davidiana* and *Populus tremula*. *Front Plant Sci*. 2020;**11**:1103

39. Shen L, Chen XY, Zhang X. *et al.* Genetic variation of *Ginkgo biloba* L. (Ginkgoaceae) based on cpDNA PCR-RFLPs: inference of glacial refugia. *Heredity*. 2005;**94**:396–401

40. Xu XX, Cheng FY, Peng LP. *et al.* Late Pleistocene speciation of three closely related tree peonies endemic to the Qinling-Daba Mountains, a major glacial refugium in Central China. *Ecol Evol*. 2019;**9**:7528–48

41. Richards TJ, Karacic A, Apuli RP. *et al.* Quantitative genetic architecture of adaptive phenology traits in the deciduous tree, *Populus trichocarpa* (Torr. and Gray). *Heredity (Edinb)*. 2020;**125**:449–58

42. Numaguchi K, Akagi T, Kitamura Y. *et al.* Interspecific introgression and natural selection in the evolution of Japanese apricot (*Prunus mume*). *Plant J*. 2020;**104**:1551–67

43. Fan L, Zheng H, Milne RI. *et al.* Strong population bottleneck and repeated demographic expansions of *Populus adenopoda* (Salicaceae) in subtropical China. *Ann Bot*. 2018;**121**:665–79

44. Guo Y-P, Zhang R, Chen C-Y. *et al.* Allopatric divergence and regional range expansion of *Juniperus sabina* in China. *J Syst Evol*. 2010;**48**:153–60

45. Wanderley AM, Machado ICS, de Almeida EM. *et al.* The roles of geography and environment in divergence within and between two closely related plant species inhabiting an island-like habitat. *J Biogeogr*. 2018;**45**:381–93

46. Evans SN, Hening A, Schreiber SJ. Protected polymorphisms and evolutionary stability of patch-selection strategies in stochastic environments. *J Math Biol*. 2015;**71**:325–59

47. Willi Y, Fracassetti M, Bachmann O. *et al.* Demographic processes linked to genetic diversity and positive selection across a species' range. *Plant Commun*. 2020;**1**:100111.

48. DeSilva R, Dodd RS. Association of genetic and climatic variability in giant sequoia, *Sequoiadendron giganteum*, reveals signatures of local adaptation along moisture-related gradients. *Ecol Evol*. 2020;**10**:10619–32

49. Capblancq T, Morin X, Gueguen M. *et al.* Climate-associated genetic variation in *Fagus sylvatica* and potential responses to climate change in the French Alps. *J Evol Biol*. 2020;**33**:783–96

50. Lobreaux S, Miquel C. Identification of *Arabis alpina* genomic regions associated with climatic variables along an elevation gradient through whole genome scan. *Genomics*. 2020;**112**:729–35

51. Nawkar GM, Kang CH, Maibam P. *et al.* HY5, a positive regulator of light signaling, negatively controls the unfolded protein response in *Arabidopsis*. *Proc Natl Acad Sci USA*. 2017;**114**:2084–9

52. Kim S, Hwang G, Lee S. *et al.* High ambient temperature represses anthocyanin biosynthesis through degradation of HY5. *Front Plant Sci*. 2017;**8**:1787.

53. Huang X, Ouyang X, Yang P. *et al.* Conversion from CUL4-based COP1-SPA E3 apparatus to UVR8-COP1-SPA complexes underlies a distinct biochemical function of COP1 under UV-B. *Proc Natl Acad Sci USA*. 2013;**110**:16669–74

54. Yu C, Hou L, Huang Y. *et al.* The multi-BRCT domain protein DDRM2 promotes the recruitment of RAD51 to DNA damage sites to facilitate homologous recombination. *New Phytol*. 2023;**238**:1073–84

55. Lovdal T, Olsen KM, Slimestad R. *et al.* Synergetic effects of nitrogen depletion, temperature, and light on the content of phenolic compounds and gene expression in leaves of tomato. *Phytochemistry*. 2010;**71**:605–13

56. Song Y, Feng J, Liu D. *et al.* Different phenylalanine pathway responses to cold stress based on metabolomics and transcriptomics in Tartary buckwheat landraces. *J Agric Food Chem*. 2022;**70**:687–98

57. Weber M, Beyene B, Nagler N. *et al.* A mutation in the essential and widely conserved DAMAGED DNA BINDING1-Cullin 4 ASSOCIATED FACTOR gene OZS3 causes hypersensitivity to zinc excess, cold and UV stress in *Arabidopsis thaliana*. *Plant J*. 2020;**103**:995–1009

58. Liang T, Yang Y, Liu H. Signal transduction mediated by the plant UV-B photoreceptor UVR8. *New Phytol*. 2019;**221**:1247–52

59. Belton JM, McCord RP, Gibcus JH. *et al.* Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods*. 2012;**58**:268–76

60. Walker BJ, Abeel T, Shea T. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*. 2014;**9**:e112963.

61. Rao SS, Huntley MH, Durand NC. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014;**159**:1665–80

62. Servant N, Varoquaux N, Lajoie BR. *et al.* HiC-pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol*. 2015;**16**:259

63. Burton JN, Adey A, Patwardhan RP. *et al.* Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat Biotechnol*. 2013;**31**:1119–25

64. Boetzer M, Henkel CV, Jansen HJ. *et al.* Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics*. 2011;**27**:578–9

65. Flynn JM, Hubley R, Goubert C. *et al.* RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci USA*. 2020;**117**:9451–7

66. Bao Z, Eddy SR. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res.* 2002;**12**: 1269–76

67. Price AL, Jones NC, Pevzner PA. *De novo* identification of repeat families in large genomes. *Bioinformatics.* 2005;**21 Suppl 1**: i351–8

68. Neumann P, Novak P, Hostakova N. *et al.* Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. *Mob DNA.* 2019;**10**:1

69. Jurka J, Kapitonov VV, Pavlicek A. *et al.* Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res.* 2005;**110**:462–7

70. Wheeler TJ, Clements J, Eddy SR. *et al.* Dfam: a database of repetitive DNA based on profile hidden Markov models. *Nucleic Acids Res.* 2013;**41**:D70–82

71. Ou S, Jiang N. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* 2018;**176**:1410–22

72. Griffiths-Jones S, Grocock RJ, van Dongen S. *et al.* miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* 2006;**34**:D140–4

73. Griffiths-Jones S, Moxon S, Marshall M. *et al.* Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* 2005;**33**:D121–4

74. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics.* 2013;**29**:2933–5

75. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 2015;**16**:157

76. Guindon S, Dufayard JF, Lefort V. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 2010;**59**: 307–21

77. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007;**24**:1586–91

78. De Bie T, Cristianini N, Demuth JP. *et al.* CAFE: a computational tool for the study of gene family evolution. *Bioinformatics.* 2006;**22**:1269–71

79. She R, Chu JS, Wang K. *et al.* GenBlastA: enabling BLAST to identify homologous gene sequences. *Genome Res.* 2009;**19**: 143–9

80. Wang Y, Tang H, Debarry JD. *et al.* MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 2012;**40**:e49.

81. Chen C, Chen H, Zhang Y. *et al.* TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant.* 2020;**13**:1194–202

82. Li H, Handsaker B, Wysoker A. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;**25**:2078–9

83. McKenna A, Hanna M, Banks E. *et al.* The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;**20**:1297–303

84. Chen K, Wallis JW, McLellan MD. *et al.* BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods.* 2009;**6**:677–81

85. Abyzov A, Urban AE, Snyder M. *et al.* CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* 2011;**21**:974–84

86. Danecek P, Auton A, Abecasis G. *et al.* The variant call format and VCFtools. *Bioinformatics.* 2011;**27**:2156–8

87. Yang J, Lee SH, Goddard ME. *et al.* Genome-wide Complex Trait Analysis (GCTA): methods, data analyses, and interpretations. *Methods Mol Biol.* 2013;**1019**:215–36

88. Gao F, Ming C, Hu W. *et al.* New software for the fast estimation of population recombination rates (FastEPRR) in the genomic era. *G3 (Bethesda).* 2016;**6**:1563–71

89. Zhang C, Dong SS, Xu JY. *et al.* PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics.* 2019;**35**:1786–8

90. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. *Nature.* 2011;**475**:493–6

91. Fumagalli M, Vieira FG, Korneliussen TS. *et al.* Quantifying population genetic differentiation from next-generation sequencing data. *Genetics.* 2013;**195**:979–92

92. Excoffier L, Dupanloup I, Huerta-Sanchez E. *et al.* Robust demographic inference from genomic and SNP data. *PLoS Genet.* 2013;**9**:e1003905.

93. Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 2012;**8**:e1002967.

94. Padalia H, Srivastava V, Kushwaha SP. How climate change might influence the potential distribution of weed, bushmint (*Hyptis suaveolens*)? *Environ Monit Assess.* 2015;**187**:210

95. Hosni EM, Nasser MG, Al-Ashaal SA. *et al.* Modeling current and future global distribution of *Chrysomya bezziana* under changing climate. *Sci Rep.* 2020;**10**:4947

96. Marques I, Draper D, Lopez-Herranz ML. *et al.* Past climate changes facilitated homoploid speciation in three mountain spiny fescues (*Festuca*, Poaceae). *Sci Rep.* 2016;**6**:36283

97. Lotterhos KE, Whitlock MC. Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Mol Ecol.* 2014;**23**:2178–92

98. Ashburner M, Ball CA, Blake JA. *et al.* Gene Ontology: tool for the unification of biology. *Nat Genet.* 2000;**25**:25–9

99. Kanehisa M, Sato Y, Kawashima M. *et al.* KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 2016;**44**:D457–62

100. Szklarczyk D, Gable AL, Lyon D. *et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;**47**:D607–13

101. Li LF, Cushman SA, He YX. *et al.* Genome sequencing and population genomics modeling provide insights into the local adaptation of weeping forsythia. *Hortic Res.* 2020;**7**:130

102. Frichot E, Schoville SD, Bouchard G. *et al.* Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol Biol Evol.* 2013;**30**:1687–99

103. Caye K, Jumentier B, Lepeule J. *et al.* LFMM 2: fast and accurate inference of gene-environment associations in genome-wide studies. *Mol Biol Evol.* 2019;**36**:852–60

104. Rellstab C, Gugerli F, Eckert AJ. *et al.* A practical guide to environmental association analysis in landscape genomics. *Mol Ecol.* 2015;**24**:4348–70

105. Forester BR, Lasky JR, Wagner HH. *et al.* Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Mol Ecol.* 2018;**27**:2215–33

106. Dauphin B, Rellstab C, Schmid M. *et al.* Genomic vulnerability to rapid climate warming in a tree species with a long generation time. *Glob Change Biol.* 2021;**27**:1181–95

107. Hoffmann AA, Weeks AR, Sgrò CM. Opportunities and challenges in assessing climate change vulnerability through genomics. *Cell.* 2021;**184**:1420–5