communications biology

A Nature Portfolio journal

Article

https://doi.org/10.1038/s42003-026-09846-x

Population genomics of *Quercus gilva* provides insights into the conservation of fengshui forests

Xiao-Long Jiang[1,2,7] ✉, Meng-Xiao WU[1,7], Yoko SAITO[3], Shi-Shi LUO[1], Xiong-Sheng LIU[4], Ze-Yi OUYANG[5], Jia-Xiang LI[1,2], Min DENG[6] & He LI[1,2] ✉

Human-driven environmental changes threaten the region's dominant trees. China's fengshui forests harbor ancient trees with cultural and ecological significance, yet their climate resilience remains uncertain. By integrating genome and resequencing data, we investigate the structural variants, demographic dynamics, and local adaptation of *Q. gilva*, a threatened East Asian oak vital to fengshui traditions. The genome structural variants are enriched in critical pathways such as stress response and genome maintenance. Divergence of *Q. gilva* populations into Chinese and Japanese lineages occurs during mid-Pliocene climatic shifts. The Chinese lineage carries a higher genetic load, including deleterious mutations in histone deacetylase-associated genes that may impair adaptability. We propose four populations (Changning, Kiyosumi, Tama, and Lianyuan) as preliminary conservation priorities due to their higher genetic diversity and lower genetic load. In contrast, the Jianou population, which contains at least 240 ancient trees, may be susceptible to inbreeding depression as a result of its low genetic diversity and high genetic load. However, enhancing population resilience to future climate change through genetic rescue will depend on further comprehensive genetic and ecological studies.

Rapid environmental changes caused by human activity have accelerated species decline in recent decades, even driving the extinction of some species[1]. The Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) report indicated that *ca.* 25% of organisms are currently at risk of extinction[2]. How to prevent or delay the effects of environmental changes on species at risk of extinction is a leading topic for conservation biologists, with current conservation biology studies focused mainly on endangered species[3]. Widely distributed trees are also increasingly threatened by habitat loss and fragmentation driven by recent human activity, but their threatened status has thus far received little attention or investigation. Forest keystone trees are critical for maintaining ecosystem function and stability; their decline or loss triggers profound, cascading ecological impacts in the ecosystems[4]. For instance, the collapse of *Castanea dentata* populations not only resulted in the loss of a key wildlife food source and valuable timber resources but also fundamentally altered forest structure and composition[5]. Similarly, the mortality of ash trees (*Fraxinus* spp.) significantly reduces biodiversity, disrupts critical ecological processes (e.g., water, nutrient, and carbon cycling), and compromises structural integrity, thereby threatening the long-term health and resilience of affected ecosystems[6,7]. Given that re-establishing populations of keystone species through artificial intervention demands significant investments of time and resources, studying these trees within a conservation and management framework is essential[8].

Species in regions with intense human activity tend to be at greater risk of extinction. That said, local people have also contributed to the long-term protection of some species in their regions, such as trees residing in culturally protected forests[9]. These culturally protected forests often consist of small remnants of forests found worldwide, especially in Africa and Asia, and include sacred groves, fengshui forests, songgye, and church forests[10]. Local people have protected them for centuries for reasons of religion, taboo, and traditional belief[9]. In addition to their cultural value, culturally protected forests are essential for maintaining ecological functions and biodiversity[11,12]. Fengshui forests, also known as fengshui woodlands, are specialized natural forests or small-scale agroforests that are culturally preserved remnants. These forest areas are typically near villages and are commonly found in southern Chinese regions. However, the adaptive

[1]College of Forestry, Central South University of Forestry and Technology, Changsha, China. [2]Yuelushan Laboratory, Changsha, China. [3]Department of Ecosystem Studies, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan. [4]Guangxi Key Laboratory of Superior Trees Resource Cultivation, Guangxi Forestry Research Institute, Nanning, China. [5]Hunan Botanical Garden, Changsha, China. [6]School of Ecology and Environmental Sciences, Yunnan University, Kunming, China. [7]These authors contributed equally: Xiao-Long Jiang, Meng-Xiao WU. ✉e-mail: xiaolongjiang1@gmail.com; lihecsuft@126.com

capacity of these centuries-old trees to persist under projected climate change remains uncertain. Therefore, evaluating their evolutionary potential is critical for informing effective conservation strategies and ensuring their long-term survival[10].

Population genomics utilizes high-throughput sequencing technologies to obtain genome-wide genetic variation data, thereby enabling high-resolution and comprehensive analysis of species' spatial genetic structure and evolutionary history[13]. By integrating analyses of adaptive evolution and genetic load, this approach enhances the development of scientifically informed and objective conservation strategies. A key objective in studying local adaptation is to identify genes undergoing rapid frequency changes under natural selection and to elucidate their biological functions[14]. This is commonly accomplished by detecting genomic signatures of selective sweeps, which are reflected in outlier values of population genetic statistics such as Tajima's D and *F*st. Another widely used method for identifying genomic variants involved in local adaptation is environmental association analysis (EAA). By modeling correlations between allele frequencies and environmental variables, EAA helps pinpoint loci associated with specific adaptive pressures. This approach often employs multivariate statistical techniques such as redundancy analysis (RDA)[15] or latent factor mixed models (LFMM)[16]. In parallel, the study of genetic load—defined as the accumulation of deleterious mutations—has gained increasing relevance in conservation genetics[17]. Genetic load provides critical insights into the severity of inbreeding depression and constraints on adaptive potential, thereby revealing hidden threats to population persistence and supporting proactive conservation planning. For instance, a population genomic study of *Rhododendron vialii* revealed that its survival is threatened not by a general lack of genetic diversity, but by the genomic vulnerability of marginal populations and their unique adaptive potentials[18]. This finding underscores the importance of precision conservation, advocating for a "one-population-one-strategy" framework.

Advances in high-throughput sequencing and the accompanying reduction in costs have also made it increasingly feasible to compare intraspecific genomes and characterize genome-wide structural variants (SVs)[19]. SVs are often associated with key phenotypic traits and play an important role in speciation and rapid environmental adaptation. For example, the rice pan-genome project revealed that extensive SVs are linked to adaptive traits, and that these variations cannot be detected by relying solely on single-nucleotide polymorphisms (SNPs) and a single reference genome[20]. Therefore, integrating the systematic analysis of both SNPs and SVs is essential for a holistic understanding of evolutionary processes in natural populations.

Quercus gilva Blume (Fagaceae) ($2n = 24$), belonging to *Quercus* section *Cyclobalanopsis*, is an ecologically and economically important oak found in the evergreen broad-leaved forests of East Asia. This species is distributed in the subtropical regions of southern China, Japan, and Korea[21]. *Quercus gilva* resides at altitudes ranging from approximately 300 to 1500 m and is highly impacted by human activities, such as the expansion of agricultural land, the construction of roads, and the production of valuable furniture. The number and size of the *Q. gilva* population have declined rapidly due to intensive land use and human disturbance. On the basis of our population survey in China, most current *Q. gilva* populations are located in fengshui forests of villages or around temples, with only limited numbers of individuals (Table 1). In Japan, *Q. gilva* is distributed from Honshu Island to Kyushu Island, with only a few individuals found in native forests[22]. In Korea, *Quercus gilva* is distributed on Jeju Island, with *ca.* 700 individuals[23]. It is listed as an endangered species by the Korean Ministry of Environment. Although the species is not listed as endangered in China or Japan, considering its significant ecological function in local evergreen broad-leaved forests (EBLFs) and its essential role in fengshui forests, *Q. gilva* offers a unique chance to investigate the conservation genetics of trees in fengshui forests.

Several studies in recent years have characterized the genetic patterns and distribution dynamics of *Q. gilva*. Using simple sequence repeats as genetic markers, the genetic diversity of Jeju Island populations was found to be similar to that of populations from southern Japan[24]. Thirteen haplotypes were detected among populations from Japan, mainland China, and Taiwan Island via chloroplast DNA sequences, with no shared haplotypes among the three geographical regions[22]. These two studies provided deep insights into the spatial genetic patterns of *Q. gilva*, although the core populations in mainland China have yet to be sampled. The most recent study using low-coverage whole genome resequencing (WGS) data did not detect significant genetic differentiation among *Q. gilva* populations from mainland China and Japan[25]. This study, however, only calculated the genetic diversity and structure of the species without analyzing the demographic history and local adaptation of the species[25]. Despite the availability of two chromosome-level genomes for *Q. gilva*, the absence of a genomic resource from western Hunan—a central part of its natural range—leaves a critical knowledge gap. This gap hinders a comprehensive understanding of the species' adaptive potential and SV landscape, underscoring the necessity of sequencing a representative from this region.

In this study, we generated a de novo chromosome-level assembly of the *Q. gilva* genome and comprehensively resequenced 55 individuals from 19 populations, including 13 populations (38 individuals) from China and 6 populations (17 individuals) from Japan (Table 1). We aimed to (1) detect the SVs of *Q. gilva* genomes; (2) infer the spatial genetic patterns and demographic history of *Q. gilva* via high-coverage resequencing data, and (3) reveal the local adaptations of *Q. gilva* to environmental changes. Our study provides deep insights to aid in species conservation within fengshui forests under future climate change.

Table 1 | Genetic diversity statistics for 19 *Q. gilva* populations

PopID	Location	θ_π*1000	*He*	Genetic load	F_{ROH}	Dis.
China group						
HT	Huitong, Hunan	1.011	0.2189	0.8655	0.0847	EV
TD	Tongdao, Hunan	1.000	0.2112	0.8665	0.1147	EV
JZ	Jingzhou, Hunan	0.973	0.2153	0.8678	0.1248	EV
CN	Changning, Hunan	1.062	0.1984	0.8694	0.0988	EV
SN	Suining, Hunan	0.889	0.2122	0.8658	0.1143	EV
XS	Xinshao, Hunan	0.877	0.1995	0.8858	0.1024	EV
LY	Lianyuan, Hunan	1.057	0.2032	0.8668	0.1028	EV
PJ	Pingjiang, Hunan	0.881	0.1844	0.8636	0.1217	SA
SZ	Sangzhi, Hunan	0.999	0.2086	0.8641	0.0857	EV
TT	Ningbo, Zhejiang	1.048	0.2026	0.8604	0.0937	SA
PT	Zhoushan, Zhejiang	0.955	0.2082	0.8823	0.1909	SA
MQ	Minqing, Fujian	0.924	0.1919	0.8762	0.1597	EV
JO	Jianou, Fujian	0.865	0.1893	0.8785	0.1880	EV
Japan group						
HIR	Hirai, Japan	0.976	0.2035	0.8561	0.1211	EV
KIY	Kiyosumi, Japan	1.067	0.2230	0.8550	0.1042	EV
OMU	Omura, Japan	1.046	0.2183	0.8593	0.1172	EV
TAM	Tama, Japan	1.058	0.1975	0.8595	0.1027	EV
TAN	Tano, Japan	1.032	0.2156	0.8661	0.1470	EV
TSU	Tsurugara, Japan	0.980	0.2044	0.8587	0.1255	EV

Dis. Distribution, *EV* edge of the village, *SA* scenic area.

Results

Genome assembly and annotation

The *Q. gilva* genome was assembled de novo and annotated via 41.90 gigabytes (Gb) of Illumina short reads, 106.39 Gb of Nanopore long reads (mean length: 21.99 kb), 123.24 Gb of Hi-C data, and 36.57 Gb of RNA-Seq

data (Supplementary Data 1). On the basis of *k*-mer analysis via Illumina reads, the *Q. gilva* genome size and heterozygosity rate were estimated to be approximately 870.20 Mb and 1.87%, respectively (Supplementary Fig. 1). The contig N50 of the *Q. gilva* genome assembled in this study is lower than those reported by Zhou et al. (2022)[26] and Wang et al. (2024)[27]. Genome assembly continuity is generally influenced by factors such as sequencing read length and genomic heterozygosity[28]. Specifically, the relatively lower contig N50 observed in this study may be attributed to the use of sequencing reads with shorter average lengths compared to the 25.23 kb long reads employed by Wang et al., as well as the higher genomic heterozygosity in our sample compared to the 1.16% heterozygosity rate reported by Zhou et al.[26].

The contig-level draft genome assembled from Nanopore long reads was *ca*. 1.13 Gb in size, with 1586 contigs and a contig N50 of 1.20 Mb (Supplementary Data 2). After redundant contigs were filtered, the genome size was 871.96 Mb, with 737 contigs and a contig N50 of 1.60 Mb. A total of 702 contigs with 857.55 Mb of sequence (98.34% of the assembled genome) were anchored on 12 pseudochromosomes, with a scaffold N50 of 71.93 Mb (Fig. 1c, Supplementary Fig. 2, Supplementary Data 2, and 3).

We first used BUSCO and CEGMA to evaluate the quality of the *Q. gilva* genome assembly. BUSCO analyses revealed that the assembled genome contained 95.8% (1547 out of 1614) complete BUSCO genes (Supplementary Data 4), and CEGMA analyses revealed that 83.06% of highly conserved core eukaryotic genes (206 out of 248) were present in the assembled genome (Supplementary Data 5). The mapping rates of the Illumina short reads and Nanopore long reads in the assembled genome were 96.20% and 99.96%, respectively, whereas the mapping rate of the RNA-Seq reads was 100% (Supplementary Data 1). The LAI (LTR Assembly Index) values were uniform across the genome, and the mean LAI value was 16.91 (Supplementary Fig. 3). Genome assemblies with LAI scores between 10 and 20 are of high enough quality to be considered reference genomes[29]. Together, these results show that this *Q. gilva* genome assembly is of high quality.

Repetitive elements accounted for 57.68% (503.11 Mb) of the assembled *Q. gilva* genome (Supplementary Data 6). Transposable elements (TEs) accounted for approximately 54.17% (472.48 Mb) of the genome, whereas tandem repeats (TRs) represented only a minor fraction (1.12%). Long terminal repeat (LTR) retrotransposons were the most abundant category of TEs (41.55%), followed by DNA TEs (5.47%) (Supplementary Data 6). Among the noncoding RNAs, 182 rRNAs, 754 tRNAs, 545 snRNAs, and 161 miRNAs were identified in the *Q. gilva* genome assembly (Supplementary Data 7). A total of 33,818 protein-coding genes were predicted (Supplementary Data 8). The average gene and CDS lengths were 5772 and 1140 bp, respectively, similar to those of other closely related species (Supplementary Data 8). BUSCO analysis revealed that the annotated genome contained 93.31% complete BUSCOs (Supplementary Data 4). Furthermore, 32,887 (97.25%) genes were identified in at least one of the five gene databases, and 7113 (21.03%) were identified in all five databases (Supplementary Data 9, Supplementary Fig. 4).

A total of 3992 SVs were identified between the genome assembled in this study and the reference genome by Zhou et al. (2022)[26], including 1883 duplications, 174 insertions/deletions (indels), 1843 translocations, and 142 inversions. In comparison, 4077 SVs were detected between the genomes by Wang et al. (2024)[27] and the same reference genome, consisting of 1610 duplications, 206 indels, 2,176 translocations, and 85 inversions. The two assembled genomes shared 1283 SVs, of which 566 were duplications, 42 were indels, 634 were translocations, and 41 were inversions (Supplementary Fig. 5). Gene Ontology (GO) enrichment analysis of genes associated with these SVs revealed significant enrichment in biological processes related to oxidative stress, water balance, genome stability, secondary metabolism, signal transduction, protein homeostasis, immune response, and nutrient absorption (Supplementary Data 10).

SNP identification and annotation

Genome resequencing was performed on 55 *Q. gilva* individuals from 19 populations. The amount of raw read sequence obtained from each *Q. gilva* individual ranged from 12.70 to 44.87 Gb, and filtered read sequences ranged from 11.42 to 41.90 Gb (Supplementary Data 11). The average depth and coverage of the filtered reads in the *Q. gilva* genome were 24.62 and 83.12%, respectively (Supplementary Data 11). A total of 93.81 million raw SNPs were identified, and 3,107,177 high-quality SNPs were retained after filtering. The numbers of SNPs identified in the intergenic, intronic, and CDS regions were 2,088,755 (67.22%), 605,412 (19.48%), and 159,357 (5.13%), respectively. The remaining SNPs were distributed upstream or downstream of open reading frames (8.14%) and in splicing regions (0.02%). The SNPs in the CDS regions included 77,388 synonymous, 80,112 nonsynonymous, 193 stop-loss, 1548 stop-gain, and 116 SNPs from unknown sites (Supplementary Data 12).

Genetic diversity and structure

Following linkage disequilibrium (LD) pruning, a final set of 523,681 SNPs was utilized to investigate the genetic structure of *Q. gilva*. ADMIXTURE analysis identified the optimal population structure at $K = 2$ (minimum cross-validation error; Supplementary Fig. 6), with clusters corresponding to the geographic origins of the sampled individuals. The Chinese group ($n = 38$) and Japanese group ($n = 17$) exhibited clear genetic differentiation (Fig. 2a; Supplementary Fig. 7). Kinship analysis further supported this divergence, revealing significantly greater genetic distances between the Japanese and Chinese groups than among individuals within each group (Supplementary Fig. 8). At $K = 3$, two additional subpopulations (JO and PJ) were distinctly segregated from other Chinese populations. Concordant results were obtained from principal component analysis (PCA), maximum likelihood (ML) phylogenetic tree, and neighbor-joining (NJ) network analyses, which collectively delineated two primary lineages corresponding to the China and Japan groups (Fig. 2). Notably, both the ML tree and NJ network revealed substructure within the Chinese populations, with distinct genetic clusters separating the western and eastern subregions (Fig. 2). These findings demonstrate that *Q. gilva* harbors two major lineages, reflecting its geographic distribution in Japan and China, while a finer-scale substructure exists within the Chinese population, particularly distinguishing the western and eastern subregions. The LD decay reached half of the maximum average r^2 at a distance of 10.90 kb ($r^2 = 0.17$) for the China group and 14.10 kb ($r^2 = 0.18$) for the Japan group (Supplementary Fig. 9).

The pairwise F_{ST} values of the 19 *Q. gilva* populations ranged from 0.20 (OMU vs. TAN) to 0.37 (XS vs. TAM) (Supplementary Data 13). A significant positive correlation was observed between genetic and geographic distances across all populations ($R^2 = 0.13$, $p < 0.001$); however, this pattern was not evident within the group populations ($R^2 = 0.01$, $p > 0.05$; Supplementary Fig. 10a). Conversely, genetic distance displayed no significant association with environmental divergence, regardless of whether the population pairs originated from the same or distinct groups (between groups: $R^2 = -0.009$, $p > 0.05$; within groups: $R^2 = -0.007$, $p > 0.05$; Supplementary Fig. 10b). The nucleotide diversity (θ_π) of the *Q. gilva* populations varied from 0.87×10^{-3} (JO) to 1.07×10^{-3} (KIY), with a mean value of 0.98×10^{-3} (Table 1). At the group level, the Chinese group presented greater genetic diversity (China: $\theta_\pi = 1.04 \times 10^{-3}$, $\theta_W = 9.24 \times 10^{-4}$; Japan: $\theta_\pi = 1.03 \times 10^{-3}$, $\theta_W = 8.78 \times 10^{-4}$, $p < 0.01$) and Tajima's D values (China: Tajima's $D = 0.91$; Japan: Tajima's $D = 0.55$, $p < 0.001$) than did the Japanese group (Fig. 2d, Supplementary Fig. 11a).

A total of 118,374 and 50,841 runs of homozygosity (ROHs) were identified in the China and Japan groups, respectively (Supplementary Data 14). The majority of ROHs (over 95% in each subgroup) ranged in length from 10-100 kb, while the longest ROHs (>500 kb) accounted for only 0.25% of the total ROH content. The fraction of the genome covered by ROHs (FROH) was similar between the China and Japan groups across different ROH length categories (Supplementary Fig. 11b and Supplementary Data 14).

Demographic history

Two individuals from each group were used to infer the *Ne* dynamics via the PSMC' method (Fig. 3a). The mean depths for samples from the China and

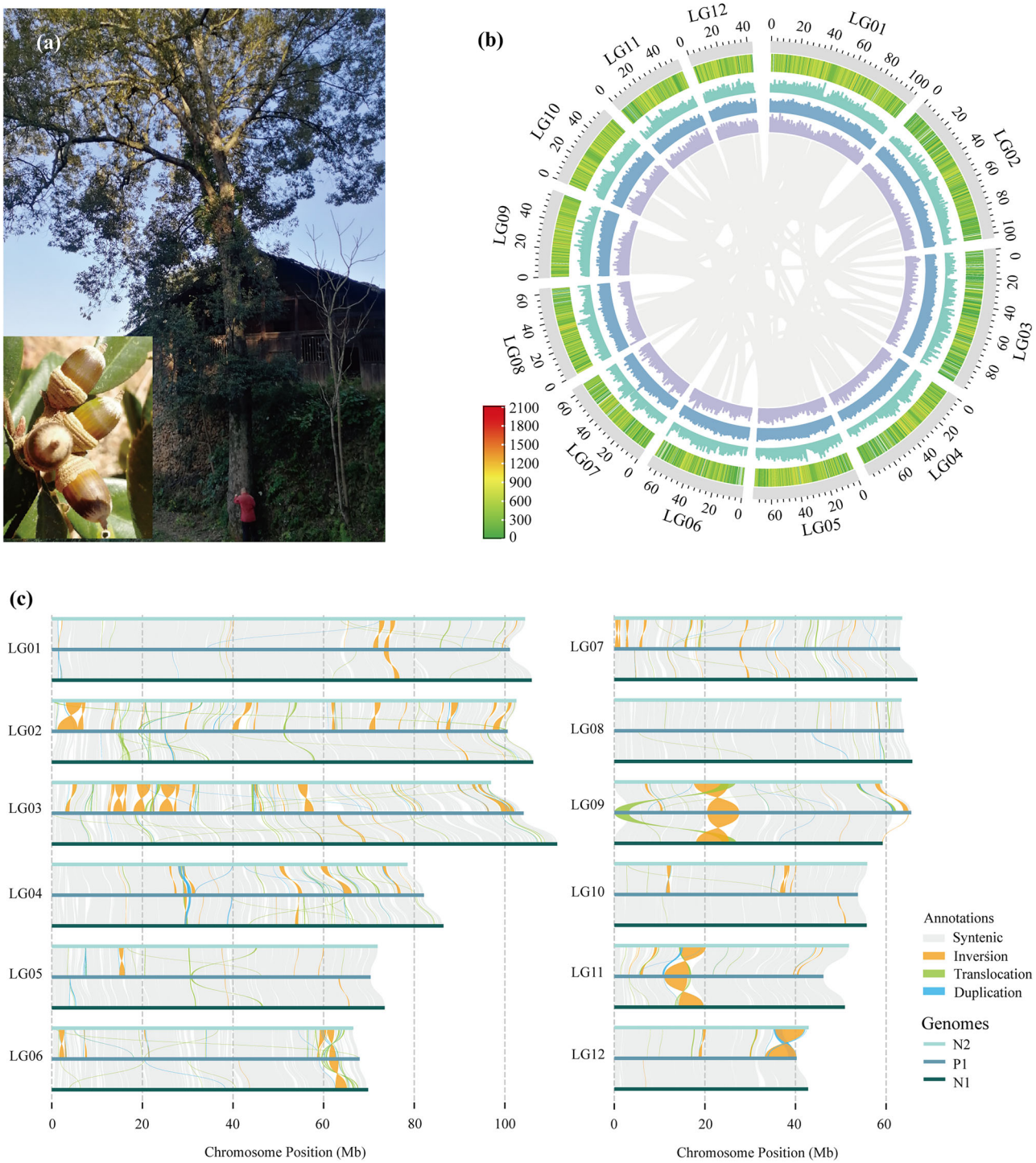

Fig. 1 | Features of the *Q. gilva* genome. a Morphological characteristics of *Q. gilva* (Photographed in Xinshao County, Hunan Province, by Xiao-long Jiang). **b** The *Q. gilva* genome circle plot (from outside to inside: 1. pseudochromosome number; 2. SNP density; 3. Nucleotide diversity (θ_π); 4. guanine-cytosine (GC) content; 5. gene density; 6. intragenome collinear blocks). **c** Chromosomal inversions in *Q. gilva* genome. The horizontal lines indicate the *Q. gilva* chromosomes. P1 represents chromosomes from a previously published genome assembly Zhou et al. (2022)[26], N1 represents chromosomes from a previously published genome assembly Wang et al. (2024)[27], and N2 represents chromosomes from the assembly in this study. The orange lines indicate inverted regions, and the gray lines indicate syntenic regions.

Japan groups were 48.65 and 32.08, respectively. After site masking and filtering, 1,760,516 and 1,661,335 heterozygous sites from Chinese and Japanese individuals, respectively, were used for the PSMC' analysis. The PSMC' results indicated that the populations from China and Japan had slightly different demographic histories. The *Ne* of both groups began to expand *ca*. 3.60 Ma, but the magnitude and duration of expansion in the Japanese group were greater than those in the Chinese group. SMC++ analysis of *Q. gilva* revealed a progressive decline in effective population size from ~1 Ma to 20 ka. Contrasting this long-term trend, a subsequent expansion since 20 ka has resulted in a current effective population size in Japan that is approximately twice the size of that in China.

Divergence and gene flow between the two groups were also evaluated via a composite likelihood approach via five models involving different gene flow scenarios (Supplementary Fig. 13). On the basis of the AIC values,

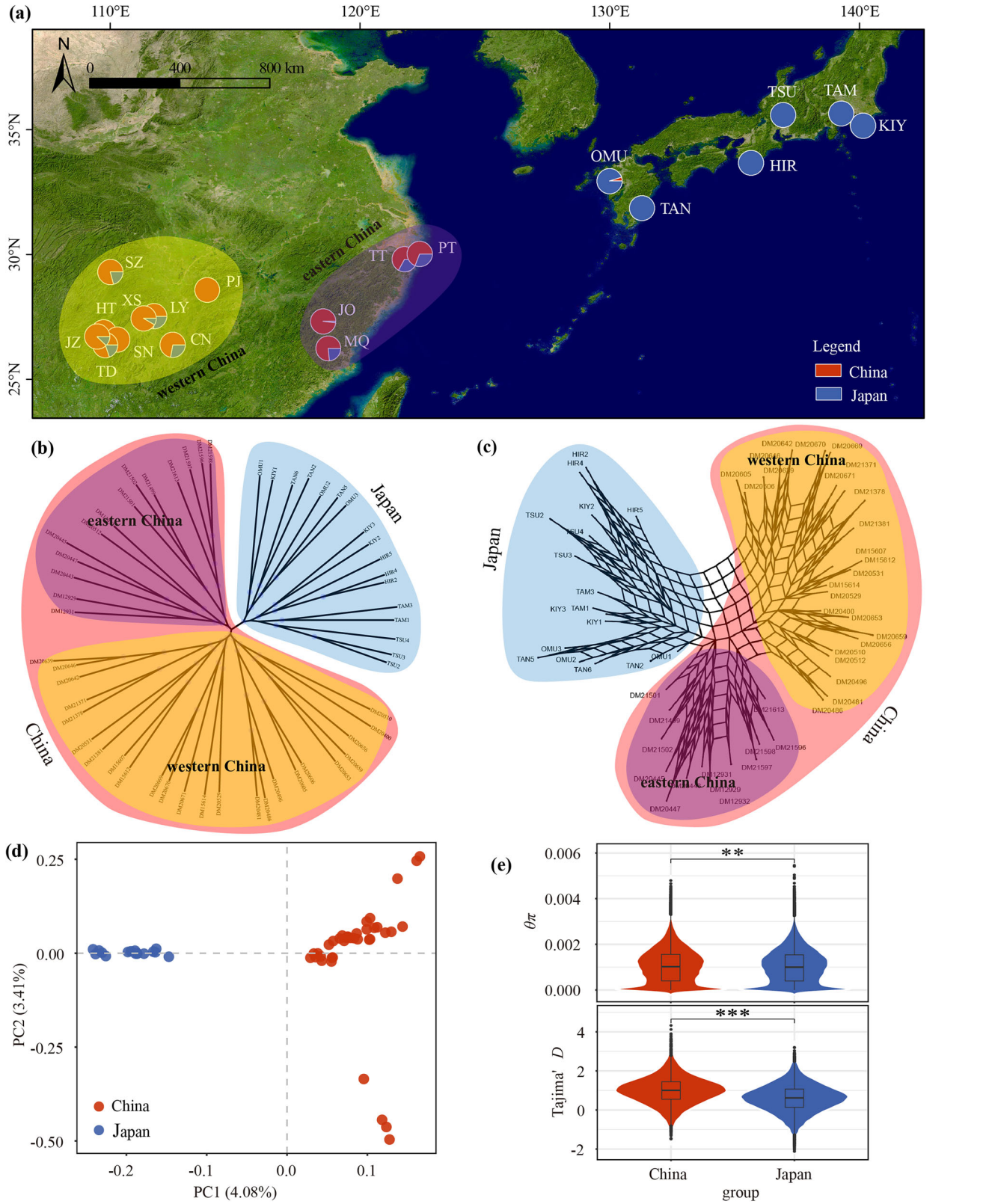

Fig. 2 | Geographic distribution and population structure of *Q. gilva*. **a** ADMIXTURE clustering of the *Q. gilva* accessions when $K = 2$. The pie charts represent locations where samples were collected, and the proportion of color within each pie chart represents ancestral components. **b** A maximum likelihood phylogenetic tree of all the *Q. gilva* samples constructed via whole-genome SNP data. **c** Neighbor-joining network of all the *Q. gilva* samples constructed via whole-genome SNP data. **d** Principal component analysis of the *Q. gilva* samples. **e** Significant divergence of the θ_π and Tajima's *D* values of accessions from the China ($n = 38$ biologically independent samples) and Japan ($n = 17$ biologically independent samples) groups.

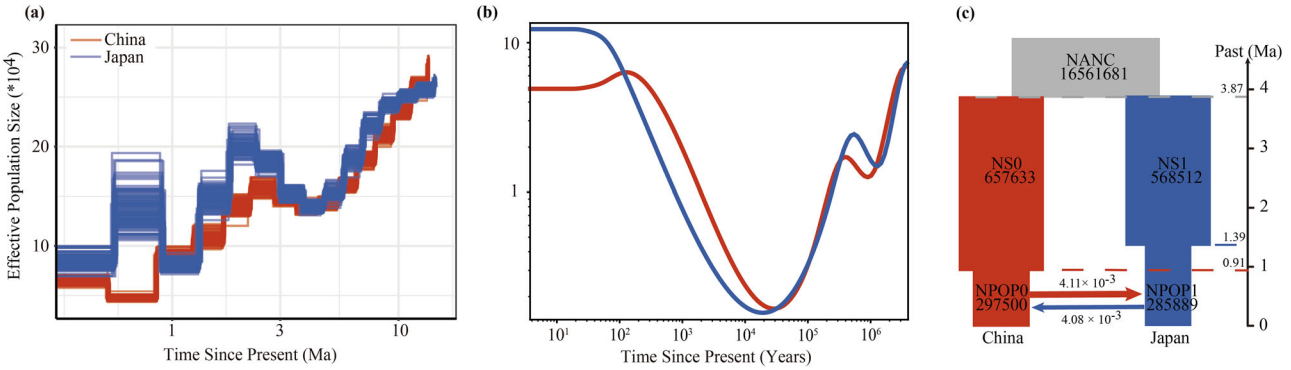

Fig. 3 | Demographic history of the China and Japan groups. a, **b** Changes in the effective population size for both geographic groups over time as inferred by the PSMC' and SMC++ models, respectively. **c** Graphical summary of the best-fitting demographic model implied by fastsimcoal2. Widths show the relative effective population sizes. The arrows indicate the direction and average number of migrants per generation in the two groups. The point estimates and 95% confidence intervals of the demographic parameters are shown in Supplementary Data 16.

Model 4, with recent gene flow between the two groups, was the best fit (Supplementary Fig. 12, 13 and Supplementary Data 15). This model suggested that the divergence between the China and Japan groups occurred in the Pliocene period (3.87 Ma, 95% CI: 3.69–4.07 Ma) (Fig. 3b, Supplementary Data 16). The effective population sizes (*Ne*) of the China lineage were slightly greater than those of the Japan lineage after their divergence ($Ne_{\text{-China}} = 657{,}633$, CI: 601,351–714,467; $Ne_{\text{-Japan}} = 568{,}512$, CI: 513,706–619,544). Both lineages subsequently experienced a bottleneck event during the early to middle Pleistocene (*c.* 1.39 Ma–0.91 Ma), and *Ne* was reduced by *ca.* 50–55%. Furthermore, low levels of recent gene flow were detected between the two lineages (China to Japan: 4.11×10^{-3}, Japan to China: 4.08×10^{-3}) (Supplementary Data 16).

Local adaptation

Regions under selection in the *Q. gilva* genome were identified via the distributions of the log2 π ratio and z score (F_{ST}) in the genome. A total of 259 nonoverlapping genomic segments were identified as regions under selection (Fig. 4, Supplementary Fig. 14, and Supplementary Data 17). Among them, 152 (708 genes) and 107 (734 genes) regions were identified in the China and Japan groups, respectively (Supplementary Data 18), with six regions overlapping between the two groups. The GO enrichment analysis for genes under selection indicated that the top 20 GO enrichment terms for the China group included metal ion transport, ubiquitin-conjugating enzyme binding, and protein homo-oligomerization (Supplementary Data 18, Fig. 15a). The top GO enrichment terms for the Japan group included proline metabolism, regulation, and intracellular signal transduction (Supplementary Data 18, Supplementary Fig. 15b).

LFMM analysis revealed 3745 SNPs significantly associated with one or more of eight environmental variables (Supplementary Fig. 16). Nearly 90% of the adaptive SNPs (3345) were associated with one of the eight environmental factors. The remaining SNPs showed associations with two (364 SNPs) or three (36 SNPs) environmental factors. The number of adaptive SNPs associated with temperature exceeded that associated with precipitation (Supplementary Fig. 16). A total of 1,039 genes included at least one adaptive SNP. GO enrichment analysis revealed that 243 genes were significantly enriched for 110 GO terms ($p < 0.05$) (Supplementary Data 19), and the top 20 GO terms included the regulation of stomatal movement, RNA metabolic process, macromolecule methylation, phosphatidic acid metabolic process, and drought recovery (Supplementary Fig. 17). Our analysis identified 168 adaptive SNPs that were located within selective sweep regions. These SNPs intersected with 46 genes. A GO enrichment analysis of these genes revealed significant enrichment for 59 GO terms, demonstrating a meaningful biological correlation between selection signals and environmental adaptation (Supplementary Data 20 and Fig. 18).

Genetic load

The genetic loads of the China and Japan groups were estimated by identifying putative deleterious loss of function (LOF) and missense variants (Fig. 5). The ratio of the number of homozygous sites to the combined number of homozygous and heterozygous sites in the China group was significantly greater than that in the Japan group (0.873 vs. 0.865 for LOF, $p < 0.01$; 0.870 vs. 0.859 for missense mutations, $p < 0.001$) (Fig. 5a). We further used the Grantham score (≥150) as an index to predict the derived deleterious missense variants in the coding regions of both groups, which also indicated a greater genetic load in the China group (0.874 ± 0.01 in the China group, 0.862 ± 0.006 in the Japan group; Wilcoxon test, $p < 0.001$) (Fig. 5a). Compared with the Chinese group, the Japanese group had a greater number of genes associated with homozygous deleterious mutations in ROHs (Fig. 5b). Therefore, inbreeding is not a major cause of increased genetic load in populations. Genes affected by homozygous putative deleterious LOF mutations were significantly enriched for GO and KEGG categories, which are primarily involved in the stress response and metabolism (Supplementary Data 21). For example, LOF mutations were enriched for histone deacetylase activity (Gene Ontology term, GO0004407).

Discussion

Structural variants in the *Q. gilva* genome

This study reports a chromosome-level genome assembly of *Q. gilva*, derived from a wild-sourced specimen collected in western Hunan, China—a central region within the species' natural distribution. By integrating the present assembly with two previously published genomes—one from a cultivated individual and another from a natural forest in Fujian—we identified an extensive catalog of SVs in the *Q. gilva* genome. Using the Zhou et al. (2022)[26] genome as a reference, 3992 SVs were detected in the current assembly, whereas 4077 were identified in the Wang et al. (2024)[27] assembly. A total of 1283 SVs were common to both assemblies, highlighting conserved genomic regions that may possess critical functional significance. In both datasets, duplications (1883 and 1610) and translocations (1843 and 2176) constituted the predominant SV types, indicating considerable genomic plasticity potentially associated with transposable element activity or polyploidization events. Such SVs are recognized drivers of gene family expansion and functional diversification. For instance, translocations have been demonstrated to contribute to phenotypic diversity and environmental adaptation in rice[30].

GO enrichment analysis of SV-associated genes revealed significant enrichment in key biological processes. Notably, terms related to oxidative stress, including "hydrogen peroxide transmembrane transport" and "cellular response to reactive oxygen species," suggest a role for SVs in modulating redox homeostasis. This is consistent with findings in *Arabidopsis*, where aquaporins such as PIPs mediate hydrogen peroxide transport and participate in stress signaling[31]. Similarly, enrichment in "water channel

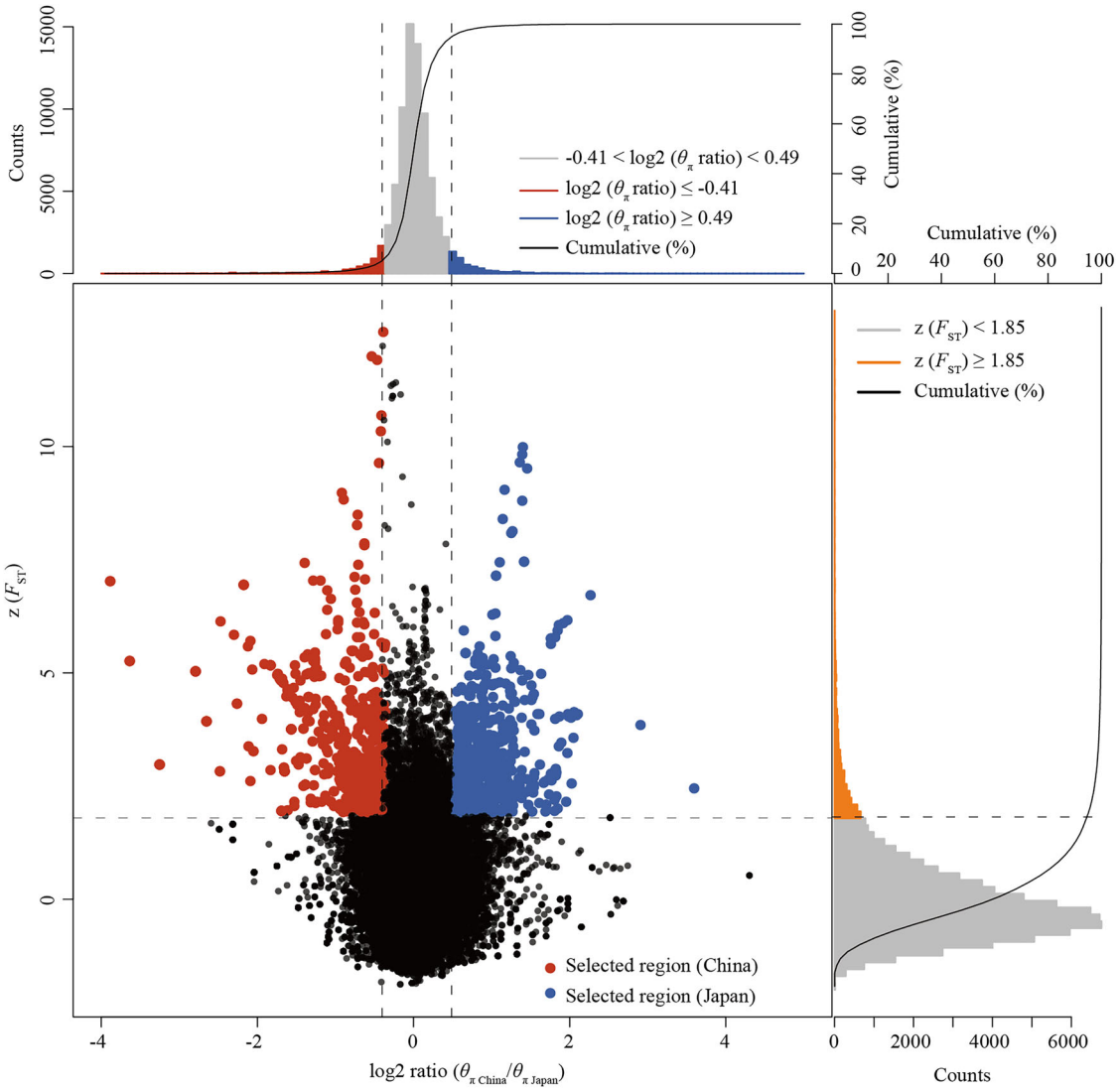

Fig. 4 | Genomic regions with signals of selective sweeps based on z(F_{ST}) and the log2 ratio ($\theta_{\pi\ China}/\theta_{\pi\ Japan}$) in the China and Japan groups. The data points located to the left and right of the left and right vertical dashed lines, respectively (corresponding to the 5% left and right tails of the empirical log2 ratio ($\theta_{\pi\ China}/\theta_{\pi\ Japan}$) distribution, where the π ratios were -0.41 and 0.49, respectively), and above the horizontal dashed line (the 5% right tail of the empirical z(F_{ST}) distribution, where z(F_{ST}) was 1.85) were identified as regions under selection for the China group (red points) and Japan group (blue points). The subplot on the top reflects the distribution of log2(θ_π ratio), and the right subplot reflects the z(F_{ST}) distribution.

activity" and "water transmembrane transporter activity" may reflect adaptations to environmental stress, as evidenced by studies in loblolly pine (*Pinus taeda*) and bald cypress (*Taxodium distichum*), where aquaporin activity underpins physiological responses to drought, flooding, and salinity[32]. SVs were also linked to "DNA mismatch repair" and "double-stranded telomeric DNA binding," indicating potential effects on genome integrity. Defects in these processes can synergistically induce telomere crisis and SVs, thereby promoting genomic instability[33]. Enrichment in secondary metabolic processes, particularly "coumarin biosynthesis," points to a potential role of SVs in chemical defense, aligning with findings in *Datura wrightii*, where tandem duplications drive the expansion of defense-related gene families involved in secondary metabolism and herbivore resistance[34]. Finally, enrichment in immune-related terms (e.g., "response to chitin") and nutrient transport (e.g., "iron ion transmembrane transporter activity") underscores the broad influence of SVs on biotic and abiotic interactions, as illustrated by polymorphisms in iron transporters affecting nutrient uptake in *Arabidopsis*[35]. Collectively, these findings highlight SVs that may shape phenotypic diversity through the modulation of genes involved in critical pathways such as stress response and genome maintenance.

Genetic patterns and demographic dynamics

Two distinct genetic lineages were detected among the *Q. gilva* populations via phylogenetic analysis, genetic structure analysis, and PCA, which corresponded to individuals from the mainland (China lineage) and continental islands (Japan lineage). Unlike oceanic islands, which are consistently separated from the mainland, continental islands are connected to adjacent continents through land bridges during periods when the climate is colder[36]. Land bridges promote species migration from the continent to the island, affecting the flora community and the genetic structure of the species[36]. PSMC' and Fastsimcoal2 analyses indicated that the divergence of the two *Q. gilva* lineages occurred during the mid-Pliocene (*ca.* 3.9 Ma). Species or lineage differentiation is generally accompanied by a sharp decrease in the effective population size, which has been observed in *Populus davidiana*[37] and *Cercidiphyllum japonicum*[38]. Our study on *Q. gilva* revealed a similar scenario. The PSMC' model revealed that the ancestral population of *Q. gilva* initially had a high effective population size (>270k) at 14 Ma, which then decreased sharply until it reached the maximum bottleneck before the divergence of the lineages (*ca.* 3.9 Ma). Radiolarian data from the Japan Sea indicated that deep cooling of the Japan Sea likely occurred

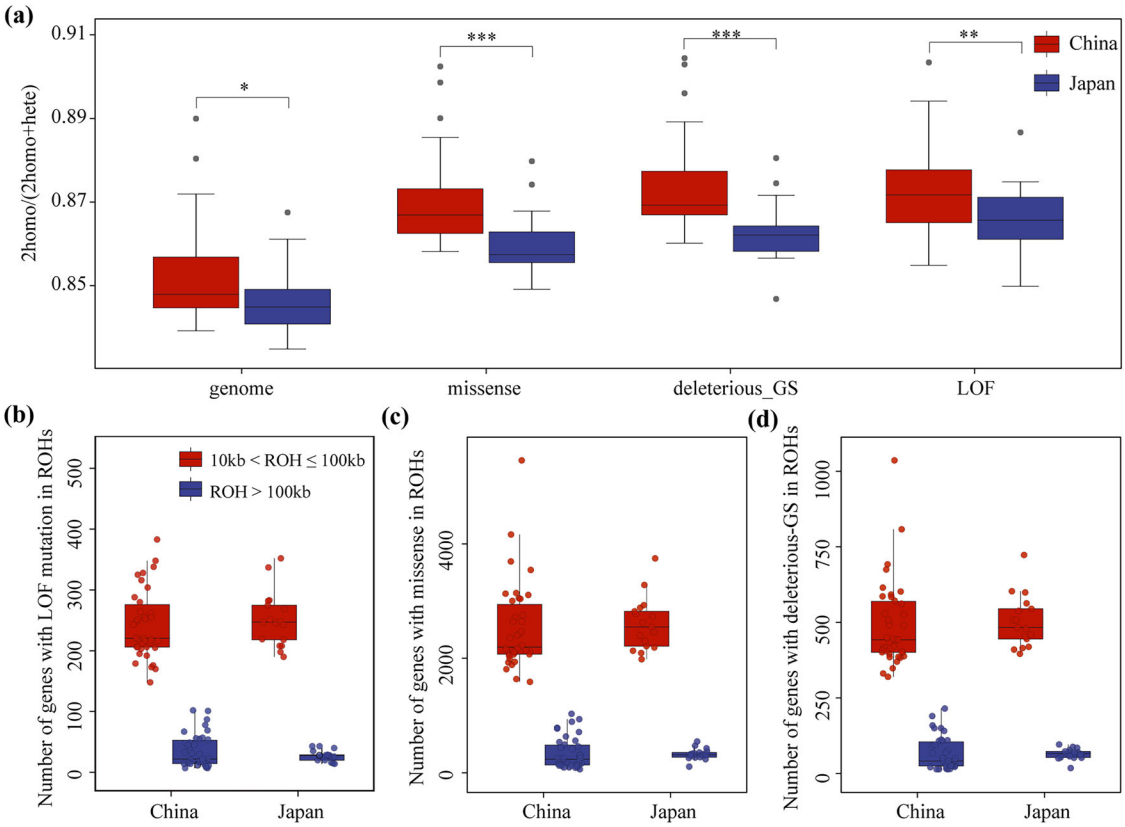

Fig. 5 | Genetic load of *Q. gilva*. **a** Genetic loads of the China ($n = 38$ biologically independent samples) and Japan ($n = 17$ biologically independent samples) groups. The symbols on the top of the error bars indicate the significance of the differences between the data for the China and Japan groups. p values of pairwise comparisons were calculated via the Wilcoxon test: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; ns, not significant. The number of genes with homozygous deleterious mutations in ROHs, including (**b**) LOF mutations, **c** missense mutations, and **d** deleterious-Grantham score mutations, in the China and Japan groups.

between *ca.* 5.0 and 3.8 Ma, followed by a warmer period[39]. The estimated divergence time of the *Q. gilva* lineages is consistent with the warming climate in Japan and surrounding areas, which suggests that climate fluctuations in the Pliocene may have contributed to its lineage differentiation. Moreover, this inference supported the significant correlation between population genetic differentiation and geographical distance. Significantly different genetic structures for populations from China and Japan have also been detected in other subtropical species, such as *Platycrater arguta*[40], *Euptelea pleiosperma*[41], and *Ligularia hodgsonii*[42]. However, the timescales of intraspecies lineage differentiation have varied. The divergence time between the Chinese lineage and Japanese lineage of *Q. gilva* was more ancient than that of *Platycrater arguta*[40] and more recent than that of *Euptelea pleiosperma*[41], but similar to that of *Cercidiphyllum japonicum*[38]. Moreover, genetic differentiation has not been detected among populations of the dominant subtropical species *Kalopanax septemlobus* across China and the Japanese islands[43]. Evolutionary history and biological characteristics, such as pollen and seed dispersal abilities, influence genetic structure[44]. The intricate evolutionary dynamics of Japanese flora suggest that a combination of geography, the environment, and local adaptation influences the spatial genetic patterns of species in these regions.

Following the initial divergence of *Q. gilva*, the Japanese Islands are known to have been reconnected to the Eurasian continent via land bridges during multiple glacial periods[45]. These recurrent connections likely facilitated the recent gene flow observed between the Chinese and Japanese lineages of *Q. gilva*. For example, a land bridge formed during the Last Glacial Maximum (LGM, ca. 22 ka) has been shown to facilitate gene flow in *Machilus thunbergii*[46]. Our SMC++ analysis indicated a substantial increase in the effective population sizes of both Chinese and Japanese *Q. gilva* lineages starting around 20 ka, a period closely associated with the LGM. This temporal correspondence suggests that post-LGM climate warming may have promoted population expansion. These results align with phylogeographic studies of other dominant East Asian evergreen broadleaved forest species, such as *Q. glauca*[47] and *Castanopsis eyrei*[48], collectively supporting the hypothesis that postglacial warming fostered range and demographic expansions in these taxa.

Within the Chinese lineage, a weak genetic structure, comprising western and eastern groups, was identified. This west-east genetic differentiation has been consistently observed in other evergreen species like *Castanopsis eyrei*[48] and *Castanopsis fargesii*[49], as well as in deciduous trees, including *Liriodendron chinense*[50] and *Fagus engleriana*[51]. The recurrent emergence of this genetic pattern across multiple species implies that allopatric genetic divergence has been a key process shaping the spatial genetic structure of plant species in subtropical China.

Genetic load and local adaptation

The F_{ROH} of *Q. gilva* (0.036 for Chinese individuals and 0.035 for Japanese individuals) was far lower than the F_{ROH} estimated in other organisms when the minimum homology length was 100 kb. For example, F_{ROH} was estimated to be 0.31–0.45 for *Ostrya rehderiana*[52], 0.30 for *Dipteronia sinensis*[53], and 0.23 for *Acer yangbiense*[54]. Even when 10 kb was used as the minimum homology length, the F_{ROH} for *Q. gilva* (0.125 for Chinese individuals and 0.121 for Japanese individuals) was still smaller than those of other species when 100 kb was used as the threshold. High levels of heterozygosity in *Q. gilva* could explain this result. The estimated heterozygosity in the *Q. gilva* genome was 1.87%, which was significantly greater than estimates from *Ostrya rehderiana* (0.26%)[52], *Dipteronia sinensis* (0.49%)[53], and *Acer yangbiense* (0.37%)[54]. High genome heterozygosity can result in a reduction in the lengths of runs of homozygosity. Inbreeding will increase the genetic load of populations[55]. However, the number of genes with LOF mutations in ROHs was similar between individuals from the Chinese and Japanese

lineages, although a significantly greater number of deleterious mutations were found in individuals from China than in individuals from Japan. This pattern aligns with the purging of highly deleterious recessive variants in the historically smaller and isolated population[56]. In such populations, an increase in inbreeding elevates the probability that strongly deleterious recessive mutations appear in homozygous states within ROHs, leading to their removal by natural selection (i.e., purging)[57]. As a consequence, the number of the most severe homozygous LOF mutations remains low. The higher overall load of deleterious alleles in the Chinese lineage likely stems from the accumulation of less severe deleterious mutations—often in heterozygous form—due to enhanced effects of genetic drift.

The extensive selective sweeps and signals of positive selection detected in the *Q. gilva* genome, particularly for genes related to species development and the abiotic stress response, indicated that adaptation was likely driven by the local environment. We also identified several homozygous putative deleterious mutations in the *Q. gilva* genome. Some of these homozygous putative deleterious mutations are associated with species development, metabolism, and stress response functions. For example, GO analysis indicated that five LOF mutations were enriched for histone deacetylase activity. Histone deacetylases (HDACs) are a supergene family with essential roles in the structural modification of chromosomes and the regulation of gene expression[58,59]. Changes in HDAC-related gene expression can affect plant growth, development, and other phenotypic changes[60,61]. Studies have indicated that downregulation of HDAC-related gene expression can lead to premature plant senescence[62], morphological defects in leaves and flowers[63], and termination of seed development[64,65]. HDACs also respond to adverse stresses and regulate growth and development[61,66]. For example, overexpression of the HDAC-related gene *PtHDT902* in poplar plants inhibits adventitious root formation and reduces salt tolerance[67]. Studies on the model plant *Arabidopsis* have shown that long-term low-temperature stress can induce the expression of the HDAC-related gene *HDA6*[68] and that mutation of this gene can increase sensitivity to low-temperature stress. In addition, HDACs respond to drought stress by regulating the expression levels of drought-responsive genes such as *RD29A*, *RD20*, and *RAP2.4*[69,70]. Therefore, LOF mutations in genes linked to the histone deacetylation process may inhibit adaptability by disrupting normal gene expression, potentially resulting in population size reduction, endangerment, or even extinction.

Implications for conservation and management

Based on genome-wide variation, we identified two distinct geographic lineages in China and Japan. The Chinese lineage overall displays higher genetic diversity and Tajima's D values compared to the Japanese lineage. In the Japanese lineage, the KIY and TAM populations show relatively high genetic diversity and low genetic load. Within the Chinese lineage, the CN and LY populations exhibit higher genetic diversity than other Chinese populations. Given that maintaining or improving genetic diversity is considered important for species conservation, these four populations may be viewed as preliminary candidates for prioritization in conservation planning. By contrast, the JO population, which is the largest in China (covering approximately 16.13 hm^2 with at least 240 individuals >50 cm in diameter), showed lower genetic diversity and higher genetic load in our dataset, a pattern potentially attributable to inbreeding depression. Its U-shaped diameter distribution[71] indicates that inbreeding depression could be a contributing factor to poor midstory recruitment in this population. In such cases, genetic rescue through the introduction of variants from populations with higher genetic diversity and lower genetic load is often discussed as a potential strategy[72,73]. However, this study is constrained by limited within-population sample sizes, which may affect the precision of genetic diversity and genetic load estimates. The findings of this study do not provide a sufficient basis for decisions regarding the introduction of non-local genotypes. Any such intervention would require more comprehensive genetic and ecological assessments—such as expanded sampling across populations and individuals, as well as controlled transplantation trials to evaluate adaptation and survival under projected climatic conditions.

Study limitations

Beyond the constraint of limited sample size per population, two additional limitations also warrant consideration in future studies. First, although a chromosome-level assembly was achieved, its lack of telomere-to-telomere completeness means that some SVs—especially those in repetitive regions—may reflect assembly artifacts rather than genuine biological variation. Resolving these uncertainties will require future sequencing using ultra-long-read technologies, such as those from Oxford Nanopore, to close remaining gaps and improve contiguity. Second, while GO enrichment analyses suggest functional roles for identified SVs and deleterious mutations in stress response and development, these associations remain speculative without experimental validation. For example, the functional redundancy within the HDAC gene family, whose members are distributed across different chromosomes, means that the phenotypic impact of a LOF mutation in any single HDAC gene requires further validation. Future work incorporating gene expression profiling, CRISPR-based genome editing, or common garden assays will be essential to verify the phenotypic effects and adaptive significance of these genetic variants.

Materials and methods

Genome sequencing and assembly

DNA from a healthy *Q. gilva* individual from Xinshao County, Hunan Province, China (111.34°E, 27.44°N; Fig. 1), was collected for reference genome sequencing. Long-read sequencing of DNA from this individual was performed on a PromethION sequencer. Libraries with sequence lengths greater than 10 kb were prepared and sequenced via the standard protocol from Oxford Nanopore Technologies. To estimate the genome size of *Q. gilva*, we constructed an Illumina library with an insert size of 400 bp. A Hi-C library was then generated following the procedures of Lieberman-Aiden et al.[74] to anchor the assembled genome onto chromosomes. To predict the structure and function of genes in the *Q. gilva* genome, four RNA-Seq libraries were constructed and sequenced from leaves, seeds, roots, and stem tissues. All Illumina libraries were sequenced on the Illumina NovaSeq 6000 platform. To obtain high-quality Illumina sequence reads, we filtered the raw data via Trimmomatic v0.39[75] with the following parameters: "LEADING:3, TRAILING:3, SLIDINGWINDOW:4:15, MINLEN:120".

The genome size of *Q. gilva* was estimated by analyzing the 17 *k*-mer distribution via the KmerFreq_AR module in SOAPdenovo2 v2.04-r241[76]. Genome heterozygosity was calculated via pIRS v2.0[77]. To reduce the read error rate, self-correction of Nanopore reads was performed via Canu v2.1.1[78]. Genome assembly was performed via SMARTdenovo v1.0[79] with default parameters. The Illumina sequencing data were used to correct the draft assembly genome. After three rounds of correction of the genome via Pilon v1.23[80], redundant contigs were removed via Purge_Haplotigs v1.1.1[81]. To anchor assembly contigs to chromosomes, the filtered Hi-C data were mapped to the assembled genome via Bowtie2 v2.3.2[82]. Reads uniquely mapped via HIC-PRO v.2.8.1[83] were then used to cluster, order, and orient scaffolds onto chromosomes via LACHESIS[84].

Four methods were applied to assess the quality of the assembled genome. First, the completeness of the genome was estimated via BUSCO v5.3.2[85] with the Embryophyta_odb10 database (https://busco.ezlab.org). Second, the accuracy and integrity of the core genes in the *Q. gilva* genome were evaluated via CEGMA v2[86]. Third, the coverage and mapping rates were calculated by realigning the Illumina, Nanopore, and RNA-Seq reads onto the genome via BWA v0.7.12[87], Minimap2 v2.24[88], and STAR v2.7.3a (Dobin et al., 2013), respectively. Finally, the integrity of the genome was evaluated via the LAI[29].

Genome annotation and collinearity analysis

Repetitive sequences include tandem repeats (TRs) and transposable elements (TEs). TRs were identified via GMATA v2.2[89] and Tandem Repeat Finder v4.07[90]. TEs were identified via both homologous alignment and de novo search methods. MITEs were identified via MITE-hunter[91]. LTR_FINDER v1.07[92] and LTRharvest[93] were used to identify LTRs. When

these results were integrated via LTR_retriever[94] to obtain an LTR retrotransposon library, the de novo identification and classification of repetitive sequences were performed via RepeatModeler v1.0.11[95] and TEclass v2.1.3[96], respectively. Finally, repetitive sequences in the *Q. gilva* genome were detected via RepeatMasker v1.331[97] through a combination of the above libraries and the Repbase database[98]. Noncoding RNAs (ncRNAs) were identified by aligning the *Q. gilva* genome to the Rfam[99] database via Infernal v1.1.2[100]. The tRNAs were detected via tRNAscan-SE v2.0[101], and the rRNAs were identified via RNAmmer v1.2[102].

Protein-coding genes were predicted via a combination of transcript-based, homology-based, and de novo strategies. For transcript-based prediction, filtered RNA-Seq reads were aligned to the *Q. gilva* genome via STAR v2.7.3a[103], and transcripts were reconstructed via StringTie v1.3.4 d[104]. Gene structures were obtained by aligning nonredundant transcripts to the genome via PASA v2.3.3[105], and GeneMark-ST v5.1[106] was used to identify open reading frames. For homology-based prediction, the protein sequences of closely related species (Supplementary Data 8) were aligned to the *Q. gilva* genome via GeMoMa v1.6.1[107]. GeneMark-ST v5.1[106] and AUGUSTUS v3.3.1[108] were used for de novo prediction. Finally, all the prediction results were integrated via EVidenceModeler v1.1.1[105]. TransposonPSI v1.0.0[109] was used to remove genes containing transposable elements. The predicted genes were functionally annotated via five publicly available databases: NCBI nonredundant (NR), Kyoto Encyclopedia of Genes and Genomes (KEGG), eukaryotic orthologous groups (KOGs), Gene Ontology (GO), and Swiss-Prot. To predict Pfam domains and GO annotations, InterProScan v5.32-71.0[110] was used to search against the InterPro database for conserved protein domains and functional sites. For other databases, BLASTP v2.7.1[111] was used to search for homology with the following parameters: "-evalue 1e-5, -max_target_seqs 1".

The SVs in the *Q. gilva* genome were identified via SyRI. First, the *Q. gilva* genome assembly in this study and from Wang et al. (2024)[27] was aligned with the Zhou et al. (2020) genome[26] via Minimap2 v2.24[88] with the parameters "-ax asm10 -eqx". The SVs between genomes were subsequently identified via SyRI v1.6[112]. Insertions, deletions, copy number variants, inversions (maximum size of 1 Mb), and translocations (minimum size of 50 bp) were retained, and the SVs containing "N" sequences were removed. Overlapping regions among the SVs were merged. Subsequently, a custom annotation library was generated using GffRead v0.12.7[113], and the functional effects of the variants were annotated with ANNOVAR[114]. GO enrichment analysis was conducted to investigate the functional roles of genes associated with SVs using clusterProfiler v3.18.0.

Resequencing and SNP calling

A total of 55 individuals from 19 populations that span the main geographic distribution of *Q. gilva* were collected and used for WGS (Fig. 2a and Table 1). For each population, individuals at least 50 m apart from each other were sampled. Genomic DNA was extracted via a DNAsecure Plant Kit (Tiangen, Beijing, China). Libraries with an insert size of 400 bp were constructed for each individual and sequenced via the Illumina NovaSeq 6000 platform in paired-end 150 bp mode. The raw reads were filtered via Trimmomatic v0.39[75] with the following parameters: "LEADING:3, TRAILING:3, SLIDINGWINDOW:4:15, MINLEN:120".

SNPs were identified from the resequencing data via Sentieon[115]. Filtered data were first mapped to the *Q. gilva* genome assembly via BWA v0.7.12[87] with default settings. Uniquely mapped and properly paired reads were retained for further analyses. After the alignment files were converted from SAM to BAM format via SAMtools v1.6[116], the alignment files were sorted via the *util* mode in Sentieon. Duplicate sequence removal, realignment, and SNP calling were then performed via *driver* mode in Sentieon. To reduce SNP false positives, we removed SNPs that met any of the following criteria: (1) site depth less than one-third or greater than twice the mean depth; (2) quality score < 30 and minor allele frequency (MAF) < 0.05; (3) allelic SNPs greater than 2 (--max-alleles 2 -min-alleles 2); (4) SNPs missing in at least one individual (-max-missing 1); or (5) SNPs within five bp of indels.

To assess the impact of genetic variation on gene structure and function, we utilized the built-in program annotate_variation.pl from the ANNOVAR package to annotate SNPs[114]. The SNPs were categorized into exonic, intronic, intergenic, upstream, downstream, or splicing site regions. On the basis of their impact on the protein-coding sequence, the SNPs in exonic regions were further classified as synonymous, nonsynonymous, stop-gain, or stop-loss.

Genetic diversity and structure

We used several methods to estimate the spatial genetic patterns of *Q. gilva*. A pruned SNP set generated via PLINK v1.90[117] with the parameters "-indep-pairwise 1000 kb 1 0.5" was used for genetic structure analysis. The kinship between individuals was calculated via PLINK with the parameters "-make-rel square". The population genetic structure was calculated via ADMIXTURE v.1.3.0[118] with default parameters and the number of clusters (K) set from 2--10. The K value was determined via tenfold cross-validation. Principal component analysis (PCA) was performed via PLINK v1.90[117] with the "*-pca*" parameter and visualized via the R package "ggplot2"[119]. The maximum likelihood (ML) tree was constructed via IQ-tree v2.0.3[120], and the best model was determined via the Bayesian information criterion (BIC). Tree reliability was estimated via the ultrafast bootstrap approach (UFboot) with 1000 replicates. A phylogenetic network of *Q. gilva* individuals was inferred via SplitsTree v6.3.35[121].

Nucleotide diversity (θ_π), Watterson's estimator (θ_W), and Tajima's D statistics were calculated for each group via VCFtools v0.1.16[122] with 100 kb nonoverlapping windows. The θ_π for each population was also calculated using the same parameters. Linkage disequilibrium (LD) was analyzed for each group via PopLDdecay v3.42[123]. The average r^2 value was measured in a 300 kb window size. LD decay measures the distance at which Pearson's correlation efficiency (r^2) decreases to half of the maximum. To minimize bias due to sample size variation, an identical number of individuals ($n = 17$) was used in each group.

The effects of geographic and environmental factors on the population's genetic differentiation were estimated via the Mantel test. Genetic differentiation (F_{ST}) was calculated via the R package hierfstat[124]. The genetic distances between populations were calculated via the formula $D_{genetic} = F_{ST}/(1\text{-}F_{ST})$. Geographic distances among populations were calculated via the Vincenty (ellipsoid) method in the R package geosphere v1.5-18[125]. For environmental distance, we first downloaded 19 bioclimatic variables from WorldClim (https://www.worldclim.org). PCA of eight bioclimatic variables with correlation coefficients less than 0.8 ($|r| < 0.8$, Supplementary Fig. 19) was performed in R. The Euclidean distance of the first principal component (PC1) between populations represents the environmental distance. The Mantel test was performed via the R package vegan v2.6-4[126], with significance determined via 9999 permutations.

Runs of homozygosity (ROH) were estimated via PLINK v1.90[117] with the following parameters: -homozyg-kb 10, -homozyg-window-het 1, -homozyg-window-snp 20, and -homozyg-window-missing 5. We classified ROH into three categories on the basis of length: <100 kb, 100–500 kb, and >500 kb. The inbreeding coefficient based on ROH (F_{ROH}) was calculated via the following formula[127]: $F_{ROH} = L_{ROH}/L_{AUTO}$, where L_{ROH} is the length of autosomes covered by ROHs and L_{AUTO} is the length of autosomes covered by SNPs, which was 857.55 Mb in our study.

Demographic history

Changes in the effective population size of *Q. gilva* were calculated via the pairwise sequentially Markovian coalescent (PSMC') model. One sample was selected for each group. After the filtered reads were mapped to the genome via BWA v0.7.12[87], duplicate reads were removed via Picard v2.1.1 (https://broadinstitute.github.io/picard/), and the reads around the indels were realigned via GATK v4.3.0[128]. SNP calling was performed via BCFtools v1.10.2[116]. To reduce mapping errors, we used VCFtools v0.1.16 to mask gaps, repetitive sequences, and indels in the genome. Sites with depths less than one-third or greater than twice the mean depth were removed for subsequent analysis. MSMC v2.0.0[129] was used to perform PSMC' modeling

with default parameters, except that the number of iterations was set to 200. A mutation rate per nucleotide per generation (μ) of 1.01×10^{-8}, estimated from *Q. lobata*, and a generation time (g) of 50 years were used to convert the number of generations into years[130].

We inferred the recent demographic history of *Q. gilva* using SMC++ v1.15.5[131]. To minimize the potential influences of uneven sample sizes and genetic heterozygosity on the analysis, a total of 35 individuals were selected, including 18 from China and 17 from Japan. The analysis procedure was as follows: First, the VCF file was converted into the SMC++ input format using the *vcf2smc* command. Then, the *estimate* command was employed to infer the population history, with the mutation rate set to $\mu = 1.01 \times 10^{-8}$ per generation and the generation time set to 50 years. Finally, the *plot* command was used to visualize the inference results.

We also used a coalescent simulation-based method performed by Fastsimcoal2 v2.6.0.3[132] to infer the divergence time and demographic history of the two groups of *Q. gilva*. The two-dimensional joint site frequency spectrum (2D-SFS) was calculated via easySFS[133]. Five different models were formulated to simulate the past population history of *Q. gilva* (Supplementary Fig. 12). Each model was run independently 100 times to calculate the composite likelihood with 100,000 simulations. The model exhibiting the largest Akaike's weight was chosen as the best model[134] for subsequent analyses. The same mutation rate (μ) and generation time used for PSMC' were used here. Parametric confidence intervals were obtained via 100 parametric bootstraps, with 50 independent runs on simulated data under the best model in each bootstrap.

Local adaptation

To identify potential regions under selection in the *Q. gilva* genome, we used VCFtools v0.1.16[122] to calculate average values of θ_π and F_{ST} across the genome for the two lineages with sliding windows of 100 kb. SNPs with a MAF less than 5% were removed from this analysis. The windows with the top 5% of the log2 π ratio and z score (F_{ST}) values were chosen as the regions under selection. We also used a univariate latent-factor linear mixed model (LFMM) implemented in the R package LEA v2.8.0[135] to test for correlations between environmental variables and SNP frequencies. Eight uncorrelated bioclimatic variables from WorldClim were chosen as environmental variables (Supplementary Fig. 19). On the basis of the number of ancestry clusters inferred with ADMIXTURE, we ran LFMM with two latent factors to account for the genetic structure of *Q. gilva*. Five independent MCMCs were run for each environmental variable via 3000 iterations as burn-in, followed by 100,000 iterations. SNPs with correlated p values (false discovery rate, FDR) less than 0.01 were considered adaptive SNPs. To investigate their functional roles, we performed GO enrichment analysis on genes from adaptive loci identified via both LFMM and selective sweep, employing clusterProfiler v3.18.0[136].

Genetic load

Deleterious mutations are expected to disrupt gene function or reduce species viability[137]. We used loss of function (LOF) and missense mutations to quantify the deleterious load. In accordance with the strategy of Feng and Hu et al.[138,139], major homozygous alleles in the *Q. gilva* genome (>50% of the individuals) were used to represent the ancestral genotypes. SNP annotation was performed via SnpEff v.4.3t[140]. The LOF mutations include premature stop codons (nonsense) and splice-site-disrupting single-nucleotide variants (SNVs). GO and KEGG functional enrichment of genes with homozygous LOF mutations was performed via eggNOG-mapper v2.1.12[141] and the R package clusterProfiler v3.18.0[136]. The missense mutations were identified based on the Grantham score (GS), which measures the physical and chemical properties of amino acid changes. A Grantham score greater than 150 is considered radical or deleterious[142]. The deleterious load was determined by calculating the ratio of homozygous sites (two per site) to the total number of sites (one per site) for each category[137,139].

Statistics and reproducibility

We generated a chromosome-scale genome assembly for *Q. gilva* by integrating Nanopore, Hi-C, and RNA-seq data. Sequencing included single biological replicates for Nanopore and Hi-C libraries, and four tissue-specific replicates (leaf, seed, root, stem) for RNA-seq. For population genomic analyses, we resequenced 55 individuals from 19 natural populations (2–4 per population), with sampled individuals spaced ≥50 m apart within populations to ensure independence.

Statistical analyses followed established bioinformatic workflows and population genetic models as described in Methods. To mitigate bias from unequal sample sizes, we standardized individual numbers: linkage disequilibrium decay analysis included 17 individuals each from China and Japan, and SMC++ analysis used 18 from China and 17 from Japan. Putatively selected regions were defined as outliers in the top 5% of both log2 π ratio and z score (F_{ST}). For genotype–environment association analysis (LFMM), SNPs with FDR < 0.01 were considered significant.

All raw sequencing data are publicly archived to support reproducibility (Data Availability). Computational reproducibility was further ensured through standardized bioinformatics pipelines, explicit documentation of key parameters—including read filtering, variant calling thresholds, and demographic model settings—and the use of default software parameters unless otherwise specified. Software versions, command-line arguments, and reference databases are detailed in Methods.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw sequencing data for *Q. gilva* used in this study—including Nanopore (Accession: CRR1169155), Hi-C (Accession: CRR1169154), transcriptome (Accession: CRR1169156—CRR1169159), and resequencing data (Accession: CRX1068264—CRX1068318)—have been deposited in the Genome Sequence Archive (GSA, https://ngdc.cncb.ac.cn/gsa/) database under BioProject ID PRJCA026221. The assembled genome sequences were deposited in the NCBI (https://static.pubmed.gov/portal/portal.fcgi/) database under Accession JBVBUK000000000 and in the Genome Warehouse (GWH, https://ngdc.cncb.ac.cn/gwh/) under Accession GWHEUVJ00000000. The source data underlying the figures are available on Figshare[143].

Received: 9 April 2025; Accepted: 2 March 2026;
Published online: 13 March 2026

References

1. Román-Palacios, C. & Wiens, J. J. Recent responses to climate change reveal the drivers of species extinction and survival. *Proc. Natl. Acad. Sci.* **117**, 4211–4217 (2020).
2. Watson, R. et al. Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. *IPBES Secretariat*, 22–47 (2019).
3. Jing, Z., Cheng, K., Shu, H., Ma, Y. & Liu, P. Whole genome resequencing approach for conservation biology of endangered plants. *Biodivers. Sci.* **31**, 22679 (2023).
4. Ellison, A. M. et al. Loss of foundation species: consequences for the structure and dynamics of forested ecosystems. *Front. Ecol. Environ.* **3**, 479–486 (2005).
5. Dalgleish, H. J. & Swihart, R. K. American Chestnut Past and Future: Implications of Restoration for Resource Pulses and Consumer Populations of Eastern U.S. Forests. *Restor. Ecol.* **20**, 490–497 (2012).
6. Klooster, W. S. et al. Ash (*Fraxinus* spp.) mortality, regeneration, and seed bank dynamics in mixed hardwood forests following invasion by emerald ash borer (*Agrilus planipennis*). *Biol. Invasions* **16**, 859–873 (2014).
7. Herms, D. A. & McCullough, D. G. Emerald Ash Borer Invasion of North America: History, Biology, Ecology, Impacts, and Management. *Annu. Rev. Entomol.* **59**, 13–30 (2014).

8. Boonman, C. C. F. et al. More than 17,000 tree species are at risk from rapid global change. *Nat. Commun.* **15**, 166 (2024).
9. Hu, L., Li, Z., Liao, W. B. & Fan, Q. Values of village fengshui forest patches in biodiversity conservation in the Pearl River Delta, China. *Biol. Conserv* **144**, 1553–1559 (2011).
10. Wang, Z. F. et al. Endangered but genetically stable—*Erythrophleum fordii* within Feng Shui woodlands in suburbanized villages. *Ecol. Evol.* **9**, 10950–10963 (2019).
11. Coggins, C. et al. in *Cultural and Spiritual Significance of Nature in Protected Areas* 225–236 (2018).
12. Avtzis, D. et al. Quantifying the conservation value of Sacred Natural Sites. *Biol. Conserv.* **222**, 95–103 (2018).
13. Holliday, J. A. et al. Advances in ecological genomics in forest trees and applications to genetic resources conservation and breeding. *Mol. Ecol.* **26**, 706–717 (2017).
14. Holliday, J. A., Ritland, K. & Aitken, S. N. Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (*Picea sitchensis*). *N. Phytol.* **188**, 501–514 (2010).
15. Capblancq, T., Luu, K., Blum, M. G. B. & Bazin, E. Evaluation of redundancy analysis to identify signatures of local adaptation. *Mol. Ecol. Resour.* **18**, 1223–1233 (2018).
16. Frichot, E., Schoville, S. D., Bouchard, G. & François, O. Testing for Associations between Loci and Environmental Gradients Using Latent Factor Mixed Models. *Mol. Biol. Evol.* **30**, 1687–1699 (2013).
17. Bertorelle, G. et al. Genetic load: genomic estimates and applications in non-model animals. *Nat. Rev. Genet.* **23**, 492–503 (2022).
18. Chang, Y. et al. Conservation genomics of a threatened subtropical *Rhododendron* species highlights the distinct conservation actions required in marginal and admixed populations. *Plant J.* **122**, e70175 (2025).
19. Jiao, C. et al. Pan-genome bridges wheat structural variations with habitat and breeding. *Nature* **637**, 384–393 (2025).
20. Qin, P. et al. Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell* **184**, 3542–3558.e3516 (2021).
21. Lin, L., Jiang, X. L., Guo, K. Q., Byrne, A. & Deng, M. Climate change impacts the distribution of *Quercus* section *Cyclobalanopsis* (Fagaceae), a keystone lineage in East Asian evergreen broadleaved forests. *Plant Divers.* **45**, 552–568 (2023).
22. Sugiura, N., Tang, D., Kurokochi, H., Saito, Y. & Ide, Y. Genetic structure of *Quercus gilva* Blume in Japan as revealed by chloroplast DNA sequences. *Botany* **93**, 873–880 (2015).
23. Hyun, H. J., Song, K. M., Choi, H. S. & Chan-Soo, K. Dynamics and Distribution of *Quercus gilva* Blume Population in Korea. *J. Environ. Prot. Ecol.* **24**, 385–392 (2014).
24. Han, E. K. et al. A Disjunctive Marginal Edge of Evergreen Broad-Leaved Oak (*Quercus gilva*) in East Asia: The High Genetic Distinctiveness and Unusual Diversity of Jeju Island Populations and Insight into a Massive, Independent Postglacial Colonization. *Genes* **11**, 1114 (2020).
25. Song, Y. G. et al. Population survey combined with genomic-wide genetic variation unravels the endangered status of *Quercus gilva*. *Diversity* **15**, 230 (2023).
26. Zhou, X. et al. A chromosome-scale genome assembly of *Quercus gilva*: Insights into the evolution of *Quercus* section *Cyclobalanopsis* (Fagaceae). *Front. Plant Sci.* **13**, 1012277 (2022).
27. Wang, T.-R. et al. Genomic insights into ecological adaptation of oaks revealed by phylogenomic analysis of multiple species. *Plant Divers.* **47**, 53–67 (2025).
28. Kajitani, R. et al. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* **24**, 1384–1395 (2014).
29. Ou, S., Chen, J. & Jiang, N. Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Res.* **46**, e126 (2018).
30. Song, X. & Cao, X. Transposon-mediated epigenetic regulation contributes to phenotypic diversity and environmental adaptation in rice. *Curr. Opin. Plant Biol.* **36**, 111–118 (2017).
31. Bienert, G. P. et al. Specific aquaporins facilitate the diffusion of hydrogen peroxide across membranes. *J. Biol. Chem.* **282**, 1183–1192 (2007).
32. Domec, J.-C. et al. Aquaporins, and not changes in root structure, provide new insights into physiological responses to drought, flooding, and salinity. *J. Exp. Bot.* **72**, 4489–4501 (2021).
33. Sakellariou, D. et al. MutSβ regulates G4-associated telomeric R-loops to maintain telomere integrity in ALT cancer cells. *Cell Rep.* **39**, 110602 (2022).
34. Goldberg, J. K. et al. A de novo long-read genome assembly of the sacred datura plant (*Datura wrightii*) reveals a role of tandem gene duplications in the evolution of herbivore-defense response. *BMC Genomics* **25**, 15 (2024).
35. Vert, G. et al. IRT1, an Arabidopsis transporter essential for iron uptake from the soil and for plant growth. *Plant Cell* **14**, 1223–1233 (2002).
36. Jiang, X. L., Gardner, E. M., Meng, H. H., Deng, M. & Xu, G. B. Land bridges in the Pleistocene contributed to flora assembly on the continental islands of South China: Insights from the evolutionary history of *Quercus championii*. *Mol. Phylogenet. Evol.* **132**, 36–45 (2019).
37. Hou, Z., Li, A. & Zhang, J. Genetic architecture, demographic history, and genomic differentiation of *Populus davidiana* revealed by whole-genome resequencing. *Evolut. Appl.* **13**, 2582–2596 (2020).
38. Zhu, S. et al. Genomic insights on the contribution of balancing selection and local adaptation to the long-term survival of a widespread living fossil tree, *Cercidiphyllum japonicum*. *N. Phytol.* **228**, 1674–1689 (2020).
39. Matsuzaki, K. M., Itaki, T., Tada, R. & Kamikuri, S. I. Paleoceanographic history of the Japan Sea over the last 9.5 million years inferred from radiolarian assemblages (IODP Expedition 346 Sites U1425 and U1430). *Prog. Earth Planet Sci.* **5**, 1–33 (2018).
40. Qi, X. S., Yuan, N., Comes, H. P., Sakaguchi, S. & Qiu, Y. X. A. strong 'filter'effect of the East China Sea land bridge for East Asia's temperate plant species: inferences from molecular phylogeography and ecological niche modelling of *Platycrater arguta* (Hydrangeaceae). *BMC Evol. Biol.* **14**, 1–16 (2014).
41. Cao, Y. N. et al. Genomic insights into historical population dynamics, local adaptation, and climate change vulnerability of the East Asian Tertiary relict *Euptelea* (Eupteleaceae). *Evol. Appl.* **13**, 2038–2055 (2020).
42. Wang, J. F., Gong, X., Chiang, Y. C. & Kuroda, C. Phylogenetic patterns and disjunct distribution in *Ligularia hodgsonii* Hook.(Asteraceae). *J. Biogeogr.* **40**, 1741–1754 (2013).
43. Sakaguchi, S. et al. Climate oscillation during the Quaternary associated with landscape heterogeneity promoted allopatric lineage divergence of a temperate tree *Kalopanax septemlobus* (Araliaceae) in East Asia. *Mol. Ecol.* **21**, 3823–3838 (2012).
44. Xu, L. L. et al. Different rates of pollen and seed gene flow cause branch-length and geographic cytonuclear discordance within Asian butternuts. *N. Phytol.* **232**, 388–403 (2021).
45. Kimura, M. Land connections between Eurasian continent and Japanese Islands–related to human migration. *Migr. Diffus.* **4**, 14–33 (2003).
46. Jiang, K. et al. Shifting roles of the East China Sea in the phylogeography of red nanmu in East Asia. *J. Biogeogr.* **48**, 2486–2501 (2021).
47. Song, Y. et al. Ensemble species distribution modeling and multilocus phylogeography provide insight into the spatial genetic patterns and distribution dynamics of a keystone forest species, *Quercus glauca*. *BMC Plant Biol.* **24**, 168 (2024).
48. Shi, M.-M., Michalski, S. G., Welk, E., Chen, X.-Y. & Durka, W. Phylogeography of a widespread Asian subtropical tree: genetic

east–west differentiation and climate envelope modelling suggest multiple glacial refugia. *J. Biogeogr.* **41**, 1710–1720 (2014).
49. Sun, Y., Hu, H., Huang, H. & Vargas-Mendoza, C. F. Chloroplast diversity and population differentiation of *Castanopsis fargesii* (Fagaceae): a dominant tree species in evergreen broad-leaved forest of subtropical China. *Tree Genet. Genomes* **10**, 1531–1539 (2014).
50. Yang, A. et al. New insight into the phylogeographic pattern of *Liriodendron chinense* (Magnoliaceae) revealed by chloroplast DNA: east-west lineage split and genetic mixture within western subtropical China. *PeerJ* **7**, e6355 (2019).
51. Lei, M. et al. Molecular phylogeography of *Fagus engleriana* (Fagaceae) in subtropical China: limited admixture among multiple refugia. *Tree Genet. Genomes* **8**, 1203–1212 (2012).
52. Yang, Y. et al. Genomic effects of population collapse in a critically endangered ironwood tree *Ostrya rehderiana*. *Nat. Commun.* **9**, 5449 (2018).
53. Feng, Y. et al. Genome sequences and population genomics provide insights into the demographic history, inbreeding, and mutation load of two 'living fossil'tree species of *Dipteronia*. *Plant J.* **117**, 177–192 (2024).
54. Ma, Y. et al. Demographic history and identification of threats revealed by population genomic analysis provide insights into conservation for an endangered maple. *Mol. Ecol.* **31**, 767–779 (2022).
55. Hohenlohe, P. A., Funk, W. C. & Rajora, O. P. Population genomics for wildlife conservation and management. *Mol. Ecol.* **30**, 62–82 (2021).
56. Charlesworth, D. & Willis, J. H. The genetics of inbreeding depression. *Nat. Rev. Genet.* **10**, 783–796 (2009).
57. Glémin, S. How are deleterious mutations purged? Drift versus nonrandom mating. *Evolution* **57**, 2678–2687 (2003).
58. Struhl, K. Histone acetylation and transcriptional regulatory mechanisms. *Gene Dev.* **12**, 599–606 (1998).
59. Grunstein, M. Histone acetylation in chromatin structure and transcription. *Nature* **389**, 349–352 (1997).
60. Dong, Y. R., Du, J. X., Chen, C. J., Zhao, D. X. & WANG, Z. H. Histone deacetylases and its roles in plants. *Biotechnol. Bull.* **32**, 44–49 (2016).
61. Ma, X., Lv, S., Zhang, C. & Yang, C. Histone deacetylases and their functions in plants. *Plant Cell Rep.* **32**, 465–478 (2013).
62. Chen, Z. J. & Tian, L. Roles of dynamic and reversible histone acetylation in plant development and polyploidy. *Biochim Biophys. Acta* **1769**, 295–307 (2007).
63. Rossi, V. et al. Maize histone deacetylase hda101 is involved in plant development, gene transcription, and sequence-specific modulation of histone modification of genes and repeats. *Plant Cell* **19**, 1145–1162 (2007).
64. Wu, K., Tian, L., Malik, K., Brown, D. & Miki, B. Functional analysis of *HD2* histone deacetylase homologues in *Arabidopsis thaliana*. *Plant J.* **22**, 19–27 (2000).
65. Zhou, C. et al. Expression and function of HD2-type histone deacetylases in *Arabidopsis* development. *Plant J.* **38**, 715–724 (2004).
66. Kumar, V., Thakur, J. K. & Prasad, M. Histone acetylation dynamics regulating plant development and stress responses. *Cell Mol. Life Sci.* **78**, 4467–4486 (2021).
67. Ma, X. et al. Histone deacetylase gene *PtHDT902* modifies adventitious root formation and negatively regulates salt stress tolerance in poplar. *Plant Sci.* **290**, 110301 (2020).
68. To, T. K. et al. *Arabidopsis HDA6* is required for freezing tolerance. *Biochem. Biophys. Res. Commun.* **406**, 414–419 (2011).
69. Nguyen, N. H., Jung, C. & Cheong, J. J. Chromatin remodeling for the transcription of type 2C protein phosphatase genes in response to salt stress. *Plant Physiol. Biochem.* **141**, 325–331 (2019).
70. Kim, J. M. et al. Transition of chromatin status during the process of recovery from drought stress in Arabidopsis thaliana. *Plant Cell Physiol.* **53**, 847–856 (2012).
71. Qin, Z. K. Research on phenotypic traits of germplasm resources and superior tree selection of *Quercus gilva*. Central South University of Forestry and Technology 25-28 (2023).
72. Supple, M. A. & Shapiro, B. Conservation of biodiversity in the genomics era. *Genome Biol.* **19**, 131 (2018).
73. Whiteley, A. R., Fitzpatrick, S. W., Funk, W. C. & Tallmon, D. A. Genetic rescue to the rescue. *Trends Ecol. Evol.* **30**, 42–49 (2015).
74. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
75. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
76. Luo, R. et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18 (2012).
77. Hu, X. et al. pIRS: Profile-based Illumina pair-end reads simulator. *Bioinformatics* **28**, 1533–1535 (2012).
78. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
79. Liu, H., Wu, S., Li, A. & Ruan, J. SMARTdenovo: a de novo assembler using long noisy reads. *Gigabyte* **2021**, 1–9 (2021).
80. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963 (2014).
81. Roach, M. J., Schmidt, S. A. & Borneman, A. R. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *Bmc Bioinforma.* **19**, 1–10 (2018).
82. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
83. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 1–11 (2015).
84. Burton, J. N. et al. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* **31**, 1119–1125 (2013).
85. 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 (2015).
86. Parra, G., Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061–1067 (2007).
87. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
88. Li, H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* **32**, 2103–2110 (2016).
89. Wang, X. & Wang, L. GMATA: an integrated software package for genome-scale SSR mining, marker development and viewing. *Front Plant Sci.* **7**, 1350 (2016).
90. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
91. Han, Y. & Wessler, S. R. MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res.* **38**, e199 (2010).
92. Zhao, X. & Hao, W. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res* **35**, W265–W268 (2007).
93. Ellinghaus, D., Kurtz, S. & Willhoeft, U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinforma.* **9**, 1–14 (2008).
94. Ou, S. & Jiang, N. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat Retrotransposons. *Plant Physiol.* **176**, 1410–1422 (2018).

95. Flynn, J. M. et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci.* **117**, 9451–9457 (2020).
96. Abrusán, G., Grundmann, N., DeMester, L. & Makalowski, W. TEclass—a tool for automated classification of unknown eukaryotic transposable elements. *Bioinformatics* **25**, 1329–1330 (2009).
97. Bedell, J. A., Korf, I. & Gish, W. MaskerAid: a performance enhancement to RepeatMasker. *Bioinformatics* **16**, 1040–1041 (2000).
98. Jurka, J. et al. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res.* **110**, 462–467 (2005).
99. Griffiths-Jones, S. et al. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33**, D121–D124 (2005).
100. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).
101. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).
102. Lagesen, K. et al. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* **35**, 3100–3108 (2007).
103. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
104. Kovaka, S. et al. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* **20**, 1–13 (2019).
105. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, 1–22 (2008).
106. Tang, S., Lomsadze, A. & Borodovsky, M. Identification of protein coding regions in RNA transcripts. *Nucleic Acids Res.* **43**, e78 (2015).
107. Keilwagen, J. et al. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res.* **44**, e89 (2016).
108. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008).
109. Urasaki, N. et al. Draft genome sequence of bitter gourd (*Momordica charantia*), a vegetable and medicinal plant in tropical and subtropical regions. *DNA Res.* **24**, 51–58 (2017).
110. Zdobnov, E. M. & Apweiler, R. InterProScan–an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847–848 (2001).
111. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinforma.* **10**, 1–9 (2009).
112. Goel, M., Sun, H., Jiao, W. B. & Schneeberger, K. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol.* **20**, 1–13 (2019).
113. Pertea, G. & Pertea, M. GFF utilities: GffRead and GffCompare. *F1000Res* **9**, 304 (2020).
114. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
115. Freed, D., Aldana, R., Weber, J. A. & Edwards, J. S. The Sentieon Genomics Tools–A fast and accurate solution to variant calling from next-generation sequence data. *BioRxiv*, https://doi.org/10.1101/115717 (2017).
116. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
117. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
118. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
119. Ginestet, C. ggplot2: elegant graphics for data analysis. *J. R. Stat. Soc. A* **174**, 245–246 (2011).
120. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
121. Huson, D. H. & Bryant, D. The SplitsTree App: interactive analysis and visualization using phylogenetic trees and networks. *Nat. Methods* **21**, 1773–1774 (2024).
122. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
123. Zhang, C., Dong, S. S., Xu, J. Y., He, W. M. & Yang, T. L. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **35**, 1786–1788 (2018).
124. Goudet, J. & Jombart, T. hierfstat: estimation and tests of hierarchical F-statistics. *R package version 0.04-22* **10** (2015).
125. Hijmans, R. J., Williams, E. & Vennes, C. Geosphere: Spherical Trigonometry. *R. Package Version* **1**, 5–10 (2019).
126. Oksanen, J. et al. Package 'vegan'. *Community Ecol.* **2**, 1–295 (2019).
127. McQuillan, R. et al. Runs of homozygosity in European populations. *Am. J. Hum. Genet* **83**, 359–372 (2008).
128. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
129. Schiffels, S. & Wang, K. MSMC and MSMC2: The Multiple Sequentially Markovian Coalescent. *Methods Mol. Biol.* **2090**, 147–166 (2020).
130. Sork, V. L. et al. High-quality genome and methylomes illustrate features underlying evolutionary success of oaks. *Nat. Commun.* **13**, 2047 (2022).
131. Terhorst, J., Kamm, J. A. & Song, Y. S. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.* **49**, 303–309 (2017).
132. Excoffier, L. et al. fastsimcoal2: demographic inference under complex evolutionary scenarios. *Bioinformatics* **37**, 4882–4885 (2021).
133. Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* **5**, e1000695 (2009).
134. Akey, J. M. et al. Robust demographic inference from genomic and SNP Data. *PLoS Genet* **9**, e1003905 (2013).
135. Frichot, E. & François, O. L. E. A. An R package for landscape and ecological association studies. *Methods Ecol. Evolution* **6**, 925–929 (2015).
136. Yu, G. C., Wang, L. G., Han, Y. Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* **16**, 284–287 (2012).
137. Robinson, J. A. et al. Genomic flatlining in the endangered island fox. *Curr. Biol.* **26**, 1183–1189 (2016).
138. Hu, J. Y. et al. Genomic consequences of population decline in critically endangered pangolins and their demographic histories. *Natl. Sci. Rev.* **7**, 798–814 (2020).
139. Feng, S. et al. The genomic footprints of the fall and recovery of the crested ibis. *Curr. Biol.* **29**, 340–349 (2019).
140. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).
141. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).
142. Li, W. H., Wu, C. I. & Luo, C. C. Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. *J. Mol. Evol.* **21**, 58–71 (1984).
143. Jiang, X. L. et al. Data Repository (Figshare): Population genomics of *Quercus gilva* provides insights into the conservation of fengshui forests. https://doi.org/10.6084/m9.figshare.31190212 (2026).

Acknowledgements
We would like to thank Nami Sugiura at the University of Tokyo for her assistance with the sampling. This work was supported by grants from the National Natural Science Foundation of China (32572096 and 32201589) and the Yuelushan Laboratory Breeding Project (YLS-2025-ZY02055).

Author contributions
X.L.J., H.L., and M.D. designed the study. X.L.J., Y.S., M.D., and Z.Y.O.Y. collected the field samples. M.X.W., S.S.L., and X.L.J. performed the data analysis. X.L.J., M.X.W. and X.S.L. wrote the draft manuscript. All the authors edited and approved the final manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-026-09846-x.

Correspondence and requests for materials should be addressed to Xiao-Long Jiang or He LI.

Peer review information *Communications Biology* thanks Austin Thomas and the other anonymous reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Jorge Duitama & Rosie Bunton-Stasyshyn. A peer review file is available.