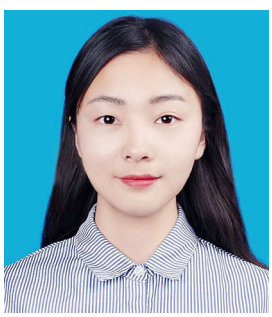
JIPB Journal of Integrative Plant Biology

Research Article
https://doi.org/10.1111/jipb.70194

Dissecting the genetic basis of climatic adaptation in wild relatives (*Malus baccata*) for climate-resilient apple breeding[oo]

Ying Su[1,2], Yani Hao[2], Xuejing Cao[3], Lin Wang[2,4], Zhiqi Xu[2,5], Fan Zhang[2], Zhiyao Ma[2], Xu Wang[2], Jiacui Li[2], Tianrong Fan[2], Ruoyan Zhao[2], Zhongqi Liu[2], Wenrui Wang[2], Yingchun Zhang[3], Xuanwen Yang[2], Sifan Yang[2], Dajiang Wang[4], Kun Wang[4], Simiao Sun[4], Zichen Li[4], Wen Tian[4], Yanming Sun[4], Zhao Liu[4], Yanshuai Xu[3], Hua Xiao[3], Yanling Peng[3], Xiaodong Xu[2], Ruirui Liu[6], Xinmin Tian[1,7*], Yongfeng Zhou[2,3*] and Yuan Gao[4*]

1. Xinjiang Key Laboratory of Biological Resources and Genetic Engineering, College of Life Science and Technology, Xinjiang University, Urumqi 830046, China
2. State Key Laboratory of Tropical Crop Breeding, Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, China
3. State Key Laboratory of Tropical Crop Breeding, Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Agricultural Genomics Institute at Shenzhen, Ministry of Agriculture and Rural Affairs, Chinese Academy of Agricultural Sciences, Shenzhen 518000, China
4. Key Laboratory of Horticultural Crop Germplasm Resources Utilization, Research Institute of Pomology, Ministry of Agriculture and Rural Affairs of the People's Republic of China, Chinese Academy of Agricultural Sciences, Xingcheng 125100, China
5. State Key Laboratory of Aridland Crop Science, Gansu Agricultural University, Lanzhou 730070, China
6. Gansu Key Laboratory of Conservation and Utilization of Biological Resources and Ecological Restoration in Longdong, School of Agriculture and Bioengineering, Longdong University, Qingyang 745000, China
7. Key Laboratory of Ecology of Rare and Endangered Species and Environmental Protection (Ministry of Education) & Guangxi Key Laboratory of Landscape Resources Conservation and Sustainable Utilization in Lijiang River Basin, Guangxi Normal University, Guilin 541006, China
*Correspondences: Yuan Gao (gaoyuan02@caas.cn, Dr. Gao is fully responsible for the distribution of all materials associated with this article); Xinmin Tian (tianxm333333@foxmail.com); Yongfeng Zhou (zhouyongfeng@caas.cn)



Ying Su

Yuan Gao

ABSTRACT

Climate change poses an increasing threat to global biodiversity and food security. As a wild relative of cultivated apples, *Malus baccata* exhibits broad environmental adaptability and robust stress tolerance. However, its effective utilization in breeding is constrained by the absence of a complete reference genome and insufficient population-level genomic characterization. In this study, we assembled a haplotype-resolved, telomere-to-telomere genome for *M. baccata*, providing unprecedented resolution for a wild apple reference genome. Population genomic analyses revealed four distinct genetic clusters. Among these, the Hebei Group 2 harbors the highest genetic diversity and heterozygosity, alongside the lowest runs of homozygosity, suggesting a complex history of genetic admixture in this population. By integrating population genomics with genotype-environment association analyses, we identified a series of climate-associated single-nucleotide polymorphisms and structural variants. A substantial proportion of these adaptive variants is localized within the coding and regulatory regions of candidate genes, providing a genomic basis for their roles in environmental adaptation. Notably, *DREB1A/D* and *NAC6* are associated with temperature seasonality and annual precipitation, respectively. Furthermore, future climate projections indicate that the Northeastern (NE) clusters face the highest risk of maladaptation, especially under

high-emission scenarios. Collectively, these findings provide critical insights into the genetic basis of climatic adaptation in wild apples, establishing a solid foundation for the conservation of crop wild relatives and the breeding of climate-resilient cultivars.

Keywords: climate-resilient apples, crop wild relatives, genomic offset, local adaptation, *Malus baccata*

Su, Y., Hao, Y., Cao, X., Wang, L., Xu, Z., Zhang, F., Ma, Z., Wang, X., Li, J., Fan, T., et al. (2026). Dissecting the genetic basis of climatic adaptation in wild relatives (*Malus baccata*) for climate-resilient apple breeding. J. Integr. Plant Biol. **00:** 1–22.

INTRODUCTION

Global environmental changes are intensifying, with rising temperatures, frequent flooding, and extreme weather events increasingly threatening biodiversity and global food security (Mace et al., 2008; Urban, 2015; Malanoski et al., 2024). Natural populations typically respond to shifting environments through migration, phenotypic plasticity, selection on standing genetic variation, or *de novo* mutations (Aitken et al., 2008; Fior et al., 2025). Fruit trees, as long-lived perennial woody species, are characterized by slow generation times and limited dispersal abilities (Long et al., 2025). These biological traits constrain their capacity to track rapid climatic change, rendering standing genetic variation essential for long-term persistence (Hereford, 2009; Aitken and Bemmels, 2016; Jia et al., 2020). Therefore, it is imperative to elucidate and quantify the genetic basis of local climatic adaptation in fruit trees to predict their resilience under future scenarios and to guide the development of climate-resilient cultivars (Zhou et al., 2014; Gougherty et al., 2021). Despite significant progress in identifying adaptive genes in various annual crops, climate-resilient breeding in perennial fruit trees remains an emerging field (Rivero et al., 2022; Xiong et al., 2024).

Crop wild relatives (CWRs) are indispensable reservoirs of adaptive variation, often retaining valuable alleles that have been lost during domestication, such as tolerance to drought, salinity, and cold, as well as resistance to diverse pathogens (Butchart et al., 2010; Long et al., 2024; Su et al., 2024; Zhang et al., 2025a, 2025b). The integration of these alleles into breeding pools has already enhanced crop resilience (Bohra et al., 2022), including heat tolerance in tomato (Gonzalo et al., 2021), drought and salt tolerance in grapevine (Carrasco et al., 2022; Cui et al., 2023), and cold tolerance in apple (Chen et al., 2019). These findings collectively demonstrate the substantial value of wild genetic resources for crop improvement. To fully exploit this potential and ensure the continued availability of these genetic resources, the conservation and comprehensive genomic characterization of CWRs are essential (Castañeda-Álvarez et al., 2016). Despite recent advances in plant genomics that have accelerated the study and conservation of several species, such as alfalfa (Zhang et al., 2024a), bamboo (Hou et al., 2024), citrus (Wang et al., 2023), kiwifruit (Zhang et al., 2023; Jiang et al., 2025), and grape (Zhang et al., 2024b), the majority of fruit tree wild relatives remain poorly characterized at the genomic level, especially regarding adaptive potential and conservation strategies (Gaut et al., 2018).

Among CWRs of temperate fruit tree, the genus *Malus* comprises numerous species with significant economic and ecological values, and represents one of the most globally widespread and widely cultivated fruit crops (Robinson et al., 2001; Li et al., 2025). However, wild apple populations are increasingly threatened by habitat loss and climate change. Their status as globally prioritized CWRs underscores the urgency of systematic collection and conservation efforts to safeguard their unique genetic resources (Khoury et al., 2013; Vincent et al., 2013). *M. baccata* (L.) Borkh is one of the most widely distributed and morphologically diverse apple species (Chen et al., 2023). Due to its exceptional cold tolerance, disease resistance, and broad environmental adaptability, it has been extensively utilized as a rootstock in high-latitude regions (Wang et al., 2019). Multiple resistance loci against major apple pathogens, including apple scab and powdery mildew, have been identified in *M. baccata* and its hybrids, further highlighting its practical value for breeding (Hemmat et al., 2003; Gygax et al., 2004; Chen et al., 2019; Wang et al., 2019). Although it has agricultural and ecological importance, the adaptive potential of *M. baccata* remains poorly understood at the genomic level. Current genome assemblies of *M. baccata* are incomplete, and population-level studies connecting genetic variation with environmental adaptation are largely lacking (Chen et al., 2019; Li et al., 2025). Recent advances in long-read sequencing technologies and population genomics offer new opportunities to address these knowledge gaps (Shi et al., 2023). By resolving highly repetitive regions and complex structural variations (SVs), telomere-to-telomere (T2T) genome assemblies in crops such as grape (Shi et al., 2023) and mango (Ahmad et al., 2025) have improved genome completeness and accuracy. These references enable precise variant identification and establish a solid foundation for genetic and breeding studies (Garg et al., 2024; Li and Durbin, 2024).

Numerous studies have demonstrated that local adaptation to climate is driven by polygenic selection across heterogeneous environments, highlighting the importance of elucidating the genetic basis of such adaptation at the genomic level (Sang et al., 2022; Long et al., 2025). However, the adaptive genetic basis of natural populations of *M. baccata* and their responses to current and future climate change remain poorly understood. Integrating landscape genomic approaches with genetic offset modeling provides a powerful framework for dissecting the molecular basis of local adaptation in apple (Aitken et al., 2024; Feng et al.,

2024; Lazic et al., 2024; Lind et al., 2024). These approaches facilitate the identification of signatures of local adaptation, as well as quantify the genomic vulnerability in natural populations in response to climate change.

In this study, we present the first haplotype-resolved T2T genome of *M. baccata*, which reveals genetic and functional divergence within its diploid genome. We further combined whole-genome resequencing of natural populations in China with environmental data to investigate the genetic structure and climate adaptation of *M. baccata*. By integrating genotype-environment association (GEA) analysis and gradient forest modeling, we identified candidate adaptive loci related to key environmental factors, including temperature, precipitation, radiation, and soil characteristics. Moreover, by projecting these models onto future climate scenarios, we quantified genomic offsets and assessed the vulnerability of populations to anticipated environmental changes. Our findings provide new insights into the mechanisms of climatic adaptation in *M. baccata* and offer valuable genomic resources to facilitate climate-resilient breeding of apples.

RESULTS

A haplotype-resolved telomere-to-telomere reference genome for *M. baccata*

A total of 20.94 Gb (~33.65×) of PacBio high-fidelity (HiFi) reads, 31.80 Gb (~51.11×) of Hi-C reads, and 16.33 Gb (~26.24×) of Oxford Nanopore Technologies (ONT) reads were generated for *M. baccata* collected from Jingpo Lake, China (Table S1). Genomic surveys using HiFi reads revealed a heterozygosity of approximately 1.1% for the *M. baccata* genome (Figure S1), indicating a highly heterozygous diploid genome and underscoring the necessity of a haplotype-resolved assembly strategy. The initial haplotype-resolved assemblies generated were 629.27 Mb and 628.71 Mb in size, with contig N50 values of 34.23 Mb and 35.02 Mb for haplotypes 1 and 2, respectively (Table S2). Based on the Hi-C contact map (Figure S2), we anchored the two haplotype assemblies to pseudochromosomes, achieving an average anchoring rate of 99.0% (Table S2). By manually resolving the gaps, two haplotypes were completely assembled with total lengths of 622.22 Mb and 623.11 Mb, and designated as MbaT2T_hap1 and MbaT2T_hap2, respectively. The two haplotypes exhibited low switch (~0.71%) and Hamming error rates (~1.44%), indicating high phasing accuracy. In addition, the high HiFi read mapping rate (average ~99.99%) and BUSCO score exceeding ~98.5% further confirmed the quality and completeness of the assembled genomes (Table S2).

We predicted 28 and 30 telomeres in MbaT2T_hap1 and MbaT2T_hap2, respectively (Figure 1A). The centromeric regions were characterized by the distribution of HODOR elements and distinct abundance patterns of specific transposable elements (TEs) (Figure S3). For TE annotation, repetitive sequences accounted for 60.45% (~376.15 Mb) and 60.26% (~375.50 Mb) of the haplotypes 1 and 2, respectively. Long terminal repeat (LTR) retrotransposons were the predominant class, with *Gypsy* elements representing 16.57% and 19.08% of haplotypes 1 and 2, respectively (Figure 1C; Table S3). For gene annotation, 42,657 and 42,649 protein-coding genes were predicted in two haplotypes, with high completeness evidenced by BUSCO scores of 98.6% and 98.3%, respectively. Functional annotation indicated that approximately 94.8% and 94.7% of genes were assigned to at least one database, confirming the accuracy and comprehensiveness of the gene models (Table S2).

Although draft and chromosome-level assemblies of the *M. baccata* genome have been reported previously, these assemblies are neither telomere-to-telomere nor haplotype-resolved. To assess the extent of improvement achieved by our assembly, we compared both haplotypes with the *M. baccata* genomes released in 2019 (Chen et al., 2019) and 2025 (Li et al., 2025). The results demonstrate that the current assembly achieves accurate haplotype phasing and superior performance across multiple quality metrics, specifically in the annotation of TEs, telomere and centromere identification, and gene annotation. Furthermore, comparative genomic analyses revealed extensive SVs among these assemblies (Table S4). Using MbaT2T_hap1 as the reference, we identified 7,889 deletion events in the *M. baccata* assembly released in 2025, affecting a total of 1,875 genes. Functional enrichment analysis revealed that these affected genes were significantly enriched in photosynthesis-related biological processes (Figure S4).

We further evaluated genomic divergence between the MbaT2T_hap1 and MbaT2T_hap2, identifying a total of 27,897 deletions (DEL), 23,314 insertions (INS), 58 duplications (DUP), 124 inversions (INV), and 3,002 translocations (TRANS) (Table S5). Furthermore, 2,632 hemizygous genes were characterized, accounting for 6.2% of all annotated protein-coding genes (Table S6). This proportion is consistent with findings reported in other *Malus* species (Peng et al., 2025). Comparative analysis of SVs between wild *M. baccata* and the cultivated apple "Golden Delicious" revealed the set of SVs (27,460 DEL, 31,794 INS, 222 DUP, 305 INV, and 4,146 TRANS) (Figure S4; Table S7). Presence/absence variation (PAV) analysis identified 7,685 genes specific to *M. baccata* (Table S8). Functional enrichment analysis showed that these genes were significantly enriched in biological processes, including photosynthesis, flavonoid metabolism, and carbohydrate catabolic processes (Figure 1D). These results collectively demonstrate that our haplotype-resolved T2T assemblies represent a high-quality genomic resource for *M. baccata*, providing a solid foundation for comparative genomics, population genetics, and studies on environmental adaptation.

Population structure and genetic diversity of *M. baccata*

To investigate the genetic dynamics of wild apple (*M. baccata*), we collected short-read sequence data from 138 individuals.

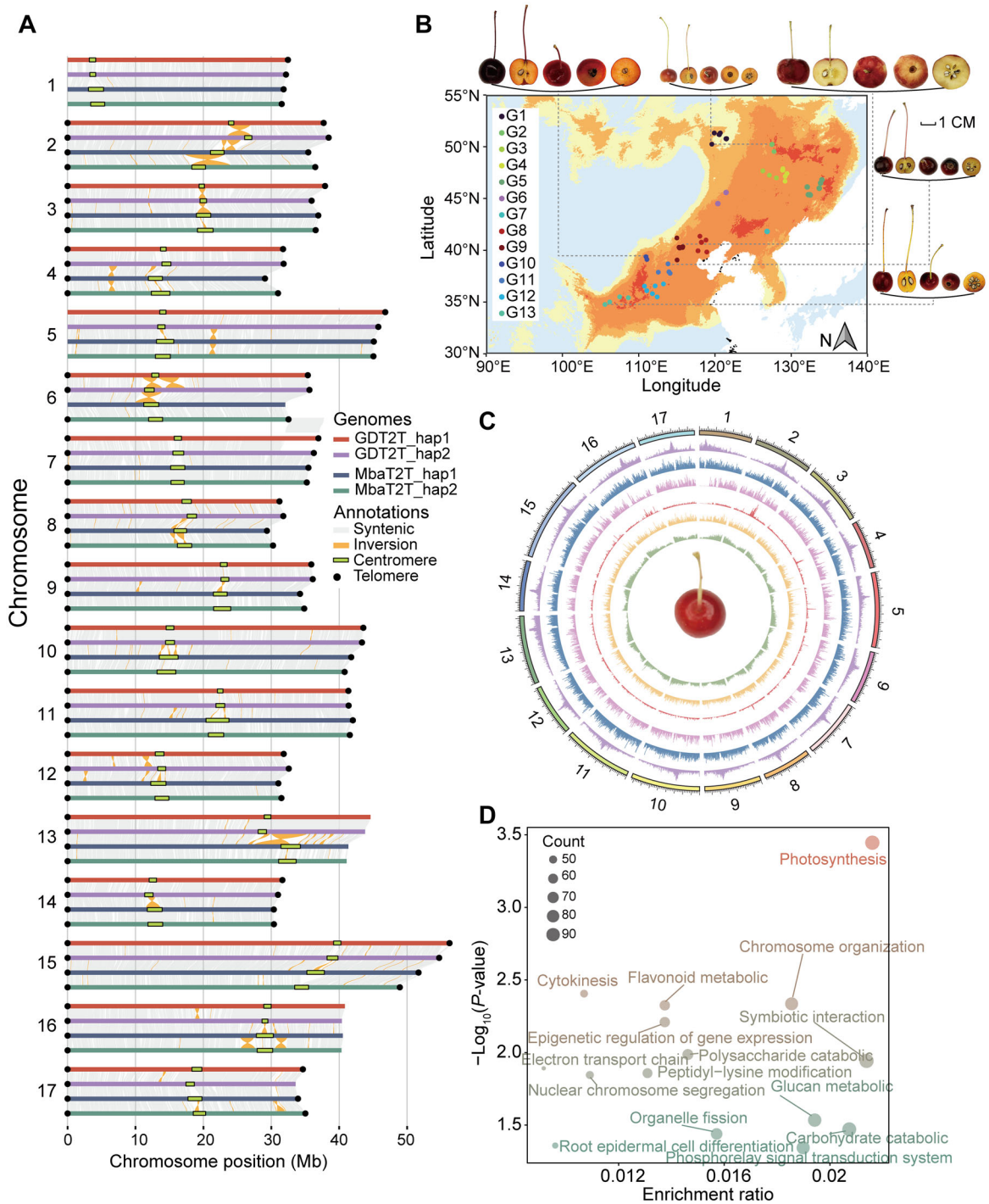


Figure 1. Genomic and phenotypic landscapes of *Malus baccata*

(A) Synteny analysis illustrating genomic conservation and divergence among the four genomes GDT2T_hap1, GDT2T_hap2, MbaT2T_hap1, and MbaT2T_hap2. GD, "Golden Delicious" assembly; Mba, *Malus baccata* assembly. **(B)** The map displays the geographical distribution of *M. baccata* accessions across East Asia, with representative fruit phenotypes shown for each region. The scale bar indicates ~1 cm. **(C)** Circos plot of the MbaT2T_hap1 assembly highlighting genomic features across the 17 chromosomes. From the innermost to the outermost tracks: gene density, structural variations (SVs), insertions and deletions (InDels), single nucleotide polymorphisms (SNPs), *Copia* transposable elements (TEs), and *Gypsy* TEs. **(D)** Gene Ontology (GO) enrichment analysis of *M. baccata* lineage-specific genes relative to the "Golden Delicious" genome.

The dataset includes 117 newly sequenced accessions with an average coverage of ~17.3×, 111 of which have precise geographic coordinates representing 13 geographically distinct natural populations. This high-resolution sampling provides a robust foundation for subsequent spatial genetic and environmental association analyses. One apple (*Malus angustifolia*) and five pear accessions were used as an outgroup for population analysis (Table S9). The 111 accessions were collected from

diverse regions across China to capture a comprehensive spectrum of genetic diversity (Figure 1B). We identified 46,877,012 single nucleotide polymorphisms (SNPs) by aligning reads to the complete MbaT2T_hap1 reference genome. Following linkage disequilibrium (LD) pruning, a subset of 604,968 SNPs was retained for population genetic analyses. ADMIXTURE was employed to infer ancestral components and individual admixture proportions, with cross-validation indicating an optimal number of populations at $K = 4$–6. The corresponding clustering results for $K = 2$–6 are illustrated in the figure (Figure 2A). Maximum-likelihood phylogenetic reconstruction and principal component analysis (PCA) consistently resolved four genetic clusters within *M. baccata*, corresponding to the Northwestern (NW, 25 individuals), Northeastern (NE, 60 individuals), Hebei Group 1 (HB1, 14 individuals), and Hebei Group 2 (HB2, 33 individuals) populations (Figures 2B, S5).

To characterize patterns of genetic variation across the four populations, we calculated nucleotide diversity (π) and genetic differentiation (F_{ST}) value. HB2 exhibited the highest nucleotide diversity ($\pi = 5.35 \times 10^{-3}$), followed by NW ($\pi = 4.60 \times 10^{-3}$), while NE and HB1 showed lower levels ($\pi = 3.30 \times 10^{-3}$). NE was highly differentiated from the two southern populations (NE vs. NW: $F_{ST} = 0.225$; NE vs. HB2: $F_{ST} = 0.206$) (Figure 2C). Notably, despite its geographic proximity to NW and HB2, the HB1 group exhibited phylogenetic and differentiation patterns more similar to NE (NE vs. HB1: $F_{ST} = 0.061$). Analyses of identity-by-descent (IBD) blocks revealed stronger signals in NE and HB1 than in other groups, indicating more extensive haplotype sharing. This pattern, consistent with their reduced nucleotide diversity, suggests recent common ancestry or smaller effective population sizes (Figure S6). Analysis of runs of homozygosity (ROH; > 500 kb) and genome-wide heterozygosity further revealed more extensive homozygous tracts and higher inbreeding in NE and HB1 compared to NW and HB2 (Figure 2D), which may result from limited genetic exchange in northern regions or selective pressures related to local environmental conditions (López-Goldar and Agrawal, 2021). Given the potential for genetic exchange to shape diversity, we further examined introgression, the transfer of alleles from a donor to a recipient population via hybridization and backcrossing, which can influence both fitness and genomic architecture (Harrison, 1993; Edelman and Mallet, 2021; Brauer et al., 2023). The *fb* statistic revealed significant gene flow into HB1 from both NW and HB2 (Figure 2E). These results suggest extensive hybridization and gene flow between HB1 and other wild species, although their role in recent adaptation remains unclear.

Deleterious and adaptive variations in *M. baccata*

Deleterious alleles are defined as variants that may determine fitness (Charlesworth and Charlesworth, 1998). We investigated the accumulation of deleterious alleles. These results revealed that under an additive model, the HB2 and NW subgroups, characterized by higher hybridization backgrounds, harbored significantly more deleterious SNPs and SVs than the NE and HB1 subgroups. Interestingly, deleterious SNPs exhibited contrasting patterns between the heterozygous and recessive models (Figure 3A). Although HB2 and NW carried a higher heterozygous load, their recessive deleterious mutations were significantly reduced. This observation suggests that interspecific hybridization in *M. baccata* facilitates the purging of recessive deleterious SNPs while simultaneously accumulating hybrid load. Notably, HB2 and NW maintained elevated deleterious SV loads under three models, indicating that SNPs and SVs follow distinct response mechanisms to evolutionary selection. This discrepancy likely arises from the stronger genetic burden and recombination suppression associated with SVs, which hinders the efficient purging of deleterious SVs compared to the more effective selection acting on individual SNP sites.

The landscape genomic frameworks exhibit strong effectiveness in detecting adaptive signals along environmental gradients (Lotterhos, 2023; Lazic et al., 2024). To investigate the genetic basis of local adaptation in *M. baccata*, we performed GEA analysis using latent factor mixed models (LFMM) for all bioclimatic variables (Figure S7). This approach effectively evaluates the associations between genetic variants and environmental factors while accounting for confounding population structure. In total, we identified 7,619 SNPs and 766 SVs that exhibited significant associations with at least one environmental variable (Table S10). Based on the variable importance rankings from gradient forest (GF) analysis and pairwise Spearman correlation coefficients ($|r| < 0.8$), we selected six uncorrelated bioclimatic variables to minimize multicollinearity in subsequent analyses (Figures S8, S9). The selected variables comprised three temperature-related variables: mean diurnal range (BIO2), temperature seasonality (BIO4), and maximum temperature of the warmest month (BIO5), and three precipitation-related variables: annual precipitation (BIO12), precipitation of the wettest month (BIO13), and precipitation of the driest quarter (BIO17) (Figure S9). Using these six selected variables, redundancy analysis (RDA) was performed to identify genetic loci potentially associated with bioclimatic gradients, ultimately detecting 45,842 SNPs and 1,144 SVs (Table S11). Based on the RDA results, we calculated an adaptive index to quantify genetic similarity and divergence across the landscape (Capblancq and Forester, 2021). The first RDA axis (RDA1) captured the majority of adaptive variation, and pronounced spatial heterogeneity in the adaptive index was observed (Figure 3B, C). These six selected bioclimatic variables largely explained a substantial portion of the environment-associated genomic variation in *M. baccata*. Specifically, genetic variants in the HB2 population showed signatures of local adaptation associated with the maximum temperature of the warmest month (BIO5) and precipitation of the wettest month (BIO13), as indicated by positive RDA1 scores (Figure 3B). In contrast, variants in NE populations were most strongly associated with temperature seasonality (BIO4) and precipitation of the driest quarter (BIO17), reflected by negative RDA1 scores (Figure 3C).

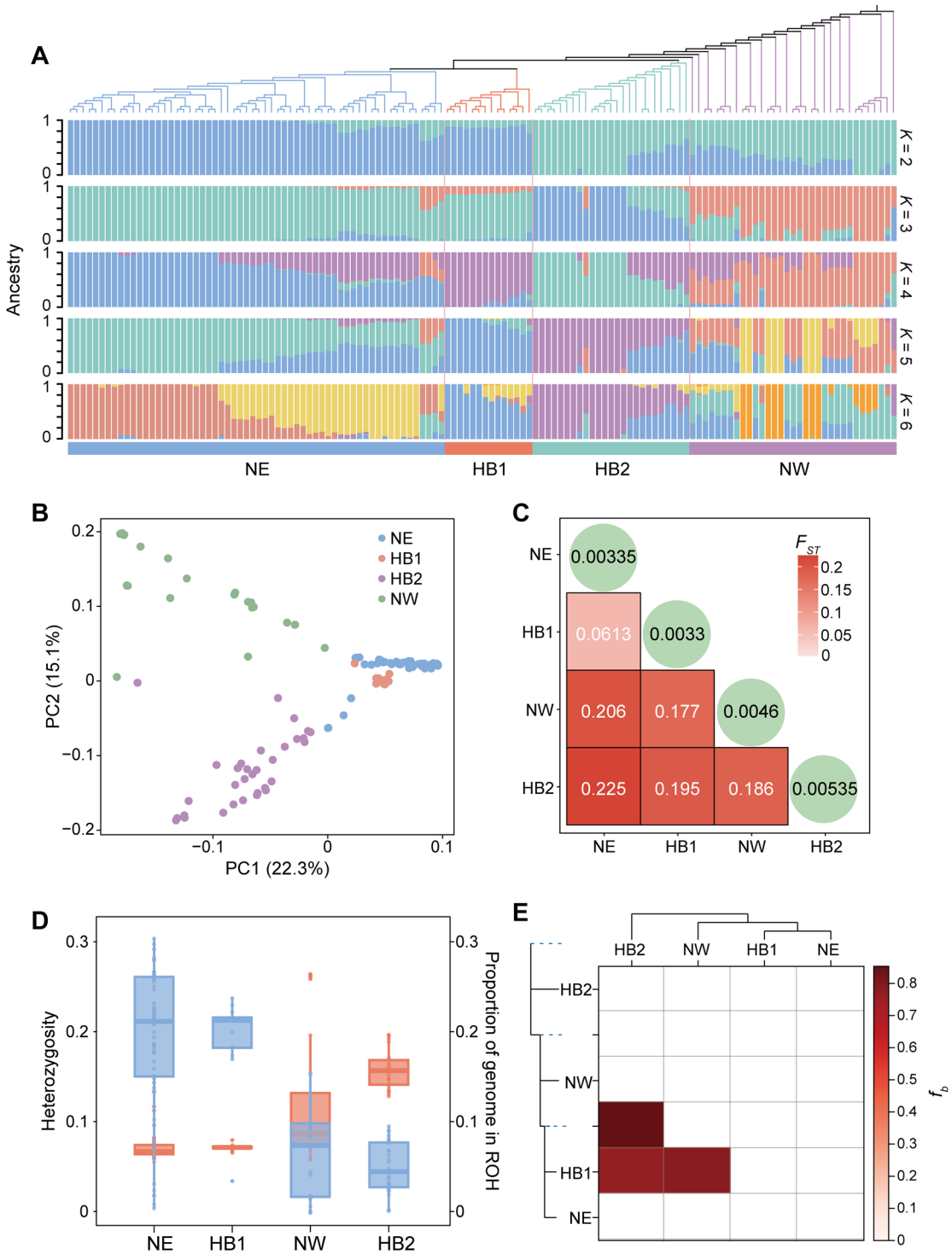


Figure 2. Population genetic structure and diversity analyses of *M. baccata*

(A) Admixture analysis displaying the ancestry components of *M. baccata* accessions across different *K* values. The upper panel exhibits a phylogenetic tree, reconstructed without the outgroup. **(B)** Principal component analysis (PCA) plot illustrating the genetic clustering of the four *M. baccata* populations: NW, Northwestern; NE, Northeastern; HB1, Hebei Group 1; and HB2, Hebei Group 2. **(C)** Heatmap representation of pairwise F_{ST} values among the four *M. baccata* populations. The circles indicate nucleotide diversity (π). **(D)** Boxplots representing heterozygosity (pink) and the proportion of the genome in runs of homozygosity (blue) across the four populations. For each box plot, the horizontal line signifies the median, the box boundaries represent the first and third quartiles, and the whiskers denote the data range (min to max). Sample sizes for each population are as follows: NE ($n = 60$), HB1 ($n = 14$), HB2 ($n = 33$), and NW ($n = 25$). **(E)** Branch introgression (*fb*) values indicating the strength of potential gene flow between groups. Darker red corresponds to higher *fb* values.

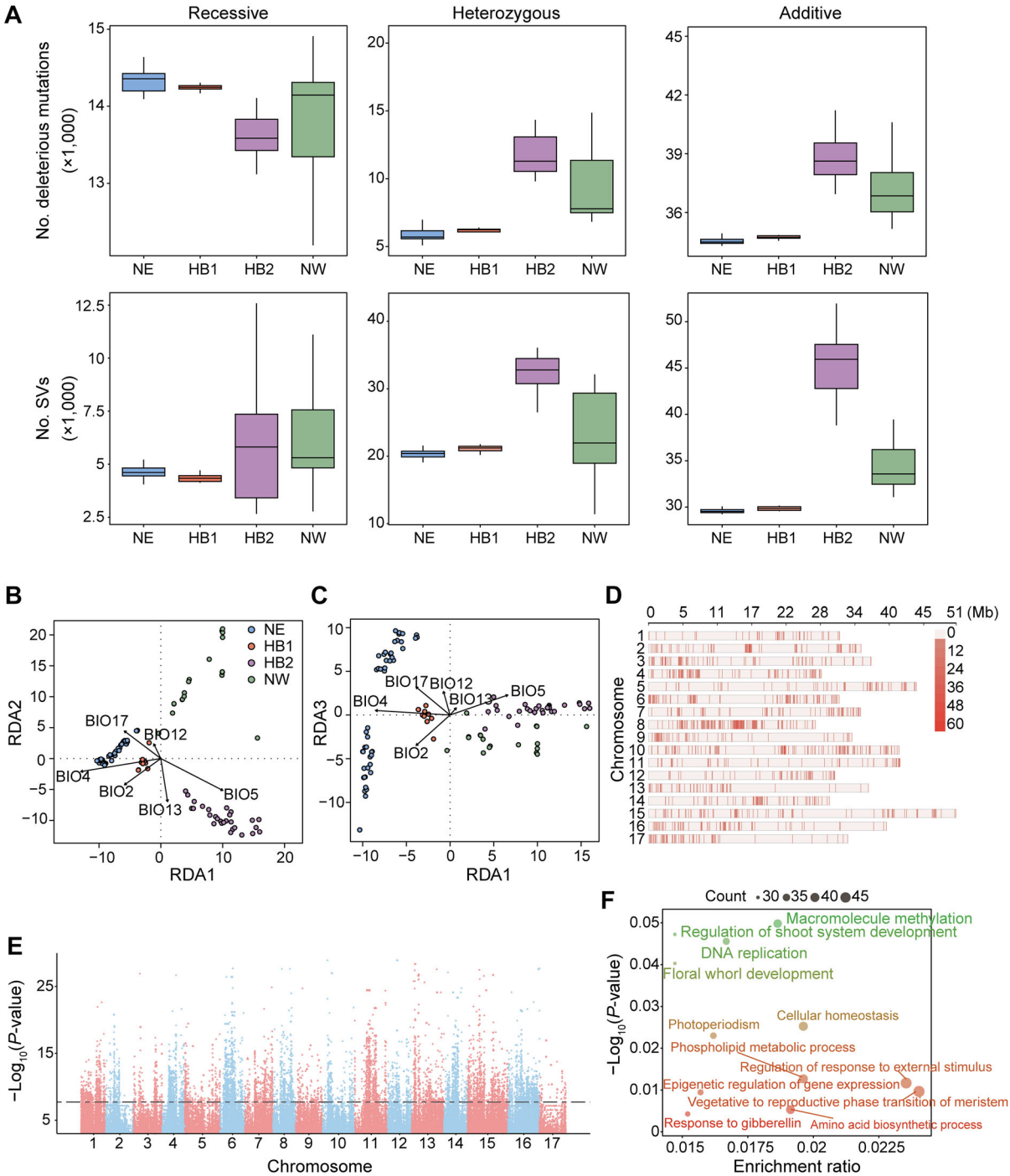


Figure 3. Genome-wide profiling of deleterious mutations and adaptive variations in *M. baccata*

(A) Boxplots showing the accumulation of deleterious mutations and structural variations (SVs) under recessive, heterozygous, and additive genetic models across the NW, NE, HB1, and HB2 populations. For each box plot, the horizontal line signifies the median, the box boundaries represent the first and third quartiles, and the whiskers denote the data range (min to max). Sample sizes for each population are as follows: NE ($n = 60$), HB1 ($n = 14$), HB2 ($n = 33$), and NW ($n = 25$). **(B, C)** Redundancy analysis (RDA) plots illustrating the relationship between genomic variants and key bioclimatic variables for the four populations. **(D)** Genomic distribution of core adaptive single nucleotide polymorphisms (SNPs) and SVs across the genome. **(E)** Manhattan plot showing the results of the PCAdapt analysis, identifying outlier SNPs putatively under selection and potentially involved in local adaptation. The horizontal dashed line represents the significance threshold (Bonferroni-corrected $P < 0.05$). **(F)** Bubble plot illustrating the GO terms significantly enriched among the outlier genes identified by PCAdapt.

Finally, the candidate environment-associated loci were defined by overlapping the results of LFMM and RDA. This consensus approach jointly identified 3,671 SNPs and 214 SVs, which were designated as putative "core variants" underlying climate adaptation (Figure 3D; Table S12; Figure S12). These high-confidence loci were distributed across the entire genome. GO enrichment analysis revealed that these "core variants" are involved in diverse biological processes, including shoot system morphogenesis, isoprenoid metabolic process, and response to hydrogen peroxide (Figure S10). These variants were primarily located in upstream cis-regulatory regions (33.18%) and intergenic regions (26.70%) (Figure S11). The high proportion of variants in cis-regulatory regions suggests that regulatory changes may contribute to climate adaptation, although further functional validation is required to confirm their effects. Only 4.60% of the core adaptive SNPs were non-synonymous, and 6.29% were synonymous. To complement the GEA results, we employed the PCAdapt algorithm to identify genomic outliers deviating from the overall population structure. This analysis revealed 7,628 outlier SNP windows and 746 SV windows, involving 3,143 candidate genes (Figure 3E; Table S13). GO enrichment analysis of genes potentially affected by these outlier loci revealed a primary enrichment in biological processes related to regulation of response to external stimulus, phospholipid metabolic process, and amino acid biosynthetic process (Figure 3F).

Local adaptation to temperature in *M. baccata*

For each environmental variable, we identified key genes associated with significant loci, in which allele frequencies showed strong correlations with both geographic distribution and environmental gradients (Figure S13; Table S14). The majority of SNPs (~70%) were associated with a single environmental variable, whereas a smaller fraction was linked to multiple variables. For example, ~12.4% of SNPs were associated with four environmental variables, suggesting that a subset of loci may be subject to consistent selective pressures across multiple ecological gradients (Figure S14). We identified numerous previously reported climate-adaptive genes, many of which are associated with temperature-related traits. Notably, two tandem duplicated homologs of *DREB* transcription factors (*DREB1A* and *DREB1D*) (Yang et al., 2011; Chen et al., 2019) were consistently detected in association with multiple bioclimatic variables, specifically temperature seasonality (BIO4) and mean temperature of the coldest quarter (BIO11) (Figure 4A, B). One structural variant (DEL00116767) and one SNP (4_8215096) were observed in the upstream regions of *DREB1D* and *DREB1A*, respectively (Figure 4C). The SNP associated with *DREB1A* was linked to six BIOs (BIO1, BIO3, BIO4, BIO6, BIO9, and BIO11), highlighting its broad environmental responsiveness. Homologs of these two genes have been previously characterized in *M. baccata*, with their roles in drought stress response validated in *Arabidopsis* (Vonapartis et al., 2022). The geographic distribution of both variants revealed clear differences in allele frequency (Figure 4D, F). Regarding the SV (DEL00116767), the allele carrying the deletion was mainly observed in southern regions characterized by lower temperature seasonality, while the non-deletion allele was almost fixed in northeastern regions with higher temperature seasonality. Allele frequencies of both variants exhibited significant correlations with climatic gradients (Figure 4E, G). In addition, compared with neutral reference variants, these adaptive alleles exhibit distinct allele frequency patterns across the BIO4 gradient (Figure S15). To further validate the functional implications of these variants, we analyzed publicly available transcriptomic data under cold stress. Both *DREB1A* and *DREB1D* showed significant upregulation in cold-treated samples compared with controls, which is consistent with their roles in cold stress responses and further underscores the adaptive significance of the identified genomic variants (Figure S16; Table S15).

Local adaptation to precipitation in *M. baccata*

We also identified numerous genes associated with precipitation (Figure S13; Table S14). A representative example is the transcription factor *NAC6*, which showed a significant association with annual precipitation (BIO12) (Figure 5A). In rice, the homolog gene *OsNAC6* has been reported to enhance drought tolerance by promoting lateral root formation (Lee et al., 2017). A critical candidate SNP (SNP_10_22531962) was identified ~1 kb upstream of *NAC6* within the putative cis-regulatory region, suggesting a potential role in modulating *NAC6* expression under climatic stress (Figure 5C). Geographical distribution analysis revealed that the G allele predominated in populations inhabiting regions with higher precipitation, whereas the A allele frequency increased in populations from drier regions (Figure 5B). Allele frequency at this locus exhibited a strong positive correlation with annual precipitation (Figure 5D). Additionally, the G allele, prevalent in high-precipitation regions, displayed moderately higher extended haplotype homozygosity (EHH) values compared to the A allele (Figure 5E), suggesting a weak signal of positive selection. Nevertheless, the absence of pronounced selective sweep signals further indicates that climate adaptation in *M. baccata* is primarily shaped by a polygenic architecture involving many loci of small effect. Consistent with these patterns, analysis of publicly available transcriptomic data revealed that *NAC6* expression was significantly downregulated following flooding treatment, supporting its responsiveness to precipitation-related stress (Figure S1; Table S15).

Several other genes were also implicated in precipitation-related adaptation. The candidate gene associated with precipitation of the driest month (BIO14) corresponds to *FAX4*, whose loss-of-function mutation disrupts fatty acid transport and membrane stability, thereby affecting osmotic regulation under drought stress (Li et al., 2020). The gene associated with precipitation of the wettest quarter (BIO16) corresponds to *GRXC2*, a homolog strongly induced by

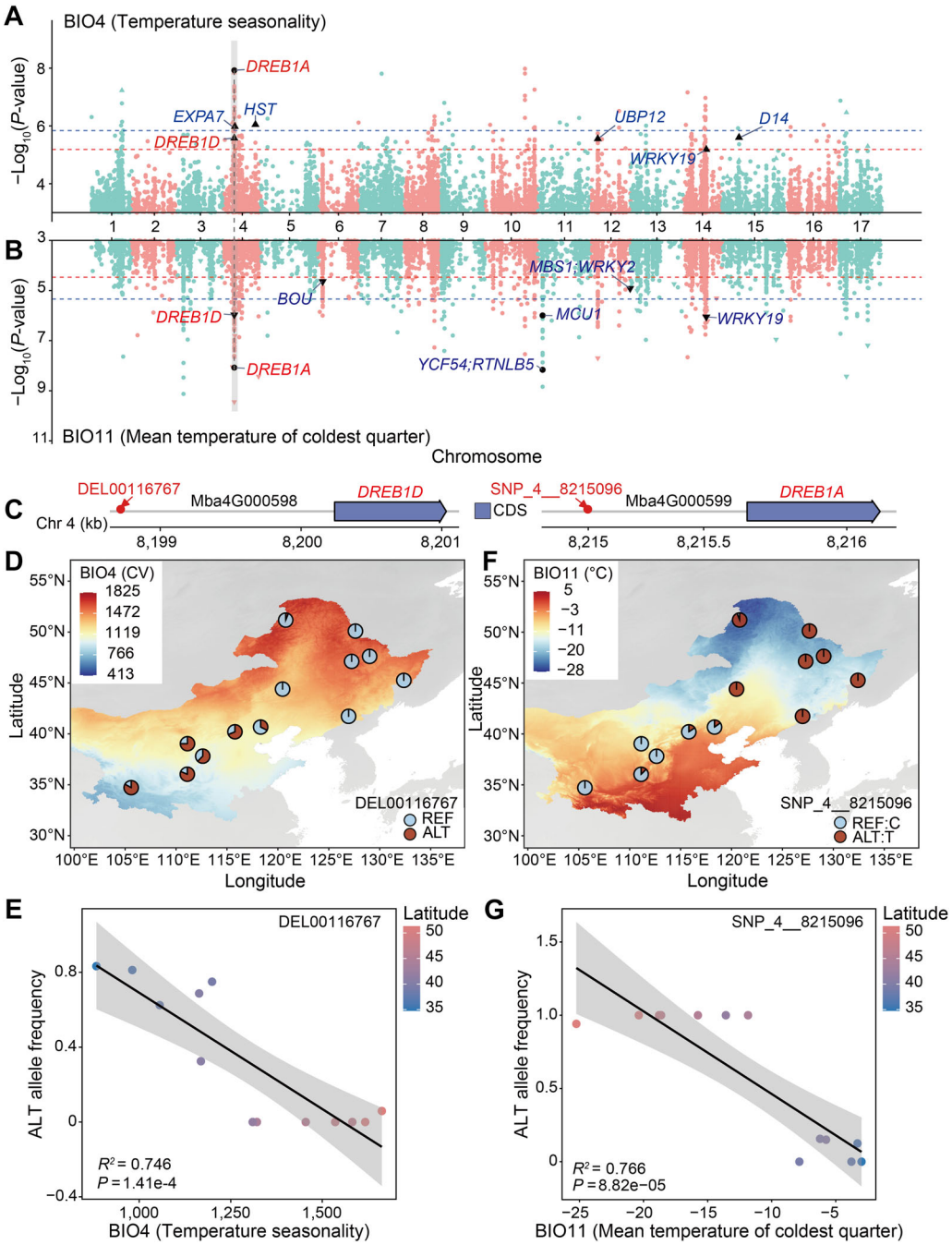


Figure 4. Genome-wide identification of candidate loci associated with local adaptation to temperature

(A, B) Manhattan plots showing the significant genomic signals associated with temperature seasonality (BIO4, top panel) and the mean temperature of the coldest quarter (BIO11, bottom panel). SNPs and SVs are represented by points and triangles, respectively. Horizontal dashed lines indicate significance thresholds (FDR ≤ 0.05) for SNPs (blue) and SVs (red), with key candidate genes labeled. **(C)** Diagrams of the *DREB1D* and *DREB1A* genomic loci. The candidate adaptive SNPs are highlighted. **(D, F)** Maps showing the geographical distributions and relative frequencies of the DEL00116767 **(D)** or SNP_4__8215096 **(F)** variant (reference [REF] and alternative [ALT] alleles) overlaid onto the local values for BIO4 and BIO11, respectively. **(E, G)** Scatterplots showing the correlation between the frequency of the ALT allele for DEL00116767 **(E)** or SNP_4__8215096 **(G)** and the corresponding bioclimate values across latitudes.

osmotic perturbation in *Arabidopsis* (Belin et al., 2015). Additionally, *ADH2* was significantly associated with precipitation of the warmest quarter (BIO18), and its soybean ortholog (*GmAdh2*) has been validated to enhance germination under flooding stress (Figure S13) (Komatsu et al., 2011; Tougou et al., 2012).

We further conducted LFMM analyses on seven additional environment variables, including monthly precipitation

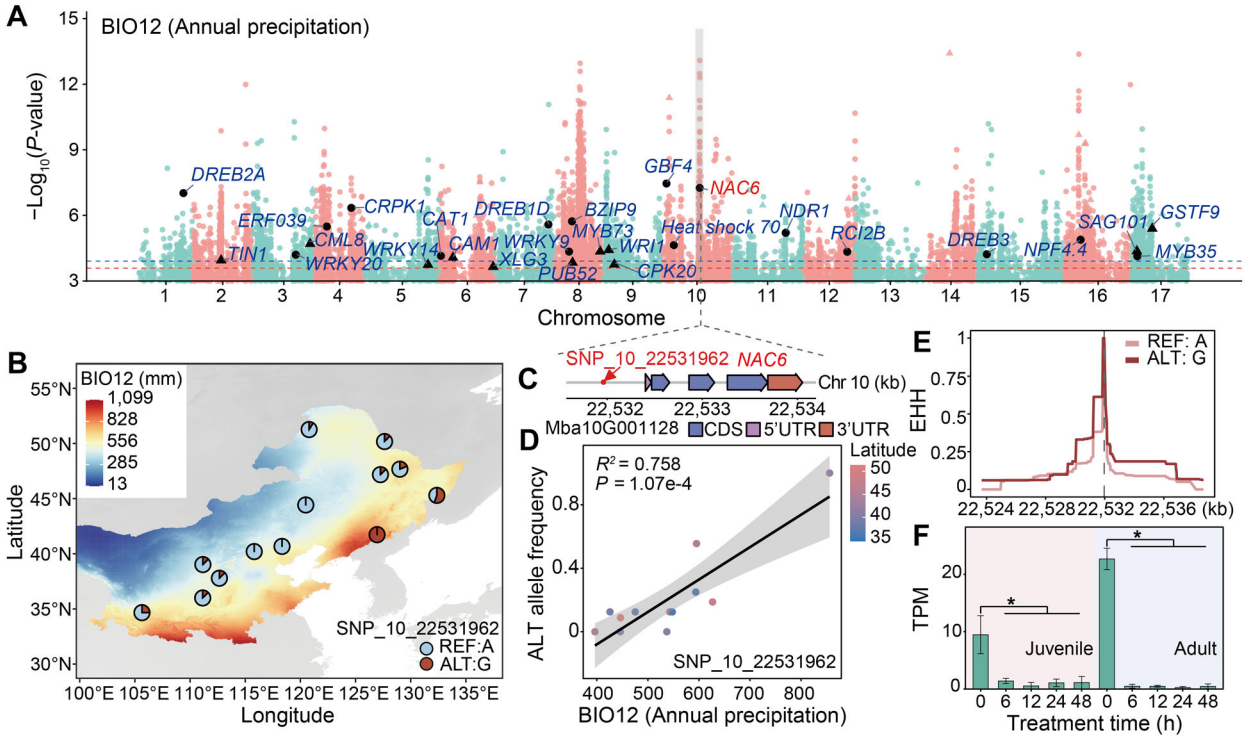


Figure 5. Genome-wide identification of candidate loci associated with local environmental adaptation analyses related to precipitation
(A) Manhattan plot showing the significant genomic signals associated with the annual precipitation (BIO12). SNPs and SVs are represented by points and triangles, respectively. Horizontal dashed lines represent the significance thresholds (FDR ≤ 0.05) for SNPs (blue) and SVs (red), with key candidate genes labeled. **(B)** Map showing the geographical distribution and relative frequencies of SNP_10__22531962 variants (REF and ALT alleles) overlaid on BIO12. **(C)** Diagram of the *NAC6* genomic locus. The candidate adaptive SNP is highlighted. **(D)** Scatterplot showing the correlation between the frequency of the ALT allele and bioclimate values (BIO12) across latitudes. **(E)** Decay of extended haplotype homozygosity (EHH) for the REF and ALT alleles at the significant SNP_10__22531962 of *NAC6*. **(F)** Bar plot illustrating the relative expression levels of *NAC6* in *M. baccata* var. *xiaojinensis* under control and water-stress conditions. Transcriptomic data are presented as transcripts per million (TPM), retrieved from NCBI (BioProject PRJNA392909). *P*-values were adjusted using the Benjamini–Hochberg method (*$P < 0.05$).

(prec6, prec12), solar radiation (srad6, srad12), vapor pressure (vapr6, vapr12), and elevation (Table S16). These complementary datasets provide valuable resources and novel insights into the genomic basis of local adaptation. We found that numerous candidate genes were significantly associated with the climatic variables, including 2,160 SNPs and 467 SVs, which involve 965 annotated genes, some of which have been previously validated, such as *CBF/DREB* transcription factors, *HSP70* (Jung et al., 2013; Usman et al., 2017), and *NAC* family genes (Table S16). Collectively, our analyses of candidate genes under selection reveal that natural populations of *M. baccata* exhibit polygenic adaptation to local climatic factors, particularly temperature and precipitation gradients. These adaptive loci elucidate the genetic basis of local adaptation and represent key genomic indicators for predicting the responses of *M. baccata* to future climate change.

Local adaptation to saline-alkali soil in *M. baccata*

Soil properties, such as fertility, nutrient availability, and cation exchange capacity, are pivotal determinants of plant growth and survival, particularly under challenging environmental conditions, such as salinity and alkalinity (Liang et al., 2024). To investigate the genetic basis of soil adaptation in *M. baccata*, we applied LFMM analyses and identified 5,034 SNPs and 1,730 SVs significantly associated with seven soil factors. These variants correspond to 2,202 candidate genes, including previously validated genes, such as *CIPK11*, *TCP20*, *NRAMP7.2*, *NLP2*, and *LYK3* (Table S17). Among these soil factors, soil cation exchange capacity (CEC_SOIL) represents a critical indicator of soil fertility, nutrient availability, and buffering capacity under saline–alkaline conditions (Gao et al., 2024). At loci significantly associated with CEC_SOIL variation, we identified 750 genes involved in biological processes significantly enriched salt–alkaline stress tolerance, including alcohol metabolic process, reactive oxygen species metabolic process, response to toxic substance, detoxification, and protein autophosphorylation (Figure 6A, B). Interestingly, a SNP located in the coding region of the *NPF* gene showed a significant positive correlation between allele frequency and CEC_SOIL levels (Figure 6C, D). Haplotype analysis further revealed that haplotypes carrying the high CEC_SOIL allele (G) exhibited faster EHH decay compared with those carrying the alternative allele

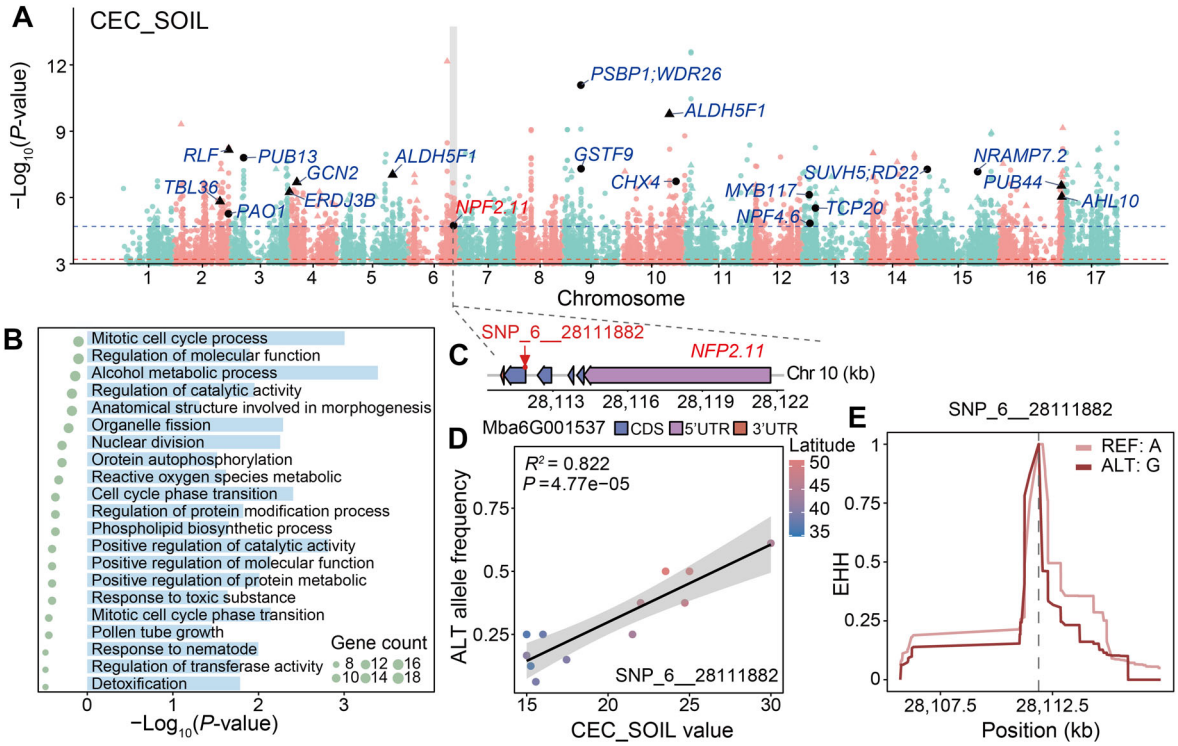


Figure 6. Genome-wide identification of candidate loci associated with local adaptation to soil cation exchange capacity
(A) Manhattan plot showing the significant genomic signals associated with soil cation exchange capacity (CEC_SOIL). SNPs and SVs are represented by points and triangles, respectively. Horizontal dashed lines represent the significance thresholds (FDR ≤ 0.05) for SNPs (blue) and SVs (red), with key candidate genes labeled. **(B)** GO term enrichment analysis of candidate genes significantly associated with CEC_SOIL. **(C)** Diagram of the *NFP2.11* genomic locus. The candidate adaptive SNP is highlighted. **(D)** Scatterplot showing the correlations between the frequency of the ALT allele and CEC_SOIL values across latitudes. **(E)** Extended haplotype homozygosity (EHH) decay for the REF and ALT alleles at the significant SNP (SNP_6__28111882).

(A) (Figure 6E). These findings highlight the genomic basis of soil adaptation and provide insights into the mechanisms enabling *M. baccata* to thrives in marginal soils. In summary, our association analyses uncovered adaptive loci and candidate genes linked to key environmental and climatic factors, deepening our understanding of the genetic basis of local adaptation in wild apple and offering critical information for predicting species' responses to future climate change.

Genomic offset reveals genetic vulnerability under future climate change

Given the accelerating pace of climate change, it is essential to assess how natural populations will respond to future environmental conditions. Here, we employed genomic offset analyses based on core adaptive variants to predict the responses of *M. baccata* populations to projected climatic changes. Specifically, we quantified climatic vulnerability using GF models, which estimate the magnitude of genomic offset required for populations to maintain local adaptation as the climate shifts. Forward and reverse offsets were further utilized to capture potential mismatches under migration and in situ scenarios, highlighting populations and regions at greatest conservation risk (Gougherty et al., 2021). To account for temporal and climatic uncertainty, projections were based on three global circulation models (ACCESS-CM2, CMCC-ESM2, and GISS-E2-1-G) under two Shared Socioeconomic Pathways (SSP245 and SSP585) for the 2060s (2041–2060), 2080s (2061–2080), and 2100s (2081–2100). We used 19 bioclimatic variables (BIO1–BIO19) to estimate the genomic offset across the geographic distribution of the species. Notably, the estimated local, forward, and reverse offsets exhibited high consistency across all models (Figure S18), supporting the robustness of our adaptive predictions.

Specifically, genomic offsets exhibited an increasing trend under higher emission scenarios, indicating that genetic mismatches between populations and their future environments will intensify as greenhouse gas emissions rise (Figure 7A, B). According to the multi-model averages, populations in the eastern and southeastern parts of the distribution, especially in coastal areas near the Korean Peninsula, consistently showed larger local offsets than those in the west (Figures S19, S20). These results suggest that the eastern edge of the range may be a hotspot for climatic vulnerability. Forward and reverse offsets provided complementary insights into the adaptive landscape, with distinct spatial patterns revealed by the ternary color maps (Figure 7C, D). Extensive areas in northeastern China

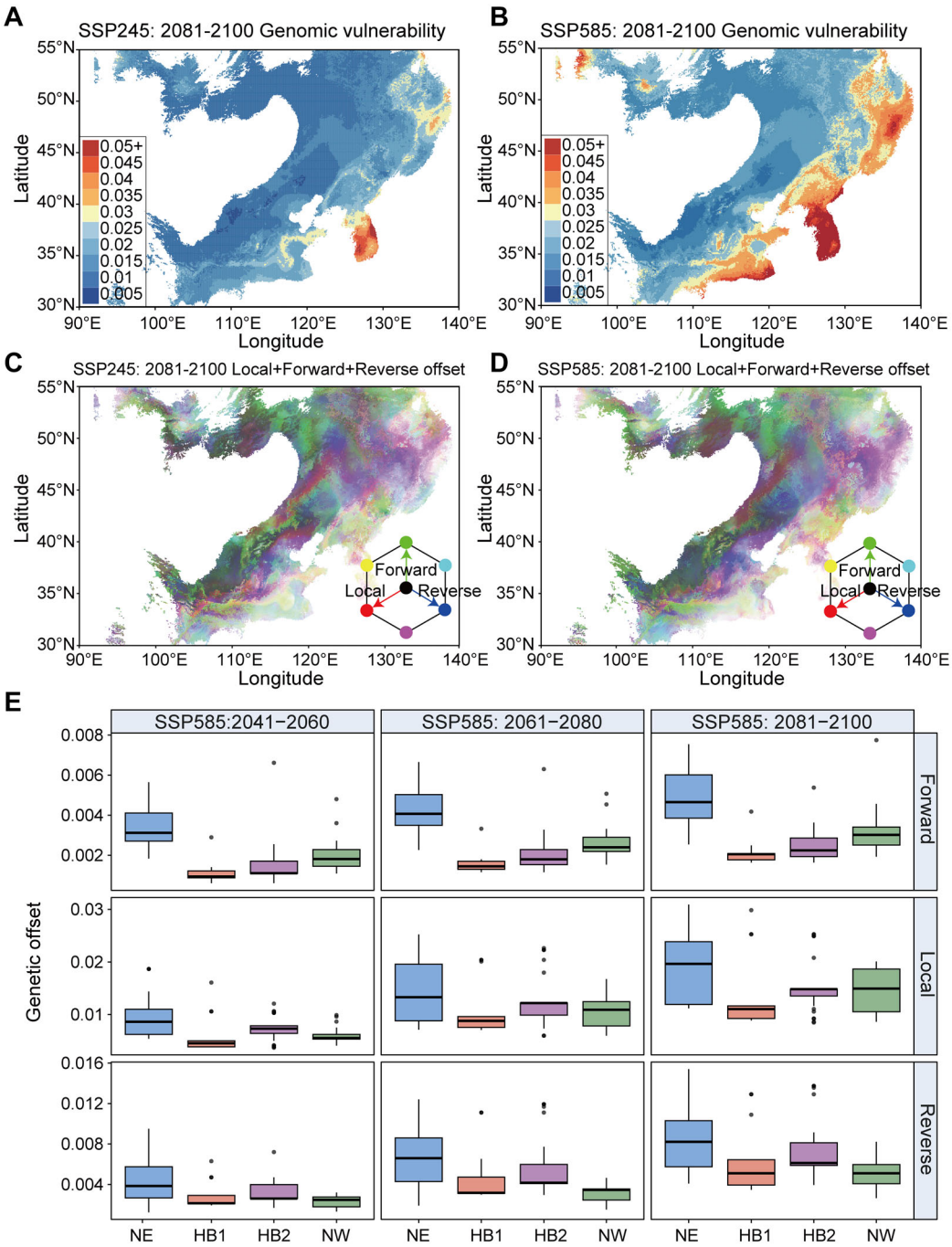


Figure 7. Projections of genomic offset under future climate scenarios for the 2081–2100 period

(A, B) Spatial distribution of local genomic offset across the adaptive range of *M. baccata* (~78,000 grid cells). Genomic offset values represent the average across three climate models under **(A)** the intermediate emission scenario (SSP245) and **(B)** the high emission scenario (SSP585) for 2081–2100 period. Blue, lower genomic offset values; red, higher genomic offset values. **(C, D)** RGB composite maps showing the relative contributions of local (red), forward (green), and reverse (blue) genetic offsets under the SSP245 **(C)** and SSP585 **(D)** models. **(E)** Local, forward, and reverse genetic offsets across the four *M. baccata* populations for three future time intervals (2041–2060, 2061–2080, and 2081–2100) under the SSP585 scenario. For each box plot, the horizontal line signifies the median, the box boundaries represent the first and third quartiles, and the whiskers denote the data range (min to max). Sample sizes for each population are as follows: NE ($n = 49$), HB1 ($n = 13$), HB2 ($n = 29$), and NW ($n = 20$).

simultaneously exhibit high levels of both local and reverse genomic offset. This spatial pattern indicates that populations within the region not only experience significant local environmental fluctuations, but also face difficulties in obtaining suitable "donor" sources from other parts of the species' distribution range. In contrast, populations along the eastern coastline are characterized by a yellow-green coloration on the genomic landscape maps, reflecting the dual

pressure of increasing local environmental maladaptation and the necessity to migrate outward to track suitable habitats. Meanwhile, inland regions are dominated by dark green values, suggesting that adaptive pressure in these areas is primarily driven by forward genomic offset. However, the overall magnitude of this pressure is substantially lower than that recorded at the margins of the species' range (Figure 7C, D).

We compared local, forward, and reverse genomic offsets among the four populations under different climate scenarios and carbon emission pathways. The NE population consistently displayed the highest levels of genetic offset, with this offset continuing to increase over time. This pattern suggests that the NE population is more susceptible to environmental changes and confronts more pronounced adaptive challenges (Figure 7E). To evaluate the potential fitness consequences of these offsets, we examined whether populations predicted to experience greater genomic offsets also harbor a higher burden of deleterious mutations. Our analyses revealed no significant correlations between predicted genetic offsets (including local and reverse offsets) and the load of deleterious variants ($R^2 = 0.0018$, $P = 0.12$) (Figure S21). This lack of association suggests that future climatic vulnerability and accumulated genetic load may represent decoupled evolutionary pressures on *M. baccata* populations. Nevertheless, a subtle yet significant correlation was observed between forward genomic offset and deleterious variation ($R^2 = 0.23$, $P = 4e{-}9$). Notably, populations harboring a higher burden of homozygous deleterious mutations may face heightened adaptive barriers when attempting to track their shifting climatic niches.

DISCUSSION

CWRs represent a critical reservoir of genetic diversity for climate-resilient breeding, harboring alleles that enhance stress tolerance and crop resilience under rapid environmental change (Hilton and Gaut, 1998; Jarvis et al., 2008; Le Corre et al., 2020). Within the genus *Malus*, *M. baccata* is particularly distinguished by its extensive geographic distribution across heterogeneous environments (Li et al., 2023). Most existing T2T reference genomes in *Malus* have been derived from cultivated varieties, such as “Golden Delicious” (Malinsky et al., 2021), which restricts their utility for population-scale studies of wild relatives. Aligning resequencing data from wild apples to a domesticated reference genome often introduces bias in variant calling, results in the loss of wild-specific genomic information, and consequently leads to an underestimation of structural and functional diversity (Ahmad et al., 2025; Akopyan et al., 2025). To address this limitation, we present a haplotype-resolved T2T genome for *M. baccata*, which provides a high-resolution reference for downstream comparative and population genomics analyses. This assembly facilitates the precise identification of SNPs and SVs, revealing lineage-specific adaptive variants (Exposito-Alonso et al., 2019; Fagny and Austerlitz, 2021; Metheringham et al., 2025).

The absence of reproductive isolation among the genus *Malus* facilitates recurrent interspecific hybridization, resulting in extensive introgression across divergent populations (Whitehead et al., 2021; Ha et al., 2022). Our population genomic analyses reveal that the NE population exhibits reduced nucleotide diversity, extended ROH, and elevated inbreeding rates. These genomic signatures are characteristic of populations undergoing bottlenecks or localized isolation, mirroring patterns observed in other long-lived woody perennials, such as *Quercus robur* (Leroy et al., 2020). The observations also suggest that the NE population may harbor unique adaptive alleles of potential conservation value. Interestingly, HB1 showed a closer genetic affinity to NE despite its geographic proximity to NW and HB2. Furthermore, the *fb* statistics confirmed robust gene flow between the HB1 and NE populations (Figure 2E). These results demonstrate that the population structure in pollinated plants is shaped not only by geographic isolation but also by complex evolutionary histories and gene exchanges (Liang et al., 2025). The pattern of spatial discordance between geography and genetic structure is also observed in other perennial woody plants (Zhou et al., 2019; Ding et al., 2022), underscoring the influence of historical admixture on contemporary genomic diversity. The HB2 and NW populations exhibit higher genetic diversity while carrying a heavier burden of deleterious alleles, predominantly in the heterozygous state (Figure 3). In *M. baccata*, frequent interspecific hybridization and introgression appear to act as a double-edged sword; they introduce adaptive alleles that may facilitate local adaptation, while simultaneously increasing heterozygous genetic load (Liu et al., 2022; Robinson et al., 2023). While this heterozygous genetic load may be temporarily masked, its potential segregation in subsequent generations poses a significant challenge for climate-resilient breeding. Consequently, leveraging genomic selection and molecular design breeding to purge deleterious mutations while preserving adaptive alleles emerges as a critical priority for future research.

Although classical common garden or reciprocal transplant experiments remain the gold standard for validating local adaptation, their implementation in long-lived woody perennials is often constrained by extended generation times (Lowry and Willis, 2010; Lind et al., 2024; Lotterhos, 2024). By integrating LFMM, RDA, and PCAdapt, we reliably detected candidate adaptive genes associated with temperature, precipitation, and other environmental variables. This integrative strategy serves as a robust and practical alternative to traditional experimental designs, enabling high-confidence detection of selection signatures in the absence of long-term phenotypic trials. (Lasky et al., 2015; Wang et al., 2018; Kang et al., 2023). Among the identified loci, approximately 30% of adaptive loci were associated with one or more climate factors. The allele frequency spectra of these loci exhibited

strong correlations with geographic distribution, highlighting their potential contribution to local adaptation. Our analysis is entirely based on statistical inference. Therefore, we cannot determine whether the effects of the putative adaptive loci on phenotypic traits are mediated through changes in protein function or through regulation of gene expression. For the candidate genes *DREB1A/D* and *NAC6*, our current inferences are supported by multiple lines of indirect evidence. These include significant genotype-climate associations, functional annotations from other species, and stress-induced transcriptomic profiles from homologous species. Collectively, these results strongly support that these loci play a crucial role in regulating the adaptive stress response of *M. baccata*. In fact, direct demonstration of whether these variants affect gene expression under climatic stress would require targeted functional validation or population-scale expression mapping using allele-specific expression analysis, which could provide definitive evidence of their adaptive roles. Consequently, we emphasize that functional validation of adaptive loci and genes is crucial for elucidating the adaptive basis of this species.

Projecting GF models onto future climate scenarios further revealed substantial spatial heterogeneity in genomic vulnerability (Bay et al., 2018). Our projections reveal that the NE population and populations in the eastern coastal regions exhibit the highest genomic offsets, suggesting a heightened risk of maladaptation under future climate change (Figure 7). This geographic pattern is consistent with observations in other species, where peripheral populations were found to face disproportionately higher risks of maladaptation (Wang et al., 2023; Fiscus et al., 2025). These results underscore the integrated effects of genomic background, interspecific hybridization, and future climate change on the adaptive landscape of *M. baccata*. While these findings provide critical insights into the genomic basis of environmental adaptation in *M. baccata*, functional validation through controlled experiments remains essential to confirm their adaptive significance (Waldvogel et al., 2020; Lazic et al., 2024). Our findings provide a genomic and population framework that establishes a theoretical basis for climate-resilient apple breeding through the strategic utilization of wild alleles, and simultaneously offer a generalizable model for investigating adaptation in other perennial fruit crops (Zhao et al., 2023; Nyine et al., 2025).

From a conservation perspective, *in situ* and *ex situ* strategies should be considered for genetically diverse populations, such as HB2 and NW, which serve as critical reservoirs of alleles for buffering against environmental change. Conversely, vulnerable populations (NE) require urgent monitoring and may benefit from assisted gene flow, introducing pre-adaptive alleles from more resilient populations (Browne et al., 2019; Xiang et al., 2024; Anderson et al., 2025; Long et al., 2025). While gene introgression can introduce beneficial adaptive variation, it may simultaneously increase the burden of deleterious alleles, emphasizing the importance of carefully evaluating its costs and benefits in conservation and breeding programs (Huang et al., 2022; Fiscus et al., 2025). Climate-resilient breeding strategies in apples can be informed by targeting populations with complementary genetic features. For instance, the HB2 population offers high genetic diversity, while the NE population provides specific stress-adapted alleles, collectively enhancing the adaptive potential of apple cultivars (Fiscus et al., 2025). In some cases, preserving locally adapted populations may necessitate limiting excessive hybridization to maintain genomic integrity (Tobias et al., 2025). Furthermore, integrating long-term genomic monitoring with predictive offset modeling will facilitate the development of adaptive management frameworks (van der Valk and Dalén, 2024). Although our data and identified adaptive loci provide a valuable foundation for future breeding, they remain insufficient. To fully harness the adaptive potential of this species, future research should expand apple genomic resources to include wild and cultivated germplasm, as well as genotypic, phenotypic, and environmental datasets from diverse regions. Furthermore, integrating multidisciplinary fields, such as physiology, evolutionary biology, and population and quantitative genetics, will also enable breeders to more accurately predict environmental challenges and develop apple varieties with improved stress resistance and adaptability.

MATERIALS AND METHODS

Plant materials, sample collection, and genome resequencing

For genome assembly, *M. baccata* was collected from Jingpo Lake, Heilongjiang Province, China, and had been preserved in the National Germplasm Repository of Pear and Apple (Xingcheng, China). Fresh and healthy leaf tissues were immediately frozen in liquid nitrogen and sent to a commercial provider for high-molecular-weight DNA extraction, library preparation, and sequencing. High-molecular-weight genomic DNA was isolated using the DNeasy Plant Mini Kit (Qiagen, Germany) following the manufacturer's instructions. HiFi reads were generated on the PacBio Revio platform. For ONT sequencing, libraries containing DNA fragments > 150 kb were constructed and sequenced on the PromethION platform. Hi-C libraries were prepared using the DpnII restriction enzyme and sequenced on the NovaSeq. 6000 to facilitate chromosome-level scaffolding.

To obtain representative materials reflecting the native genetic diversity of *M. baccata*, field surveys were conducted between 2011 and 2015 across its natural distribution range. A total of 117 wild accessions were collected, including 111 individuals representing 13 natural populations. These samples predominantly originated from Heilongjiang, Jilin, Hebei, Shanxi, Gansu, and Shaanxi. The collected wild materials were subsequently propagated asexually and maintained for long-term conservation in the National Germplasm Repository of Pear and Apple (Xingcheng). Samples used in this study consisted of leaf tissues directly collected from repository-maintained plants. In addition, 16 previously published *Malus* and five pear (*Pyrus*) accessions (used as outgroups) were retrieved from the NCBI

database (Table S9). For genome annotation, publicly available RNA-seq datasets (SRR8156047, SRR8156048, SRR8156049, SRR8156050) were obtained from NCBI.

Genome assembly, quality assessment, and annotation

To estimate genome size and heterozygosity, *K*-mer frequency distributions were generated from HiFi reads using GenomeScope (v2.0) (Ranallo Benavidez et al., 2020). Haplotype-resolved assemblies were then constructed using hifiasm (v0.16) (Cheng et al., 2021) by integrating HiFi, ONT, and Hi-C data under default parameters. Hi-C data were processed with the Juicer (v1.5) (Durand et al., 2016b) pipeline, and chromosomal scaffolds were built and ordered with 3D-DNA (v201008) (Dudchenko et al., 2017). Manual editing of the interactive Hi-C contact maps for both haplotypes was performed using Juicebox (v1.11.08) (Durand et al., 2016a). Gap regions in the pseudomolecules were identified using custom Python scripts and corrected with overlapping contigs and ONT reads (Shi et al., 2023). The polished sequences were aligned to the assembly with Minimap2 (v2.24) (Li, 2018) and visualized in Integrative Genomics Viewer (IGV) (v2.13.1) (Robinson et al., 2011) to verify read support. Finally, the most consistent reads were used to replace gap sequences, yielding the completed assembly.

Assembly quality and completeness were evaluated with BUSCO (v5.2.2) using the "eudicots_odb10" dataset (Simão et al., 2015). Switch and Hamming errors in the phased assemblies were evaluated with the calc_switchErr pipeline (github.com/tangerzhang/calc_switchErr). Assembly continuity was assessed using N50 statistics, and mapping rates were determined by aligning raw reads to the final assemblies, with the resulting data processed using SAMtools (v1.9) (Danecek et al., 2021). Chromosomal organization was further examined through Hi-C contact maps visualized in Juicebox. Telomeric repeats were identified with the TIDK (v0.2.0) (https://github.com/tolkit/telomeric-identifier) using the *explore* and *search* functions, and summary statistics were obtained from the resulting plots. To identify and refine centromere boundaries, we inferred their positions based on the chromosomal distributions of different TE types, annotated genes, and HODOR repeats (Daccord et al., 2017), and visualized the results for both haplotypes in IGV (Daccord et al., 2017). Gene annotation was performed using an integrated approach, combining results from three pipelines: Genome-Wide-Annotation-Pipeline (https://github.com/unavailable-2374/Genome-Wide-Annotation-Pipeline), Liftoff (v1.6.1) (Shumate and Salzberg, 2021), and EviAnn (v2.0.2) (Zimin et al., 2025). InterProScan (v5.29) (Jones et al., 2014) was applied to predict potential protein domains. Functional annotation was first performed using eggNOG-mapper (v2.1.13) (Cantalapiedra et al., 2021). TEs were identified by the Extensive *de novo* TE Annotator (EDTA, v2.0.0) with default parameters (Ou et al., 2019).

Genome comparison and synteny analysis

Syntenic orthologs and structural rearrangements were identified using Synteny and Rearrangement Identifier (SyRI) (Goel et al., 2019) and visualized with Plotsr (Goel and Schneeberger, 2022). To detect SVs between cultivated and wild species, HiFi reads from GD and *M. baccata* were mapped to the MbaT2T_hap1 genome using Minimap2. Variants were subsequently called with Sniffles (v2.0.7) using default parameters (Sedlazeck et al., 2018) and filtered using the following criteria: length $\geq$ 50 bp, quality $\geq$ 60, "PASS", and "PRECISE". Genes that fully overlapped with heterozygous SVs within the two haplotypes were classified as hemizygous genes (Zhou et al., 2019). To investigate the functions of these hemizygous genes, sequences were aligned against reference databases using DIAMOND (v0.9.24) (Buchfink et al., 2015) with the parameters "blastp –k 1 –e 1e-5". Gene Ontology (GO) enrichment analysis was subsequently performed using clusterProfiler (Wu et al., 2021).

Environmental data and species distribution modeling

Environmental predictors were downloaded at a spatial resolution of 2.5 arc-min (~5 × 5 km), including 19 bioclimatic variables, monthly precipitation (prec6 and prec12), solar radiation (srad6 and srad12), vapor pressure (vapr6 and vapr12), and elevation, all from WorldClim v2 (https://worldclim.org/data/bioclim.html). In addition, seven soil variables (CEC_SOIL, CN_RATIO, ORG_CARBON, TOTAL_N, PH_WATER, CLAY, ELEC_COND) were obtained from the Harmonized World Soil Database (HWSD; https://www.fao.org/soils-portal/data-hub/en/). Species occurrence data were collected from two sources: 5,466 occurrence records of *M. baccata* downloaded from GBIF (gbif.org, accessed 2024) and 111 field-collected samples from our study.

We integrated spatial data with current environmental layers to predict the potential range of *M. baccata* surrounding our sampling sites. ENMTools (v1.3) (Warren et al., 2010) was used to reduce spatial autocorrelation by filtering duplicate occurrence records across the 19 bioclimatic variables. To reduce multicollinearity, highly correlated predictors (Spearman's $|r| > 0.8$) were excluded prior to modeling. Species distribution models were implemented in MaxEnt (Elith et al., 2011) using default settings.

Genome resequencing and variant calling

Genomic DNA from 111 wild *M. baccata* individuals was extracted using the DNeasy Plant Mini Kit (Qiagen). Libraries with an average insert size of 150 bp were sequenced on an Illumina NovaSeq. 6000 platform, generating paired-end reads per sample. Raw reads were aligned to the haplotype 1 reference genome of *M. baccata* using BWA-MEM (v0.7.17) (Li, 2013). Small variants were identified by GTX (http://www.gtxlab.com/product/cat) (v2.2.1) using the default parameters. To ensure high-quality variant calls, SNPs were filtered with VCFtools (v0.1.16) (Danecek et al., 2011) using the following criteria: QD > 2.0, QUAL > 30.0, GQ > 20.0, and missing genotype rate < 20%. SVs were detected using Delly (v1.1.6) (Rausch et al., 2012). Only "PASS" and "PRECISE" variants were retained, while sites with over 20% missing data were excluded.

Population structure and genetic statistics

Following LD pruning in PLINK (–indep-pairwise 100 50 0.2) (Purcell et al., 2007), 698,888 SNPs were retained for population structure analyses. PCA was performed using PLINK (v1.9) (Purcell et al., 2007), and individual ancestry proportions were estimated using ADMIXTURE (v1.30) with K values ranging from 2 to 12 (Alexander et al., 2009). For phylogenetic reconstruction, a thinned SNP dataset (VCFtools –thin 500) was used to infer a maximum-likelihood tree in IQ-TREE (v2.1.4) under the "GTR + I + G" model with 1,000 ultrafast bootstrap replicates (Nguyen et al., 2015). The resulting tree was visualized in iTOL (https://itol.embl.de/). Outgroup samples were included only to root the ML tree and were excluded from PCA and ADMIXTURE. The consistent results across PCA, ADMIXTURE, and phylogenetic inference allowed us to delineate four genetic groups (HB1, HB2, NE, and NW).

Population genetic statistics were computed in non-overlapping 25 kb windows. Nucleotide diversity (π) and pairwise genetic differentiation (F_{ST}) among populations were calculated using VCFtools. Genome-wide heterozygosity and ROH were estimated with PLINK. The total ROH length per individual was obtained by summing all detected ROH segments. Genomic introgression was assessed with Dsuite (v0.4) (Malinsky et al., 2021). Based on the ML phylogenetic framework, we calculated the *f*-branch statistics. The intensity of these events was subsequently visualized.

Detection of genomic selection signatures using PCAdapt

To detect loci associated with population structure, we employed PCAdapt (v4.3.3) (Luu et al., 2017), which does not require predefined population assignments. The number of principal components was determined from the scree plot. Multiple testing correction was performed using the Bonferroni method. SNPs and SVs with an adjusted P-value < 0.05 were identified as significant outliers, providing a high-confidence set of candidate loci under putative selection.

Estimation of genetic load

Genomic genetic load was estimated as the number of deleterious mutations per individual. Using the reference genome annotation, we applied the SIFT4G (v4.3.3) algorithm to predict the functional effects of SNPs (Vaser et al., 2016). Amino acid substitutions with scores ≤ 0.05 were classified as deleterious, whereas those with scores > 0.05 were considered tolerated. For each individual, deleterious alleles were quantified in terms of heterozygous, homozygous, and additive counts, with additive load calculated as the number of heterozygous deleterious alleles plus two times the number of homozygous deleterious alleles across SNPs and SVs (Zhou et al., 2019).

Identification of environment-associated genetic variants

To investigate associations between genetic variation and environmental factors, we applied latent factor mixed models (LFMM) implemented in the R package LEA (v3.10.1) (Frichot et al., 2013). Two final variant datasets were analyzed: SNPs (MAF > 0.1, LD pruned with 200 100 0.9) and SVs (MAF > 0.1). Missing SNP genotypes were imputed with Beagle (v4.1) (Browning and Browning, 2013), while SVs with missing data were removed. After filtering, 2,567,174 SNPs and 73,574 SVs were retained for downstream analyses. Before LFMM analysis, population structure was inferred using the pca() and snmf() functions in LEA, and the optimal number of latent factors was set to $K = 6$ to account for population structure confounding. LFMM models were fitted using the lfmm() function, with bioclimatic variables as predictors. P-values were extracted with lfmm.pvalues(), and significance was assessed using 5% FDR correction.

As a complementary approach, we performed redundancy analysis (RDA) to identify genetic variants strongly associated with multivariate environmental gradients (Capblancq and Forester, 2021). We first ranked the importance of the 19 bioclimatic variables using the "gradientForest" R package (Ellis et al., 2012) and reduced multicollinearity by excluding variables with pairwise correlations $|r| \geq 0.8$. Six uncorrelated predictors (BIO2, BIO4, BIO5, BIO12, BIO13, and BIO17) were retained and incorporated into RDA using the vegan package (v2.46) (Oksanen et al., 2018). Variants were considered significant if their loadings on any RDA axis exceeded three standard deviations from the mean. To further investigate whether these environmentally associated variants were driven by recent positive selection, we analyzed EHH patterns using the R package "rehh" (Gautier and Vitalis, 2012). To assess the robustness of adaptive variants, we examined allele frequency changes along environmental gradients and compared these patterns with a neutral reference set containing the same number of variants. We used the R package "gradientForest" to quantify allele frequency variation and calculate the cumulative importance as the sum of contributions from all climate variables across split points. Visualization was performed using the ggplot2 package in R (v3.4.2).

RNA-seq data processing and expression quantification

We collected publicly available RNA-seq datasets, primarily from samples subjected to various forms of abiotic stress. All clean RNA-seq reads were aligned to haplotype 1 of the *M. baccata* reference genome using STAR (v.2.7.11b) (Dobin et al., 2013). Subsequently, we performed differential expression analysis using the R package DESeq. 2 (v.1.36.0) (Love et al., 2014), with gene expression levels quantified as transcripts per million (TPM) from the raw counts generated by featureCounts.

Genomic offset assessment

To quantify genome-wide offset under future climate change, we applied the gradient forest-based offset metric. Nineteen bioclimatic variables were obtained from the WorldClim CMIP6 dataset for three future time periods (2041–2060, 2061–2080, and 2081–2100) under two emission scenarios (SSP245 and SSP585), at a spatial resolution of 2.5 arc-min

(~5 × 5 km). Three general circulation models (GCMs), including ACCESS-CM2, CMCC-ESM2, and GISS-E2-1-G, were used to capture model uncertainty. Genomic offset was estimated using the R package "gradientForest", with three complementary metrics calculated: local offset, reverse offset, and forward offset. These metrics were defined as Euclidean distances between genomic composition predicted under current and future climatic conditions. Correlations among offsets estimated under different GCMs were assessed using the "cor.test" function in R, and visualized with the corrplot package. The mean offset values across GCMs were further mapped as RGB composites, with red, green, and blue channels representing local, forward, and reverse offsets, respectively, and plotted using ggplot2 (Sang et al., 2022). Finally, we tested the associations between genomic offsets and deleterious variant accumulation by calculating Spearman's rank correlations (cor.test in R) between three offset metrics and genetic load indicators (homozygous, heterozygous, and additive) across the four genetic groups under each future scenario.

Data availability statement

All raw sequence data and the assembly of the *Malus baccata* genome are deposited in the NCBI Sequence Read Archive under project numbers PRJNA1332630, PRJNA1390481, and PRJNA1390482. These data are also available in the National Genomics Data Center (NGDC: PRJCA046864; https://ngdc.cncb.ac.cn/). The assembly and its annotation are deposited in Zenodo (https://zenodo.org/records/17205814). All analysis codes are deposited at https://github.com/suying77/Malus-Baccata-climate-adaptation.

ACKNOWLEDGEMENTS

This study was supported by the National Key R&D Program of China (No. 2023YFD1200100), the Agricultural Science and Technology Innovation Program (CAAS-ASTIP-2022-RIP-02), and the "Xingliao Talent Program" Project of Liaoning Province (XLYC2203177). It was also supported by the Project of State Key Laboratory of Tropical Crop Breeding (No. NKLTCB-RC202501, No. SKLTCBYWF202507, No. NKLTCBCXTD40, No. SKLTCBBSH202501, and No. SKLTCBBSH202502). We are grateful to all members of the Zhou Lab for valuable discussions and constructive comments.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

Y.F.Z., Y.G., and X.T. conceived and designed this project. Y.G., D.W., K.W., S.S., Z.C.L., W.T., Y.M.S., and Z.L. collected leaf tissues and recorded corresponding geographical coordinates of all samples. X.C., T.F., R.Z., Z.Q.L., W.W., S.F.Y., and Y.C.Z. were responsible for genome assembly and annotation. Y.S., X.X., Y.X., Y.H., and H.X. performed the population genetic analyses. Y.S., F.Z., and Z.M. conducted the local adaptation analyses. L.W., Z.X., X.W., Y.P., and J.L. collected and deposited the data. Y.S. wrote the manuscript. Y.F.Z., X.Y., and R.L. supervised the study. All authors participated in revising the manuscript and approved the final content.

Edited by: Yuannian Jiao, Institute of Botany, CAS, China

Received Nov. 10, 2025; **Accepted** Jan. 30, 2026; **Published** Feb. 26, 2026

OO: OnlineOpen

REFERENCES

Ahmad, B., Su, Y., Hao, Y., Razzaq, T., Arshad, R., Zhang, Y., Zhang, Y., Wang, X., Huang, G., and Su, X. (2025). Mango pangenome reveals dramatic impacts of reference bias on genomic analyses. Hortic. Res. **12:** uhaf166.

Aitken, S.N., and Bemmels, J.B. (2016). Time to get moving: Assisted gene flow of forest trees. Evol. Appl. **9:** 271–290.

Aitken, S.N., Jordan, R., and Tumas, H.R. (2024). Conserving evolutionary potential: Combining landscape genomics with established methods to inform plant conservation. Annu. Rev. Plant Biol. **75:** 707–736.

Aitken, S.N., Yeaman, S., Holliday, J.A., Wang, T., and Curtis-McLane, S. (2008). Adaptation, migration or extirpation: Climate change outcomes for tree populations. Evol. Appl. **1:** 95–111.

Akopyan, M., Genchev, M., Armstrong, E.E., and Mooney, J.A. (2025). Reference genome choice compromises population genetic analyses. Cell **188:** 6939–6952.

Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. **19:** 1655–1664.

Anderson, J.T., DeMarche, M.L., Denney, D.A., Breckheimer, I., Santangelo, J., and Wadgymar, S.M. (2025). Adaptation and gene flow are insufficient to rescue a montane plant under climate change. Science **388:** 525–531.

Bay, R.A., Harrigan, R.J., Underwood, V.L., Gibbs, H.L., Smith, T.B., and Ruegg, K. (2018). Genomic signals of selection predict climate-driven population declines in a migratory bird. Science **359:** 83–86.

Belin, C., Bashandy, T., Cela, J., Delorme-Hinoux, V., Riondet, C., and Reichheld, J. (2015). A comprehensive study of thiol reduction gene expression under stress conditions in *Arabidopsis thaliana*. Plant Cell Environ. **38:** 299–314.

Bohra, A., Kilian, B., Sivasankar, S., Caccamo, M., Mba, C., McCouch, S.R., and Varshney, R.K. (2022). Reap the crop wild relatives for breeding future crops. Trends Biotechnol. **40:** 412–431.

Brauer, C.J., Sandoval-Castillo, J., Gates, K., Hammer, M.P., Unmack, P.J., Bernatchez, L., and Beheregaray, L.B. (2023). Natural hybridization reduces vulnerability to climate change. Nat. Clim. Change **13:** 282–289.

Browne, L., Wright, J.W., Fitz-Gibbon, S., Gugger, P.F., and Sork, V.L. (2019). Adaptational lag to temperature in valley oak (*Quercus lobata*) can be mitigated by genome-informed assisted gene flow. Proc. Natl. Acad. Sci. U. S. A. **116:** 25179–25185.

Browning, B.L., and Browning, S.R. (2013). Improving the accuracy and efficiency of identity-by-descent detection in population data. Genetics **194:** 459–471.

Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat. Methods **12:** 59–60.

Butchart, S.H., Walpole, M., Collen, B., Van Strien, A., Scharlemann, J.P., Almond, R.E., Baillie, J.E., Bomhard, B., Brown, C., and Bruno, J. (2010). Global biodiversity: Indicators of recent declines. Science **328:** 1164–1168.

Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol. Biol. Evol. **38:** 5825–5829.

Capblancq, T., and Forester, B.R. (2021). Redundancy analysis: A Swiss Army Knife for landscape genomics. Methods Ecol. Evol **12:** 2298–2309.

Carrasco, D., Zhou-Tsang, A., Rodriguez-Izquierdo, A., Ocete, R., Revilla, M.A., and Arroyo-García, R. (2022). Coastal wild grapevine accession (*Vitis vinifera* L. ssp. *sylvestris*) shows distinct late and early transcriptome changes under salt stress in comparison to commercial rootstock Richter 110. Plants (Basel) **11:** 2688.

Castañeda-Álvarez, N.P., Khoury, C.K., Achicanoy, H.A., Bernau, V., Dempewolf, H., Eastwood, R.J., Guarino, L., Harker, R.H., Jarvis, A., and Maxted, N. (2016). Global conservation priorities for crop wild relatives. Nat. Plants **2:** 1–6.

Charlesworth, B., and Charlesworth, D. (1998). Some evolutionary consequences of deleterious mutations. Genetica **102:** 3–19.

Chen, X., Cornille, A., An, N., Xing, L., Ma, J., Zhao, C., Wang, Y., Han, M., and Zhang, D. (2023). The East Asian wild apples, *Malus baccata* (L.) Borkh and *Malus hupehensis* (Pamp.) Rehder., are additional contributors to the genomes of cultivated European and Chinese varieties. Mol. Ecol. **32:** 5125–5139.

Chen, X., Li, S., Zhang, D., Han, M., Jin, X., Zhao, C., Wang, S., Xing, L., Ma, J., and Ji, J. (2019). Sequencing of a wild apple (*Malus baccata*) genome unravels the differences between cultivated and wild apple species regarding disease resistance and cold tolerance. G3: Genes Genomes Genetics **9:** 2051–2060.

Cheng, H.Y., Concepcion, G.T., Feng, X.W., Zhang, H.W., and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat. Methods **18:** 170–175.

Cui, X., Zhang, P., Chen, C., and Zhang, J. (2023). VyUSPA3, a universal stress protein from the Chinese wild grape *Vitis yeshanensis*, confers drought tolerance to transgenic V. vinifera. Plant Cell Rep. **42:** 181–196.

Daccord, N., Celton, J.-M., Linsmith, G., Becker, C., Choisne, N., Schijlen, E., Van de Geest, H., Bianco, L., Micheletti, D., and Velasco, R. (2017). High-quality *de novo* assembly of the apple genome and methylome dynamics of early fruit development. Nat. Genet. **49:** 1099–1106.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics **27:** 2156–2158.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., and Davies, R.M. (2021). Twelve years of SAMtools and BCFtools. Gigascience **10:** giab008.

Ding, Y.-M., Cao, Y., Zhang, W.-P., Chen, J., Liu, J., Li, P., Renner, S.S., Zhang, D.-Y., and Bai, W.-N. (2022). Population-genomic analyses reveal bottlenecks and asymmetric introgression from Persian into iron walnut during domestication. Genome Biol. **23:** 145.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: Ultrafast universal RNA-seq aligner. Bioinformatics **29:** 15–21.

Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., and Aiden, A.P. (2017). *De novo* assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science **356:** 92–95.

Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., and Aiden, E.L. (2016a). Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. **3:** 99–101.

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016b). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. **3:** 95–98.

Edelman, N.B., and Mallet, J. (2021). Prevalence and adaptive impact of introgression. Annu. Rev. Genet. **55:** 265–283.

Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., and Yates, C.J. (2011). A statistical explanation of MaxEnt for ecologists. Diversity Distrib. **17:** 43–57.

Ellis, N., Smith, S.J., and Pitcher, C.R. (2012). Gradient forests: Calculating importance gradients on physical predictors. Ecology **93:** 156–168.

Exposito-Alonso, M., Burbano, H.A., Bossdorf, O., Nielsen, R., and Weigel, D. (2019). Natural selection on the *Arabidopsis thaliana* genome in present and future climates. Nature **573:** 126–129.

Fagny, M., and Austerlitz, F. (2021). Polygenic adaptation: Integrating population genetics and gene regulatory networks. Trends Genet. **37:** 631–638.

Feng, J., Dan, X., Cui, Y., Gong, Y., Peng, M., Sang, Y., Ingvarsson, P.K., and Wang, J. (2024) Integrating evolutionary genomics of forest trees to inform future tree breeding amid rapid climate change. Plant Commun. **5:** 101044.

Fior, S., Luqman, H., Scharmann, M., Pålsson, A., de Jonge, J., Zoller, S., Zemp, N., Gargano, D., Wegmann, D., and Widmer, A. (2025). Ancient alleles drive contemporary climate adaptation in an alpine plant. Science **390:** 59–64.

Fiscus, C.J., Aguirre-Liguori, J.A., Gaut, G.R., and Gaut, B.S. (2025). Mutational load and adaptive variation are shaped by climate and species range dynamics in *Vitis arizonica*. New Phytol. **247:** 998–1014.

Frichot, E., Schoville, S.D., Bouchard, G., and François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. Mol. Biol. Evol. **30:** 1687–1699.

Gao, G., Yan, L., Tong, K., Yu, H., Lu, M., Wang, L., and Niu, Y. (2024). The potential and prospects of modified biochar for comprehensive management of salt-affected soils and plants: A critical review. Sci. Total Environ. **912:** 169618.

Garg, V., Bohra, A., Mascher, M., Spannagl, M., Xu, X., Bevan, M.W., Bennetzen, J.L., and Varshney, R.K. (2024). Unlocking plant genetics with telomere-to-telomere genome assemblies. Nat. Genet. **56:** 1788–1799.

Gaut, B.S., Seymour, D.K., Liu, Q., and Zhou, Y. (2018). Demography and its effects on genomic variation in crop domestication. Nat. Plants **4:** 512–520.

Gautier, M., and Vitalis, R. (2012). rehh: An R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics **28:** 1176–1177.

Goel, M., and Schneeberger, K. (2022). plotsr: Visualizing structural similarities and rearrangements between multiple genomes. Bioinformatics **38:** 2922–2926.

Goel, M., Sun, H., Jiao, W., and Schneeberger, K. (2019). SyRI: Finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome Biol. **20:** 1–13.

Gonzalo, M.J., Nájera, I., Baixauli, C., Gil, D., Montoro, T., Soriano, V., Olivieri, F., Rigano, M.M., Ganeva, D., and Grozeva-Tileva, S. (2021). Identification of tomato accessions as source of new genes for improving heat tolerance: From controlled experiments to field. BMC Plant Biol. **21:** 345.

Gougherty, A.V., Keller, S.R., and Fitzpatrick, M.C. (2021). Maladaptation, migration and extirpation fuel climate change risk in a forest tree species. Nat. Clim. Change **11:** 166–171.

Gygax, M., Gianfranceschi, L., Liebhard, R., Kellerhals, M., Gessler, C., and Patocchi, A. (2004). Molecular markers linked to the apple scab resistance gene *Vbj* derived from *Malus baccata* jackii. Theor. Appl. Genet. **109:** 1702–1709.

Ha, Y.-H., Gil, H.-Y., Kim, S.-C., Choi, K., and Kim, J.-H. (2022) Genetic structure and geneflow of Malus across the Korean Peninsula using genotyping-by-sequencing. Sci. Rep. **12:** 16262.

Harrison, R.G. (1993). *Hybrid zones and the evolutionary process*. New York, NY: Oxford University Press.

Hemmat, M., Brown, S., Aldwinckle, H., Mehlenbacher, S., and Weeden, N. (2003). Identification and mapping of markers for resistance to apple scab from 'Antonovka' and 'Hansen's baccata# 2'. Acta Hortic. **622:** 153–161.

Hereford, J. (2009). A quantitative survey of local adaptation and fitness trade-offs. Am. Nat. **173:** 579–588.

Hilton, H., and Gaut, B.S. (1998). Speciation and domestication in maize and its wild relatives: Evidence from the globulin-1 gene. Genetics **150:** 863–872.

Hou, Y., Gan, J., Fan, Z., Sun, L., Garg, V., Wang, Y., Li, S., Bao, P., Cao, B., and Varshney, R.K. (2024). Haplotype-based pangenomes reveal genetic variations and climate adaptations in moso bamboo populations. Nat. Commun. **15:** 8085.

Huang, K., Ostevik, K.L., Elphinstone, C., Todesco, M., Bercovich, N., Owens, G.L., and Rieseberg, L.H. (2022). Mutation load in sunflower inversions is negatively correlated with inversion heterozygosity. Mol. Biol. Evol. **39:** msac101.

Jarvis, A., Lane, A., and Hijmans, R.J. (2008). The effect of climate change on crop wild relatives. Agric., Ecosyst. Environ. **126:** 13–23.

Jia, K.H., Zhao, W., Maier, P.A., Hu, X.G., Jin, Y., Zhou, S.S., Jiao, S.Q., El-Kassaby, Y.A., Wang, T., and Wang, X.R. (2020). Landscape genomics predicts climate change-related genetic offset for the widespread *Platycladus orientalis* (*Cupressaceae*). Evol. Appl. **13:** 665–676.

Jiang, Q., Shen, Y., Wu, L., Jiang, Z., and Yao, X. (2025). Genomic signatures of local adaptation to precipitation and solar radiation in kiwifruit. Plant Divers. **47:** 733–745.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., and Nuka, G. (2014). InterProScan 5: Genome-scale protein function classification. Bioinformatics **30:** 1236–1240.

Jung, K.-H., Gho, H.-J., Nguyen, M.X., Kim, S.-R., and An, G. (2013). Genome-wide expression analysis of *HSP70* family genes in rice and identification of a cytosolic *HSP70* gene highly induced under heat stress. Funct. Integr. Genomics **13:** 391–402.

Kang, M., Wu, H., Liu, H., Liu, W., Zhu, M., Han, Y., Liu, W., Chen, C., Song, Y., and Tan, L. (2023). The pan-genome and local adaptation of *Arabidopsis thaliana*. Nat. Commun. **14:** 6259.

Khoury, C.K., Greene, S., Wiersema, J., Maxted, N., Jarvis, A., and Struik, P.C. (2013). An inventory of crop wild relatives of the United States. Crop Sci. **53:** 1496–1508.

Komatsu, S., Thibaut, D., Hiraga, S., Kato, M., Chiba, M., Hashiguchi, A., Tougou, M., Shimamura, S., and Yasue, H. (2011). Characterization of a novel flooding stress-responsive alcohol dehydrogenase expressed in soybean roots. Plant Mol. Biol. **77:** 309–322.

Lasky, J.R., Upadhyaya, H.D., Ramu, P., Deshpande, S., Hash, C.T., Bonnette, J., Juenger, T.E., Hyma, K., Acharya, C., and Mitchell, S.E. (2015). Genome-environment associations in sorghum landraces predict adaptive traits. Sci. Adv. **1:** e1400218.

Lazic, D., Geßner, C., Liepe, K.J., Lesur-Kupin, I., Mader, M., Blanc-Jolivet, C., Gömöry, D., Liesebach, M., González-Martínez, S.C., and Fladung, M. (2024). Genomic variation of European beech reveals signals of local adaptation despite high levels of phenotypic plasticity. Nat. Commun. **15:** 8553.

Le Corre, V., Siol, M., Vigouroux, Y., Tenaillon, M.I., and Délye, C. (2020). Adaptive introgression from maize has facilitated the establishment of teosinte as a noxious weed in Europe. Proc. Natl. Acad. Sci. U. S. A. **117:** 25618–25627.

Lee, D.K., Chung, P.J., Jeong, J.S., Jang, G., Bang, S.W., Jung, H., Kim, Y.S., Ha, S.H., Choi, Y.D., and Kim, J.K. (2017). The rice *OsNAC6* transcription factor orchestrates multiple molecular mechanisms involving root structural adaptions and nicotianamine biosynthesis for drought tolerance. Plant Biotechnol. J. **15:** 754–764.

Leroy, T., Louvet, J.M., Lalanne, C., Le Provost, G., Labadie, K., Aury, J.M., Delzon, S., Plomion, C., and Kremer, A. (2020). Adaptive introgression as a driver of local adaptation to climate in European white oaks. New Phytol. **226:** 1171–1182.

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

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics **34:** 3094–3100.

Li, H., and Durbin, R. (2024). Genome assembly in the telomere-to-telomere era. Nat. Rev. Genet. **25:** 658–670.

Li, N., Meng, H., Li, S., Zhang, Z., Zhao, X., Wang, S., Liu, A., Li, Q., Song, Q., and Li, X. (2020). Two plastid fatty acid exporters contribute to seed oil accumulation in *Arabidopsis*. Plant Physiol. **182:** 1910–1919.

Li, W., Chu, C., Zhang, T., Sun, H., Wang, S., Liu, Z., Wang, Z., Li, H., Li, Y., and Zhang, X. (2025). Pan-genome analysis reveals the evolution and diversity of *Malus*. Nat. Genet. **57:** 1274–1286.

Li, X., Ma, Z., Song, Y., Shen, W., Yue, Q., Khan, A., Tahir, M.M., Wang, X., Malnoy, M., and Ma, F. (2023). Insights into the molecular mechanisms underlying responses of apple trees to abiotic stresses. Hortic. Res. **10:** uhad144.

Liang, X., Li, J., Yang, Y., Jiang, C., and Guo, Y. (2024). Designing salt stress-resilient crops: Current progress and future challenges. J. Integr. Plant Biol. **66:** 303–329.

Liang, Y.-Y., Liu, H., Lin, Q.-Q., Shi, Y., Zhou, B.-F., Wang, J.-S., Chen, X.-Y., Shen, Z., Qiao, L.-J., and Niu, J.-W. (2025). Pan-genome analysis reveals local adaptation to climate driven by introgression in oak species. Mol. Biol. Evol. **42:** msaf088.

Lind, B.M., Candido-Ribeiro, R., Singh, P., Lu, M., Obreht Vidakovic, D., Booker, T.R., Whitlock, M.C., Yeaman, S., Isabel, N., and Aitken, S.N. (2024). How useful is genomic data for predicting maladaptation to future climate? Glob. Change Biol. **30:** e17227.

Liu, S., Zhang, L., Sang, Y., Lai, Q., Zhang, X., Jia, C., Long, Z., Wu, J., Ma, T., and Mao, K. (2022). Demographic history and natural selection shape patterns of deleterious mutation load and barriers to introgression across Populus genome. Mol. Biol. Evol. **39:** msac008.

Long, Q., Cao, S., Huang, G., Wang, X., Liu, Z., Liu, W., Wang, Y., Xiao, H., Peng, Y., and Zhou, Y. (2024). Population comparative genomics discovers gene gain and loss during grapevine domestication. Plant Physiol. **195:** 1401–1413.

Long, Z., Sang, Y., Feng, J., Zhang, X., Shi, T., Zhang, L., Mao, K., Rieseberg, L.H., Liu, J., and Wang, J. (2025). Evolutionary genomics unravels the responses and adaptation to climate change in a key alpine forest tree species. Mol. Biol. Evol. **42:** msaf116.

López-Goldar, X., and Agrawal, A.A. (2021). Ecological interactions, environmental gradients, and gene flow in local adaptation. Trends Plant Sci. **26:** 796–809.

Lotterhos, K.E. (2023). The paradox of adaptive trait clines with nonclinal patterns in the underlying genes. Proc. Natl. Acad. Sci. U. S. A. **120:** e2220313120.

Lotterhos, K.E. (2024). Principles in experimental design for evaluating genomic forecasts. Methods Ecol. Evol. **15:** 1466–1482.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq. 2. Genome Biol. **15:** 550.

Lowry, D.B., and Willis, J.H. (2010). A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. PLoS Biol. **8:** e1000500.

Luu, K., Bazin, E., and Blum, M.G. (2017). pcadapt: An R package to perform genome scans for selection based on principal component analysis. Mol. Ecol. Resour. **17:** 67–77.

Mace, G.M., Collar, N.J., Gaston, K.J., Hilton-Taylor, C., Akçakaya, H.R., Leader-Williams, N., Milner-Gulland, E.J., and Stuart, S.N. (2008). Quantification of extinction risk: IUCN's system for classifying threatened species. Conserv. Biol. **22:** 1424–1442.

Malanoski, C.M., Farnsworth, A., Lunt, D.J., Valdes, P.J., and Saupe, E.E. (2024). Climate change is an important predictor of extinction risk on macroevolutionary timescales. Science **383:** 1130–1134.

Malinsky, M., Matschiner, M., and Svardal, H. (2021). Dsuite-fast D-statistics and related admixture evidence from VCF files. Mol. Ecol. Resour. **21:** 584–595.

Metheringham, C.L., Plumb, W.J., Flynn, W.R., Stocks, J.J., Kelly, L.J., Nemesio Gorriz, M., Grieve, S.W., Moat, J., Lines, E.R., and Buggs, R.J. (2025). Rapid polygenic adaptation in a wild population of ash trees under a novel fungal epidemic. Science **388:** 1422–1425.

Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. **32:** 268–274.

Nyine, M., Davidson, D., Adhikari, E., Clinesmith, M., Wang, H., Akhunova, A., Fritz, A., and Akhunov, E. (2025). Genomic signals of ecogeographic adaptation in a wild relative are associated with improved wheat performance under drought stress. Genome Biol. **26:** 35.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'hara, R., Simpson, G.L., Solymos, P., Stevens, M.H.H., and Wagner, H. (2018). vegan: Community ecology package. R package version **2:** 4–6.

Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J.R., Hellinga, A.J., Lugo, C.S.B., Elliott, T.A., Ware, D., and Peterson, T. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. **20:** 1–18.

Peng, Y., Wang, Y., Liu, Y., Fang, X., Cheng, L., Long, Q., Su, D., Zhang, T., Shi, X., and Xu, X. (2025). The genomic and epigenomic landscapes of hemizygous genes across crops with contrasting reproductive systems. Proc. Natl. Acad. Sci. U. S. A. **122:** e2422487122.

Purcell, S., Neale, B., Todd Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., and Daly, M.J. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. **81:** 559–575.

Ranallo Benavidez, T.R., Jaron, K.S., and Schatz, M.C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat. Commun. **11:** 1432.

Rausch, T., Zichner, T., Schlattl, A., Stuetz, A.M., Benes, V., and Korbel, J.O. (2012). DELLY: Structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics **28:** I333–I339.

Rivero, R.M., Mittler, R., Blumwald, E., and Zandalinas, S.I. (2022). Developing climate-resilient crops: Improving plant tolerance to stress combination. Plant J. **109:** 373–389.

Robinson, J., Kyriazis, C.C., Yuan, S.C., and Lohmueller, K.E. (2023). Deleterious variation in natural populations and implications for conservation genetics. Annu. Rev. Anim. Biosci. **11:** 93–114.

Robinson, J.P., Harris, S.A., and Juniper, B. (2001). Taxonomy of the genus *Malus* Mill.(*Rosaceae*) with emphasis on the cultivated apple, *Malus domestica* Borkh. Plant Syst. Evol. **226:** 35–58.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. **29:** 24–26.

Sang, Y., Long, Z., Dan, X., Feng, J., Shi, T., Jia, C., Zhang, X., Lai, Q., Yang, G., and Zhang, H. (2022). Genomic insights into local adaptation and future climate-induced vulnerability of a keystone forest tree in East Asia. Nat. Commun. **13:** 6541.

Sedlazeck, F.J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., Von Haeseler, A., and Schatz, M.C. (2018). Accurate detection of complex structural variations using single-molecule sequencing. Nat. Methods **15:** 461–468.

Shi, X., Cao, S., Wang, X., Huang, S., Wang, Y., Liu, Z., Liu, W., Leng, X., Peng, Y., and Wang, N. (2023). The complete reference genome for grapevine (*Vitis vinifera* L.) genetics and breeding. Hortic. Res. **10:** uhad061.

Shumate, A., and Salzberg, S.L. (2021). Liftoff: Accurate mapping of gene annotations. Bioinformatics **37:** 1639–1643.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics **31:** 3210–3212.

Su, Y., Yang, X., Wang, Y., Li, J., Long, Q., Cao, S., Wang, X., Liu, Z., Huang, S., and Chen, Z. (2024). Phased telomere-to-telomere reference genome and pangenome reveal an expansion of resistance genes during apple domestication. Plant Physiol. **195:** 2799–2814.

Tobias, J.A., Bullock, J.M., Dicks, L.V., Forester, B.R., and Razgour, O. (2025). Biodiversity conservation requires integration of species-centric and process-based strategies. Proc. Natl. Acad. Sci. U. S. A. **122:** e2410936122.

Tougou, M., Hashiguchi, A., Yukawa, K., Nanjo, Y., Hiraga, S., Nakamura, T., Nishizawa, K., and Komatsu, S. (2012). Responses to flooding stress in soybean seedlings with the alcohol dehydrogenase transgene. Plant Biotechnol. (Tokyo) **29:** 301–305.

Urban, M.C. (2015). Accelerating extinction risk from climate change. Science **348:** 571–573.

Usman, M.G., Rafii, M.Y., Martini, M.Y., Yusuff, O.A., Ismail, M.R., and Miah, G. (2017). Molecular analysis of *Hsp70* mechanisms in plants and their function in response to stress. Biotechnol. Genet. Eng. Rev. **33:** 26–39.

van der Valk, T., and Dalén, L. (2024). From genomic threat assessment to conservation action. Cell **187:** 1038–1041.

Vaser, R., Adusumalli, S., Leng, S.N., Sikic, M., and Ng, P.C. (2016). SIFT missense predictions for genomes. Nat. Protoc. **11:** 1–9.

Vincent, H., Wiersema, J., Kell, S., Fielder, H., Dobbie, S., Castañeda-Álvarez, N.P., Guarino, L., Eastwood, R., León, B., and Maxted, N. (2013). A prioritized crop wild relative inventory to help underpin global food security. Biol. Conserv. **167:** 265–275.

Vonapartis, E., Mohamed, D., Li, J., Pan, W., Wu, J., and Gazzarrini, S. (2022). *CBF4/DREB1D* represses *XERICO* to attenuate ABA, osmotic and drought stress responses in *Arabidopsis*. Plant J. **110:** 961–977.

Waldvogel, A.-M., Feldmeyer, B., Rolshausen, G., Exposito-Alonso, M., Rellstab, C., Kofler, R., Mock, T., Schmid, K., Schmitt, I., and

Bataillon, T. (2020). Evolutionary genomics can improve prediction of species' responses to climate change. Evol. Lett **4:** 4–18.

Wang, J., Ding, J., Tan, B., Robinson, K.M., Michelson, I.H., Johansson, A., Nystedt, B., Scofield, D.G., Nilsson, O., and Jansson, S. (2018). A major locus controls local adaptation and adaptive life history variation in a perennial plant. Genome Biol. **19:** 72.

Wang, N., Cao, S., Liu, Z., Xiao, H., Hu, J., Xu, X., Chen, P., Ma, Z., Ye, J., and Chai, L. (2023). Genomic conservation of crop wild relatives: A case study of citrus. PLoS Genet. **19:** e1010811.

Wang, Y., Li, W., Xu, X., Qiu, C., Wu, T., Wei, Q., Ma, F., and Han, Z. (2019). Progress of apple rootstock breeding and its use. Hortic. Plant J. **5:** 183–191.

Warren, D.L., Glor, R.E., and Turelli, M. (2010). ENMTools: A toolbox for comparative studies of environmental niche models. Ecography **33:** 607–611.

Whitehead, S.R., Wisniewski, M.E., Droby, S., Abdelfattah, A., Freilich, S., and Mazzola, M. (2021). The apple microbiome: Structure, function, and manipulation for improved plant health. In *The Apple Genome Compendium of Plant Genomes*. S.S. Korban, ed, (Cham: Springer), pp. 341–382.

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., and Zhan, L. (2021) clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb) **2:** 100141.

Xiang, X., Zhou, X., Zi, H., Wei, H., Cao, D., Zhang, Y., Zhang, L., and Hu, J. (2024). Populus cathayana genome and population resequencing provide insights into its evolution and adaptation. Hortic. Res. **11:** uhad255.

Xiong, W., Reynolds, M.P., Montes, C., Crossa, J., Snapp, S., Akin, B., Mesut, K., Ozdemir, F., Li, H., and He, Z. (2024). New wheat breeding paradigms for a warming climate. Nat. Clim. Change **14:** 869–875.

Yang, W., Liu, X.-D., Chi, X.-J., Wu, C.-A., Li, Y.-Z., Song, L.-L., Liu, X.-M., Wang, Y.-F., Wang, F.-W., and Zhang, C. (2011). Dwarf apple *MbDREB1* enhances plant tolerance to low temperature, drought, and salt stress via both ABA-dependent and ABA-independent pathways. Planta **233:** 219–229.

Zhang, F., Long, R., Ma, Z., Xiao, H., Xu, X., Liu, Z., Wei, C., Wang, Y., Peng, Y., and Yang, X. (2024a). Evolutionary genomics of climatic adaptation and resilience to climate change in alfalfa. Mol. Plant **17:** 867–883.

Zhang, M., Xu, X., Zhang, T., Liu, Z., Wang, X., Shi, X., Peng, W., Wang, X., Chen, Z., and Zhao, R. (2025a). The dynamics of wild Vitis species in response to climate change facilitate the breeding of grapevine and its rootstocks with climate resilience. Hortic. Res. **12:** uhaf104.

Zhang, T., Peng, W., Xiao, H., Cao, S., Chen, Z., Su, X., Luo, Y., Liu, Z., Peng, Y., and Yang, X. (2024b). Population genomics highlights structural variations in local adaptation to saline coastal environments in woolly grape. J. Integr. Plant Biol. **66:** 1408–1426.

Zhang, X., Guo, R., Shen, R., Landis, J.B., Jiang, Q., Liu, F., Wang, H., and Yao, X. (2023). The genomic and epigenetic footprint of local adaptation to variable climates in kiwifruit. Hortic. Res. **10:** uhad031.

Zhang, Z., Liu, W., Zhang, T., Chen, Z., Zhang, M., Yao, J., Liu, Y., Li, J., Fan, T., and Zhao, R. (2025b). Population genomics of *Vitis pseudoreticulata* reveals the genetic basis of fungal resistance in grapevine. Hortic. Plant J., https://doi.org/10.1016/j.hpj.2025.01.007

Zhao, X., Guo, Y., Kang, L., Yin, C., Bi, A., Xu, D., Zhang, Z., Zhang, J., Yang, X., and Xu, J. (2023). Population genomics unravels the Holocene history of bread wheat and its relatives. Nat. Plants **9:** 403–419.

Zhou, Y., Minio, A., Massonnet, M., Solares, E., Lv, Y., Beridze, T., Cantu, D., and Gaut, B.S. (2019). The population genetics of structural variants in grapevine domestication. Nat. Plants **5:** 965–979.

Zhou, Y., Zhang, L., Liu, J., Wu, G., and Savolainen, O. (2014). Climatic adaptation and ecological divergence between two closely related pine species in Southeast China. Mol. Ecol. **23:** 3504–3522.

Zimin, A.V., Puiu, D., Pertea, M., Yorke, J.A., and Salzberg, S.L. (2025). Efficient evidence-based genome annotation with EviAnn. bioRxiv, 2025.05.07.652745. https://doi.org/10.1101/2025.05.07.652745

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article: http://onlinelibrary.wiley.com/doi/10.1111/jipb.70194/suppinfo

Figure S1. Assessment of *Malus baccata* genome size and heterozygosity by *k*-mer
Figure S2. Genome-wide Hi-C contact heatmap of chromatin interaction frequencies
Figure S3. Genomic distribution of genes, HODOR sequences, and transposable elements across haplotypes
Figure S4. Comparison between *M. baccata* haplotype 1 and other genomes
Figure S5. Phylogenetic tree generated using the GTR + I + G model
Figure S6. Identity-by-descent (IBD) heatmap across the four groups (NE, HB1, HB2, and NW)
Figure S7. Frequency distributions of environmental variables
Figure S8. Correlation matrix of 19 bioclimatic variables
Figure S9. The ranked importance of 19 environmental variables was determined by gradient forest analysis
Figure S10. Gene Ontology (GO) enrichment analysis of candidate adaptive genes
Figure S11. Distribution of core adaptive variants across genomic regions
Figure S12. Identification of adaptive variants using RDA and LFMM
Figure S13. Genome-wide variation associated with various environmental variables
Figure S14. Distribution and proportion of genetic variants detected across multiple bioclimatic variables
Figure S15. Cumulative importance curve distinguishing adaptive and neutral SNPs along the temperature seasonality (BIO4) gradient
Figure S16. Expression of adaptive genes under cold treatment
Figure S17. Expression dynamics of adaptive genes under water treatment
Figure S18. Pairwise correlations between traditional gradient forest-based genetic offsets and three migration-integrated offset types (forward, local, and reverse offsets)
Figure S19. Projections of genomic offset under future climate scenarios in 2041–2060
Figure S20. Projections of genomic offset under future climate scenarios in 2061–2080
Figure S21. Relationship between genetic load and genomic offsets under future climate scenarios
Table S1. Statistics of raw sequencing data utilized for the *M. baccata* genome assembly
Table S2. Comparative summary of genomic features between previously published and newly assembled *M. baccata* genomes
Table S3. Summary of annotated transposable elements in the two haplotype-resolved assemblies of *M. baccata*
Table S4. Summary of SVs identified through comparative genomic analysis between published *M. baccata* genomes and the MbaT2T_hap1 reference
Table S5. Summary of SVs between MbaT2T_hap1 and MbaT2T_hap2
Table S6. Identification of hemizygous genes in the *M. baccata* genome
Table S7. Summary of SVs identified between the MbaT2T_hap1 and "Golden Delicious" genome
Table S8. Identification of genes unique to *M. baccata* through presence/absence variation (PAV) analysis
Table S9. Summary of sampling information, geographic coordinates, and whole-genome resequencing statistics for the *M. baccata* accessions
Table S10. SNPs and SVs in *M. baccata* are significantly associated with 19 bioclimatic factors as identified using LFMM
Table S11. SNPs and SVs are significantly associated with six environmental predictors identified by Redundancy Analysis (RDA) in *M. baccata*
Table S12. Candidate environment-associated SNPs and SVs in *M. baccata* cross-validated by LFMM and RDA

Table S13. Identification of outlier SNP windows and SVs in *M. baccata*
Table S14. Adaptive SNPs and all SVs and their associated candidate genes were identified with 19 bioclimatic factors in *M. baccata*
Table S15. Statistical significance of differential expression levels in adaptive genes across different treatments
Table S16. SNPs and SVs are significantly associated with six additional environmental variables (prec6, prec12, srad6, srad12, vapr6, vapr12) and the elevation of *M. baccata*
Table S17. Summary of SNPs, SVs, and candidate genes significantly associated with seven soil factors in *M. baccata*

Scan the QR code to view JIPB on WeChat (WeChat: **jipb1952**)

Scan the QR code to view JIPB on X (X: **@JIPBio**)

Scan the QR code to view JIPB on Bluesky (Bluesky: **jipb.bsky.social**)