Natural population re-sequencing detects the genetic basis of local adaptation to low temperature in a woody plant

Yanmin Hu[1], Xianjun Peng[1], Fenfen Wang[1], Peilin Chen[1], Meiling Zhao[1], and Shihua Shen[1]

[1]Institute of Botany Chinese Academy of Sciences

July 22, 2020

Abstract

Local adaptation to temperature is essential for woody plants to against changeable climate and safely survive the winter. To uncover the specific molecular mechanism of low temperature adaptation in woody plants, we performed selective-sweep analysis and genome-wide association study (GWAS) on a wild woody plant naturally distributed in different climate zones and latitudes. We sequenced a core collection of 134 accessions selected from 494 paper mulberry (Broussonetia papyrifera L.), phenotyped the accessions in high latitudes of 40º N for two overwintering traits. We further performed genome-phenotype and genome-environment associations, and genome-wide scans for temperature selection. The population structure analysis indicated that accessions showed forceful geographic distribution patterns because of the adaptation to local climate. We detected 75 selective regions possibly undergone temperature selection and identified 14 trait-associated SNPs corresponded to 16 candidate genes. Meanwhile, low temperature adaptation was also supported by other three SNPs with values lower than threshold but harboring different primary genotype among geographic groups. Overall, we propose a possible network of cold signal perception and responses in woody plants, some genes are considered unique to woody plants while others have been studied in herbs, which highlighting a key hit for studying the specific molecular mechanism of low temperature adaptation or overwintering in woody plants.

1 | INTRODUCTION

Local adaptation is a critical evolutionary process that allows plants to grow better to have large geographic ranges by generating adaptive genetic variation (Lobo et al., 2018; Lasky et al., 2015). In the natural world, cold stress frequently governs the growth, development and distribution of plants (Yang et al., 2019). By long generation times, plants have adapted to local low-temperature and generated temperature-dependent variation of cold tolerance (Pluess et al., 2016). Previous studies have shown that the low-temperature signals can be perceived by some cold sensors, such as COLD1 (chilling tolerance divergence 1), transmitted downstream to initiate multiple responses, and then induce the expression of *CBFs*(C-repeat binding factor) and cold-responsive genes (*CORs*) (Ding, Shi, & Yang, 2020). However, the most of the researches of plant cold tolerance are focus on model plants or crops, such as Arabidopsis, rice and maize (Gao, Ma, Wu, Zhou, & Zhang, 2019). Nevertheless, the molecular regulatory mechanism of cold responses between woody plants and herbs are significantly different because of the existing of cold acclimation, dormancy and overwintering (Wisniewski, Nassuth, & Arora, 2018), but there is lack of systematic study on the responses to low temperature in woody plants.

Tree species cover approximately 30% of land surface (Bonan, 2008), having great economic value and vast ecological importance, which can substantively impact the global carbon cycle (Alberto et al., 2013). As long-lived organisms, woody plants have evolved the ability to tolerate cold winters, but the biophysical mechanisms responsible for low-temperature tolerance are still poorly understood (Kalberer, Leyva-Estrada, Krebs, & Arora, 2007). The main researches are focus on studying the genetic regulation of dormancy status, spring phenology and the*ICE-CBF* pathway, but unlike rice or Arabidopsis, we still unclear the

details of *CBF* regulation in woody plants, not to mention other signaling pathways (Wisniewski et al., 2018). To understand the complex processes of low temperature response in woody plants, we should resolve the formation mechanism of low temperature adaptation during evolution. GWAS analysis as an integrated approach should be taken into account, which has been applied to elucidate the genes responding to cold stress in model plants and common crops, such as rice (Shakiba et al., 2017), maize (Huang et al., 2013), winter faba bean (Sallam, Arbaoui, El-Esawi, Abshire, & Martsch, 2016) and so on, but GWAS analysis has not been applied to elucidate the genes responding to low-temperature in woody plants. Recently, the GWAS analysis between SNP and the original environment (EnvGWAS) also be used to illuminate the genetic basis of adaptive evolution (Li et al., 2019), which can help us reveal the genetic basis of local temperature adaptation. Therefore, to expose the molecular basis of low temperature adaptation in woody plants, GWAS analysis should be used to identify significant variations associated with low temperature adaptive traits of woody plants.

Low temperature seriously affects the growth and geographic distribution of woody plants, but we have insufficient knowledge of the genomic basis of low temperature adaptation in woody plants. It's because the research on low temperature adaptation is restricted by the living areas of many woody plants, for example, both populus and oaks are grown in temperate and subtropical regions of the world (Kutsokon, Jose, & Holzmueller, 2015; Rellstab et al., 2016), which are no significant difference in cold tolerance. Furthermore, geographical and climatic factors of sampling sites can strongly influence plant cold tolerance (Xie et al., 2019), and some researches show that certain key genes play nonredundant roles in adaptation to temperature in different woody plants (Yeaman et al., 2016). Therefore, a tree species collected from different latitudes, having extensive temperature adaptability, will help us to explore the molecular mechanism of local adaptation to low temperature in woody plants. We found that paper mulberry from high latitudes can survive in extreme low temperature below -30 without winter injury, also adjust to the tropical heat, the geographic distribution extends from temperate to tropics, covering a wider range of climate than populus and oaks, which making paper mulberry become an ideal material to explore genetic basis of local adaptation to low temperature. As a diploid, dioecious (Peñailillo et al., 2016), and woody species in the family Moraceae (Pi, Zhao, Peng, & Shen, 2017), paper mulberry is native to East Asia and mainland Southeast-Asia, propagated across the Pacific by Austronesian-speaking voyagers for making barkcloth and cordage with practical and ceremonial or ritual purposes (Chang et al., 2015). As a novel model woody plant (Peng & Shen, 2018), paper mulberry has a long history of applying to papermaking, medicine, and livestock breeding (Hu, Peng, & Shen, 2018), having important economic and ecological values.

Thus, this work aims to uncover the specific molecular mechanism of low temperature adaptation in woody plants. We collected 479 paper mulberry accessions from 24 provinces in China, one from Korea and fourteen from America, ranged from 19º N to 41º N (latitude), 81º W to 122º E (longitude) and 2 m to 2358 m above sea level (altitude), distributed across a broad geographical range and spanned tropics, subtropics and temperate. Based on the useful reference genome (Peng et al., 2019), we performed analyses of genome-phenotype associations, genome-environment associations and genome-wide scans for the local low-temperature dependent variations. Finally, a series of significant selective regions and candidate genes were identified, and we discussed the potential molecular mechanism of local adaptation to low temperature in woody plants.

2 | MATERIALS AND METHODS

2.1 | Materials and field trials

In this review, a total of 494 paper mulberry accessions were collected from different geographic origins (Table S1). Then, a core collection of 134 accessions selected from 494 accessions were sown in greenhouse (Table S6), each accession was planted fifty (3 m in raw, 2 m in row) in Beijing, China (40°12' N, 116deg45' E) in June with a complete randomized block design, during the growing period, all plants were equally managed. Phenotypic data for plant death rate and plant withering rate were determined by counting the samples fail to overwinter successfully and the samples all the branches above-ground withering, then calculated the percentage between the samples death or withering and the total samples after the overwintering process. The average frost-free

period for each sampling site and the other climatic information were got from http://data.cma.cn..

2.2 | DNA preparation and sequencing

Genomic DNA was extracted from young fresh leaves of each accession, which collected from various locations and stored at -80, using the DNAsecure plant kit (Tiangen, Beijing) following the instructions. The quantity of DNA was determined on a NanoDrop (Thermo Fisher Scientific, Waltham, MA, USA) and agarose gel electrophoresis had been used to evaluate DNA quality.

To make the RAD-seq libraries, the restriction enzyme *Fok* I (NEB, Beijing, China) was used to digest the genomic DNA, and then ligate the adapter (6 nucleotides) to digested genomic DNA. Then all DNA fragments with adapter were purified and amplified in a 50 uL reaction, iTru5 and iTru7 as the PCR amplification primers. The PCR products were run out on 2% agarose gel and DNA fragments 300 bp to 500 bp were excised, then purified using a Gel Extraction Kit (Transgene, Beijing, China). After libraries detecting, all DNA libraries were sequenced on the Illumina platform. The raw data have been deposited in NCBI database under BioProject code: PRJNA635706.

Then, we selected 134 accessions constructing the paired-end genomic libraries with insert sizes of approximately 400 bp for the whole genome re-sequencing, using the NEB Next Ultra DNA Library Prep Kit (NEB, USA) following the manufacturer's instructions, and sequenced on an Illumina HiSeq X Ten platform. The raw data have been deposited in NCBI database under BioProject code: PRJNA635453.

2.3 | SNP calling

Raw data from RAD-seq and re-sequencing were filtered by removing low-quality reads, adaptor sequences, generating clean data for subsequent analyses. After quality filtering, the remaining clean reads were aligned to paper mulberry reference genome using BWA (ver. 0.7.10-r789) with the default settings (Li & Durbin, 2009), and only retained the uniquely mapped reads. Following mapping, SNP calling was performed following best practices workflow recommended by GATK (ver. 3.1) (Mckenna et al., 2010). The program ANNOVAR (Yang & Wang, 2015) was used to perform SNPs and InDels annotation according to the paper mulberry genome, the whole-genome SNPs were divided into two categories: transition and transversion, then the Ts/Tv ratio was counted. We further filtered SNPs with missing rate $< 50\%$ and minor allele frequency (MAF) $> 5\%$, in the end, a total of 8747 SNPs from RAD-seq and 2,936,477 SNPs from re-sequencing were retained for downstream phylogenetic and population structure analyses. The SNP density across each chromosome was calculated with 100 kb sliding windows using VCFtools (Danecek et al., 2011).

2.4 | Population genetics analysis and LD analysis

A subset of 2,936,477 SNPs was used to construct the phylogenetic tree with 100 bootstrap replications using the software PHYLIP (Felsenstein, 1989), and MEGA7 (Kumar, Stecher, & Tamura, 2016) was used to display the tree, the tool Evolview (https://www.evolgenius.info//evolview) was used to color the phylogenetic tree. The population structure of the 134 accessions was investigated using the program ADMIXTURE (ver. 1.3.0) (Alexander, Novembre, & Lange, 2009), to explore the most likely group number, the input parameter K in ADMIXTURE software was ranged from 1 to 10 following the default methods and settings. Principal component analysis (PCA) was performed with the Genome-wide Complex Trait Analysis (GCTA, ver. 1.92.0) (Yang, Lee, Goddard, & Visscher, 2011), and the program Origin was used to draw the figures of first two PCs. In advance, the phylogenetic tree, population structure and principal component analysis (PCA) were investigated using the 8,747 SNPs from 494 accessions for selecting samples following the mentioned above.

Linkage disequilibrium decay over distance was calculated based on the pairwise correlation coefficient (r^2) between each pair of SNPs using Haploview (Barrett, Fry, Maller, & Daly, 2005).

2.5 | Genetic diversity and population differentiation

The nucleotide diversity (π) was calculated by VCFtools for three geographic groups (Danecek et al., 2011), using 100-kb sliding windows with a step size of 10 kb. The VCFtools also was used to calculate genetic

differentiation (F_{ST}) among different geographic groups using 100-kb sliding windows with a step size of 10 kb.

2.6 | Analysis of selective sweeps

The average F_{ST} of all the windows was the value of the whole genome across different geographic groups. Whole-genome screening of selective sweeps were also performed by comparing the genetic diversity of samples from north of China and samples from south of China ($\pi N/\pi S$) in 100-kb sliding windows with a step size of 10 kb. The windows with the top 1% ofF_{ST} values or top 1% of $\pi N/\pi S$ values were selected as the selective regions, the genes located in these regions were considered as selected candidate genes. The gene ontology (GO) term enrichment for candidate genes was performed by AgriGO analysis toolkit (Du, Zhou, Ling, Zhang, & Su, 2010). The functional enrichment analyses of genes were accomplished by Kyoto Encyclopedia of Genes and Genomes (KEGG) databases.

2.7 | Genome-wide association study

In this study, a total of 2,936,477 SNPs (MAF $>$ 5% and missing data $<$ 50%) from 134 accessions were used to detect the relationships between genotypes and phenotypes using two different methods. (1) Mixed Linear Model, MLM was used to minimize the false positives and increasing the statistical power by considering PCA results and the Kinship value, conducted in GAPITR package (Lipka et al., 2012). (2) The Fixed and random model Circulating Probability Unification (Liu, Huang, Fan, Buckler, & Zhang, 2016), FarmCPU separately estimated a fixed effect and a random effect, having superior statistical power and efficient computing time. Manhattan plots and Q-Q plots were fulfilled through the R software package, the threshold was estimated to be approximately $P=10^{-8}$.

2.8 | Identification and annotation of candidate genes

Based on the LD decay distance, we selected the genes directly hit by the associated SNPs, and the genes in linkage disequilibrium (LD) with the associated SNPs as candidate genes. The paper mulberry reference genome was used to map the loci and genetic annotation.

3 | R ESULTS

3.1 | RAD sequencing and SNP calling

To enable large-scale genetic analysis, we firstly constructed RAD sequencing of 494 diverse accessions, which were collected from different geographic origins (Figure S1; Table S1). We applied an improved method to construct the RAD-seq library, only used*Fok* I to digest the genomic DNA and just ligated one adapter (Figure S2), which could save steps, experiment reagent and work time than traditional method (Díaz-Arce, Arrizabalaga, Murua, Irigoien, & Rodríguez-Ezpeleta, 2016). The other advantage was that we could detect more SNPs by using the restriction enzyme *Fok* I, which had more recognition sites on chromosomes. By sequencing, 240 Gb cleaned sequences were generated (Table S2), a total of 8,747 high-quality single nucleotide polymorphisms (SNPs) (Table S3) and 261 small insertions and deletions (InDels) (Table S4) were obtained after mapping, calling and filtering. A total of 2377 (27.2%) SNPs and 47 InDels (18.0%) were located within coding regions of genes (Figure S3; Table S5), the ratio of nonsynonymous-to-synonymous substitutions was 0.79, which was lower than pigeonpea (1.18) (Varshney et al., 2017) and mei (1.30) (Zhang et al., 2018).

3.2 | Population genetic analyses by RAD-seq data

To obtain a broad view of the genetic relationship and population structure of paper mulberry, we analyzed the phylogeny and population structure using the 8,747 high-quality SNPs. The 494 accessions were divided into three clusters in the neighbor joining (NJ) tree (Figure S4), with the geographical information of sample site (Table S1), we found that cluster I, cluster II and cluster III mainly contain accessions from east, north and south of China, respectively (Figure S4). The population structure analysis indicated that three ancestral types best explained the current population structure (Figure S5), and the principal component

analysis (PCA) further illustrated a similar pattern (Figure S6). Therefore, these results suggested that the 494 accessions were classified into three distinct clusters, corresponding to their geographical distributions.

3.3 | The whole genomere-sequencing and variation calling

Based on the geographical distributions and population genetic analyses of 494 accessions, we selected 134 accessions preformed the whole genome re-sequencing (Table S6), which having significant differences in freezing tolerance (Figure 1). According to the geographical and climatic information of the sample sites, the 134 accessions can be divided into three geographic groups (Figure 2a; Table S6). The north group consisted of 58 accessions from north, northwest and central of China, the east group included 36 accessions from east of China, and the south group comprised 40 accessions from tropic and south subtropics zone in the south of China. We generated a total of 582 Gb cleaned sequences, with an average sequencing depth of 11.4x (Table S7; Figure S7). Then identified a total of 2,936,477 high-quality SNPs, SNP density along chromosome was calculated to reveal the distribution of SNPs (Figure S8). We also analyzed the SNP mutation types (Figure S9; Table S8) and calculated Ts/Tv along chromosomes (Figure S10), the transition/transversion ratio (Ts/Tv) is 2.01.

3.4 | Population structure, nucleotide diversity, population divergence and LD decay

Using the 2,936,477 high-quality SNPs identified, we calculated the genetic distances from the genotypes, constructed a neighbor-joining (NJ) tree of the 134 accessions re-sequenced. The NJ tree supported the classification of the accessions into three clusters (Figure 2b). Almost all the accessions from east of China were classified into cluster I, the most of the accessions from north of China were roughly clustering together, which belong to cluster II, and 82.5% of the accessions from south of China were clearly clustered into cluster III. Overall, the NJ tree showed that the most of the samples from north, south and east of China well clustered together while some accessions were scattered in the other groups, indicating that the accessions from different zones might have experienced introgression or gene flow during natural evolution. The principle component analysis (PCA) suggested that the pattern was similar to the phylogenetic tree (Figure 2c; Figure S11). The observation was further supported this pattern by the population structure analysis (Figure 2d), indicated that three populations (when K = 3, the cross-validation error was minimum) explained the best model for these 134 accessions (Figure S12). But in population structure analysis, the accessions native to the Yunnan-Guizhou Plateau had the similar ancient components with accessions from the northern of China, rather than the accessions from southern of China (Table S6, S9), different with the results of NJ tree and PCA, suggesting that these accessions might harbor admixed ancestry.

Based on the SNP data, we evaluated the genetic diversity of the three geographic groups. The genome-wide nucleotide diversity (π) of south group (2.31×10^{-3}) and east group (2.33×10^{-3}) were slightly lower than that of north group (2.68×10^{-3}) (Figure 2e). These results reflected that the north group most likely to generate a number of genetic diversities during the nature selection to adapt to the climate of high latitudes. Compared to other major crops, the nucleotide diversity in paper mulberry is higher than that of cotton (1.3×10^{-3}) (Wang et al., 2017) and peach (1.3×10^{-3}) (Li et al., 2019), but relatively low compared with date palm (9.2×10^{-3}) (Hazzouri et al., 2015), wild mei (2.8×10^{-3}) (Zhang et al., 2018). Besides, we also evaluated the pairwise genome-wide fixation index (F_{ST}) values, the F_{ST}between north group and south group was 0.102, slightly greater than 0.046 (the F_{ST}between north group and east group) and 0.077 (theF_{ST} between south group and east group) (Figure 2e), reflected that the relationship of north group and south group was further than the other. The little genetic differentiation and similar level of nucleotide diversity suggest that there might exit a weaker selection bottleneck during the natural evolution, as suggested in peach (Li et al., 2019).

We further investigated the linkage disequilibrium (LD) decay pattern and found that the LD rapidly decayed to half of R^2 at 90 bp, decaying to $R^2 = 0.2$ at 150 bp and $R^2 = 0.1$ at ˜ 7 kb (Figure 2f), which was faster than previous reports in black cottonwood (decaying to $R^2 = 0.2$ within 3-6 kb) (Slavov et al., 2012) and wild European grapevines (decaying to R^2 = half of maximum average at 2.9 kb) (Liang et al., 2019). As a dioecious plant, the natural paper mulberry populations experienced long-term hybridization and had a relatively rapid LD decay, illuminating that the mating systems could affect linkage disequilibrium (LD) in

populations, which was consistent with a previous study (Rafalski & Michele, 2004).

3.5 | Genome-wide scans for potential selective signals of low temperature adaptation

To determine the potential selective signals of low temperature adaptive occurring along the history of natural evolution, we performed genome-wide scans using genetic differentiation (F_{ST}) and nucleotide diversity $\pi N/\pi S$ between the north group and south group, which having a significant difference in freezing tolerance, and the genomic windows with top 1% ofF_{ST} values and top 1% of $\pi N/\pi S$ values were considered as selective sweeps.

We detected a total of 324 selective sweeps with F_{ST} values greater than 0.35 (0.102 at the whole-genome level) (Figure 3a; Table S10), and identified 7.1Mb genomic regions harboring 275 genes (Table S11). Among the 324 selective sweeps, 274 selective sweeps (84.6%) located in chromosome 2 (Figure S13), indicated that the genes in chromosome 2 maybe experienced stronger selection. Moreover, there were 123 selective sweeps with $\pi N/\pi S$ values greater than 2 (1.2 at the whole-genome level) (Figure 3d; Table S12), and 4.8 Mb genomic regions (containing 217 genes) were identified (Table S13). In general, selective-sweep analyses revealed that 75 selective regions have occurred during natural evolution, the regions and genes identified here are crucial for further study of low temperature adaptation.

To classify the functions of these predicted genes, we performed Gene Ontology (GO) and KEGG enrichment analyses. The 275 genes located in top 1% selective sweeps of F_{ST} can be categorized into 39 GO terms and 24 KEGG pathways (Figure S14, S16). We noticed that a high percentage of genes from "metabolic process and cellular process (belong to biological processes)" and "binding and catalytic activity (belong to molecular functions)". The top 4 enriched pathways were "metabolic pathways", "biosynthesis of secondary metabolite", "diterpenoid biosynthesis", and "ribosome biogenesis in eukaryotes". In addition, the 217 genes located in top 1% selective sweeps of $\pi N/\pi S$ can be classified into 31 GO terms and 31 KEGG pathways (Figure S15, S17), "metabolic process and cellular process (belong to biological processes)" and "binding and catalytic activity (belong to molecular functions)" were the top four terms. The difference was that the mainly pathways included "biosynthesis of secondary metabolites", "metabolic pathways", "pyruvate metabolism". "Response to stimulus" was one of the enriched ontology term, low temperature as a stimulus signal may be related to these genes (Shakiba et al., 2017). The KEGG analyses found that several genes participated in mitogen-activated protein kinase (MAPK) signaling pathway (*Bp11g0585* , *Bp13g0782*) and lipid metabolism (*Bp06g0565*), and environmental adaptation (*Bp06g1381* , *Bp02g0231* , *Bp09g0200* ,*Bp01g0880*), which were thought to play a major role in low temperature adaptation in woody plants.

What's more, we identified one genomic region on chromosome 9, having higher F_{ST} value (0.35) as well as significantly $\pi N/\pi S$ value (2.1) (Figure 3), and detected 9 genes in this region (Table 1). Among them, *Bp09g0252* and *Bp09g0253* both encodeed SMARCA3 (SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily A, member 3) -like proteins, which is a helicase-like transcription factor (HLTF). Previous research shows that *Sh1* (encoding a SMARCA3-like factor) is related to local adaptation and under selection during domestication in wild cucumber (Bo et al., 2016), indicating that *Bp09g0252* and *Bp09g0253* are possibly associated with local adaptation. *Bp09g0254* encoded a bromodomain PHD finger transcription factor (BPTF), which playing an important role in chromatin remodeling (Zhao et al., 2019). Besides, *Bp09g0256* was a *GA3ox* gene involved in gibberellin metabolism and related to cold tolerance. These results indicate that some loci probably have undergone a temperature selection during the history of evolution.

3.6 | Genome-wide association analyses

In this study, we phenotyped the accessions in high latitudes (40^{o}N) for two traits about overwintering ability based on a long-term field experiment, including plant death rate and plant withering rate after overwintering process, also collected a bioclimatic variable (frost-free period of the sampling sites). As expected, we observed strongly variation in the natural wintering capability among different accessions, the plant death rate and plant withering rate ranged from 0 to 100%, the frost-free period of the sampling site ranged from 157 to 365 (Figure S18; Table S14-15), and the three traits were highly correlated with each other (Table S16).

There were significant negative correlations among plant death rate, plant withering rate and latitude. Besides, plant death rate, plant withering rate were significantly positive correlated with average annual temperature, accumulative temperature and frost-free period (Table S17). The correlation analysis indicated that the ability of overwintering was significantly correlated with climatic and geographical data of each collection site, reflecting the existence of local adaptation to low temperature in woody plants.

To identify genes associated with local adaptation to low temperature, we performed GWAS analyses on plant death rate and plant withering rate with 2,936,477 SNPs. For GWAS analyses, we tested two models: MLM and FarmCPU, only the FarmCPU model was adopted, because MLM missing some potentially important discoveries failed to detect association signals (Figure S19). According to the FarmCPU model, the association analysis for plant death rate detected four significant markers, distributed on chromosome 1, 2, 8, and one significant signal on chromosome 8 associated with plant withering rate (Figure 4; Table S18). The results of GWAS analyses were exhibited by Manhattan plots, and the P-value of QQ-plots suggested that association signals had high reliability.

Given that LD decay distance was about 7 kb ($R^2 = 0.1$), we searched candidate genes on downstream and upstream 7 kb range of the significantly association SNPs. Five annotated genes related to plant death rate were predicted (Table1; Table S18). Among them, the candidate gene *Bp08g1155* , which associated with 8:9642341, encoded to a leucine-rich repeat receptor-like serine/threonine kinase (LRR-RLK). We found that the candidate gene *Bp01g1321* was associated with 1:17805729, encoded a peroxidase belonging to plant-specific class III peroxidases, previous research shows that overexpression of *AtPrx* can increase the cold tolerance without growth inhibition in Arabidopsis (Kim, Kim, & Nam, 2012). What's more,*Bp02g1805* was associated with SNP 2:30599497, encoding a two-component response regulator (ARR12). Besides, the candidate gene *Bp02g1806* was associated with 2:30599497, encoding a hydroxyacid/fatty alcohol hydroxycinnamoyl transferase (FHT). In addition, only one significant SNP 8:21407459 associated with plant withering rate was revealed,*Bp08g2102* was predicted as the candidate gene (Table 1; Table S18), encoding an arabinosyltransferase (ARAD1) that was localized to glycosyltransferases (GT) superfamily (Harholt et al., 2012). A study in Arabidopsis demonstrates that the overexpression of *UGT79B2/B3* can significantly enhance cold tolerance, which directly controlled by C-repeat binding factor (CBF) (Li et al., 2017).

To further investigate the local adaptation to low temperature, we also conducted EnvGWAS on a bioclimatic variable (frost-free period of the sampling sites). A total of 8 significant markers were detected and 10 candidate genes were identified (Figure 5a; Table 1; Table S18). The candidate gene*Bp01g1842* was associated with 1:26086393, encoding a polyadenylate-binding protein-interacting protein (PAIP). The SNP 2:33247134 was located close to *Bp02g2091* encoding a phosphatase 2A (PP2A) protein, which participated in cell cycle and signal transduction (Luo, Shen, Yan, He, & Zhang, 2010). The candidate gene*Bp04g1554* , which associated with 4:28298004, encoded to a monodehydroascorbate reductase (MDHAR). The significant SNP 4:30942806 was located 3 kb away from *Bp04g1852* , encoding an E3 ubiquitin-protein ligase, which playing an important role in cold resistance (Fan, Chen, Kuang, Lu, & Shan, 2017). We also detected a gene (*Bp10g1488*) encoding a mitochondrial chaperone BCS1 (bc1 synthesis) associated with SNP 10:17145593, which belong to AAA (ATPase associated with diverse cellular activities) ATPase family. Besides, the other significant SNP 13:17681302 was located within the first exon of *Bp13g1480* , which encoding a late embryogenesis abundant (LEA) protein. LEA proteins are involved in the response to various abiotic stresses of plants, including cold, salinity and so on (Wang et al., 2019). We also detected candidate genes *Bp13g0698* encoding a helicase SEN1 (splicing endonuclease) and*Bp04g1851* encoding a TBC1 (Tre-2/Bub2/Cdc16) domain family member.

Most importantly, we detected three markers harboring different primary genotype among three geographic groups, which providing evidence for local adaptation to low temperature in woody plants. SNP 1:38895333 associated with frost-free period, 3:1782695 and 11:16274932 associated with plant death rate, while the P-values of the three markers slightly lower than threshold value (Table 1; Table S18). Among the 134 paper mulberry accessions, 63 accessions carrying 1:38895333-GG, the most of them from high-latitude, having smaller plant death rate and plant withering rate than accessions carrying 1:38895333-AA from low-latitude (Figure 5c, 5d, 5e). We discovered that SNP 1:38895333 was located within the first exon of*Bp01g2734*

encoding an AAA ATPase BCS1 (Table 1; Table S18). More interestingly, the genetic region of *Bp01g2734* had significantlyF_{ST} value (0.36) (Figure 5b). The other maker 3:1782695 was located in the promoter region of *Bp03g0251* encoding a γ-irradiation and mitomycin C induced 1 (GMI1) -like protein (Table 1; Table S18). We found that accessions carrying 3:1782695-TT were mainly collected from low latitude regions (Figure S20), having significantly weaker cold-resistant capacity than those carrying 3:1782695-CC (Figure 5d, 5e). Besides, SNP 11:16274932 was located 3 kb away from *Bp11g1556* which encoding ethylene response factor 1 (ERF1) (Table 1; Table S18), the accessions carrying 11:16274932-TT were mainly collected from the south of China (Figure S21), having poor low temperature resistance compared with the accessions carrying 11:16274932-CC (Figure 5d, 5e).

4 | DISCUSSION

In this study, to uncover the specific mechanism of low temperature adaptation in woody plants, we phenotyped the accessions in high latitude (40^{o} N) for two overwintering traits and performed genome-phenotype association, also collected a bioclimatic variable for genome-environment association. We found that the overwintering ability of woody plant was correlated with local temperature of sampling sites, showing that the accessions have adapted to local temperatures. We detected significant selective regions and candidate genes through genome-wide scans and GWAS analyses, especially the genes encoding LRR-RLK, BCS1 ATPase, PP2A and FHT, which were considered as the crucial molecular basis of low temperature adaptation in woody plants.

4.1 | Different geographic origins having adapted to local temperature

The overwintering ability of accessions were correlated with local temperatures, with the northern accessions (collected from higher latitudes) having better overwintering ability than the southern accessions (Table S14, S17), as described in weedy rice (Xie et al., 2019). The accessions of north group mainly collected from high-latitude temperate zone, characterized by cold and dry winter, these accessions had strong ability of enduring low temperature because of long-term natural selection of cold winter, which having better adaptation to low temperature and normally overwintering in northern regions. The south group was consisted of the accessions from tropics and southern subtropics, such climate as warm winter and long frost-free period making accessions sensitive to low temperature, and the field experimentation of overwintering proved that the accessions from south China couldn't overwinter in northern regions. Therefore, the overwintering ability of woody plant is strongly correlated with geographical and climatic data of the sampling sites, demonstrating that woody plants from different geographic origins have adapted to local temperature.

4.2 | Natural woody population showing forceful geographic distribution patternsbecause of the adaptation to local climate

This research suggested that the accessions from natural population usually were clustered into several groups corresponding to their geographical origin because of the adaptation to local climate environments, which showing forceful geographic distribution patterns. For example, the loblolly pine samples are clustered into three major genetic clusters named east, west and center (De La Torre, Wilhite, & Neale, 2019). Whereas, the population structure analyses in many crops support the classification of wild, landrace and improved groups on account of long-term domestication, such as soybean (Zhou et al., 2015) and pigeonpea (Varshney et al., 2017). The populations including semi-wild accessions are generally divided into cultivated, wild, and mixed subgroups, the mixed subgroup including wild accessions and semi-wild accessions (Wang, Li, & Liu, 2012). As natural woody population, paper mulberry also be divided into three genetic clusters, which having markedly different overwintering capability in northern region, proving that the accessions have adapted to the local temperature and indicating that paper mulberry is an ideal material to explore genetic basis of local adaptation to low temperature in woody plants. The classification results were largely congruent with geographical distributions, corresponding to the natural overwintering capability of accessions in high-latitude, low-temperature areas, suggested that the accessions have adapted to their distinct local temperatures.

4.3 | The specific molecular mechanism of local adaptation to low temperature in woody plants

Through selective sweep analyses and GWAS analyses, significant candidate genes were identified, which were thought to have undergone natural selection to adapt to the low temperature of sampling sites, and the genes encoding transmembrane protein LRR-RLK, BCS1 ATPase, PP2A and FHT were proposed to be important for low temperature adaptation of woody plants. We proposed that the LRR-RLK-MAPK mediated signaling system and ABA-PP2A pathway were involved in sensing cold signal and regulating cold responses in woody plants, and the BCS1 ATPase probably participated in cold tolerance through regulating mitochondrial respiration (MR), the above three pathways were considered as the specific molecular mechanism of local adaptation to low temperature in woody plants (Table 1; Figure 6). Besides, we inferred that the transmembrane protein LRR-RLK (Bp08g1155) was a unique cold sensor in woody plants.

The transduction of cold signal is crucial for plant cold tolerance, sensing low temperature is the most important step. Transmembrane protein COLD1 is involved in sensing cold stress, which localized at the plasma membrane (Ma et al., 2015), and an unknown transmembrane protein RLK (receptor-like kinase) may be also involved in sensing cold signal and phosphorylates CRPK1 in regulating the CBF-signaling pathway in Arabidopsis (Liu et al., 2017). Importantly, we detected a gene encoding a transmembrane protein LRR-RLK, which was thought as a unique cold sensor in woody plants. LRR-RLKs may play crucial roles in sensing external signals to regulate gene expression (Wang et al., 2018), which contain three functional domains: an extracellular domain perceiving signals, a transmembrane domain and an intracellular kinase domain transduced the signals downstream (Liu, Du, Huang, Gao, & Yu, 2017). Other researches show that LRR-RLK proteins can improve drought and salt stress tolerance and modulate salt tolerance in rice (Lin et al., 2020). It is likely that LRR-RLKs participate in regulating MAPK signal pathway, Na+/K+ ratio and reactive oxygen species (ROS) scavenging system (Lin et al., 2020).*Bp11g0585* and*Bp13g0782* were found through selective sweep analyses, which participating in MAPK signaling pathway. The MAPK signal pathway plays important roles in cold signal transduction (Moustafa, AbuQamar, Jarrar, Al-Rajab, & Trémouillaux-Guiller, 2014), so *Bp11g0585* and*Bp13g0782* may be regulated by LRR-RLK (Bp08g1155), thus realize cold signals transmission and regulate woody plants adapt to low temperature (Figure 6). LRR-RLK proteins also play a critical role in resistance to cold stress and maybe respond to low temperature by regulating the expression of*CBFs* and their target genes in *Glycine soja* (Yang et al., 2014). The *LEA* gene*Bp13g1480* , *peroxidase* gene *Bp01g1321* , and *GT*gene *Bp08g2102* were thought to be regulated by CBFs and participated in the LRR-RLK mediated signaling system (Figure 6). Though lacking correlational studies about LRR-RLKs sensing cold signal and regulating cold responses of woody plants, we inferred that the LRR-RLK (Bp08g1155) was the crucial factor of local adaptation to low temperature in woody plants, so further study could focus on how LRR-RLKs perceive cold signals and regulate cold stress signaling pathways.

Moreover, hormonesalicylic acid (SA) and abscisic acid (ABA) also take part in cold signals transduction. In this study, we detected candidate genes *Bp01g2734* and *Bp10g1488* encoding BCS1 proteins. BCS1 is an AAA ATPase, involved in assembly of cytochrome bc1, which is a key component of mitochondrial respiratory chain and highly responsive to SA (Cui, Smith, Fox, Khalimonchuk, & Winge, 2012). Interestingly, the genetic region of gene *Bp01g2734* having significantlyF_{ST} value (0.36) (Figure 5b), indicating that*Bp01g2734* had undergone a temperature selection during the evolution. More importantly, the different primary genotype among geographic groups also demonstrated that *Bp01g2734* was absolutely crucial to local adaptation to low temperature in woody plants (Figure 5c). An Arabidopsis AAA ATPase gene *At1g64110* can be inducible by abiotic stress such as cold and drought, the expression level of*At1g64110* is up-regulated when plants suffer cold stress (Ali et al., 2013). So *BCS1* genes (*Bp01g2734* , *Bp10g1488*) were thought to be involved in low temperature adaptation of woody plants through regulating mitochondrial respiration (MR) and participating in SA mediated cold signaling pathway (Figure 6). Therefore, we were tempted to consider that *Bp01g2734* had great research values to elucidate the SA-BCS1-MR pathway in low temperature adaptation of woody plants.

We considered that ABA-PP2A pathway played a major role in low temperature adaptation of woody plants (Figure 6). One *PP2A* gene (*Bp02g2091*) was detected through EnvGWAS analysis, which playing a crucial role in signal transduction. Some studies show that ABA receptor PYLs binding ABA and inhibit the

activity of PP2A, thereby modulating auxin transport and regulating root developmental under stress (Li et al., 2020). Previous research has suggested that PP2A is a crucial element in the low-temperature signaling pathway in Arabidopsis, and E3 ubiquitin ligase may function upstream of PP2A under low temperature (Luo et al., 2010). PP2A phosphatases also regulate ROS signaling network of plants, and the ROS metabolism is associated with plant cold tolerance (Mathe, Garda, Freytag, & Mhamvas, 2019; Wang et al., 2019). Under stress, monodehydroascorbate (MDHA) is reduced to ascorbic acid (AsA) through monodehydroascorbate reductase (MDHAR) for scavenging ROS (Shin, Kim, Kim, Park, & Yoon, 2013), so we inferred that *MDHAR* gene*Bp04g1554* might be involved in ROS scavenging system together with PP2A in woody plants. Thus, elucidating the ABA-PP2A signaling pathway would help us understand the local adaptation to low temperature in woody plants. Besides, the *FHT* gene*Bp02g1806* might take part in the ABA-induced suberin biosynthesis. ABA induces production of MYBs (myeloblastosis) and ABFs (ABA-responsive element binding factors), then activates the downstream *FHT* gene promoting suberin biosynthesis (Figure 6), as described in previous research (Wei et al., 2020). We proposed that *PP2A* gene *Bp02g2091* and *FHT* gene*Bp02g1806* were involved in ABA-mediated cold stress signaling network, playing an important role in local adaptation to low temperature in woody plants.

In addition, hormonal can regulate plant growth when responding to cold stress, we thought that*GA3ox* gene *Bp09g0256* and *ARR12* gene*Bp02g1805* could influence low temperature adaptation of woody plants through regulating plant growth (Figure 6). Previous studies suggest that CBF can enhance cold tolerance and restrain growth by repressed the expression of *GA3ox* genes, while gibberellin 3-oxidase can convert the inactive gibberellin to promote plant growth and weaken cold induced CBF expression (Zhou et al., 2014). So GA3ox gene *Bp09g0256* located in selective sweeps reflected that some genes had experienced low-temperature selection during evolution. What's more, *ARR12* gene *Bp02g1805* was considered to participate in the reduction of auxin accumulation to inhibit root growth when plants responding to low temperature (Figure 6), as illustrated in previous study (Zhu et al., 2015). Therefore,*Bp09g0256* and *Bp02g1805* were thought to play a major role of cold tolerance through regulating the level of bioactive GA or auxin, thereby controlling plant growth, increasing cold tolerance and safely overwintering. Besides, helicase-like proteins SEN1 and SMARCA3 are related to regulating transcription (Chen et al., 2017; Bo et al., 2016), and PAIP1 interacts with PABP involved in regulating translation (Ivanov et al., 2019), so the related genes we identified might take part in responses to cold stress by regulating transcription or translation, providing valuable information for understanding the mechanism of low temperature adaptation in woody plants.

In conclusion, our results suggested that GWAS analysis was very suitable for detecting the local temperature-dependent variation in woody plants. The significant selective regions and candidate genes were identified as the important genetic basis of low temperature adaptation in woody plants. Most importantly, LRR-RLK, BCS1 ATPase and PP2A were thought to mediate distinctive cold signaling transduction and regulated pathways of woody plants. Meanwhile, the similar ICE-CBF-COR pathway shared with Arabidopsis and rice was also identified. Besides, the transmembrane protein LRR-RLK (Bp08g1155) was considered as a unique cold sensor in woody plants. Consequently, we put forward a possible network of cold signal perception, transduction and responses in woody plants, which providing significant insights into the research of low temperature adaptation or overwintering in woody plants.

ACKNOWLEDGEMENTS

We thank Hui Chen, Zhi Pi, Feng Tang, Xiaokang Zhang, Jie Hu, Jinshan Wang, Songwei Li and Pu Shu for their help of sample collection and cultivation in this study. This work was supported by the National Natural Science Foundation of China (31870247, 31770360) and the Beijing Natural Science Foundation (6202026).

AUTHOR CONTRIBUTIONS

Y.H. and X.P. performed the research, collected genetic and phenotypic data, analyzed the data, and drafted the manuscript. F.W., M.Z. and P.C. contributed to the acquisition of the climate and common garden data. S.S. conceived and designed the experiments. S.S. and X.P. provided funding. All authors approved the final

manuscript.

DATA AVAILABILITY STATEMENT

The RAD-seq sequences underlying this study have been deposited in NCBI database under BioProject code: PRJNA635706. The re-sequencing sequences underlying this study have been deposited in NCBI database under BioProject code: PRJNA635453.

REFERENCES

Alberto, F. J., Aitken, S. N., Alía, R., González-Martínez, S. C., Hänninen, H., Kremer, A.,... Savolainen, O. (2013). Potential for evolutionary responses to climate change-evidence from tree populations. *Global Change Biology, 19* (6), 1645-1661. https://doi.org/10.1111/gcb.12181

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research, 19* (9), 1655-1664. https://doi.org/10.1101/gr.094052.109Q

Ali, M., Plattner, S., Radakovic, Z., Wieczorek, K., Elashry, A., Grundler, F. M., ... Bohlmann, H. (2013). An Arabidopsis ATPase gene involved in nematode-induced syncytium development and abiotic stress responses. *Plant Journal, 74* (5), 852-866. https://doi.org/10.1111/tpj.12170

Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics, 21* (2), 263-265. https://doi.org/10.1093/bioinformatics/bth457

Bo, K., Wang, H., Pan, Y., Behera, T., Pandey, S., Wen, C., ... Weng, Y. (2016). *SHORT HYPOCOTYL1* encodes a SMARCA3-like chromatin remodeling factor regulating elongation. *Plant physiology, 172* (2), 1273-1292. https://doi.org/10.1104/pp.16.00501

Bonan, G. B. (2008). Forests and climate change: forcings, feedbacks, and the climate benefits of forests. *Science, 320* (5882), 1444-1449. https://doi.org/10.1126/science.1155121

Chang, C. S., Liu, H. L., Moncada, X., Seelenfreund, A., Seelenfreund, D., & Chung, K. F. (2015). A holistic picture of Austronesian migrations revealed by phylogeography of Pacific paper mulberry. *Proceedings of the National Academy of Sciences of the United States of America, 112* (44), 13537-13542. https://doi.org/10.1073/pnas.1503205112

Chen, X., Poorey, K., Carver, M. N., Muller, U., Bekiranov, S., Auble, D. T., & Brow, D.A. (2017). Transcriptomes of six mutants in the Sen1 pathway reveal combinatorial control of transcription termination across the *Saccharomyces cerevisiae* genome. *Plos Genetics, 13* (6), e1006863. https://doi.org/10.1371/journal.pgen.1006863

Cui, T. Z., Smith, P. M., Fox, J. L., Khalimonchuk, O., & Winge, D. R. (2012). Late-stage maturation of the rieske Fe/S protein: Mzm1 stabilizes Rip1 but does not facilitate its translocation by the AAA ATPase Bcs1. *Molecular and Cellular Biology, 32* (21), 4400-4409. https://doi.org/10.1128/MCB.00441-12

Díaz-Arce, N., Arrizabalaga, H., Murua, H., Irigoien, X., & Rodríguez-Ezpeleta, N. (2016).RAD-seq derived genome-wide nuclear markers resolve the phylogeny of tunas. *Molecular Phylogenetics and Evolution, 102* , 202-207. https://doi.org/10.1016/j.ympev.2016.06.002

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... Durbin,R. (2011). The variant call format and VCFtools. *Bioinformatics, 27* (15), 2156-2158. https://doi.org/https://doi.org/10.1093/bioinformatics/btr330

De La Torre, A. R., Wilhite, B., & Neale, D. B. (2019). Environmental genome-wide association reveals climate adaptation is shaped by subtle to moderate allele frequency shifts in loblolly pine. *Genome Biology and Evolution, 11* (10), 2976-2989. https://doi.org/10.1093/gbe/evz220

Ding, Y., Shi, Y., & Yang, S. (2020). Molecular regulation of plant responses to environmental temperatures. *Molecular Plant, 13* , 544-564. https://doi.org/10.1016/j.molp.2020.02.004

Du, Z., Zhou, X., Ling, Y., Zhang, Z., & Su, Z. (2010). agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Research, 38* , W64-W70. https://doi.org/10.1093/nar/gkq310

Fan, Z. Q., Chen, J. Y., Kuang, J. F., Lu, W. J., & Shan, W. (2017). The banana fruit SINA ubiquitin ligase MaSINA1 regulates the stability of MaICE1 to be negatively involved in cold stress response.*Frontiers in Plant Science, 8* , 995. https://doi.org/10.3389/fpls.2017.00995

Felsenstein, J. (1989). PHYLIP: phylogeny inference package (version 3.2). *Cladistics, 5* , 164-166.

Gao, F., Ma, P., Wu, Y., Zhou, Y., & Zhang, G. (2019). Quantitative proteomic analysis of the response to cold stress in jojoba, a tropical woody crop. *International Journal of Molecular Sciences, 20* (2), 243. https://doi.org/10.3390/ijms20020243

Harholt, J., Jensen, J. K., Verhertbruggen, Y., Søgaard, C., Bernard, S., Nafisi, M., ... Scheller, H. V. (2012). ARAD proteins associated with pectic Arabinan biosynthesis form complexes when transiently overexpressed in planta. *Planta, 236* (1), 115-128. https://doi.org/10.1007/s00425-012-1592-3

Hazzouri, K. M., Flowers, J. M., Visser, H. J., Khierallah, H. S. M., Rosas, U., Pham, G.M., ... Purugganan, M. D. (2015). Whole genome re-sequencing of date palms yield sinsights into diversification of a fruit tree crop. *Nature Communications, 6* , 8824. https://doi.org/10.1038/ncomms9824

Hu, Y. M., Peng, X. J., & Shen, S. H. (2018). Multiculture and values of the paper mulberry from the perspective of ancient China's science and technology. *Journal of Beijing Forestry University, 17* (2), 38-43. https://doi.org/10.13931/j.cnki.bjfuss.2017089

Huang, J., Zhang, J., Li, W., Hu, W., Duan, L., Feng, Y., ... Yue, B. (2013). Genome-wide association analysis of ten chilling tolerance indices at the germination and seedling stages in maize. *Journal of Integrative Plant Biology, 55* (8), 735-744. https://doi.org/10.1111/jipb.12051

Ivanov, A., Shuvalova, E., Egorova, T., Shuvalov, A., Sokolova, E., Bizyaev, N., ... Alkalaeva, E. (2019). Polyadenylate-binding protein-interacting proteins PAIP1 and PAIP2 affect translation termination. *Journal of Biological Chemistry, 294* (21), 8630-8639. https://doi.org/10.1074/jbc.RA118.006856

Kalberer, S. R., Leyva-Estrada, N., Krebs, S. L., & Arora, R. (2007). Frost dehardening and rehardening of floral buds of deciduous azaleas are influenced by genotypic biogeography. *Environmental and Experimental Botany, 59* (3), 264-275. https://doi.org/10.1016/j.envexpbot.2006.02.001

Kim. B. H., Kim, S. Y., & Nam, K. H. (2012). Genes encoding plant-specific class III peroxidases are responsible for increased cold tolerance of the *brassinosteroid-insensitive1* mutant.*Molecules and Cells, 34* (6), 539. https://doi.org/10.1007/s10059-012-0230-z

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets.*Molecular Biology and Evolution, 33* (7), 1870-1874. https://doi.org/10.1093/molbev/msw054

Kutsokon, N. K., Jose, S., & Holzmueller, E. (2015). A global analysis of temperature effects on *Populus* plantation production potential. *American Journal of Plant Sciences, 06* (01), 23-33. https://doi.org/10.4236/ajps.2015.61004

Lasky, J. R., Upadhyaya, H. D., Ramu, P., Deshpande, S., Hash, C. T., Bonnette, J., ... Mitchell, S. E. (2015). Genome-environment associations in sorghum landraces predict adaptive traits. *Science Advances, 1* (6), e1400218. https://doi.org/10.1126/sciadv.1400218

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics, 25* (14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, J., Chen, G. B., Rasheed, A., Li, D., Sonder, K., Espinosa, C. Z., ... Li, H. (2019). Identifying loci with breeding potential across temperate and tropical adaptation via EigenGWAS and EnvGWAS.*Molecular Ecology, 28* (15), 3544-3560. https://doi.org/10.1111/mec.15169

Li, P., Li, Y. J., Zhang, F. J., Zhang, G. Z., Jiang, X. Y., Yu, H. M., & Hou, B. K. (2017). The Arabidopsis UDP-glycosyltransferases UGT79B2 and UGT79B3, contribute to cold, salt and drought stress tolerance via modulating anthocyanin accumulation. *The Plant Journal, 89* , 85-103. https://doi.org/10.1111/tpj.13324

Li, Y., Cao, K., Zhu, G., Fang, W., Chen, C., Wang, X., ... Wang, L. (2019). Genomic analyses of an extensive collection of wild and cultivated accessions provide new insights into peach breeding history. *Genome Biology, 20* (1), 36. https://doi.org/https://doi.org/10.1186/s13059-019-1648-9

Li, Y., Wang, Y., Tan, S., Li, Z., Yuan, Z., Glanc, M., ... Zhang, J. (2020). Root growth adaptation is mediated by PYLs ABA receptor-PP2A protein phosphatase complex. *Advanced Science, 7* (3), 1901455. https://doi.org/10.1002/advs.201901455

Liang, Z. C., Duan, S. C., Sheng, J., Zhu, S. S., Ni, X. M., Shao, J. H., ... Dong, Y. (2019). Whole-genome resequencing of 472 *Vitis*accessions for grapevine diversity and demographic history analyses. *Nature Communications, 10* , 1190. https://doi.org/10.1038/s41467-019-09135-8

Lin, F., Li, S., Wang, K., Tian, H., Gao, J., Zhao, Q., & Du, C. (2020). A Leucine-rich repeat receptor-like kinase, *OsSTLK* , modulates salt tolerance in rice. *Plant Science, 296* , 110465. https://doi.org/10.1016/j.plantsci.2020.110465

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., ... Zhang, Z. (2012).GAPIT: genome association and prediction integrated tool. *Bioinformatics, 28* (18), 2397-2399. https://doi.org/10.1093/bioinformatics/bts444

Liu, P., Du, L., Huang, Y., Gao, S., & Yu, M. (2017). Origin and diversification of leucine-rich repeat receptor-like protein kinase (*LRR-RLK*) genes in plants. *BMC Evolutionary Biology, 17* (1), 47. https://doi.org/10.1186/s12862-017-0891-5

Liu, X., Huang, M., Fan, B., Buckler, E. S., & Zhang, Z. (2016). Iterative usage of fixedand random effect models for powerful and efficient genome-wide association studies. *Plos Genetics, 12* (3), e1005767. https://doi.org/10.1371/journal.pgen.1005767

Liu, Z., Jia, Y., Ding, Y., Shi, Y., Li , Z., Guo, Y., ... Yang, S. (2017). Plasma membraneCRPK1-mediated phosphorylation of 14-3-3 proteins induces their nuclear import to fine-tune CBF signaling during cold response. *Molecular Cell, 66* (1), 117-128. https://doi.org/10.1016/j.molcel.2017.02.016

Lobo, A., Hansen, O. K., Hansen, J. K., Erichsen, E. O., Jacobsen, B., & Kjaer, E. D. (2018). Local adaptation through genetic differentiation in highly fragmented *Tilia cordata* populations.*Ecology and Evolution, 8* (12), 5968-5976. https://doi.org/10.1002/ece3.4131

Luo, J., Shen, G., Yan, J., He, C., & Zhang, H. (2010). AtCHIP functions as an E3 ubiquitin ligase of protein phosphatase 2A subunits and alters plant response to abscisic acid treatment. *Plant Journal, 46* (4), 649-657. https://doi.org/10.1111/j.1365-313X.2006.02730.x

Ma, Y., Dai, X., Xu, Y., Luo, W., Zheng, X., Zeng, D., ... Zhang, D. (2015). *COLD1* confers chilling tolerance in rice. *Cell, 160* (6), 1209-1221. https://doi.org/10.1016/j.cell.2015.01.046

Mathe, C., Garda, T., Freytag, C., & Mhamvas, M. (2019). The role of serine-threonine protein phosphatase PP2A in plant oxidative stress signaling-facts and hypotheses. *International Journal of Molecular Sciences, 20* (12), 3028. https://doi.org/10.3390/ijms20123028

Mckenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., ... Depristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research, 20* (9), 1297-1303. https://doi.org/10.1101/gr.107524.110

Moustafa, K., AbuQamar, S., Jarrar, M., Al-Rajab, A. J., & Tremouillaux-Guiller, J. (2014).MAPK cascades and major abiotic stresses. *Plant cell reports, 33* (8), 1217-1225. https://doi.org/10.1007/s00299-014-1629-0

Penailillo, J., Olivares, G., Moncada, X., Payacan, C., Chang, C. S., Chung, K. F., ... Seelenfreund, D. (2016). Sex distribution of paper mulberry (*Broussonetia papyrifera*) in the Pacific. *Plos One, 11* (8), e0161148. https://doi.org/10.1371/journal.pone.0161148

Peng, X. J., Liu, H., Chen, P. L., Tang, F., Hu, Y. M., Wang, F. F., ... Shen, S. H. (2019). A chromosome-scale genome assembly of paper mulberry (*Broussonetia papyrifera*) provides new insights into its forage and papermaking usage. *Molecular Plant, 12* (2), 661-667. https://doi.org/10.1016/j.molp.2019.01.021

Peng, X. J., & Shen, S. H. (2018). The paper mulberry: a novel model system for woody plant research. *Chinese Bulletin of Botany, 53* (3), 372-381.

Pi, Z., Zhao, M. L., Peng, X. J., & Shen, S. H. (2017). Phosphoproteomic analysis of paper mulberry reveals phosphorylation functions in chilling tolerance. *Journal of Proteome Research, 16* (5), 1944-1961. https://doi.org/10.1021/acs.jproteome.6b01016

Pluess, A. R., Frank, A., Heiri, C., Lalague, H., Vendramin, G. G., & Oddou-Muratorio, S.(2016). Genome-environment association study suggests local adaptation to climate at the regional scale in *Fagus sylvatica* . *New Phytologist, 210* (2), 589-601. https://doi.org/10.1111/nph.13809

Rafalski, A., & Michele, M. (2004). Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. *Trends in Genetics, 20* (2), 103-111. https://doi.org/10.1016/j.tig.2003.12.002

Rellstab, C., Zoller, S., Walthert, L., Lesur, I., Pluess, A. R., Graf, R., ... Gugerli, F. (2016). Signatures of local adaptation in candidate genes of oaks (*Quercus* spp.) with respect to present and future climatic conditions. *Molecular Ecology, 25* (23), 5907-5924. https://doi.org/10.1111/mec.13889

Sallam, A., Arbaoui, M., El-Esawi, M., Abshire, N., & Martsch, R. (2016). Identification and verification of QTL associated with frost tolerance using linkage mapping and GWAS in winter faba bean.*Frontiers in Plant Science, 7* , 1098. https://doi.org/10.3389/fpls.2016.01098

Shakiba, E., Edwards, J. D., Jodari, F., Duke, S. E., Baldo, A. M., Korniliev, P., ... Eizenga, G. C. (2017). Genetic architecture of cold tolerance in rice (*Oryza sativa*) determined through high resolution genome-wide analysis. *Plos One, 12* (3), e0172133. https://doi.org/10.1371/journal.pone.0172133

Shin, S. Y., Kim, M. H., Kim, Y. H., Park, H. M., & Yoon, H. S. (2013). Co-expression of monodehydroascorbate reductase and dehydroascorbate reductase from *Brassica rapa* effectively confers tolerance to freezing-induced oxidative stress. *Molecules and Cells, 36* (4), 304-315. https://doi.org/10.1007/s10059-013-0071-4

Slavov, G. T., DiFazio, S. P., Martin, J., Schackwitz, W., Muchero, W., Rodgers-Melnick, E., ... Tuskan, G. A. (2012). Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest tree *Populus trichocarpa* . *New Phytologist, 196* (3), 713-725. https://doi.org/10.1111/j.1469-8137.2012.04258.x

Varshney, R. K., Saxena, R. K., Upadhyaya, H. D., Khan, A. W., Yu, Y., Kim, C., ... Datta, S. K. (2017). Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. *Nature Genetics, 49* (7), 1082. https://doi.org/doi:10.1038/ng.3872

Wang, J., Li, C., Yao, X., Liu, S., Zhang, P., & Chen, K. (2018). The Antarctic moss leucine-rich repeat receptor-like kinase (PnLRR-RLK2) functions in salinity and drought stress adaptation. *Polar Biology, 41* (2), 353-364. https://doi.org/10.1371/journal.pone.0172869

Wang, J., Lu, X., Yin, Z., Wang, D., Wang, S., Mu, M., ... Ye, W. (2019). Isolation and characterization of *GhLEA3* gene from upland cotton and its expression in response to low temperature stress.*Cotton Science, 31* (2), 89-100. https://doi.org/10.11963/1002-7807.wjjyww.20190311

Wang, K. J., Li, X. H., & Liu, Y. (2012). Fine-scale phylogenetic structure and major events in the history of the current wild soybean (*Glycine soja*) and taxonomic assignment of semi-wild type

(*Glycine gracilis Skvortz.*) within the Chinese subgenus*Soja* . *Journal of Heredity, 103* (1), 13-27. https://doi.org/10.1093/jhered/esr102

Wang, L., Yao, L., Hao, X., Li, N., Wang, Y., Ding, C., ... Yang, Y. (2019). Transcriptional and physiological analyses reveal the association of ROS metabolism with cold tolerance in tea plant.*Environmental and Experimental Botany, 160* , 45-58. https://doi.org/10.1016/j.envexpbot.2018.11.011

Wang, M., Tu, L., Lin, M., Lin, Z., Wang, P., Yang, Q., ... Zhang, X. (2017). Asymmetricsubgenome selection and *cis* -regulatory divergence during cotton domestication. *Nature Genetics, 49* (4), 579-587. https://doi.org/10.1038/ng.3807

Wei, X., Lu, W., Mao, L., Han, X., Wei, X., Zhao, X., ... Xu, C. (2020). ABF2 and MYB transcription factors regulate feruloyl transferase*FHT* involved in ABA-mediated wound suberization of kiwifruit.*Journal of experimental botany, 71* (1), 305-317. https://doi.org/10.1093/jxb/erz430

Wisniewski, M., Nassuth, A., & Arora, R. (2018). Cold hardiness in trees: a mini-review. *Frontiers in Plant Science, 9* , 1394. https://doi.org/10.3389/fpls.2018.01394

Xie, H., Han, Y., Li, X., Dai, W., Song, X., Olsen, K. M., & Qiang, S. (2019). Climate-dependent variation in cold tolerance of weedy rice and rice mediated by *OsICE1* promoter methylation. *Molecular Ecology, 29* (1), 121-137. https://doi.org/10.1111/mec.15305

Yang, H., & Wang, K. (2015). Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nature Protocols, 10* (10), 1556-1566. https://doi.org/10.1038/nprot.2015.105

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *American Journal of Human Genetics, 88* (1), 76-82. https://doi.org/10.1016/j.ajhg.2010.11.011

Yang, L., Wu, K., Gao, P., Liu, X., Li, G., & Wu, Z. (2014). GsLRPK, a novel cold-activated leucine-rich repeat receptor-like protein kinase from *Glycine soja* , is a positive regulator to cold stress tolerance. *Plant Science, 215* (3), 19-28. https://doi.org/10.1016/j.plantsci.2013.10.009

Yang, X., Wang, R., Hu, Q., Li, S., Mao, X., Jing, H., ... Liu, C. (2019). *DlICE1* , a stress-responsive gene from *Dimocarpus longan* , enhances cold tolerance in transgenic *Arabidopsis* .*Plant Physiology and Biochemistry, 142* , 490-499. https://doi.org/10.1016/j.plaphy.2019.08.007

Yeaman, S., Hodgins, K. A., Lotterhos, K. E., Suren, H., Nadeau, S., Degner, J. C., ... Aitken, S. N. (2016). Convergent local adaptation to climate in distantly related conifers. *Science, 353* (6306), 1431-1433. https://doi.org/10.1126/science.aaf7812

Zhang, Q., Zhang, H., Sun, L., Fan, G., Ye, M., Jiang, L., ... Bao, F. (2018). The geneticarchitecture of floral traits in the woody plant*Prunus mume* . *Nature Communications, 9* (1), 1702. https://doi.org/10.1038/s41467-018-04093-z

Zhao, X., Zheng, F., Li, Y., Hao, J., Tang, Z., Tian, C., ... Deng, W. (2019). BPTF promotes hepatocellular carcinoma growth by modulating hTERT signaling and cancer stem cell traits. *Redox biology, 20* , 427-441. https://doi.org/10.1016/j.redox.2018.10.018

Zhou, M., Xu, M., Wu, L., Shen, C., Ma, H., & Lin, J. (2014).*CbCBF* from *Capsella bursa-pastoris* enhances cold tolerance and restrains growth in *Nicotiana tabacumby* antagonizing with gibberellin and affecting cell cycle signaling. *Plant Molecular Biology, 85* (3), 259-275. https://doi.org/10.1007/s11103-014-0181-1

Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., ... Tian, Z. (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nature Biotechnology, 33* (4), 408-414. https://doi.org/10.1038/nbt.3096

Zhu, J., Zhang, K. X., Wang, W. S., Gong, W., Liu, W. X., Chen, H. G., ... Lu, Y. T. (2015). Low temperature inhibits root growth by reducing auxin accumulation via *ARR1/12* . *Plant and Cell Physiology, 56* (4), 727-736. https://doi.org/10.1093/pcp/pcu217

FIGURE LEGENDS

FIGURE 1 The phenotypic changes of accessions after overwintering process. The 134 accessions overwintered in Beijing (40o N) and suffered freezing stress for three months (daily minimum temperature below -5), then the samples were divided into freezing sensitive, moderate freezing sensitive and freezing insensitive (freezing tolerant) types according to the significant differences in spring bud. (a) The samples collected from the south temperate zone in high latitudes are divided into freezing insensitive (freezing tolerant) type, reflected that all the branches bud in the spring. (b) The samples collected from north subtropical zone in the middle latitudes are divided into moderate freezing sensitive type, reflected that stalk part of the plants frozen to death and generate new shoots from rhizome in the spring. (c) The samples collected from south subtropical zone and tropical zone in low latitudes are divided into freezing sensitive type, reflected that the whole plant frozen to death.

FIGURE 2 Geographic distribution, phylogeny and population genetics of 134 accessions used for the whole genome re-sequencing. (a) The geographic locations of the 134 accessions. Color codes indicate three geographic groups, blue, red and green colors indicate accessions from north, south, and east of China, respectively. (b) Neighbor-joining tree of the 134 accessions, the accessions were divided into three clusters. Blue, red and green labels indicate accessions from north, south, and east of China, respectively. (c) PCA analysis was conducted using the first two components (PC1 and PC2). Blue, red and green colors indicate accessions from north, south, and east of China, respectively, the three geographic groups were clearly clustered. (d) Population structure analysis of the 134 accessions estimated by ADMIXTURE (K = 3 explained the best model for these 134 accessions), the x axis represents the different accessions, the y axis quantifies ancestry membership, and the colors represent ancestral populations. (e) Summary of nucleotide diversity (π) and population divergence (F_{ST}) across the three geographic groups. Values in circles represent nucleotide diversity, and values between pairs indicate population divergence (F_{ST}). (f) Decay of linkage disequilibrium (LD) on the the whole genome of the 134 accessions, the rate of LD decay is displayed by the black curves, LD rapidly decayed to 0.2 at ˜ 200 bp.

FIGURE 3 Genome-wide scans of selection sweeps during natural selection. (a) Population differentiation between north group and south group measured by population divergence (F_{ST}), and the top 1% of F_{ST} scores were considered to be candidate sweeps (red bar). (b-c) One genomic region on chromosome 9, having higher F_{ST} value (0.35) as well as significantly $\pi N/\pi S$ value (2.1), the red line indicates the threshold. (d) The genome-wide selective signals associated with $\pi N/\pi S$ between north group and south group, and the top 1% of $\pi N/\pi S$ scores were considered to be candidate sweeps.

FIGURE 4 Genome-wide association studies (GWAS) on two important overwintering traits. (a) The manhattan plot of GWAS on plant death rate in the panel of 134 accessions using FarmCPU algorithm, the red horizontal dashed line indicates the significance threshold of GWAS (1.79E-08) and the spectrum column represents the SNP density along chromosome. (b) The QQ plot of GWAS on plant death rate. (c) The manhattan plot of GWAS on plant withering rate in the panel of 134 accessions using FarmCPU algorithm, the red horizontal dashed line indicates the significance threshold of GWAS (1.79E-08) and the spectrum column represents the SNP density along chromosome. (d) The QQ plot of GWAS on plant withering rate.

FIGURE 5 Identification and analysis of SNP 1:38895333 associated with frost-free period. (a) The manhattan plot of GWAS on frost-free period of the sampling sites in the panel of 134 accessions using FarmCPU algorithm, the red horizontal dashed line indicates the significance threshold of GWAS (1.79E-08) and the spectrum column represents the SNP density along chromosome, 1:38895333 was located within the first exon of gene *Bp01g2734* . (b) The selective signals on chromosomes1 that overlapped with SNP 1:38895333, the red line indicates the threshold. (c) Distribution of genotypes of 1:38895333 in three geographic groups of north, east and south. (d) Box plots for plant death rate based on the different genotypes of 1:38895333, 3:1782695 and 11:16274932. (e) Box plots for plant withering rate based on the different genotypes of 1:38895333, 3:1782695 and 11:16274932.

FIGURE 6 A potential cold stress signaling pathway in woody plants based on the candidate genes in this

paper. LRR-RLK is a possible cold signal sensor, involved in perception of cold stimulus in woody plants, then MAPK is activated by LRR-RLK and regulate the expression of*CBFs* and their target genes (LEAs, GT, Peroxidases). Also, CBFs represses the expression of *GA3ox* genes, thus reduce bioactive GA levels and restrains plant growth, while *CBFs* are also regulated by ERF1. ABA binds to PYLs (ABA receptor) and inhibits the activity of PP2A, thereby modulating auxin transport and regulating root developmental under stress. Also, ABA induces production of MYBs and ABFs, which activating the downstream FHT gene and promoting suberin biosynthesis. The E3 ubiquitin ligase may function upstream of PP2A under low temperature and suppress the transcriptional activation of ICE1 through ubiquitinating. The ARRs are involved in the reduction of auxin accumulation to inhibit root growth under low temperature. BCS1 ATPase responds to SA and probably response to cold stress through regulating mitochondrial respiration (MR). Under stress, monodehydroascorbate (MDHA) is reduced to ascorbic acid (AsA) by monodehydroascorbate reductase (MDHAR) for scavenging ROS. Helicase-like proteins SEN1 and SMARCA3 are involved in regulating transcription, PAIP1 (polyadenylate-binding protein-interacting protein) interacts with PABP (polyadenylate-binding protein) regulating translation. Red font indicates the genes or proteins detected via GWAS or selection sweep analysis, solid lines represent direct or indirect regulation, and dashed lines represent that mechanism is unclear.

SUPPORTING INFORMATION

Figure S1 The geographic distribution of the 494 accessions used for RAD-seq.

Figure S2 The improved workflow used in this study for RAD-seq libraries compared with traditional method.

Figure S3 Summary of genome-wide variations within various genomic regions based on RAD-seq data.

Figure S4 A neighbor-joining tree constructed with 8,747 SNPs from RAD-seq.

Figure S5 Population structure of the 494 accessions performed RAD-seq (K = 3).

Figure S6 PCA plot of the first two components constructed based on 8,747 SNPs from RAD-seq.

Figure S7 Sequencing depth of the 134 accessions re-sequenced.

Figure S8 The density of the SNPs on 13 chromosomes of the 134 re-sequenced samples within 1Mb window size.

Figure S9 The summary of SNP mutation types in the whole genome based on the re-sequencing data.

Figure S10 The distribution of Ts/Tv along chromosomes based on the re-sequencing data.

Figure S11 The eigenvalues of the first 10 eigenvectors (Ev1 to Ev10).

Figure S12 The cross-validation error estimate plot used for selecting the optimal K value of the population structure.

Figure S13 The selective sweeps of F_{ST}between north group and south group in chromosome 2.

Figure S14 GO annotation of genes located in the selected regions (with the top 1% ofF_{ST} values) between the north group and south group.

Figure S15 GO annotation of genes located in the selected regions (with the top 1% of $\pi N/\pi S$ values) between the north group and south group.

Figure S16 KEGG annotation of genes located in the selected regions (top 1%) of F_{ST} between north group and south group.

Figure S17 KEGG annotation of genes located in the selected regions (top 1%) of $\pi N/\pi S$ between north group and south group.

Figure S18 Frequency distribution of variation in three traits used for genome-wide association study (GWAS).

Figure S19 Manhattan and Q-Q plots resulting from genome-wide association studies through MLM algorithm.

Figure S20 The summary of genotypes of 3:1782695 in three geographic groups of north, east and south.

Figure S21 The summary of genotypes of 11:16274932 in three geographic groups of north, east and south.

Table S1 Geographical distribution pattern of 494 accessions for RAD-seq.

Table S2 Summary of RAD-seq data from 494 accessions.

Table S3 Distribution of SNPs within various genomic regions based on RAD-seq data.

Table S4 Distribution of InDels within various genomic regions based on RAD-seq data.

Table S5 Summary of the variations identified in the genomes of 494 accessions.

Table S6 The name, origin, geographical distribution and bioclimatic variables for 134 accessions.

Table S7 The whole genome re-sequencing information of 134 accessions.

Table S8 The whole genome SNP mutation types and the ration of transition/transversion (Ts/Tv).

Table S9 The details of genetic groups within 134 accessions, population structure (Q group) analyzed by ADMIXTURE.

Table S10 Selected sweeps by top 1% highestF_{ST} between north group and south group.

Table S11 The list of genes in the selected regions (top 1%) of F_{ST} between north group and south group.

Table S12 Selected sweeps by top 1% highest πN/πS between north group and south group.

Table S13 The list of genes in the selected regions (top 1%) of πN/πS between north group and south group.

Table S14 The phenotypic data used for genome-wide association study (GWAS).

Table S15 Descriptive statistics of the traits used for GWAS analysis.

Table S16 The paired correlation coefficients among the three traits investigated for GWAS analysis.

Table S17 The paired correlation coefficients between three traits investigated for GWAS analysis and bioclimatic variables of sampling sites.

Table S18 The detail information of the significant loci and candidate genes identified by GWAS analysis.

Table 1 Summary of candidate genes identified through GWAS analysis and selective sweep analysis.

Gene ID	Start	End	SNP	P v
GWAS analysis	**GWAS analysis**	**GWAS analysis**	**GWAS analysis**	**GW**
Bp01g1321	17817043	17818808	1:17805729	1.5
Bp02g1805	30595868	30600263	2:30599497	9.4
Bp02g1806	30600783	30610348	2:30599497	9.4
Bp08g1155	9643161	9654292	8:9642341	9.3
Bp08g1595	14221905	14228438	8:14222246	1.3
Bp08g2102	21400436	21406554	8:21407459	1.0
Bp01g1842	26078704	26084418	1:26086393	2.3
Bp02g2091	33248418	33254388	2:33247134	9.4
Bp04g1554	28293593	28297888	4:28298004	9.3
Bp04g1851	30937659	30944110	4:30942806	3.4

Bp04g1852	30945722	30948624	4:30942806	3.4
Bp05g1281	22451334	22451858	5:22442064	5.3
Bp10g1488	17147030	17148853	10:17145593	3.3
Bp12g1206	12928682	12929350	12:12925009	2.0
Bp13g1480	17681004	17682842	13:17681302	2.5
Bp13g0698	10470907	10477099	13:10465150	1.6
Bp03g0251	1783380	1795958	3:1782695	3.6
Bp11g1556	16277719	16278444	11:16274932	6.6
Bp01g2734	38894888	38897675	1:38895333	3.2
Selective sweep analysis	**Selective sweep analysis**	**Selective sweep analysis**	**Selective sweep analysis**	**Sel**
Bp09g0248	3368381	3374190	-	-
Bp09g0249	3396433	3399661	-	-
Bp09g0250	3408182	3410031	-	-
Bp09g0251	3414811	3418145	-	-
Bp09g0252	3435005	3438558	-	-
Bp09g0253	3440604	3448303	-	-
Bp09g0254	3483884	3501526	-	-
Bp09g0255	3509092	3512987	-	-
Bp09g0256	3519132	3521729	-	-