



Genomic Differentiation and Demographic Histories of Two Closely Related Salicaceae Species

Zhe Hou* and Ang Li

College of Landscape Engineering, Suzhou Polytechnic Institute of Agriculture, Suzhou, China

Populus alba (*P. alba*) and *Populus davidiana* (*P. davidiana*) are important plant species for answering a variety of issues on species evolution due to their wide distribution and ability to adapt to a variety of environments and climates. Even though *P. alba* and *P. davidiana* belong to ecologically and economically important forest trees in the Northern Hemisphere, little is known about their genomic landscape and genome divergence during speciation. We re-sequenced 20 and 19 members of *P. davidiana* and *P. alba*, respectively, and found that the D_{xy} value between *P. alba* and *P. davidiana* was 0.2658, whereas the F_{ST} values were 0.2988, indicating that the genetic divergence was fairly clear. *Populus davidiana* and *P. alba* diverged from the ancestor in the middle Pleistocene, c. 0.80 Ma (95% HPD: 0.79–0.81 Ma). The population sizes of *P. davidiana* increased ~20,000 years ago after a considerable long-term decline following divergence. However, after differentiation, the effective population size of *P. alba* expanded slightly before experiencing a long-term bottleneck effect. According to the expectation of allopatric speciation, we found a significant number of genomic differentiation sites in both species' speciation events, and the majority of these genomic differentiation regions can be attributed to neutral evolutionary processes. Nevertheless, the regions with extreme divergence exist in abundance, indicating that natural selection has had an impact. Positive selection can be found in highly differentiated regions, while long-term balancing selection traits can be easily observed in low differentiated regions. According to these findings, climate differences over the Quaternary, as well as variance in linked selection and recombination, all contributed significantly to genomic divergence during allopatric speciation of the two aspens.

Keywords: whole-genome sequencing, genetic diversity, demographic history, adaptation, natural selection

INTRODUCTION

One of the fundamental goals of evolutionary genomics is to understand the relative contributions of diverse evolutionary factors in the generation and shaping of genetic diversity within and among species (Nordborg et al., 2005). Determining the evolutionary forces affecting patterns of genome-wide variation has been a central goal in evolutionary biology over the past several decades (Seehausen et al., 2014). An important aspect of understanding speciation is the study of variation in levels of differentiation among closely related species (Mark et al., 2018). Under strict neutrality, differentiation is expected to accumulate as a result of the stochastic fixation of polymorphisms by genetic drift (Hellmann et al., 2005). Demographic fluctuations, including population

OPEN ACCESS

Edited by:

Daniel Pinero,
National Autonomous University of
Mexico, Mexico

Reviewed by:

Alejandra Moreno-Letelier,
National Autonomous University of
Mexico, Mexico
Alejandra Vázquez-Lobo,
Universidad Autónoma del Estado de
Morelos, Mexico

*Correspondence:

Zhe Hou
houzhedota@163.com

Specialty section:

This article was submitted to
Plant Systematics and Evolution,
a section of the journal
Frontiers in Plant Science

Received: 02 April 2022

Accepted: 06 May 2022

Published: 07 June 2022

Citation:

Hou Z and Li A (2022) Genomic
Differentiation and Demographic
Histories of Two Closely Related
Salicaceae Species.
Front. Plant Sci. 13:911467.
doi: 10.3389/fpls.2022.911467

expansions and bottlenecks, can decelerate or accelerate the rate of differentiation and genome-wide variation in various ways (Li and Durbin, 2011). Historical, geographic and climatic events can trigger population differentiation by affecting the demographic history of the populations (Sanna et al., 2008). In the absence of gene flow, differentiation between populations increases gradually as a result of natural selection and/or random processes like mutation and genetic drift (Begun et al., 2007).

For those genome regions that are affected by natural selection, the linked sites can also be affected in several ways (Via, 2009). Moreover, as the speciation process progresses, numerous forms of selection can impact the patterns of development of genomic divergence and genetic variation. Various types of natural selection (purifying, balancing, and positive selection) are adequate in these settings to produce varied genomic differentiation patterns and even speciation (Turner et al., 2005; Noor and Bennett, 2009). Regardless of the role of natural selection, demographic dynamics, stochastic genetic drift, and recombination rate can all alter genome-wide patterns of diversity and genomic architecture of adaptation (Nosil et al., 2009; Campagna et al., 2015). The hypotheses outlined previously are generally not mutually exclusive, and a complete analysis of these hypotheses requires a thorough understanding of the speciation process, such as the period of speciation and the demographic and geographic environment in which it occurred (Nosil and Feder, 2012). As it becomes increasingly feasible to generate whole genome resequencing data from closely related species, the importance of conserved genomic features in shaping the topography of the genomic landscape of speciation has increasingly been highlighted (Wang et al., 2020).

Forest trees provide a promising resource to address adaptive evolution and patterns of genome variation because most of them are undomesticated, possess a variety of phenotypic and genetic variations, and can adapt to a variety of climates (Neale and Antoine, 2011). Here, we focus on two *Populus* species (*Populus alba* L and *Populus davidiana* Dode) to evaluate the evolutionary forces affecting patterns of genome-wide variation. The two species are all from the section *Leuce Duby* of the genus *Populus*. *Populus alba* and *P. davidiana* are the most ecologically important forest trees in China, as well as ecologically important members of mesic forest ecosystems that were isolated by Pleistocene glaciations (Jianchao et al., 2018). Initial investigations based on complete plastid genomic data (Lei et al., 2018) and resequenced genomes (Wang et al., 2020) demonstrated that *P. alba* and *P. davidiana* have the closest relationship. *Populus davidiana* and *P. alba* are sibling aspen species based on their close phylogenetic relationship and morphological similarity belonging to the section of *Leuce (Populus)* (Eckenwalder, 1977). *Populus davidiana* in this study is mainly distributed in the northeastern part of China, which is very cold, but *P. davidiana* is a strong positive species, tolerant of cold and dry and infertile soil (Zong et al., 2019). *Populus alba* is mainly distributed in the Xinjiang region of China, which has a very arid climate and little rainfall, but *P. alba* is wind-resistant and tolerant of dry climate (Honglei et al., 2017). The taxonomy of these two aspens has been controversial with respect to their extreme morphological congruence with only minor

differences in leaf shape (Löve and Löve, 1975). However, they have significant genetic and ecological differences, indicating that a strong divergent ecological selection has influenced them (Honglei et al., 2017). *Populus alba* and *P. davidiana* have close genetic affinity based on previous phylogenetic analyses (**Supplementary Figure 2**) (Liu et al., 2016). *Populus alba* and *P. davidiana* are sister species, according to a recent study (Zong et al., 2019) based on morphological analysis and whole chloroplast genome sequences (**Supplementary Figure 3**).

Both *P. davidiana* and *P. alba* have wide geographic distribution, high intraspecific polymorphism, adaptability to different environments, phenotypic diversity, combined with a relatively small genome size (Liu et al., 2019), and further facilitation has taken from the accessibility of reference data of a good caliber from the genome of *Populus trichocarpa* Torr. & A. Gray ex. Hook. reference genome, version 3 (Tuskan, 2006), the two species provide an ideal system to study heterogeneous genomic differentiation, genetic diversity, demographic history, adaptation, and gene flow during the process of speciation. In the current work, we collected and sequenced 39 individuals from *P. davidiana* and *P. alba* populations in China. We specifically aimed to (i) explore their population structure; (ii) estimate the historical demographic processes and species divergence time points; (iii) infer the overall patterns of heterogeneous genomic differentiation and fine-scale genomic landscapes of diversity; (iv) identify signatures of long-term balancing selection and positive selection over the entire genome. Specifically, the main purpose of this study is to understand and disentangle how the patterns of genetic variation are shaped by the multitude of evolutionary forces within and among species during the process of speciation.

MATERIALS AND METHODS

Population Sample Collection, Sequencing, and Genotype Calling

Silica-gel dried leaves of 19 *P. alba* and 20 *P. davidiana* individuals covering their geographical distributions in China were collected for DNA extraction (**Table 1**). Genomic DNA from the leaf of each individual was extracted using the CTAB method (Pahlich and Gerlitz, 1980) and Qubit dsDNA BR assay (Life Technologies, Carlsbad, CA, USA) was employed to assess the quantity and quality of DNA. Paired-end sequencing libraries with an insert size of 600 bp were constructed using the standard Illumina HiSeq 2000 platform protocol. Later on, every *P. alba* and *P. davidiana* specimen was sequenced on the Illumina HiSeq 2000 platform. All samples were sequenced to a target coverage of 30×. The raw sequence data reported in this work have been submitted with the Genome Sequence Archive (Wang et al., 2017) at the BIG Data Center, Beijing Institute of Genomics (BIG), Chinese Academy of Sciences, with accession numbers CRA003302 and CRA001674 for samples of *P. alba* and *P. davidiana*, respectively. CRA003302 and CRA001674 are publicly accessible at <http://bigd.big.ac.cn/gsa>. Adapter sequences of raw sequencing reads were eliminated with Trimmomatic (Lohse et al., 2012) before reading mapping. The bases were cut off

TABLE 1 | Statistics summary of Illumina re-sequencing data per sample.

SampleID	Location	Latitude	Longitude	Mapping rate (%)	Data volume(Gb)
<i>P. davidiana</i>					
<i>P. davidiana</i> 1	Shuangyashan	E 131°15'	N 46°64'	92.36	6.8
<i>P. davidiana</i> 2	Shuangyashan	E 131°15'	N 46°64'	91.25	6.5
<i>P. davidiana</i> 3	Huanan	E 130°55'	N 46°23'	92.08	8.2
<i>P. davidiana</i> 4	Huanan	E 130°55'	N 46°23'	91.78	6.3
<i>P. davidiana</i> 5	Heihe	E 127°52'	N 50°24'	91.25	6.9
<i>P. davidiana</i> 6	Heihe	E 127°52'	N 50°24'	92.36	6.1
<i>P. davidiana</i> 7	Heihe	E 127°52'	N 50°24'	92.08	6.2
<i>P. davidiana</i> 8	Heihe	E 127°52'	N 50°24'	91.98	6.6
<i>P. davidiana</i> 9	Heihe	E 127°52'	N 50°24'	92.05	7.0
<i>P. davidiana</i> 10	Lianhuashan	E 124°49'	N 44°47'	93.25	7.2
<i>P. davidiana</i> 11	Lianhuashan	E 124°49'	N 44°47'	91.28	6.8
<i>P. davidiana</i> 12	Lianhuashan	E 124°49'	N 44°47'	91.89	6.3
<i>P. davidiana</i> 13	Lianhuashan	E 124°49'	N 44°47'	91.21	7.3
<i>P. davidiana</i> 14	Shuangshan	E 123°88'	N 43°68'	93.25	6.0
<i>P. davidiana</i> 15	Shuangshan	E 123°88'	N 43°68'	92.08	6.1
<i>P. davidiana</i> 16	Weichang	E 117°76'	N 41°93'	91.23	6.4
<i>P. davidiana</i> 17	Wutai	E 113°25'	N 38°72'	91.78	7.5
<i>P. davidiana</i> 18	Hualin	E 116°40'	N 39°90'	91.96	7.1
<i>P. davidiana</i> 19	Wutai	E 113°25'	N 38°72'	91.98	6.9
<i>P. davidiana</i> 20	Hualin	E 116°40'	N 39°90'	92.25	6.2
<i>P. alba</i>					
<i>P. alba</i> 1	Xinjiang	E 131°15'	N 46°64'	92.36	5.9
<i>P. alba</i> 2	Xinjiang	E 131°15'	N 46°64'	91.25	6.2
<i>P. alba</i> 3	Xinjiang	E 130°55'	N 46°23'	92.08	7.3
<i>P. alba</i> 4	Xinjiang	E 130°55'	N 46°23'	92.78	6.3
<i>P. alba</i> 5	Xinjiang	E 127°52'	N 50°24'	91.25	6.9
<i>P. alba</i> 6	Xinjiang	E 127°52'	N 50°24'	92.36	6.2
<i>P. alba</i> 7	Xinjiang	E 127°52'	N 50°24'	91.08	6.1
<i>P. alba</i> 8	Xinjiang	E 127°52'	N 50°24'	91.98	6.1
<i>P. alba</i> 9	Xinjiang	E 127°52'	N 50°24'	92.05	6.8
<i>P. alba</i> 10	Xinjiang	E 124°49'	N 44°47'	93.25	6.5
<i>P. alba</i> 11	Xinjiang	E 124°49'	N 44°47'	91.28	6.7
<i>P. alba</i> 12	Xinjiang	E 124°49'	N 44°47'	91.89	6.0
<i>P. alba</i> 13	Xinjiang	E 124°49'	N 44°47'	91.21	6.9
<i>P. alba</i> 14	Xinjiang	E 123°88'	N 43°68'	93.25	6.6
<i>P. alba</i> 15	Xinjiang	E 123°88'	N 43°68'	92.08	6.2
<i>P. alba</i> 16	Xinjiang	E 123°88'	N 43°68'	91.78	6.0
<i>P. alba</i> 17	Xinjiang	E 123°88'	N 43°68'	92.58	6.3
<i>P. alba</i> 18	Xinjiang	E 126°66'	N 51°72'	91.32	6.5
<i>P. alba</i> 19	Xinjiang	E 126°66'	N 51°72'	93.25	6.0

and trimmed from the beginning or the ending of reads when the base quality was ≤ 20 . Reads were completely discarded if there were fewer than 36 bases remaining after trimming. The remaining reads were mapped via the BWA-MEM algorithm employing the default parameters in BWA v0.7.8 (Li et al., 2009) against *Populus trichocarpa* Torr. & A. Gray ex. Hook. reference genome, version 3 (Tuskan, 2006). We further employed several filtering steps to minimize the influence of mapping bias before

genotype calling. First, the genome regions around deletions or insertions had some misalignment of bases, and we corrected the bias by IndelRealigner and RealignerTargetCreator in GATK v3.8.0 (Depristo et al., 2011) with default parameters. In order to account for the occurrence of PCR duplicates introduced during library construction, we used MarkDuplicates in Picard (<http://picard.sourceforge.net>) to remove reads with identical external coordinates and insert lengths. For downstream analyses, we only retained the reads with the highest summed base quality. We further filtered the reads that could potentially cause mapping bias according to three additional standards: (1) those with extreme read coverage (<4 or higher than twice of the mean coverage); (2) RepeatMasker detecting overlapping known repeated elements (Tarailo-Graovac and Chen, 2009); (3) covered by greater than two reads of mapping score equaling zero per individual.

After filtering, two complementary methods were implemented for downstream analysis. First, a series of population genetic analyses was carried out using ANGSD v0.917. The pipeline of ANGSD v0.917 relies on allele frequency spectrum and genotype probabilities but does not take genotype calls into account (Korneliussen et al., 2014). Second, the HaplotypeCaller of the GATK v3.8.0 was implemented to perform genotype calling in each individual for downstream analysis requiring accurate genotype calls. For re-annotation and re-genotyping of the newly merged VCF, we used GenotypeGVCFs to unify multi-sample records from the two species together (Depristo et al., 2011). We further performed a set of filtering steps to reduce genotype calling bias, such that the downstream analyses were carried out by retaining only high-quality single nucleotide polymorphisms (SNPs): (1) removed all SNPs sites that had not passed previously the filtering criterion; (2) reserved only 2 alleles with a distance ≥ 5 bp away from any indels; (3) genotypes with read depth (DP) <5 or with a genotype quality score (GQ) <10 were categorized as missing, all SNPs with a genotype missing rate over 10% were discarded.

Population Structure, Genetic Relationship, and Principal Components Analysis

NGSadmix was implemented to deduce the genetic structure of the *P. alba* and *P. davidiana* population, and only used sites with $\leq 10\%$ of missing data (Skotte et al., 2013). The SAMTools model in ANGSD (Li et al., 2009) was employed to assess the genotype likelihoods and later a beagle file was generated to infer the population genetic structure with a likelihood ratio test ($P < 10^{-6}$) (Wang et al., 2016), and the count of genetic clusters (K) ranging from 1 to 6. The program of smartpca in PCAngsd software (<http://www.popgen.dk/software/index.php/PCAngsd>) was used to perform the Principal component analysis (PCA) with a Tracy–Widom test to ascertain the significance level of eigenvectors. The neighbor-joining (NJ) model with TreeBest software (<http://treesoft.sourceforge.net/treebest.shtml>) was used for the construction of the phylogenetic tree.

Demographic History

We used individuals with high sequencing coverage ($>35\times$) from each population and applied the Pairwise Sequentially Markovian Coalescent (PSMC) model to reconstruct demographic history (Li and Durbin, 2011). The parameter was set to the default parameter during the calculating procedure. Based on these parameters to covert scaled population sizes and time to actual sizes and time, the mutation rate was set as 3.75×10^{-8} per base per generation, and we adopted a generation time of 15 years. We carried out 100 bootstrapping simulations to estimate the variance fluctuation of population size (Koch et al., 2000).

We further used Fastsimcoal 2.6.1 software (Excoffier et al., 2013) to infer past demographic histories, the divergence time, and gene flow patterns of *P. alba* and *P. davidiana*. Allele frequencies in the 39 samples were calculated using the realSFS module in ngsTools software so as to construct the required two-dimensional joint site frequency spectrum (2D-SFS), which was estimated with 100,000 coalescent simulations in each model. Alternative models of historical events were fitted to the joint site frequency spectra data (Fumagalli et al., 2014). We simulated 18 different models and all these models proceed with the divergence of an ancestral species into *P. alba* and *P. davidiana* lineage. Various models were made to model the earlier demographic and speciation histories of *P. alba* and *P. davidiana* that varied with respect to (1) No Migration and no population expansion; (2) Asymmetric Migration but no population expansion; (3) Asymmetric Migration along with population expansion; (4) complicated model, comprising a bottleneck in *P. alba*; (5) complicated model, containing a bottleneck in *P. davidiana* (Supplementary Figure 1). The two-dimensional joint SFS data had a good fit with the alternative demographic history models. We obtained global maximum likelihood estimates for the 18 different models from 50 individual runs, with 40 conditional rounds of the likelihood maximization algorithm. The criteria for comparison of these models was the maximal value of likelihood generated by 50 independent runs on the basis of calculated Akaike's weight (Excoffier et al., 2013). The model with the highest Akaike's weight estimate was deemed the best of the models. Parametric bootstrapping was carried out with 100 bootstrap cycles and 50 independent runs in every single bootstrap and was used to calculate confidence intervals. We also considered 15 years as a generation (Tuskan, 2006) and the mutation rate of each generation was 3.75×10^{-8} (Ingvarsson, 2008) for the conversion of estimates to units of individuals and years.

Genetic Diversity and Divergence

The software VCFtools with the default parameters was used to calculate the levels of genetic differentiation (Weir and Cockerham mean F_{ST}), nucleotide diversity (π), and Tajima's D for individual species. We then took an average of (within each 10 Kbp non-overlapping window) F_{ST} , π , and Tajima's D values of all sites (Petr et al., 2011).

Genome-Wide Patterns of Differentiation and Detection in Outlier Windows

In order to examine thresholds for detection of outlier windows that may have been targets of natural selection, we conducted coalescent simulations to compare observed patterns of genetic differentiation (F_{ST}) to those expected under different demographic models (see Results). All simulations were performed using the program *msms* v3.2rc based on demographic parameters derived from the best-fitting model inferred by *fastsimcoal2.5.1*. Population-scaled recombination rates ρ were assumed to be between 1 and 5 Kbp⁻¹ given the large variation. We simulated genotypes corresponding to a 10 Kbp region with the same sample size as the real data for 100,000 replications, from where we simulate genotype likelihoods using the program *msToGlf* in ANGSD (Korneliussen et al., 2014) by assuming a mean sequencing depth of $20\times$ and an error rate of 0.5%. We estimated two summary statistics, nucleotide diversity π and Tajima's D, from sample allele frequency likelihoods in ANGSD for all simulation replicates to test whether the simulated data matches the observed data. To assess whether any of the observed windows display F_{ST} values deviating significantly from neutral expectations, we determined the conditional probability (P -value) of observing more extreme inter-specific F_{ST} values among simulated data sets than among the observed data. The determination of significance was based on running 500,000 coalescent simulations of the most acceptable demographic null model. We then corrected for multiple testing by using the False Discovery Rate (FDR) adjustment, and only windows with FDR lower than 1% were considered as candidate regions affected by selection.

Population Genetic Analysis and Molecular Signatures in Outlier Regions

In both *P. alba* and *P. davidiana* species, multiple population genetic analyses were contrasted with the remaining regions of the genome in order to infer the molecular signatures of selection in outlier regions manifesting either extremely low or high degree of differentiation. ANGSD was first employed to assess the probabilities of sample allele frequency between populations of the *P. alba* and *P. davidiana* over non-overlapping 10 Kbp windows for the estimation of $\theta\pi$, F_u and F_L (1993) and Fay and Wu (2000). Further, to calculate the correlation coefficients (r^2) for evaluating the levels of LD within each 10 Kbp window, we used VCFtools v0.1.12b (Danecek et al., 2011). The recombination rates (ρ) within each 10 Kbp window were calculated using FastEPRR software (Gao et al., 2016). Finally, ngsStat (Fumagalli et al., 2014) was used to estimate a number of additional population genetic parameters of genetic differentiation: (1) the proportion of derived alleles fixed in either *P. alba* and *P. davidiana* and the proportionate number of inter-specific shared polymorphisms between all segregating sites; (2) dividing the d_{xy} of the *P. alba* and *P. davidiana* species by the d_{xy} between the *P. alba* population and *P. tremula* to estimate the relative node depth (RND); (3) the posterior probability of the sample allele frequency at each locus and averaged over each

10 Kbp window was used to calculate dxy. Wilcoxon ranked-sum tests were used to analyze the genome-wide mean for all population genetic data presented, as well as the significance of differences among outlier windows.

Gene Ontology (GO) and KEGG Pathways Enrichment

GO enrichment analysis was used to test if the highly differentiated regions had any overrepresentation of functional classes of genes, and a Fisher's exact test was carried out using agriGO's Term Enrichment tool (<http://bioinfo.cau.edu.cn/agriGO/index.php>) (Du et al., 2010). In an attempt to rectify the *P*-value corresponding to Fisher's exact test, numerous tests were carried out by employing the Benjamini-Hochberg error detection rate. Taking *P*-value under 0.05 as standard, we identified significantly enriched GO terms. The KEGG pathways were analyzed using the KOBAS system (Mao et al., 2005) and the FDR method was applied to correct the different comparisons.

RESULTS

We generated whole-genome resequencing data for 19 *P. alba* and 20 *P. davidiana*. *Populus alba* and *P. davidiana* were found to have a high degree of conservation in their genomes (Pakull et al., 2009), in a way that roughly 91.08% (Table 1) of all sequences of *P. alba* and *P. davidiana* can be conveniently mapped to the *P. trichocarpa* reference genome (Tuskan, 2006) following a quality control process. In the mapped reads of *P. alba* and *P. davidiana* samples, the mean coverage of individual site reached 31.86 (Table 1). SNP and genotype data were reliably obtained by two different yet complementary methods: (1) High-quality site-frequency-spectrum (SFS) data were obtained using ANGSD software (Korneliussen et al., 2014) for determining the genetic parameters of the population without calling genotypes (Nielsen, 2005). (2) Multi-sample SNP and genotype calling was implemented in GATK v3.2.2 with HaplotypeCaller and GenotypeGVCFs (Danecek et al., 2011). Following filtration and thorough quality control, the 19 *P. alba* and 20 *P. davidiana* samples yielded 5,035,016 and 5,158,369 high-quality SNP sites, respectively.

Structure of the Population

NGSadmix was employed to infer individual ancestry based on genotype likelihoods, which takes the uncertainty of genotype calling into account. When *K* was set as 2, the majority of the members were categorized into two different genetic clusters. When *K* = 3, *P. alba* showed additional sub-structuring of the population, but the bulk of *P. alba* members were deduced to be a product of two genetic components mixing, demonstrating very minor clinal variation with increasing latitude. When *K* = 4, hybridization and introgression can be found between *P. alba* and *P. davidiana* (Figure 1A). We constructed a neighbor-joining tree that further reinforced these patterns, with the different geographical placement of *P. alba* and *P. davidiana* reflecting the grouping of populations (Figure 1B). The results were also reinforced by PCA since there were two distinct

populations of the 39 *P. alba* and *P. davidiana* specimens taken from different regions (Figure 1C). Fixed differences accounted for 3% of polymorphisms in the two populations, whereas shared polymorphism accounted for 12%, and private polymorphic loci accounted for 42% and 43% polymorphism in the populations of *P. alba* and *P. davidiana*, respectively (Figure 1D).

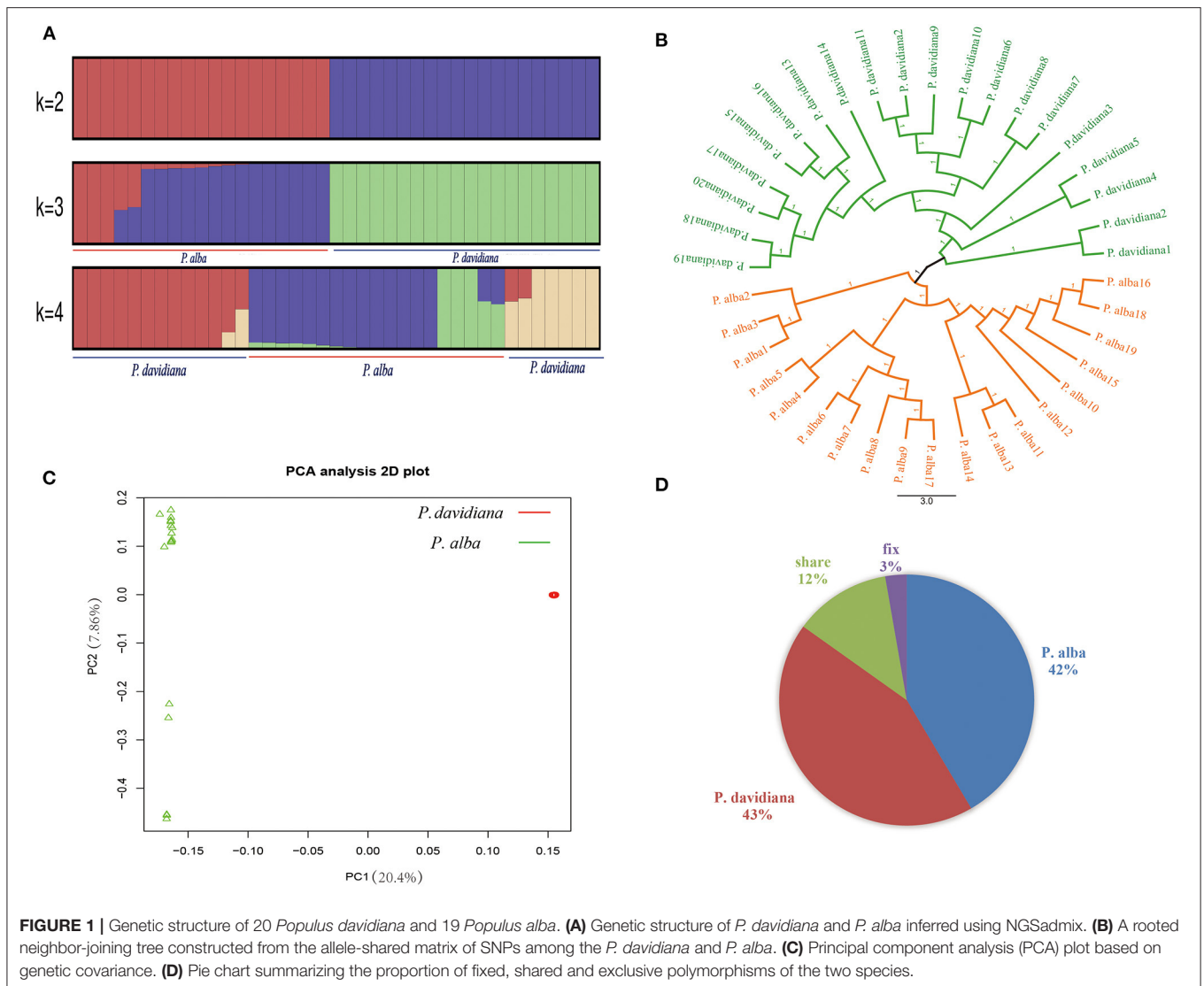
Demographic Histories

The demographic histories and timings of divergence of the two species were further estimated by employing a coalescent simulation-based technique carried out in Fastsimcoal 2.6.1 (Excoffier et al., 2013). There were 18 models developed that focused on the divergence of *P. alba* and *P. davidiana* (Supplementary Figure 1). A comparison of the AIC values for all four scenarios established that the complex isolation-with-migration model (model 4; Figure 2A; Table 2) with the highest likelihood and lowest AIC value is the best-fitting model. As revealed by the estimates of parameter acquired from the best model, it was found that following the divergence of the two species, a stepwise change in the population size occurred in *P. davidiana*, however *P. alba* underwent exponential growth (Figure 2A). Table 3 lists the details regarding differentiation time point, effective population size, and gene flow of *P. alba* and *P. davidiana*, displaying the 95% confidence interval (CIs) for the associated variables alongside. Around 0.80 million years ago (Mya), the progenitors of the two aspens underwent divergence into *P. alba* and *P. davidiana* populations (bootstrap range [BR]: 0.79–0.81 Mya). Currently, the functional population sizes (N_e) of *P. davidiana* ($N_{e-P.davidiana}$) and *P. alba* ($N_{e-P.alba}$) are 1,882,692 (BR:1,863,658–1,891,256) and 2,663,695 (BR:2,651,235–2,678,958) respectively. However, both species have a significantly low effective population sizes in comparison to their common ancestor ($N_{e-ANC} = 7,826,134 [7,652,251-7,962,365]$). The rate of migration (*m*) also distinguishes between the two species, the least generation migration rate (*m*) between *P. alba* and *P. davidiana* (3.97×10^{-8} and 9.16×10^{-7}), which is not accidental as the two populations are separated by a significant geographical distance.

PSMC (Li and Durbin, 2011) was used to estimate demographic history which depicted that a dramatic decline was experienced by both species in terms of population size around 100 Ma (Figure 2B). In comparison to the population of *P. alba*, which only recovered moderately about 800 kya, the population sizes of *P. davidiana* increased around 20,000 years ago, after a long period of decrease following divergence (Figure 2B).

Genome Differentiation and Identification of Outlier Regions

We have previously shown that linkage disequilibrium (LD) decays within 10 kilobases (Kbp) in both *P. alba* and *P. davidiana*, and thus 10,000 bp windows were used to investigate patterns of genomic differentiation between species. As a standard parameter evaluating genetic differentiation, the fixation index F_{ST} is sensitive to any event or process that leads to an alteration in interspecific variation (Cruickshank and Hahn, 2014). The current work involves the calculation of the genetic



differentiation coefficient F_{ST} for the two species. Since the mean F_{ST} value between the two species was 0.2988 (Table 5), genetic differentiation was found to be evident among the two populations. Total sequence differentiation among the populations which stands as a universal criterion for evaluating interspecific differentiation and is referred to as d_{xy} was also calculated. The two populations were found to have evident sequence differentiation among them, with the mean value of d_{xy} between *P. davidiana* and *P. alba* being 0.2658 (Table 5 and Supplementary Table 1).

Comparing the empirical distribution of inter-specific F_{ST} with that obtained from simulations based on the best-fitting demographic model, we identified 379 and 1,156 outlier windows exhibiting significantly (False Discovery Rate < 0.01) high and low interspecific F_{ST} compared to the expected null distribution obtained from the coalescent simulations (Supplementary Tables 2, 3).

Population Genetic Analysis

The parameters $\theta\pi$, θ_w , H_E , ρ , and Tajima's D were calculated for the two species to estimate the patterns of genetic diversity throughout the entire genome. The lowest H_E was manifested by individuals of *P. davidiana* (Table 5), whereas the highest H_E was harbored by individuals of *P. alba*. In light of the results of H_E , relatively lower genetic diversity ($\pi = 0.0089$) is exhibited by *P. davidiana* in comparison to *P. alba* ($\pi = 0.0095$; Table 5). *P. alba* possesses had a slower LD decay and a slightly higher recombination rate (Table 5) compared to *P. davidiana* (Figure 3). In comparison to *P. davidiana* (Tajima's D = 0.13), *P. alba* (Tajima's D = -0.06) has a lower Tajima's D-value. The physical distance after the average LD coefficient r^2 decays to half of the maximal value is generally referred to as the LD decay distance. Different LD decay curves are manifested by the *P. alba* and *P. davidiana* populations (Figure 3), suggesting diverse demographic backgrounds of both species, population reduction,

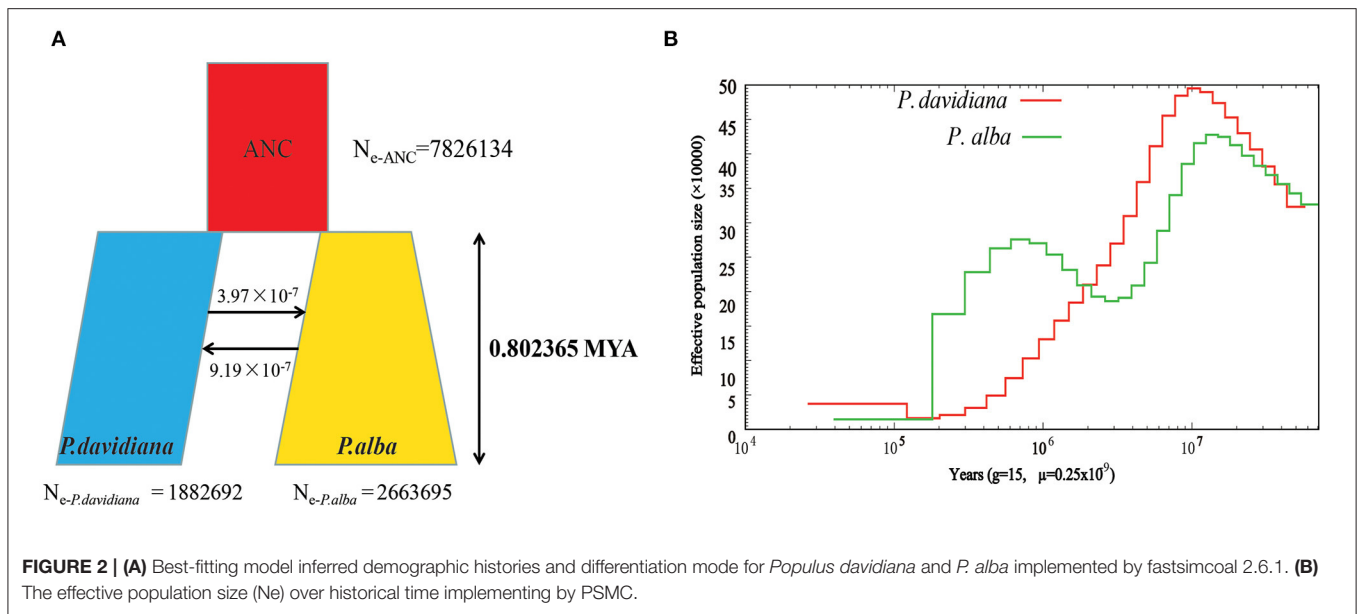


TABLE 2 | Relative likelihood of the different models.

Model	Max(log10(Lhood)) ^a	AIC _i ^b	Δi^b	Model normalized relative likelihood (w_i) ^b
Model 1	-752381556.3	356121231.2	3569618.436	~0
Model 2	-766668258.6	355678656.1	3658352.526	~0
Model 3	-769356470.5	354588633.5	2636451.235	~0
Model 4	-707411635.8	356531580.6	0	~1
Model 5	-738200638.82	355259547.5	2065856.026	~0
Model 6	-742253556.3	355955172.2	216486.022	~0
Model 7	-742377887.6	358863315.6	1868753.036	~0
Model 8	-741803292.7	354568532.2	658541.035	~0
Model 9	-72078163.5	353426226.1	1522688.015	~0
Model 10	-737524429.5	353223549.6	1478223.020	~0
Model 11	-737955129.5	352793852.6	4583631.025	~0
Model 12	-740129825.3	353659253.0	2587456.015	~0
Model 13	-743961354.6	353336542.2	5214852.061	~0
Model 14	-743705455.5	356823658.2	136586.099	~0
Model 15	-741996176.5	353338885.3	1253589.018	~0
Model 16	-740495214.920	356852515.6	2659565.498	~0
Model 17	-745052395.561	359596326.3	2568742.259	~0
Model 18	-739893331.173	355944120.3	375961.568	~0

^aBased on the best likelihood among the 50 independent runs for each model.
^bThe calculation of AIC_i, Δi , and w_i are according to the methods shown in Excoffier et al. (2013).

or genetic differentiation may alter the LD pattern of the genome. The *P. davidiana* population has the fastest decay rate and the smallest LD value, whereas the *P. alba* population had the slowest decay rate and the largest LD value (Figure 3).

Signatures of Selection in Outlier Regions

As F_{ST} is a relative measure of differentiation and is thus sensitive to any processes that alter intra-species genetic variation, we

TABLE 3 | Demographic parameters and confidence interval of the best model.

Parameters	Point estimation	95% CI ^a	
		Lower bound	Upper bound
N_{e-ANC}	7826134	7652251	7962365
$N_{e-P.davidiana}$	1882692	1863658	1891256
$N_{e-P.alba}$	2663695	2651235	2678958
$m_{P.davidiana \rightarrow P.alba}$	3.97×10^{-7}	3.50×10^{-7}	4.15×10^{-7}
$m_{P.alba \rightarrow P.davidiana}$	9.19×10^{-7}	8.99×10^{-7}	9.98×10^{-7}
TDIV	802365	793658	812546

Parameters are defined in Figure 2A. N_{e-ANC} , N_{e-N} , $N_{e-P.davidiana}$, and $N_{e-P.alba}$ indicate the effective population sizes of ancestral, *P. davidiana* and *P. alba* population respectively, $m_{P.davidiana \rightarrow P.alba}$ indicates the per generation migration rate from *P. davidiana* to *P. alba*, $m_{P.alba \rightarrow P.davidiana}$ indicates the per generation migration rate from *P. alba* to *P. davidiana*, TDIV indicates the estimated divergence time between *P. davidiana* and *P. alba* populations.

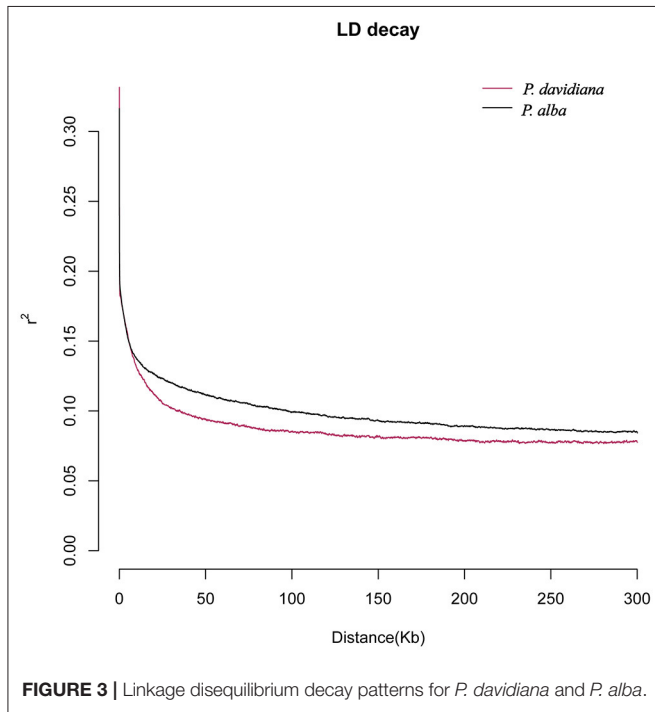
^aParametric bootstrap estimates obtained by parameter estimation from 100 data sets simulated according to the overall maximum composite likelihood estimates shown in point estimation columns. Estimations were obtained from 100,000 simulations per likelihood.

quantified and compared inter-specific genetic differentiation between two unions of outlier windows and the rest of the genome using several additional approaches (Table 4). Compared to the genomic background averages, both dxy and RND revealed significantly greater divergence between the two species in regions of high differentiation (Figures 5A,B). For both populations, the highly differentiated regions manifested distinct positive selection characteristics (Nielsen, 2005). As an example, both *P. alba* and *P. davidiana* populations had exceptionally low levels of polymorphism (π) (Figures 5A,B). The more negative Fay & Wu's H values reflected frequently appearing derived alleles (Figures 5C,D), whereas the more

negative Tajima's D values reflected frequently appearing rare alleles (Figures 5E,F). A more visible characteristic is the regions with high differentiation having powerful signals of linkage disequilibrium (LD) (Figures 5G,H) ($P < 0.001$, Mann-Whitney U test). Inter-specific shared polymorphisms among the two populations as well as the alleles fixed in the *P. alba* and *P.*

davidiana populations were also compared. The results indicated that in the regions with high differentiation, the ratio of inter-specific shared polymorphisms was very low (Figure 4C) and in both the *P. alba* and *P. davidiana* populations, the proportion of fixed differences was notably high (Figures 5K,L).

Nevertheless, long-term balancing selection characteristics are exhibited by regions of low differentiation (Charlesworth et al., 1995). For instance, less differentiation was exhibited by the *dxy* and RND values of the regions with low differentiation between both populations in comparison to the regions manifesting high differentiation (Figures 4A,B), both *P. alba* and *P. davidiana* populations exhibiting a notably high level of polymorphism (π) (Figures 5A,B). Frequently appearing intermediate-frequency alleles were represented by the higher Tajima's D and Fay & Wu's H values (Figures 5C-F), with LD levels lower in comparison to the corresponding levels in the regions with high differentiation (Figures 5G,H), possibly under the influence of recombination (Lee et al., 2011). In the low differentiated regions, the ratio of inter-specific shared polymorphisms was rather high (Figure 4C) whereas the proportion of fixed differences was found to be trivial in both the *P. alba* and *P. davidiana* populations (Figures 5K,L).



Influence of the Rate of Recombination on Genome Differentiation

Another important factor affecting the differentiation of the genome is the rate of recombination. Recombination rates were calculated using FastEPRR software ($\rho = 4N_e c$) over a 10,000 bp window size. Since $\rho = 4N_e c$, a decline in N_e in an area linked to selection decreases the local estimate of ρ . In an attempt to get rid of this impact, the influence of recombination rate upon genomic differentiation was evaluated by estimating ρ/θ_π in regions

TABLE 4 | Statistics summary comparing regions displaying extreme genetic differentiation with the rest of the genomic regions in both *P. davidiana* and *P. alba* (the values are shown as mean \pm standard deviation).

Parameters	Species	Regions displaying high differentiation	Regions displaying low differentiation	Background
θ_π	<i>P. davidiana</i>	0.0062(± 0.0036)***	0.0346(± 0.0253)***	0.0125(± 0.0056)
	<i>P. alba</i>	0.0089(± 0.0045)***	0.0359(± 0.0151)***	0.0159(± 0.0023)
Tajima's D	<i>P. davidiana</i>	-0.9175(± 0.5101)***	-0.0650(± 0.5126)***	-0.2988(± 0.5026)
	<i>P. alba</i>	-1.5966(± 0.3958)***	-0.6123(± 0.5622)***	-1.2113(± 0.4562)
Fay&Wu's H	<i>P. davidiana</i>	-0.5995(± 0.3658)***	-0.1169(± 0.2569)***	-0.3898(± 0.2759)
	<i>P. alba</i>	-0.5136(± 0.3042)***	-0.1154(± 0.2368)***	-0.3125(± 0.2032)
r^2	<i>P. davidiana</i>	0.2876(± 0.1513)***	0.1969(± 0.0926)	0.2102(± 0.1312)
	<i>P. alba</i>	0.2428(± 0.1368)***	0.1659(± 0.0960)*	0.1536(± 0.1123)
ρ/θ_π	<i>P. davidiana</i>	0.1655(± 0.2643)***	0.1142(± 0.1199)**	0.2332(± 0.3502)
	<i>P. alba</i>	0.2588(± 0.3535)***	0.2066(± 0.2356)**	0.5399(± 0.5201)
Fixed (%)	<i>P. davidiana</i>	0.0565(± 0.0315)***	~ 0 (± 0.0000)***	0.0055(± 0.0099)
	<i>P. alba</i>	0.0431(± 0.0291)***	~ 0 (± 0.0000)***	0.0037(± 0.0076)
Shared (%)		0.0788(± 0.0338)***	0.3466(± 0.0933)***	0.1675(± 0.0528)
F_{ST}		0.7305(± 0.0410)***	0.1125(± 0.0179)***	0.3806(± 0.1615)
d_{xy}		0.3036(± 0.0025)***	0.0422(± 0.0188)***	0.0252(± 0.0121)
RND		0.7261(± 0.2365)***	0.6523(± 0.1887)***	0.4998(± 0.1501)

Asterisks designate significant differences between the regions displaying exceptionally genetic differentiation and the rest of genomic regions by Mann-Whitney U test ($P < 0.05$; ** $P < 1e-4$; *** $P < 2.2e-16$).

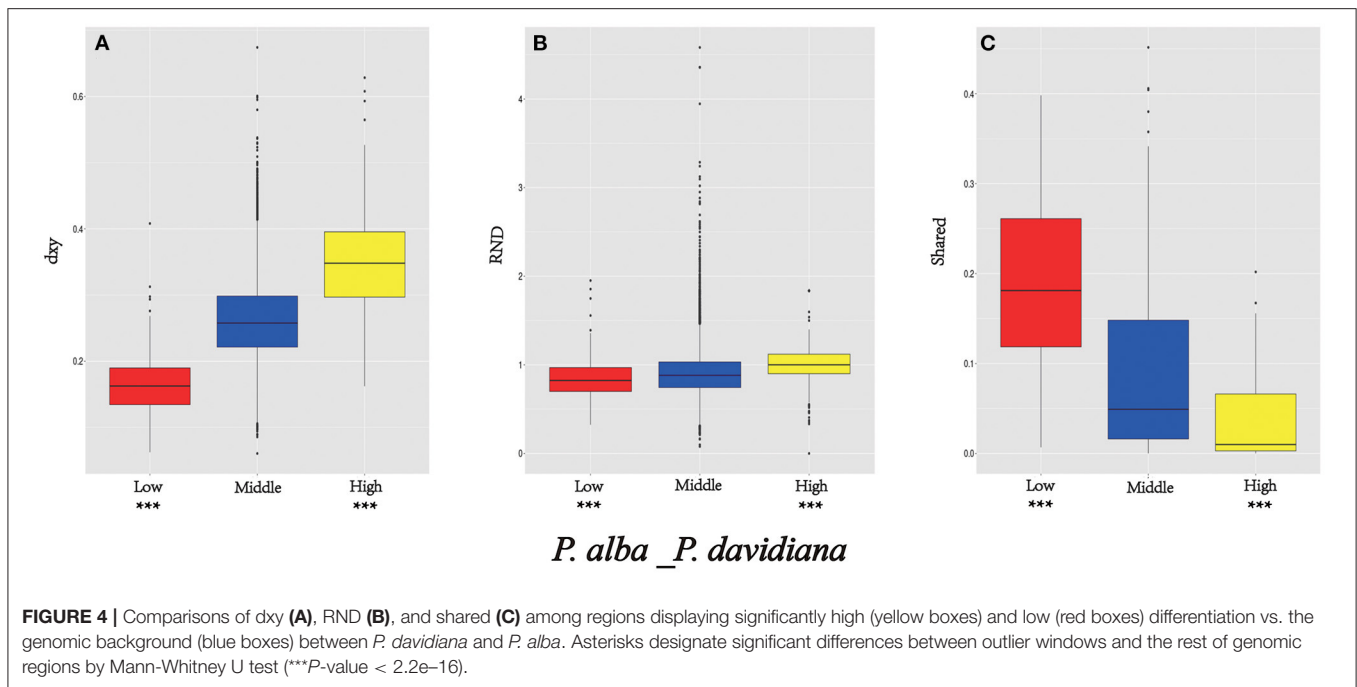


TABLE 5 | Mean (\pm standard deviation) values of population genomic statistics (θ_{π} , θ_w , H_E , Tajima's D, ρ , F_{ST} and d_{xy}) comparisons between *P. davidiana* and *P. alba* population.

Species	θ_{π}	θ_w	H_E	Tajima's D	ρ	F_{ST}	d_{xy}
<i>Populus davidiana</i>	0.0089	0.0156	0.0125	0.13	1.22	0.2988	0.2658
<i>Populus alba</i>	0.0095	0.0175	0.0235	-0.06	2.36		

with high or low differentiation. Particularly, a notable negative correlation was found between the rate of recombination and F_{ST} . Highly differentiated genomic regions exhibited a very low rate of recombination, whereas a low differentiated genomic region had a much higher rate of recombination (Figures 5I,J). This outcome suggests that in the *P. alba* and *P. davidiana* populations, the rate of recombination has an indispensable role in the genomic differentiation process.

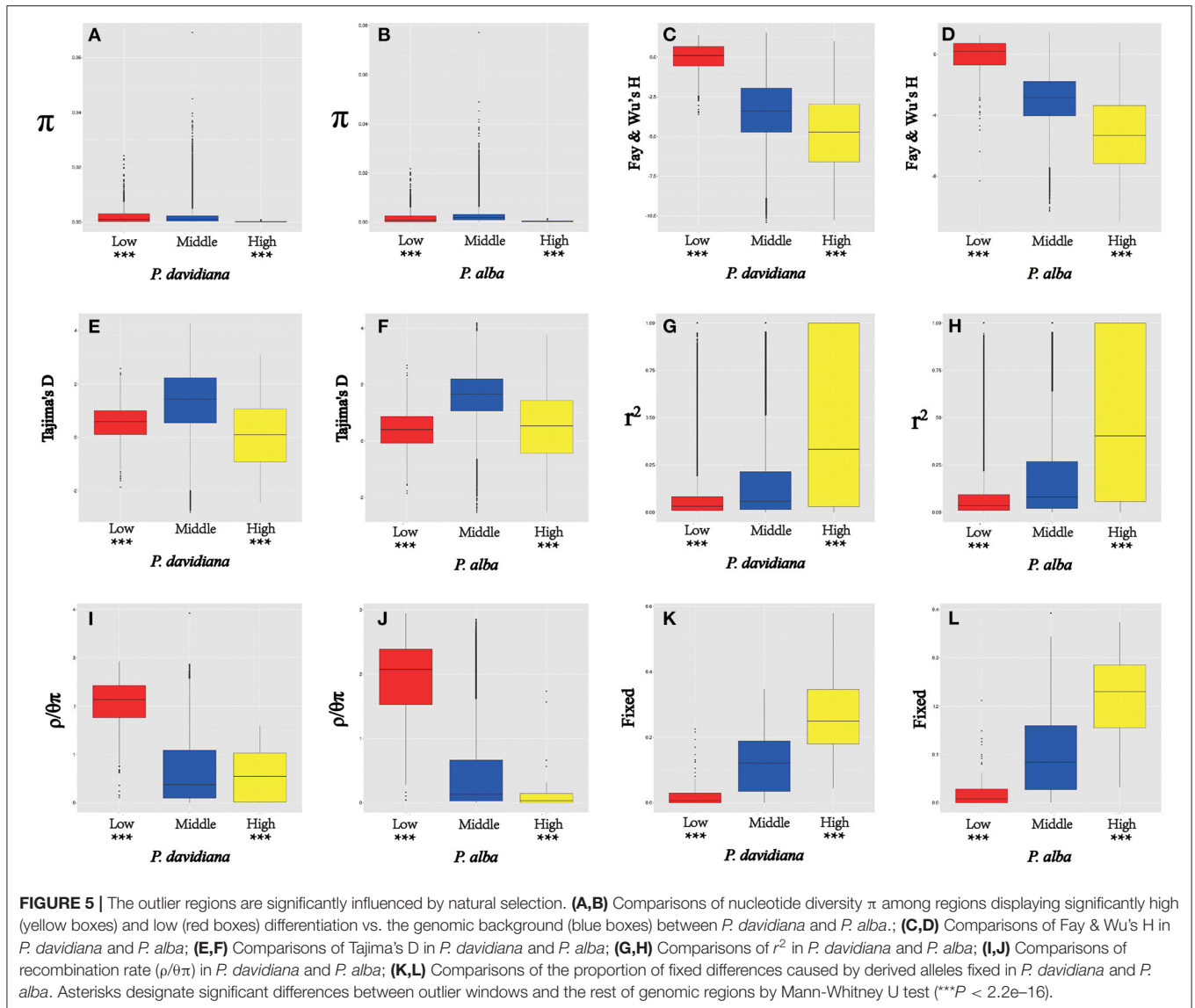
Genes Under Selection

Functional annotations on regions manifesting high differentiation in *P. davidiana* and *P. alba* were performed using annotation of the reference genome of *P. alba*. The present study identified 556 selected genes in total. The differential enrichment of candidate genes was investigated using gene ontology (GO), and 34 significantly overrepresented GO terms were found for genes located in zones with high genetic differentiation. The majority of these GO categories were associated with stimulus-response, catalytic activity, and positive regulation of biological processes (Figure 6). The KEGG pathway enrichment method (using the KOBAS system) also found that functional genes related to cell growth and death, signal transduction, environmental adaptation, and immune

system functions were all significantly enriched (Figure 7). Thus, the results demonstrate that the functional genes associated with environmental adaptation are majorly responsible for the differentiation in *P. davidiana* and *P. alba* populations. Presumably, genes with similar functions were subjected to strong selection pressure, resulting in significant evidence of increased genetic differentiation and selection in *P. davidiana* as compared to its sister species, *P. alba*, however more functional studies are needed to confirm this. In this work, genes related to the adaptation to environmental factors or photosynthesis have been identified. Such genes could be used as contenders in higher-level studies of the mechanisms governing important features, and they could be useful in preserving the populus in the face of a variety of challenges.

DISCUSSION

We use a population genomic approach to resolve the evolutionary histories of two widespread and closely related forest tree species, and to highlight how genome-wide patterns of differentiation have been influenced by a variety of evolutionary processes. We estimated d_{xy} and F_{ST} values across the genome and discovered distinct genetic differentiation between *P. alba* and *P. davidiana*. In order to understand how multiple diversified evolutionary forces driving the demographic history and differentiation of the two distinct populations, a whole-genome resequencing method was applied on the basis of a large count of impartial SNPs dispersed through the entire genome. The findings suggest that Quaternary climate change leading to geographic isolation may have aided in the initial divergence of *P. alba* and *P. davidiana*, and that natural selection plays an essential role in preserving species distinctiveness after gene flow.



Demographic History of the Two Aspen Species

Demographic histories of *P. davidiana* and *P. alba* were inferred by employing a coalescent simulation-based approach in fastsimcoal 2.6.1 (Excoffier et al., 2013). According to the best-fitting demographic model conducted from Fastsimcoal2, the early divergence of two species was estimated to initialize ~802,365 years ago (Figure 2A), the glaciation events and geological uplift observed for the Qinghai-Tibetan Plateau (QTP) in the MLP coincide with the time of differentiation (Zheng et al., 2002). About 0.6–1.2 Ma on the QTP, the Kun-Huang diastrophisms took place (Hsü, 1982). At the turn of the early-late and MLP, a powerful regional uplift took place around the Yellow River drainage, leading to the glaciation of the QTP (Zheng et al., 2002). That precisely was the Naynayxungla Glaciation, which also took place around 0.8 Ma (Zheng et al., 2002). Amidst this cold period, a large part of the QTP was covered with ice sheets which could potentially fragment the

habitats belonging to ancestral populations, thereby limiting gene flow among populations. When the Naynayxungla Glaciation was at its peak, the ice sheet covered a region that was five to seven times greater than what it covers now (Wu et al., 2001). When the Kun-Huang diastrophisms and the Naynayxungla Glaciation arrived, the climate in China did not allow warm adapted plants to grow and the subtropical vegetation moved from north to south of Qinling Mountains (Zheng et al., 2002). Because gene flow was practically impeded by the species' geographical isolation, a large-scale succession of plant communities occurred, and geological uplift and glacier events may have caused the differentiation of the ancestors of *P. davidiana* and *P. alba* in China. The divergence time was consistent with *Cupressus chengiana* on QTP (Li et al., 2020), which diverged at 0.8 Ma and was influenced by the Kun-Huang diastrophisms and the Naynayxungla Glaciation.

Geographic isolation between various populations was an obvious consequence of regional uplift and quaternary climate

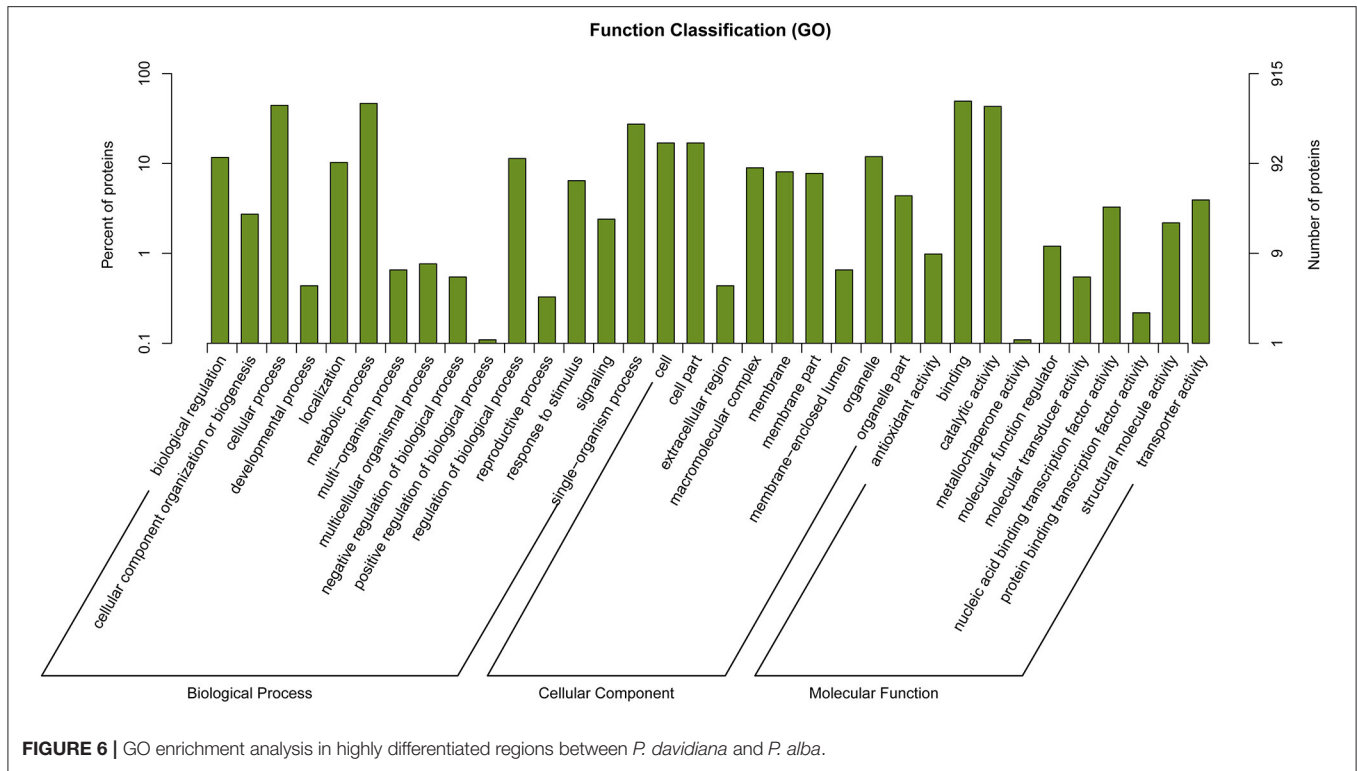


FIGURE 6 | GO enrichment analysis in highly differentiated regions between *P. davidiana* and *P. alba*.

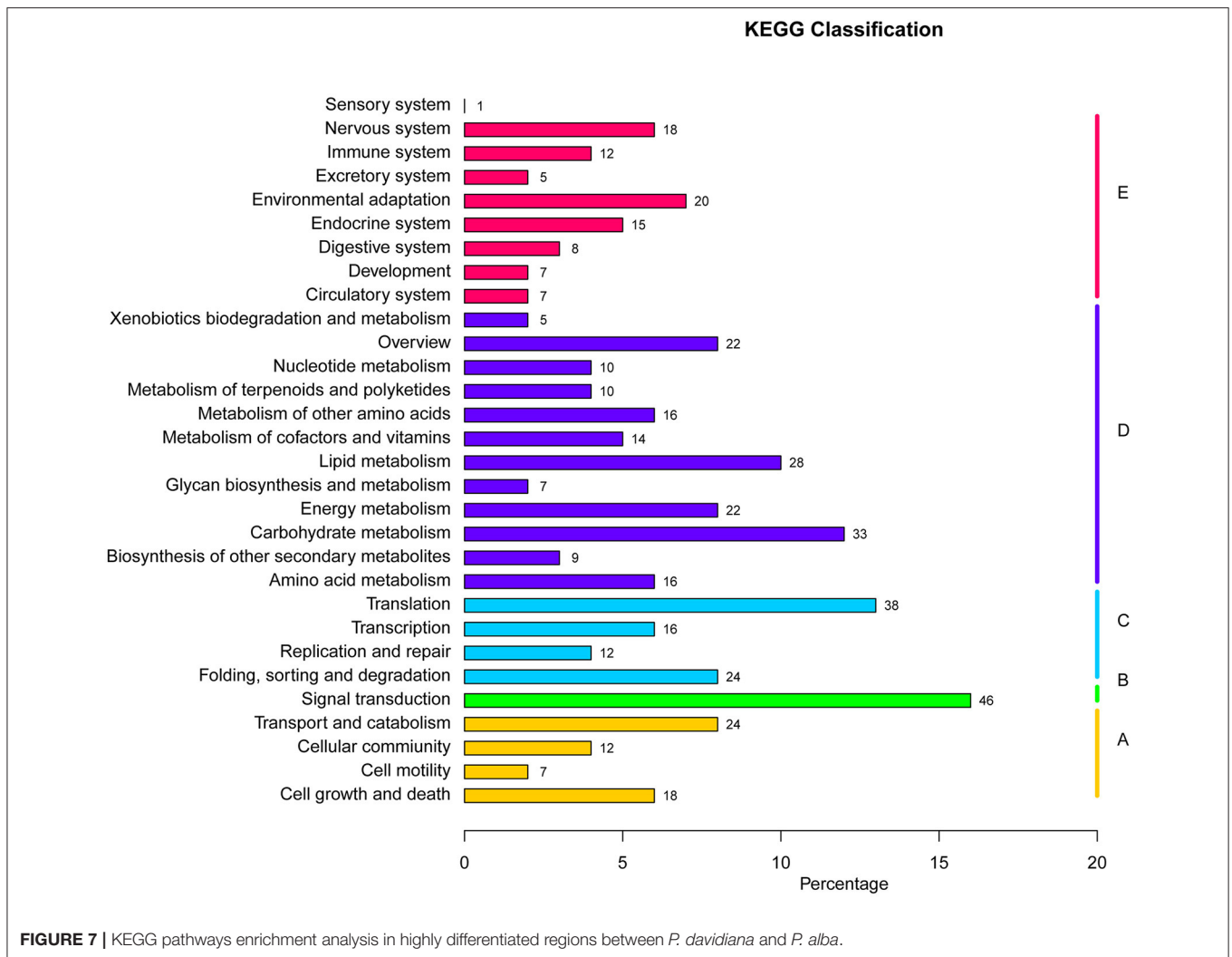
fluctuations (Han et al., 2017). The geographic barriers theory was validated by our research, and the discontinuous distribution pattern of *P. davidiana* and *P. alba* could have been the result of geographic barriers where their ancestral populations started to differentiate due to vicariance. Due to previous geological events such as the Kun-Huang diastrophism and the Naynayxungla Glaciation, as well as significant climate oscillations, geographic barriers confined *P. davidiana* and *P. alba* into separate continental regions, effectively preventing gene flow between the two species. For instance, we discovered obvious genetic differentiation among both populations under consideration, the gene flow between *P. davidiana* and *P. alba* was considerably low (3.97×10^{-8} and 9.16×10^{-7}) and the F_{ST} values corresponding to *P. davidiana* and *P. alba* were around 0.2988 (Table 5) and (Figure 2A), which further proved that the existence of geographic barriers significantly obstructed the gene flow between *P. davidiana* and *P. alba* to rather insignificant levels. Moreover, for *P. davidiana* and *P. alba*, the distribution of ancestral population could have further fragmented due to Naynayxungla Glaciation and Kun-Huang diastrophism (for instance, the existence of glacial refugia), thus driving the divergence of species. Gene flow had been impeded among populations due to geographical isolation (Hancock and Bergelson, 2011). Simultaneously, owing to the differing selection pressures of various populations, the isolated populations slowly gathered variation, leading to the differentiation between *P. alba* and *P. davidiana*. Earlier studies have reported a noteworthy impact on the biodiversity of the QTP as a result of Quaternary climatic oscillation (Liu and Harada, 2014), which led to the current species diverging in interspecific (Xu et al., 2019)

and intraspecific (Guangpeng et al., 2017) fashion. Therefore, Quaternary climatic changes may have cast the primary influence on the divergence that initially took place among *P. davidiana* and *P. alba*. Our findings for these two aspen species support an allopatric model of speciation, given their current geographic isolation, extraordinarily low rates of gene flow, and disjunct distribution (Morjan and Rieseberg, 2015).

PSMC indicated that a significantly long-term bottleneck was encountered by these two aspens post divergence, with population expansion initiating after the conclusion of the last glacial maximum (LGM) that took place around 20,000 years ago (Figure 2B). A variety of other forest trees found in Eurasia corroborate well with this demographic data (Hewitt, 2000, 2004).

Heterogeneous Genomic Differentiation of *P. alba* and *P. davidiana*

Consistent with the expectation for allopatric speciation, loss of gene flow, and stochastic genetic drift resulted in the build-up of interspecific differentiation (Chen et al., 2018). Quite a few regions of genomic differentiation were detected between the two populations. Though neutral processes can effectively explain the majority of these in the two populations (Strasburg et al., 2012), natural selection has had a significant impact on a few outlier regions (Nielsen, 2005). If natural selection has truly been one of the dominant evolutionary forces shaping patterns of genetic differentiation between the two species, regions of low recombination would be expected to show increased F_{ST} values (Noor and Bennett, 2009), because neutral variation is most often removed by a natural selection including background



selection and selective sweeps, particularly in areas having a very low recombination rate (Begun et al., 2007). As a consequence, relative measures of divergence (e.g., F_{ST}) that rely on within-species diversity are expected to be higher in regions with restricted recombination (Noor and Bennett, 2009; Nachman and Payseur, 2012). A notable negative correlation between recombination rate (ρ) and F_{ST} was found in the populations of both *P. alba* (Figure 5I) and *P. davidiana* which is in good agreement with the above observations (Figure 5J) (Keinan and Clark, 2012). Consequently, our results indicate that ρ and linked selection were key factors affecting genomic differentiation among the populations of *P. davidiana* and *P. alba* (Turner et al., 2005; Cruickshank and Hahn, 2014).

Rather than being physically clustered into just a few large, discrete genomic “islands” as expected when species diverge in the presence of gene flow, differentiation islands in our study system are distributed throughout the genome, being narrowly defined and mostly located in regions with substantially suppressed recombination (Turner et al., 2005; Cruickshank and Hahn, 2014). Neutral variation linked to a

deleterious mutation is lost through background selection, and neutral variation linked to a positive mutation is fixed by the genetic hitchhiking of positive selection (Turner et al., 2005; Noor and Bennett, 2009; Cruickshank and Hahn, 2014). Hence, several population genetic parameters were evaluated to understand how the whole genome differentiation in *P. davidiana* and *P. alba* populations was driven by diverse evolutionary forces and how genomic variation took place during the sequence of population differentiation (Figure 6). We found that significant positive selection characteristics were exhibited in the two populations by the region having high differentiation (Nielsen, 2005). Both *P. davidiana* and *P. alba* populations, for example, manifest a very low level of polymorphism (π) (Figures 5A,B). The powerful signals of linkage disequilibrium, higher RND, and dxy values were a more obvious feature of the regions having high differentiation (Figures 4A,B and Figures 5G,H), indicating absolute interspecific divergence. It has also been highlighted by quite a few recent studies that in a selective sweep model, genetic variants associated with useful mutations are exposed to positive selection and attain

a high frequency by hitchhiking along (Kaplan et al., 1989; Mei et al., 2018). Selection due to factors such as local ecological adaptation, particularly in the absence of gene flow, can indirectly inflate F_{ST} by leading to reduced within-population diversity (Cruickshank and Hahn, 2014). Therefore, although a contribution of background selection to the observed patterns cannot be completely discounted, the independent action of positive selection in both *P. davidiana* and *P. alba* is expected to be the dominant driver for the evolution of reduced diversity and increased differentiation in most islands of differentiation. Although genetic diversity underwent a significant decrease under the influence of positive selection, there was an increase in interspecific differentiation. However, because assessing variance in these locations with high differentiation and low genetic diversity is difficult, extra caution should be exercised when interpreting the functional features of the overrepresented genes identified in this example. As a result, more research into these functional genes is required in order to understand how forest tree species respond to the stimulus of climate change throughout adaptive evolution.

While the highly differentiated regions exhibited features of positive selection, the regions with low differentiation in both populations were also identified to manifest long-term balancing selection (Charlesworth et al., 1995). Long-term balancing selection supports maintaining beneficial polymorphisms for quite many generations as compared to positive purifying and selection, resulting in regions of the genome with reduced F_{ST} and increased genetic diversity (Guerrero and Hahn, 2017). In this study, in comparison to the highly differentiated regions, absolute interspecific divergence (d_{xy} and RND values) was found to be lower (Figures 4A,B). Both *P. davidiana* and *P. alba* populations had a notably high genetic diversity (π) (Figures 5A,B). Balancing selection may be identified due to its influence upon neutral sites lying nearby when the same alleles persist for a long time. The population genetics of balancing selection shows that, despite maintaining diversity at the selected sites (usually distinct amino acids), there is a concurrent rise in diversity at closely connected neutral sites (Mei et al., 2018). The frequent appearance of intermediate-frequency alleles is indicated by higher Tajima's D and Fay & Wu's H values (Figure 5F), with levels of LD lower in comparison to the highly differentiated regions, which may have been impacted by recombination (Lee et al., 2011). The proportion of inter-specific shared polymorphisms was found to be higher in low differentiated regions (Figure 5C) in both the *P. davidiana* and *P. alba* populations, the proportion of fixed differences was found to negligible (Figures 5K,L). While selection reduces diversity at selected loci and linked gene regions in general, balancing selection frequently works to maintain diversity at such loci, allowing distinct beneficial variations to be retained over long periods (Begun et al., 2007). Our results, thus, suggest that long-term balancing selection, presumably regulated by "recycling polymorphism" (Holub, 2001) or "trench warfare" (Stahl et al., 1999) of co-evolutionary interaction between natural enemies and hosts, may have caused an increase in genetic diversity over extended periods (Salvaudon et al., 2008).

CONCLUSIONS

Here we provide insights into the speciation and recent evolutionary histories of two closely related forest tree species, *P. davidiana* and *P. alba*. The genetic differentiation between the two species was found to be obvious, with an F_{ST} value of 0.2988 for *P. davidiana* and *P. alba*. Coalescent simulations suggest that the divergence of the two species occurred in the middle Pleistocene, with only minimal gene flow after the divergence. Following divergence, the two species encountered a significant long-term bottleneck, with population increase beginning roughly 20,000 years ago, at the end of the last glacial maximum. Though neutral processes explain the vast majority of regions of genomic differentiation between the two species, a few outlier regions that are noticeably influenced by natural selection have also been investigated. The highly differentiated regions of both species showed substantial positive selection characteristics, and long-term balancing selection was also observed in the regions with limited differentiation in both species. Our results point out that Natural selection and Quaternary climate changes both played a role in the divergence of the *P. davidiana* and *P. alba* populations.

DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the Genome Sequence Archive repository (<http://bigd.big.ac.cn/gsa>), accession number CRA003302 and CRA001674.

AUTHOR CONTRIBUTIONS

ZH performed the experiments and wrote the study. AL designed the research. All authors contributed to the article and approved the submitted version.

FUNDING

Financial support for this research was provided by the Fundamental Research Funds of Sichuan Provincial Science Program Project (22NSFSC3363), China West Normal University (Grant No. 20B007 and 19E044), the Open Fund of Ecological Security and Protection Key Laboratory of Sichuan Province, Mianyang Normal University (Grant No. ESP2004), and the Open Fund of MOE Key Laboratory of Biodiversity and Ecology Engineering, Beijing Normal University (Grant No. K202001).

ACKNOWLEDGMENTS

We thank Li Mingzhi for his helpful suggestions on data analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.911467/full#supplementary-material>

REFERENCES

- Begun, D. J., Holloway, A. K., Stevens, K., Hillier, L. D. W., Poh, Y. P., Hahn, M. W., et al. (2007). Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol.* 5:e310. doi: 10.1371/journal.pbio.0050310
- Campagna, L., Gronau, I., Silveira, L. F., Siepel, A., and Lovette, I. J. (2015). Distinguishing noise from signal in patterns of genomic divergence in a highly polymorphic avian radiation. *Mol. Ecol.* 24, 4238–4251. doi: 10.1111/mec.13314
- Charlesworth, D., Charlesworth, B., and Morgan, M. T. (1995). The pattern of neutral molecular variation under the background selection model. *Genetics* 141, 1619–1632. doi: 10.1093/genetics/141.4.1619
- Chen, C., Wang, H., Liu, Z., Chen, X., Tang, J., Meng, F., et al. (2018). Population genomics provide insights into the evolution and adaptation of the eastern honey bee (*Apis cerana*). *Mol. Biol. Evol.* 35, 2260–2271. doi: 10.1093/molbev/msy130
- Cruickshank, T. E., and Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* 23, 3133–3157. doi: 10.1111/mec.12796
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498. doi: 10.1038/ng.806
- Du, Z., Zhou, X., Ling, Y., Zhang, Z., and Su, Z. (2010). agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* 38, W64–W70. doi: 10.1093/nar/gkq310
- Eckenwalder, J. E. (1977). North American cottonwoods (populus, Salicaceae) of sections Abaso and aigeiros. *J. Arnold. Arbor.* 58, 193–208. doi: 10.5962/bhl.part.29239
- Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., and Foll, M. (2013). Robust demographic inference from Genomic and SNP data. *PLoS Genet.* 9:e1003905. doi: 10.1371/journal.pgen.1003905
- Fay, J. C., and Wu, C. I. (2000). Hitchhiking under positive Darwinian selection. *Genetics* 155, 1405–1413. doi: 10.1093/genetics/155.3.1405
- Fu, Y. X., and Li, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics* 133, 693–709. doi: 10.1093/genetics/133.3.693
- Fumagalli, M., Vieira, F. G., Linderth, T., and Nielsen, R. (2014). ngsTools: methods for population genetics analyses from next-generation sequencing data. *Bioinformatics* 30, 1486–1487. doi: 10.1093/bioinformatics/btu041
- Gao, F., Ming, C., Hu, W., and Li, H. (2016). New software for the Fast Estimation of Population Recombination Rates (FastEPRR) in the genomic era. *G3 GenesGenetics* 6, 1563–1571. doi: 10.1534/g3.116.028233
- Guangpeng, R., Rubén, G., and Mateo, J. N. (2017). *Genetic Consequences of Quaternary Climatic Oscillations in the Himalayas: Primula tibetica as a Case Study Based on Restriction Site-Associated DNA Sequencing*. New Phytologist.
- Guerrero, R. F., and Hahn, M. W. (2017). Speciation as a sieve for ancestral polymorphism. *Mol. Ecol.* 26, 5362–5368. doi: 10.1111/mec.14290
- Han, F., Lamichhane, S., Grant, B. R., Grant, P. R., Andersson, L., and Webster, M. T. (2017). Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. *Genome Res.* 27, 1004–1015. doi: 10.1101/gr.212522.116
- Hancock, A. M., and Bergelson, J. (2011). Adaptation to climate across the Arabidopsis thaliana genome. *Science* 334, 83–86. doi: 10.1126/science.1209244
- Hellmann, I., Prüfer, K., Ji, H., Zody, M. C., Pääbo, S., Ptak, S. E. (2005). Why do human diversity levels vary at a megabase scale? *Genome Res.* 15, 1222–1231. doi: 10.1101/gr.3461105
- Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature* 405, 907–913. doi: 10.1038/35016000
- Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 359, 183–195. doi: 10.1098/rstb.2003.1388
- Holub, E. B. (2001). The arms race is ancient history in Arabidopsis, the wildflower. *Nat. Rev. Genet.* 2:516–27. doi: 10.1038/35080508
- Honglei, Z., Liqiang, F., and Richard, I. Y. (2017). Species delimitation and lineage separation history of a species complex of aspens in China. *Front Plant Sci.* (2017) 8:375. doi: 10.3389/fpls.2017.00375
- Hsü, J. (1982). *The Uplift of the Qinghai-Xizang (Tibet) Plateau in Relation to the Vegetational Changes in the Past*. Acta Phytotaxonomica Sinica.
- Ingvansson, P. R. K. (2008). Multilocus patterns of nucleotide polymorphism and the demographic history of *Populus tremula*. *Genetics* 180, 329–340. doi: 10.1534/genetics.108.090431
- Jianchao, M., Dongshi, W., Bingbing, D., Xiaotao, B., Qiuxian, B., Ningning, C., et al. (2018). Genome sequence and genetic transformation of a widely distributed and cultivated poplar. *Plant Biotechnol. J.* 17, 451–460. doi: 10.1111/pbi.12989
- Kaplan, N. L., Hudson, R. R., and Langley, C. H. (1989). The “hitchhiking effect” revisited. *Genetics* 123:887. doi: 10.1093/genetics/123.4.887
- Keinan, A., and Clark, A. G. (2012). Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science* 336, 740–743. doi: 10.1126/science.1217283
- Koch, M. A., Haubold, B., and Mitchell-Olds, T. (2000). Comparative evolutionary analysis of chalcone synthase and alcoholdehydrogenase loci in Arabidopsis, Arabis, and related genera (Brassicaceae). *Mol. Biol. Evol.* 17:1483. doi: 10.1093/oxfordjournals.molbev.a026248
- Korneliusson, T. S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. *BMC Bioinform.* 15:356. doi: 10.1186/s12859-014-0356-4
- Lee, K. M., Yong, Y. K., and Hyun, J. O. (2011). Genetic variation in populations of *Populus davidiana* Dode based on microsatellite marker analysis. *Genes Genom.* 33, 163–171. doi: 10.1007/s13258-010-0148-9
- Lei, Z., Zhenxiang, X., Mingcheng, W., Xinyi, G., and Tao, M. (2018). Plastome phylogeny and lineage diversification of Salicaceae with focus on poplars and willows. *Ecol. Evol.* 8, 7817–7823. doi: 10.1002/ece3.4261
- Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature* 475:493. doi: 10.1038/nature10231
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Li, J., Milne, R. I., Ru, D., Miao, J., and Mao, K. (2020). Allopatric divergence and hybridization within *Cupressus changiana* (cupressaceae), a threatened conifer in the northern hengduan mountains of western china. *Mol. Ecol.* 29, 1250–1266. doi: 10.1111/mec.15407
- Liu, H. Z., and Harada, K. (2014). Geographic distribution and origin of the chloroplast T/C-type in *Quercus mongolica* var. *crispula* in northeastern Japan. *Plant Species Biol.* 29, 207–211. doi: 10.1111/1442-1984.12008
- Liu, X., Wang, Z., Shao, W., Ye, Z., and Zhang, J. (2016). Phylogenetic and taxonomic status analyses of the Abaso section from multiple nuclear genes and plastid fragments reveal new insights into the north america origin of *Populus* (Salicaceae). *Front. Plant Sci.* 7:2022. doi: 10.3389/fpls.2016.02022
- Liu, Y. J., Wang, X. R., and Zeng, Q. Y. (2019). De novo assembly of white poplar genome and genetic diversity of white poplar population in Irtysh River basin in China. *Sci. China* 62:609–18. doi: 10.1007/s11427-018-9455-2
- Lohse, M., Bolger, A. M., Nagel, A., Fernie, A. R., Lunn, J. E., Stitt, M., et al. (2012). RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res.* 40(Web Server issue), 622–627. doi: 10.1093/nar/gks540
- Löve, A., and Löve, D. (1975). Nomenclatural notes on arctic plants. *Bot. Notiser* 128, 497–523. doi: 10.1007/BF02860833
- Mao, X., Tao, C. J. G. O., and Wei, L. (2005). Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* 21, 3787–3793. doi: 10.1093/bioinformatics/bti430
- Mark, R., Kohta, Y., Shuji, S., Atsushi, T., Asao, F., Jun, K., et al. (2018). The genomic landscape at a late stage of stickleback speciation: high genomic divergence interspersed by small localized regions of introgression. *PLoS Genet.* 14:e1007358. doi: 10.1371/journal.pgen.1007358
- Mei, C., Wang, H., Liao, Q., Wang, L., Cheng, G., Wang, H., et al. (2018). Genetic architecture and selection of Chinese cattle revealed by whole genome resequencing. *Mol. Biol. Evolution* 35, 688–699. doi: 10.1093/molbev/msx322
- Morjan, C. L., and Rieseberg, L. H. (2015). How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Mol. Ecol.* 13, 1341–1356. doi: 10.1111/j.1365-294X.2004.02164.x
- Nachman, M. W., and Payseur, B. A. (2012). Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philosophical Transac. R. Soc. Biol. Sci.* 367:409. doi: 10.1098/rstb.2011.0249

- Neale, D. B., and Antoine, K. (2011). Forest tree genomics: growing resources and applications. *Nat. Rev. Genet.* 12, 111–122. doi: 10.1038/nrg2931
- Nielsen, R. (2005). Molecular signatures of natural selection. *Annu. Rev. Genet.* 39, 197–218. doi: 10.1146/annurev.genet.39.073003.112420
- Noor, M. A. F., and Bennett, S. M. (2009). Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*. (2009) 103:439–44. doi: 10.1038/hdy.2009.151
- Nordborg, M., Hu, T. T., Ishino, Y., Jhaveri, J., Toomajian, C., Zheng, H., et al. (2005). The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol.* 3, 1289–1299. doi: 10.1371/journal.pbio.0030196
- Nosil, P., and Feder, J. L. (2012). Genomic divergence during speciation: causes and consequences. *Philosophical Transac. R. Soc. London* 367, 332–342. doi: 10.1098/rstb.2011.0263
- Nosil, P., Funk, D. J., and Ortizbarrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Mol. Ecol.* 18, 375–402. doi: 10.1111/j.1365-294X.2008.03946.x
- Pahlich, E., and Gerlitz, C. (1980). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemistry* 19, 11–13. doi: 10.1016/0031-9422(80)85004-7
- Pakull, B., Groppe, K., Meyer, M., Markussen, T., and Fladung, M. (2009). Genetic linkage mapping in aspen (*Populus tremula* L. and *Populus tremuloides* Michx). *Tree Genet. Genomes* 5, 505–515. doi: 10.1007/s11295-009-0204-2
- Petr, D., Adam, A., Goncalo, A., Albers, C. A., Eric, B., Depristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158.
- Salvaudon, L., Giraud, T., and Shykoff, J. A. (2008). Genetic diversity in natural populations: a fundamental component of plant–microbe interactions. *Curr Opin Plant Biol.* 11, 135–143. doi: 10.1016/j.pbi.2008.02.002
- Sanna, H., Lars, H. S., Ignatov, M. S., Nicolas, D., and Alain, V. (2008). Origin and evolution of the northern hemisphere disjunction in the moss genus *Homalothecium* (Brachytheciaceae). *Am. J. Bot.* 95, 720–730. doi: 10.3732/ajb.2007407
- Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., et al. (2014). Genomics and the origin of species. *Nat. Rev. Genet.* 15, 176–192. doi: 10.1038/nrg3644
- Skotte, L., Korneliusen, T. S., and Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. *Genetics* 195, 693–702. doi: 10.1534/genetics.113.154138
- Stahl, E. A., Dwyer, G., Mauricio, R., Kreitman, M., and Bergelson, J. (1999). Dynamics of disease resistance polymorphism at the RPM1 locus of *Arabidopsis*. *Nature* 400, 667–671. doi: 10.1038/23260
- Strasburg, J. L., Sherman, N. A., Wright, K. M., Moyle, L. C., Willis, J. H., and Rieseberg, L. H. (2012). What can patterns of differentiation across plant genomes tell us about adaptation and speciation? *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 367, 364–373. doi: 10.1098/rstb.2011.0199
- Tarailo-Graovac, M., and Chen, N. (2009). Using repeatmasker to identify repetitive elements in genomic sequences. *Curr. Protocols Bioinform.* 25, 4.10.1–4.10.14. doi: 10.1002/0471250953.bi0410s25
- Turner, T. L., Hahn, M. W., and Nuzhdin, S. V. (2005). Genomic Islands of speciation in *Anopheles gambiae*. *PLoS Biol.* 3:e285. doi: 10.1371/journal.pbio.0030285
- Tuskan, G. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr.andGray). *Science* 313, 1596–1604. doi: 10.1126/science.1128691
- Via, S. (2009). Natural selection in action during speciation. *Proc. Natl. Acad. Sci. U.S.A.* 106:9939. doi: 10.1073/pnas.0901397106
- Wang, J., Street, N. R., Park, E. J., Liu, J., and Ingvarsson, P. K. (2020). Evidence for widespread selection in shaping the genomic landscape during speciation of *Populus*. *Mol. Ecol.* 29, 1120–1136. doi: 10.1111/mec.15388
- Wang, J., Street, N. R., Scofield, D. G., and Ingvarsson, P. K. (2016). Variation in linked selection and recombination drive genomic divergence during allopatric speciation of European and American Aspens. *Mol. Biol. Evol.* 33, 1754–1767. doi: 10.1093/molbev/msw051
- Wang, Y., Song, F., Zhu, J., Zhang, S., Yang, Y., Chen, T., et al. (2017). GSA: genome sequence archive. *Genom. Proteom. Bioinform.* 15, 14–18. doi: 10.1016/j.gpb.2017.01.001
- Wu, Y., Cui, Z., Liu, G., Ge, D., Yin, J., Xu, Q., et al. (2001). Quaternary geomorphological evolution of the Kunlun Pass area and uplift of the Qinghai-Xizang (Tibet) Plateau. *Geomorphology* 36, 203–216. doi: 10.1016/S0169-555X(00)00057-X
- Xu, J., Song, X., Ruhsam, M., Liu, T., and Mao, K. (2019). Distinctiveness, speciation and demographic history of the rare endemic conifer *Juniperus erectopatens* in the eastern Qinghai-Tibet Plateau. *Conservation Genetics*. doi: 10.1007/s10592-019-01211-2
- Zheng, B., Xu, Q., and Shen, Y. (2002). The relationship between climate change and Quaternary glacial cycles on the Qinghai–Tibetan Plateau: review and speculation. *Quaternary Int.* 97, 93–101. doi: 10.1016/S1040-6182(02)00054-X
- Zong, D., Gan, P., Zhou, A., Zhang, Y., Zou, X., Duan, A., et al. (2019). Plastome Sequences help to resolve deep-level relationships of populus in the family Salicaceae. *Front. Plant Sci.* 10:5. doi: 10.3389/fpls.2019.00005

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Hou and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.