

Utah State University

DigitalCommons@USU

---

Aspen Bibliography

Aspen Research

---

7-28-2020

## Genomic Insights into Speciation History and Local Adaptation of an Alpine Aspen in the Qinghai–Tibet Plateau and Adjacent Highlands

Jia-Lang Li  
*Sichuan University*

Lin-Ling Zhong  
*Sichuan University*

Jing Wang  
*Sichuan University*

Tao Ma  
*Sichuan University*

Kang-Shan Mao  
*Sichuan University*

Follow this and additional works at: [https://digitalcommons.usu.edu/aspen\\_bib](https://digitalcommons.usu.edu/aspen_bib)



Sichuan University

Part of the [Agriculture Commons](#), [Ecology and Evolutionary Biology Commons](#), [Forest Sciences Commons](#), [Genetics and Genomics Commons](#), and the [Plant Sciences Commons](#)

---

### Recommended Citation

Li, J.-L., Zhong, L.-L., Wang, J., Ma, T., Mao, K.-S. and Zhang, L. (2020), Genomic insights into speciation history and local adaptation of an alpine aspen in the Qinghai–Tibet Plateau and adjacent highlands. *J. Syst. Evol.* <https://doi.org/10.1111/jse.12665>

This Article is brought to you for free and open access by the Aspen Research at DigitalCommons@USU. It has been accepted for inclusion in Aspen Bibliography by an authorized administrator of DigitalCommons@USU. For more information, please contact [digitalcommons@usu.edu](mailto:digitalcommons@usu.edu).





## Research Article

## Genomic insights into speciation history and local adaptation of an alpine aspen in the Qinghai–Tibet Plateau and adjacent highlands

Jia-Liang Li<sup>†</sup> , Lin-Ling Zhong<sup>†</sup>, Jing Wang, Tao Ma, Kang-Shan Mao, and Lei Zhang<sup>\*</sup>

Key Laboratory of Bio-Resource and Eco-Environment of the Ministry of Education, College of Life Sciences, Sichuan University, Chengdu 610065, China

<sup>†</sup>These authors contributed equally to this work.<sup>\*</sup>Author for correspondence. E-mail: zhangsanshi-0319@163.com

Received 14 February 2020; Accepted 14 July 2020; Article first published online 28 July 2020

**Abstract** Natural selection serves as an important agent to drive and maintain interspecific divergence. *Populus rotundifolia* Griff. is an alpine aspen species that mainly occurs in the Qinghai–Tibet Plateau (QTP) and adjacent highlands, whereas its sister species, *P. davidiana* Dode, is distributed across southwest and central to northeast China in much lower altitude regions. In this study, we collected genome resequencing data of 53 *P. rotundifolia* and 42 *P. davidiana* individuals across their natural distribution regions. Our population genomic data suggest that the two species are well delimited in the allopatric regions, but with hybrid zones in their adjacent region in the eastern QTP. Coalescent simulations suggest that *P. rotundifolia* diverged from *P. davidiana* in the middle Pleistocene with following continuous gene flow since divergence. In addition, we found numerous highly diverged genes with outlier signatures that are likely associated with high-altitude adaptation of these alpine aspens. Our finding indicate that Quaternary climatic changes and natural selection have greatly contributed to the origin and distinction maintenance of *P. rotundifolia* in the QTP.

**Key words:** aspen, gene flow, high-altitude adaptation, Qinghai–Tibet Plateau, speciation.

## 1 Introduction

Understanding how species originate and maintain is a central question in evolutionary biology (Savolainen et al., 2013). The initial divergence between populations was previously assumed to be driven by geographical isolation in the allopatric speciation (Mayr, 1963). 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. Recently, speciation with gene flow has been frequently reported based on population genomic studies from a wide range of species (Nosil, 2008; Abbott et al., 2013; Taylor & Larson, 2019; Chaturvedi et al., 2020). Despite the homogenization effect of the high level of gene flow, natural selection could play a key role in maintaining the status of closely related species (Hoskin et al., 2005; Karrenberg et al., 2019). Individuals grown in different geographic regions would be under different selective pressures and are therefore adapted to different local environmental conditions (Savolainen et al., 2013; Zhang et al., 2019a; Jia et al., 2020a, 2020b). In this case, introgression is

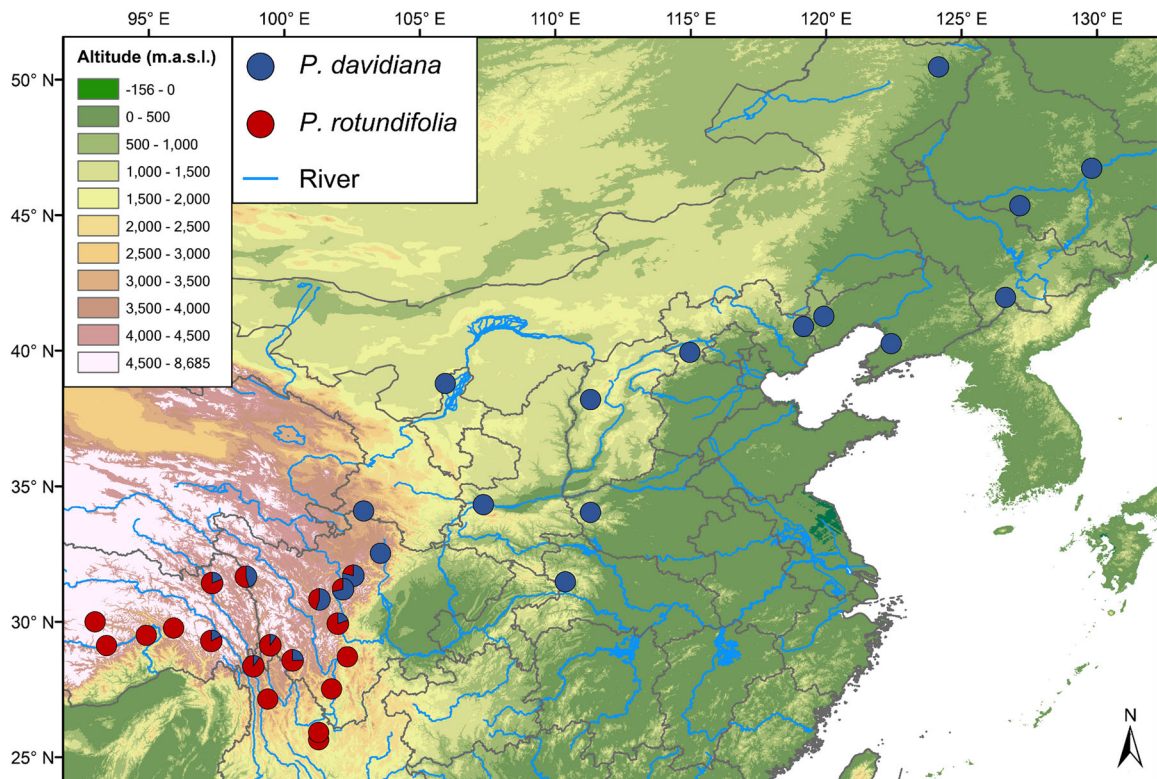
supposed to be reduced in the genomic regions containing genes involved in local adaptation (Via, 2012). Therefore, in addition to neutral evolutionary forces, natural selection also plays an important role in driving allopatric speciation and maintains their distinctiveness in contact regions (Hoskin et al., 2005). For example, natural selection was inferred to drive genomic divergence between two teosinte subspecies despite continuous migrations (Aguirre-Liguori et al., 2019). Using genomic data, Riesch et al. (2017) analyzed stick insect speciation and showed transitions from a few adaptive loci in early phases of speciation to genome-wide differentiation in species pairs diverged at different timescales and connected with various levels of gene flow. Moreover, many speciation studies have found that divergence between two closely related species was accompanied with gene flow and also observed the existence of hybrids in the parapatric distributions with indistinctly differentiated niches (Martin et al., 2013; Clarkson et al., 2014; Wang et al., 2016, 2019; Han et al., 2017; Ma et al., 2018, 2019; Zhang et al., 2019b).

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

The high alpine species diversity and exceptionally diverse endemics of the arid Qinghai–Tibet Plateau *sensu lato* (QTP) (Liang et al., 2018) provide a unique chance to investigate how species originated and adapted to high-altitude environment (Liu et al., 2014). Here we focus on two closely related aspen species: an alpine aspen distributed in the QTP and adjacent regions (*Populus rotundifolia* Griff.) and its sister species *P. davidiana* Dode distributed in the low-altitude region. Both species belong to sect. *Populus* of the genus *Populus* L. (Salicaceae), which contains approximately seven species with widespread distribution throughout the Northern Hemisphere (Zhang et al., 2018; Wang et al., 2020a, 2020b). Aspens are commonly recognized as “keystone species” in boreal and temperate forests and usually play a pioneer role following forest disturbance (Edenius et al., 2011; Berrill et al., 2017; Zheng et al., 2017; Fan et al., 2018; Rogers et al., 2020). *Populus rotundifolia* occurs mainly in the Hengduan Mountains and the Himalayas with the high-altitude preference (Zheng et al., 2016), where harsh abiotic stresses usually exist, such as low level of oxygen partial pressure (Streb et al., 2005), low precipitation or temperatures (Bliss, 1962; Li et al., 2007; Manel et al., 2012; Ma et al., 2019), and strong UV radiation (Norsang et al., 2011). In contrast, *P. davidiana* has a wider distribution, ranging from southwest and central to northeast China with minor extensions to the adjacent countries (Zheng et al., 2017; Hou et al., 2018; Rogers et al., 2020). Early studies using complete plastid genomic data (Zhang et al., 2018) and resequenced genomes

(Wang et al., 2020b) constructed phylogenetic trees of the genus *Populus* and showed that *P. rotundifolia* and *P. davidiana* are distinct species with the closest relationship of all *Populus* species. Because of subtle morphological differentiation between the two species, some researchers treated them as one species (*P. davidiana*; Hou et al., 2018, 2020; Rogers et al., 2020). However, their genetic and ecological differentiations are significant, suggesting that they could have been under strong divergent ecological selection (Zheng et al., 2017). Previous studies using the molecular dating approach showed that these two species diverged from each other in the recent past and gene flow might have continued to the current day; hybrid populations with ongoing gene flows were detected in intermediate areas between their distributions (Zheng et al., 2017; Shang et al., 2020). In fact, from a recent phylogenomic study of the genus *Populus*, there is frequent interspecific gene flow even between distantly related species across the whole genus (Wang et al., 2020b).

To better understand the speciation divergence process of these two recently diverged aspen species and adaptive patterns of *P. rotundifolia* to the high-altitude environment, we collected genomic data of 95 samples from the whole distribution of the two species. We aimed to address the following questions. When did *P. rotundifolia* diverge from *P. davidiana*? Is there gene flow accompanied the divergence or not? Were genes with signatures of positive selection in *P. rotundifolia* correlated with adaptation to the alpine environment? Both lines of evidence are



**Fig. 1.** Sample locations of sampled *Populus davidiana* and *P. rotundifolia* populations. Pie charts at sampling locations indicate the distribution of genetic groups identified by ADMIXTURE analysis when  $K = 2$ . The map was retrieved from Google Earth (<https://www.google.com/earth/>). Elevation data for the map were derived from SRTM elevation data through the WorldClim data website (<https://www.worldclim.org/>).

critical to illuminate the speciation history and high-altitude adaptation of *P. rotundifolia*.

## 2 Material and Methods

### 2.1 Sample collection, sequencing, and read mapping

Silica-gel dried leaves of 53 *Populus rotundifolia* and 42 *P. davidiana* individuals from 38 populations covering their geographical distributions in China were collected for DNA extraction (Fig. 1; Table S1). For each population, sampled individuals were at least 100 m apart. Total genomic DNA was sequenced on the HiSeq X Ten platforms (Illumina, Inc. San Diego, California, U.S.). Raw reads were filtered using TRIMMOMATIC version 0.36 (Bolger et al., 2014) with the following parameters: “SLIDINGWINDOW: 4:15, LEADING: 3, TRAILING: 3, MINLENGTH: 36.” The quality-filtered reads were mapped to the *Populus trichocarpa* Torr. & A. Gray ex Hook. reference genome version 3.0 (Tuskan et al., 2006) using BWA-MEM version 0.7.10 (Li & Durbin, 2009) with default parameters. The alignments were sorted by SAMTOOLS version 0.1.19 (Li et al., 2009). Local realignment was carried out to correct the misalignment around indels by RealignerTargetCreator and IndelRealigner in GATK version 4.0.5.1 (DePristo et al., 2011).

### 2.2 Single nucleotide polymorphism calling and filtering

Multi-sample single nucleotide polymorphism (SNP) calling was implemented in GATK version 4.0.5.1 using HaplotypeCaller and GenotypeGVCFs tools (DePristo et al., 2011). Single nucleotide polymorphisms were further filtered using the following criteria: (i) removal of sites with a quality score less than 30, base quality less than 10, or more than two alleles; (ii) removal of sites that overlapped with known repeat elements as identified by RepeatMasker (Tarailo-Graovac & Chen, 2009) or occurred in 5-bp windows around any indel; and (iii) sites with extremely low ( $<3\times$ ) or extremely high ( $>60\times$ ) coverage per sample, or genotype quality score  $<10$  were assigned as missing data. We retained SNPs with missing genotype rate  $<20\%$  and minimum allele frequency (MAF)  $>0.05$  across all sampled individuals. Finally, we used the software SNPEFF version 4.3 (Cingolani et al., 2012) to annotate the filtered SNPs.

Individuals with close kinship might result in an inaccurate estimate of population structure. Thus, we used the KING program (Manichaikul et al., 2010) to identify kinships for all samples and excluded individuals exhibiting closer than third-degree relationships (Manichaikul et al., 2010).

### 2.3 Population structure and phylogenetic analysis

We used VCFTOOLS version 0.1.15 (Danecek et al., 2011) to convert the input data and used PLINK version 1.90 (Purcell et al., 2007) to remove linked sites with the parameter set as “--indep-pairwise 50 5 0.4.” Principal component analysis was applied for all SNPs using the smartpca program in the software EIGENSOFT version 6.1.3 (Price et al., 2006). We also used the software ADMIXTURE version 1.23 (Alexander et al., 2009) to estimate population structure of these two species with a  $K$ -value setting ranging from 1 to 5. For further insight into the population structure among these two species, we undertook an identity-by-descent blocks analysis

using the algorithm from BEAGLE version 4.1 (Browning & Browning, 2007, 2013) with the following parameters: “ibdtrim = 100 window = 100 000 overlap = 10 000 ibdlod = 10”. Finally, concatenated SNPs were used to construct a neighbor-joining tree in MEGA version 6.0 (Tamura et al., 2013).

### 2.4 Genetic diversity

Because individuals with different genetic background might bias the estimates of genetic diversity for the two species, here we removed admixture individuals that were inferred from the population structure analysis. We estimated the heterozygosity rate and runs of homozygosity (ROH) for each individual using the program PLINK version 1.90 (Purcell et al., 2007). The nucleotide diversity ( $\pi$ ), levels of genetic differentiation (Weir and Cockerham mean  $F_{ST}$ ), and Tajima's  $D$  for each species were calculated in the software VCFtools. The absolute sequence divergence ( $D_{XY}$ ) between populations at each site was quantified using a perl script from Ru et al. (2018). All of the parameters except  $\pi$  were calculated using a sliding window approach (2-kb non-overlapping windows). The  $\pi$  was estimated by sum over each site divided by the number of all callable sites. To assess the effects of missing bases and rare SNPs on  $\pi$ , Tajima's  $D$ , and  $F_{ST}$ , we used four datasets to estimate these three parameters: (i) dataset 1 with a genotype missing rate  $>20\%$  and MAF  $>0.05$ ; (ii) dataset 2 with a genotype missing rate  $>20\%$  and MAF  $>0$ ; (iii) dataset 3 with a genotype missing rate  $>0\%$  and MAF  $>0.05$ ; and (iv) dataset 4 with a genotype missing rate  $>0\%$  and MAF  $>0$ .

Moreover, we measured and compared patterns of linkage disequilibrium (LD) between the two species using the program PLINK. Population-scale recombination rates ( $\rho$ ) along each chromosome were estimated using the software FastEPRR (Gao et al., 2016) with default parameters.

### 2.5 Demographic history and gene flow

We first estimated demographic changes in population size over time through pairwise sequentially Markovian coalescent (PSMC; Li & Durbin, 2011) by examining heterozygosity across the diploid genome of a selected “pure” individual. The parameter settings followed the default parameters. Based on previous studies for *Populus* species (Tuskan et al., 2006; Wang et al., 2016; Ma et al., 2018), the mutation rate was set as  $3.75 \times 10^{-8}$  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.

We further used fastsimcoal2 (Excoffier & Foll, 2011) to estimate the timing of divergence and gene flow patterns of *P. rotundifolia* and *P. davidiana*. At first, the 2-D folded site frequency spectrum was extracted from the resequenced genomes of “pure” individuals of the two species by a custom perl script (Ru et al., 2018). To examine the effects of interspecific gene flow on the estimation of demographic history, we considered five scenarios of migration: (i) strict isolation (model1; Fig. S2); (ii) isolation with constant migration (model2; Fig. S2); (iii) ancient migration (model3; Fig. S2); (iv) secondary contact (model4; Fig. S2); and (v) two-rate model of isolation with migration (model5; Fig. S2), where migration rates are assumed to be changed at some time during divergence. A total of five demographic models

were compared, and the Akaike information criterion (AIC) was used to rank all models and choose the model that best fits the data. For each model, 50 independent runs, with 100 000 coalescent simulations per likelihood estimation and 40 cycles of the likelihood maximization algorithm, were carried out to obtain a reliable global maximum likelihood estimation. To construct the confidence interval (CI), we used a parametric bootstrapping approach with 200 independent bootstrap replicates undertaken in *fastsimcoal2*. We used a mutation rate of  $3.75 \times 10^{-8}$  per base per generation and generation time of 15 years as for PSMC.

## 2.6 Signatures of positive selection

To identify candidate genes involved in high-altitude adaptation of *P. rotundifolia*, we combined three approaches to select positive selection genes (PSGs), including inter-specific differentiation index ( $F_{ST}$ ), nucleotide diversity ratio ( $\pi_{Pda}/\pi_{Pro}$ ) of the lowland species (Pda: *P. davidiana*) to highland species (Pro: *P. rotundifolia*), and Tajima's  $D$  within *P. rotundifolia*. We calculated all diversity indices ( $F_{ST}$ ,  $\pi$ , and Tajima's  $D$ ) with 2-kb sliding window in VCFtools ( $-fst$  window-size 2000  $-weir-fst-pop$ ,  $-TajimaD$  2000,  $-window-pi$  2000). If sites were under strong positive or purifying selection in the highland population, a relatively high genetic divergence and a decrease in genetic diversity will be expected compared to the lowland population (Qu et al., 2019). The windows exhibiting extremely high values of  $F_{ST}$  and  $\pi_{Pda}/\pi_{Pro}$  (using the top 5% quantile of the simulated distribution), and the negative Tajima's  $D$  were selected as PSGs. We then used the analysis tool Singular Enrichment Analysis in AGRIGO version 2.0 (Tian et al., 2017) to undertake the Gene Ontology (GO) enrichment analysis for PSGs based on the JGI version 3.1 gene annotation for *P. trichocarpa* in Phytozome (Goodstein et al., 2012). We used the  $\chi^2$ -test to calculate the statistical significance of enrichment, and  $P$ -values (adjusted by Benjamini–Yekutieli

false discovery rate) below 0.05 were recognized as significant.

## 3 Results

### 3.1 Genome resequencing and genetic variation

High-quality whole-genome resequencing data were generated for 53 *Populus rotundifolia* and 42 *P. davidiana* individuals spanning their geographical distributions in China (Table S1). We obtained an average sequencing depth of 16.66× with mean mapping ratio of 88% per individual (Table S1). After discarding individuals with closer than third-degree relationships, a total of 61 unrelated samples, comprising 19 genetic “pure” individuals of *P. rotundifolia*, 24 individuals of *P. davidiana*, and 18 individuals of putative hybrids, remained for subsequent population genetic analyses. After filtering, total of 2 959 212 high-quality SNPs was identified for the all 61 individuals (missing rate <20%, MAF >5%; Table 1). The numbers of SNPs for “pure” *P. davidiana* and *P. rotundifolia* were 2 559 484 and 2 295 512, respectively. *Populus davidiana* harbored more private SNPs (478 959) than *P. rotundifolia* (298 021), and the two species shared 2 180 759 SNPs in total (Table 1).

### 3.2 Population structure and genomic diversity

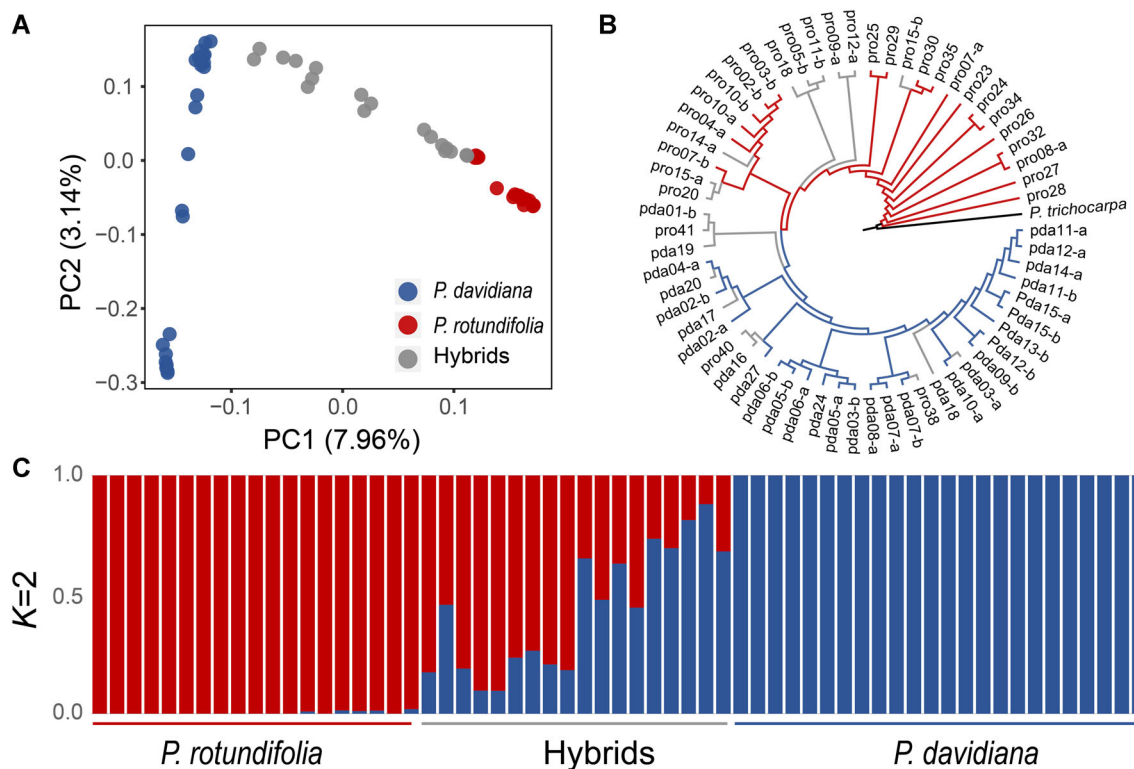
Most individuals were divided into two distinct genetic clusters when  $K$  was set as 2, which was the optimal  $K$ -value with the lowest cross-validation error, from results of ADMIXTURE analysis (Table S2). Some individuals collected in the adjacent regions of *P. rotundifolia* and *P. davidiana* in the eastern QTP showed a mixed genetic makeup of the two clusters (Figs. 1, 2C). The first principal component (PC1, variance explained = 7.96%) identified two genetic clusters for “pure” individuals of *P. rotundifolia* and *P. davidiana*. In the principal component analysis plot (Fig. 2A), some

**Table 1** Summary of genomic polymorphisms and variants in *Populus rotundifolia* (Pro) and *P. davidiana* (Pda) based four datasets

	Pda		Pro		Pda versus Pro	
	MAF >5%	MAF >0%	MAF >5%	MAF >0%	MAF >5%	MAF >0%
Missing data = 0%						
$\pi \pm SD$ ( $\times 10^{-3}$ )	5.2 $\pm$ 3.1	6.5 $\pm$ 9.4	4.6 $\pm$ 3.4	5.0 $\pm$ 9.2	-	-
Tajima's $D \pm SD$	1.08 $\pm$ 0.93	-0.32 $\pm$ 0.84	0.77 $\pm$ 1.14	0.21 $\pm$ 1.09	-	-
$F_{ST} \pm SD$	-	-	-	-	0.17 $\pm$ 0.13	0.09 $\pm$ 0.08
$D_{XY} \pm SD$ ( $\times 10^{-3}$ )	-	-	-	-	-	7.5 $\pm$ 13.4
SNPs	927 779	2 745 146	834 810	2 320 449	-	-
Private SNPs	173 056	926 412	112 227	272 805	-	-
Shared SNPs	-	-	-	-	788 042	2 231 400
Missing data <20%						
$\pi \pm SD$ ( $\times 10^{-3}$ )	7.0 $\pm$ 4.1	8.4 $\pm$ 10.8	6.2 $\pm$ 4.5	6.6 $\pm$ 10.9	-	-
Tajima's $D \pm SD$	1.16 $\pm$ 0.94	-0.24 $\pm$ 0.85	0.88 $\pm$ 1.16	0.29 $\pm$ 1.12	-	-
$F_{ST} \pm SD$	-	-	-	-	0.16 $\pm$ 0.12	0.09 $\pm$ 0.07
$D_{XY} \pm SD$ ( $\times 10^{-3}$ )	-	-	-	-	-	10 $\pm$ 13.0
SNPs	2 559 484	5 989 070	2 295 512	4 961 391	-	-
Private SNPs	478 959	2 247 263	298 021	688 517	-	-
Shared SNPs	-	-	-	-	2 180 759	4 757 124

MAF, minimum allele frequency; SD, standard deviation; SNP, single nucleotide polymorphisms.





**Fig. 2.** **A**, Principal component analysis (PCA) plots showing the first two principal components. **B**, A Neighbor-joining tree of *P. davidiana* (blue) and *P. rotundifolia* (red) and their hybrids (grey) using *P. trichocarpa* as an outgroup. **C**, Population structure bar plots. The scenarios of  $K = 2$  was shown, and  $K = 2$  is the best value according to cross-validation analysis.

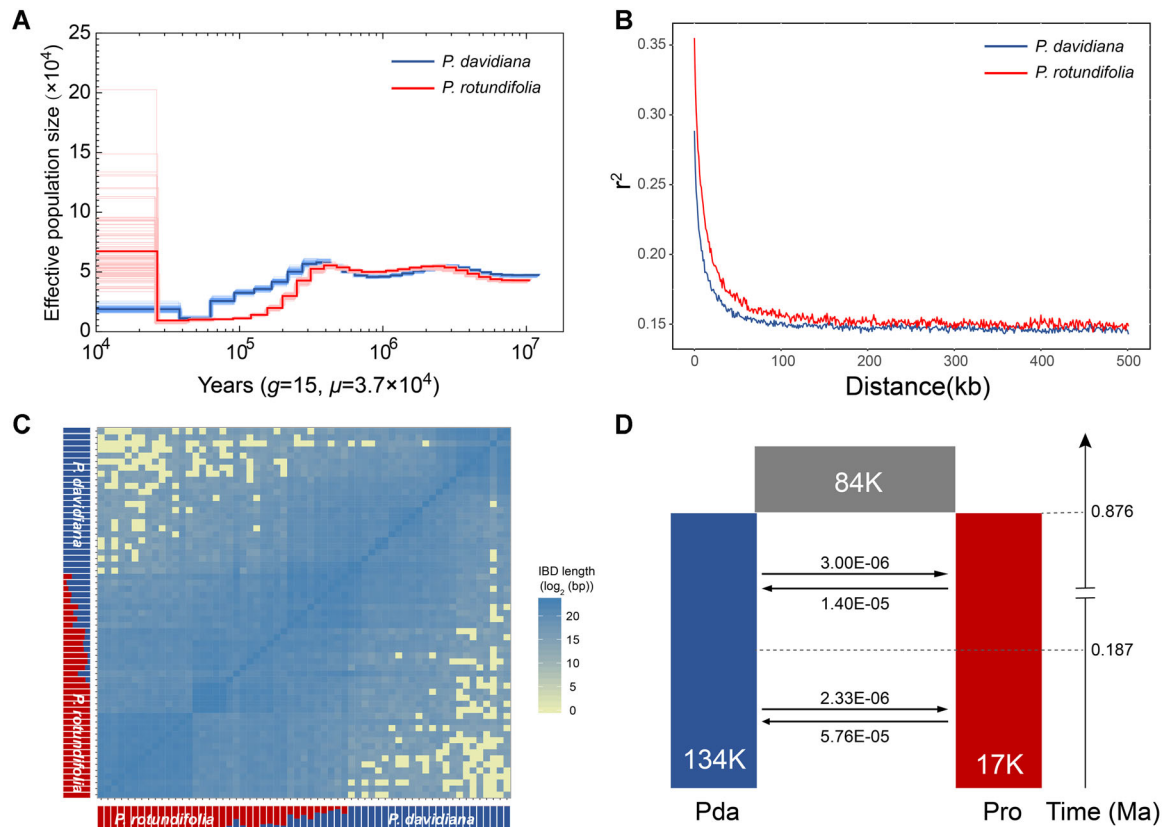
individuals showed intermediate positions between *P. rotundifolia* and *P. davidiana* (gray points, Fig. 2A). In the neighbor-joining tree, all genetically “pure” individuals of *P. rotundifolia* and *P. davidiana* were clustered separately (Fig. 2B). However, genetically mixed individuals clustered randomly with either of the two species (gray line, Fig. 2B).

To estimate patterns of genetic differentiation and genetic diversity across the genome, we excluded all genetically admixed individuals and calculated parameters of heterozygosity, ROH,  $\pi$ , LD patterns, Tajima's  $D$ ,  $F_{ST}$ , and  $D_{XY}$  for the two species based on dataset 1 (MAF >0.05, missing rate <20%; Table 1). The heterozygosity ranged from  $1.34 \times 10^{-3}$  to  $3.37 \times 10^{-3}$  across all individuals (Fig. S3). Individuals of *P. rotundifolia* harbored the lowest heterozygosity (Fig. S3), whereas individuals of *P. davidiana* exhibited the highest heterozygosity and, accordingly, the lowest level of ROH (Figs. S3, S4). In accordance with the results of heterozygosity, *P. rotundifolia* had a relatively lower genetic diversity ( $\pi = 0.0062 \pm 0.0045$  [mean  $\pm$  SD]) than *P. davidiana* ( $\pi = 0.0070 \pm 0.0041$ ; Table 1). *Populus rotundifolia* had a slightly higher recombination rate and quicker LD decay compared to *P. davidiana* (Fig. 2C). Tajima's  $D$ -value of *P. rotundifolia* (Tajima's  $D = 0.88 \pm 1.16$ ) was estimated to be lower than that in *P. davidiana* (Tajima's  $D = 1.16 \pm 0.94$ ). The Weir and Cockerham mean  $F_{ST}$  between the two species was  $0.16 \pm 0.12$  (Table 1), and the value of absolute sequence divergence ( $D_{XY}$ ) was  $0.01 \pm 0.013$ .

### 3.3 Demographic history and gene flow

Demographic history estimated by PSMC showed that both species experienced dramatic declines in population size approximately 300 kya (Fig. 3A). Compared to the population of *P. davidiana*, which experienced a slight recovery approximately 40 kya, *P. rotundifolia* expanded rapidly approximately 25 kya (Fig. 3A).

We further used a coalescent simulation-based method performed in *fastsimcoal2* to estimate the timing of divergence and demographic histories of the two species. We established five models for the divergence of *P. rotundifolia* and *P. davidiana* (Fig. S2). By comparing the AIC values for the four scenarios, the isolation with two stages of migration model (models; Fig. 2C; Table 2) with lowest AIC value and highest likelihood was the best-fitting model. Parameter estimates obtained from the best model showed that *P. rotundifolia* had a smaller population size ( $N_{Pro} = 17\,294$ ; 95% CI, 16 463–17 960; Table 3) than *P. davidiana* ( $N_{Pda} = 134\,331$ ; 95% CI, 130 408–139 030; Table 3). Assuming a mutation rate of  $3.75 \times 10^{-8}$  per base per generation and a generation time of 15 years, the divergence time between *P. rotundifolia* and *P. davidiana* was dated to 0.88 Ma (95% CI, 0.83–1.02 Ma; Table 3). The best-fit model (models; Fig. 3D; Table 2) strongly supported that gene flow between the two species was prevalent. Both the current and ancient migration rates from *P. rotundifolia* to *P. davidiana* ( $M_{Pda-Pro} = 5.76E-05$ ,  $M_{A-Pda-Pro} = 1.40E-05$ ) were



**Fig. 3.** **A**, Inferred demographic history for *Populus davidiana* and *P. rotundifolia* using the pairwise sequentially Markovian coalescent method. **B**, Patterns of linkage disequilibrium for the two species. **C**, Estimated haplotype sharing between individuals. Heatmap colors represent the total length of identity-by-descent blocks for each pairwise comparison. **D**, Maximum likelihood parameter estimates of the best fit model (model5) in *fastsimcoal2*. Pda, *P. davidiana*; Pro, *P. rotundifolia*.

much higher than that in the opposite direction ( $M_{Pro \leftarrow Pda} = 2.33E-06$ ,  $M_{A-Pro \leftarrow Pda} = 3.00E-06$ ).

### 3.4 Genome-wide selection scans for high-altitude adaptation of *P. rotundifolia*

To identify candidate genes that could be involved in high altitude adaptation of *P. rotundifolia*, we scanned the population genomic data of “pure” *P. rotundifolia* and *P. davidiana* through interspecific differentiation ( $F_{ST}$ ), nucleotide diversity ratio ( $\pi_{Pda}/\pi_{Pro}$ ), and Tajima's *D*-values within 2-kb sliding windows (Fig. 4). A total of 245 protein-coding genes in 417 windows were identified under positive selection. Although there were no significantly overrepresented GO categories detected with false discovery rate  $<0.05$ , 15 significant GO

terms were identified using the  $\chi^2$ -test. These categories were annotated to be enriched with the cellular response to stimulus (GO:0051716), the cellular response to stress (GO:0033554), DNA repair (GO:0006281), and the response to DNA damage stimulus (GO:0006974) (Table 4).

## 4 Discussion

In this study, we used population genomic data to investigate divergence history and local adaptation of *Populus rotundifolia* and its sister species *P. davidiana*. All analyses based on these population genomic data confirmed that the two aspen species were genetically distinct. The two species were

**Table 2** Comparison of parameters estimated using *fastsimcoal2* under five candidate models

Model	Scenarios	logL	k	AIC	$\Delta$ AIC	w
model5	IM2R	-1 852 843.95	12	8 532 716.98	0	1
model3	SC	-1 852 926.02	10	8 533 085.73	368.75	0
model2	IM	-1 852 933.63	8	8 533 111.55	394.57	0
model4	AM	-1 854 916.35	10	8 542 251.54	9534.56	0
model1	SI	-1 864 849.29	6	8 587 975.99	55 259.01	0

$\Delta$ AIC, AIC difference; AIC, Akaike information criterion; AM, ancient migration; k, number of estimated parameters; logL, log likelihood; IM, isolation with constant migration; IM2R, two-rate model of isolation with migration; SC, secondary contact; SI, strict isolation; w, AIC weight.

**Table 3** Inferred demographic parameters for the best-fitting demographic model for *Populus rotundifolia* (Pro) and *P. davidiana* (Pda) (shown in Fig. 3D)

Parameter	Point estimation	95% confidence intervals	
		Lower bound	Upper bound
$N_A$	83 765	77 452	87 505
$N_{Pda}$	134 331	130 408	139 030
$N_{Pro}$	17 294	16 463	17 960
$T_1$	186 840	168 690	259 440
$T_2$	876 945	833 040	1 022 280
$M_{Pda \leftarrow Pro}$	5.76E-05	5.5E-05	6.10E-05
$M_{Pro \leftarrow Pda}$	2.33E-06	1.91E-06	2.64E-06
$M_{A-Pda \leftarrow Pro}$	1.40E-05	8.65E-08	1.84E-05
$M_{A-Pro \leftarrow Pda}$	3.00E-06	2.50E-06	3.96E-06

$M$ , current migration rate per generation between species;  $M_A$ , the ancient migration rate per generation between species;  $N$ , the current effective population sizes of each species;  $N_A$ , the ancestral effective population sizes;  $T$ , estimated divergence time ( $T_2$ , years) and the time of change of migration rates ( $T_1$ , years).

estimated to start to diverge approximately 0.88 Ma (95% CI, 0.83–1.02 Ma), corresponding with the end of the Xixiabangma glaciation (0.8–1.17 Ma; Zheng et al., 2002). Coalescent simulations showed that gene flows occurred between both current and ancient populations after their divergence, and hybrids were confirmed to exist in the contacting regions. The highly diverged genes with signatures of selection are likely associated with high-altitude adaptation. Our results suggest that geographic isolation caused by Quaternary climate change might contribute to the initial divergence between *P. rotundifolia* and *P. davidiana*, and natural selection is supposed to play an important role in maintaining species distinction despite continuous migrations between them.

#### 4.1 Divergence of *P. rotundifolia* from *P. davidiana* with gene flow

Based on population genomic data, we found that these two species were well delimited in their allopatric distributions (Fig. 1). In contrast, the individuals from their contacting regions were found to be hybrids with the admixed makeup from two parental species (Fig. 2A). These findings support those of previous studies reporting the distinctions of the two species in both morphological traits and genetic composition (Zheng et al., 2017; Shang et al., 2020; Wang et al., 2020b).

According to the best-fitting demographic model conducted from *fastsimcoal2*, the early divergence of two species was estimated to initialize approximately 0.88 Ma (95% CI, 0.83–1.02 Ma), which falls within the timescale of the earliest Quaternary glaciation recognized on the QTP, the Xixiabangma glaciation (0.8–1.17 Ma; Zheng et al., 2002). During this cold period, ice sheets covered a large part of the QTP and could likely fragment the habitats of ancestral populations and therefore limit gene flow between populations. Previous studies showed

that Quaternary climatic oscillation had important effects on the biodiversity of the QTP (Liu et al., 2014), in shaping both the interspecific (Xu et al., 2019) and intraspecific (Ren et al., 2017) divergences of current species. Therefore, the initial divergence between *P. rotundifolia* and *P. davidiana* might be mainly influenced by Quaternary climatic changes.

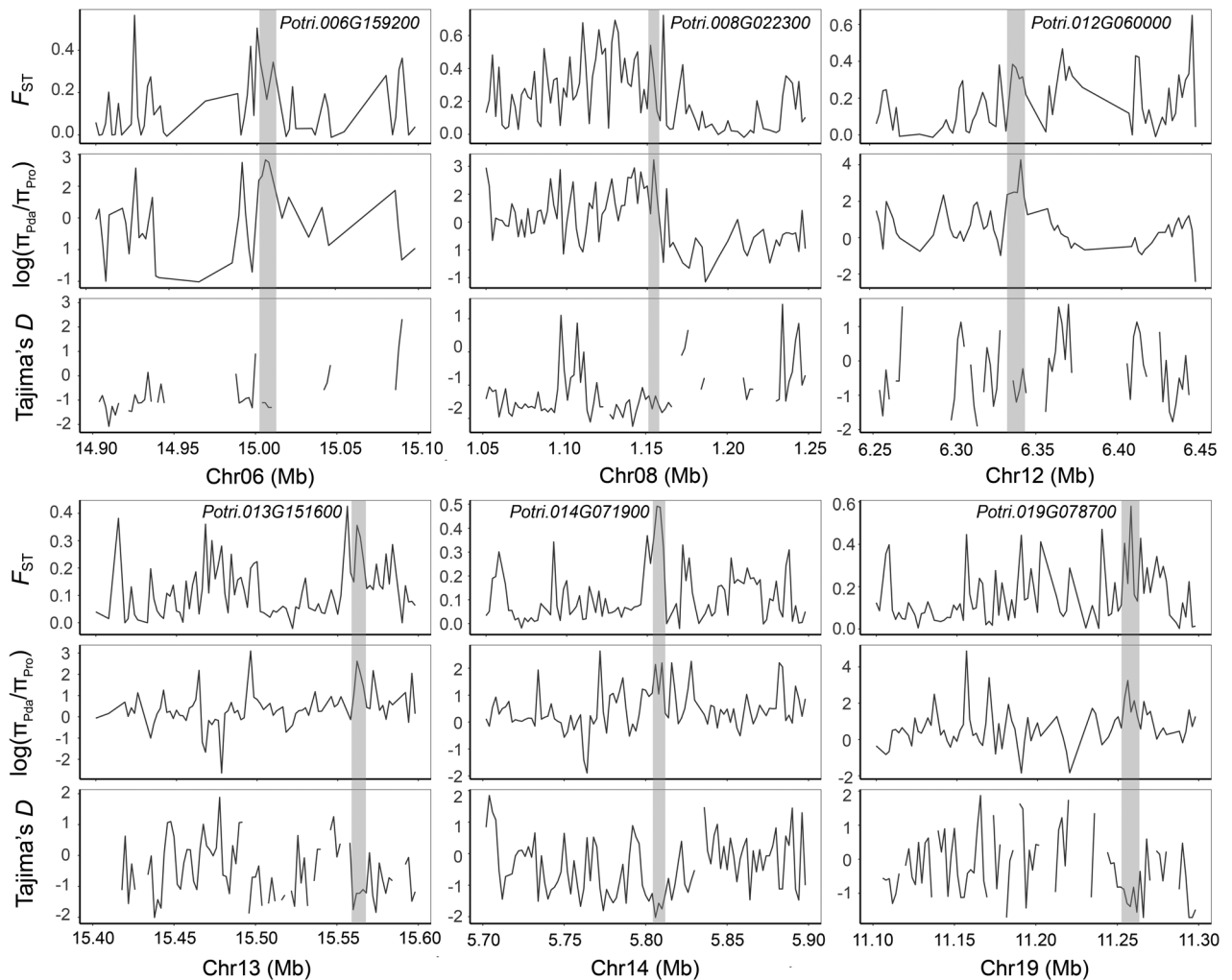
Although these two species diverged nearly 1 million years ago, reproductive isolations are incomplete between them, possibly because of the large population size and long generation time of *Populus* species (Wang et al., 2020a, 2020b). This would have resulted in the occurrence of hybrids in the contacting region of the two species in the eastern QTP. In addition, seeds and pollen of *Populus* species are wind-dispersed, which further contributes to their long-distance dispersal, and accordingly causes frequent gene flow between populations, even over great distances. Moreover, gene flow might have also occurred between allopatric populations of the two species, as suggested based on modelling of their speciation process (Fig. S2). This modelling was undertaken based on coalescent analyses of joint site frequency spectra extracted from individuals of two species when all putative hybrids from the contacting regions were excluded. In contrast to populations in the contacting regions, the effects of natural selection could work against interspecific gene flow and maintain species distinction in the allopatric distributions of two species. Therefore, ecological selection is supposed to have played a central role in restricting high levels of gene flow between populations in allopatric habitats of the two species. Taken together, gene flow seems to have been continuous during the divergence of two species while divergent ecological selection is supposed to maintain their genetic distinction in allopatric distributions. Similar findings were also observed in population genomic analyses of other aspen or poplar species in the recent past (Wang et al., 2016; Ma et al., 2018; Shang et al., 2020).

Finally, we note that some other poplars co-occur with *P. rotundifolia* and *P. davidiana* (Fang et al., 1999; Rogers et al., 2020), which suggested that there might be potential gene flow between these species. Indeed, previous studies using ABBA–BABA tests (Wang et al., 2020b) and coalescent-based simulations (Wang et al., 2020a) showed that gene flow occurs between either *P. rotundifolia* or *P. davidiana* and other poplars, such as *P. tremula* L. and *P. adenopoda* Maxim. Nevertheless, *P. davidiana* and *P. rotundifolia* are sister species to each other, sharing the most recent common ancestor that is not shared with any other poplar species, and the magnitude of gene flow between closely related species was expected to be much higher than that between species with distant relationships (Wang et al., 2020b). Therefore, it seems unlikely that the introgressed alleles from other poplars can bias our inference. Moreover, obvious gene flow in *Populus* species (Wang et al., 2020b) further indicates that speciation with gene flow would be very common in *Populus* species.

#### 4.2 Genomic footprints of high-altitude adaptation in *P. rotundifolia*

*Populus rotundifolia* is an alpine tree in the QTP and occupies the highest habitats among all species in the





**Fig. 4.** Positive selection signatures around six genes in *Populus rotundifolia* (Potri.006G159200, Potri.008G022300, Potri.012G060000, Potri.013G151600, Potri.014G071900, and Potri.019G078700) involved in the functions of “DNA repair” and “response to DNA damage stimulus”. High altitude living species *P. rotundifolia* (Pro) was compared with lowland species *P. davidiana* (Pda). The population genetic differentiation  $F_{ST}$  value, the nucleotide diversity ratio ( $\pi_{Pda}/\pi_{Pro}$ ), and Tajima's  $D$ -values of Pro are calculated within 2-kb sliding windows. Chr, Chromosome.

genus *Populus* (Zheng et al., 2016; Shang et al., 2020; Wang et al., 2020b). Species in highlands are exposed to various abiotic stresses (Körner, 2003; Wang et al., 2018), such as low level of oxygen partial pressure (Streb et al., 2005), low precipitation or temperatures (Bliss, 1962; Li et al., 2007; Manel et al., 2012; Ma et al., 2019), and high levels of UV radiation (Norsang et al., 2011). For instance, previous studies based on de novo genomes of two alpine plants suggested that a few genes involved in avoiding UV radiation had been duplicated, and further contribute to the high-altitude adaptation of these alpine plants (Guo et al., 2018; Zhang et al., 2019b).

Based on the estimates of genetic differentiation ( $F_{ST}$ ), diversity ( $\pi$ ), and Tajima's  $D$ , we searched for genes that likely experienced strong selection in response to alpine environmental stress in *P. rotundifolia*. A total of 245 protein-coding genes were identified to be under positive

selection based on such a combination of different indexes, and they are involved in the GO terms like “cellular response to stimulus,” “cellular response to stress,” “DNA repair,” and “response to DNA damage stimulus” (Table 4). Alpine plants are exposed to UV radiation, leading to cell and DNA damage (Britt, 1999). Cell and DNA repair after damage are important for alpine plants to live in highland regions with strong UV radiation, which suggested that these genes could possibly be important in high-altitude adaptation. For example, the Potri.012G060000 and Potri.006G159200 genes, which are homologous to AT3G18524 and AT3G24495 genes in *Arabidopsis thaliana* (L.) Heynh., encode DNA mismatch repair proteins MSH2 and MSH7, respectively, which form the MutS gamma (MSH2–MSH7 heterodimer) of the post-replicative DNA mismatch repair system in plants. In addition, these genes play important

**Table 4** Significantly over-represented Gene Ontology (GO) terms of positive selection genes (PSGs) in *Populus rotundifolia*

GO term	Ontology	Description	PSGs	Annotated	P-value
GO:0006281	BP	DNA repair	6	205	0.0018
GO:0006974	BP	Response to DNA damage stimulus	6	206	0.0019
GO:0051716	BP	Cellular response to stimulus	6	217	0.0031
GO:0033554	BP	Cellular response to stress	6	217	0.0031
GO:0006520	BP	Cellular amino acid metabolic process	7	310	0.0074
GO:0044106	BP	Cellular amine metabolic process	7	325	0.011
GO:0006259	BP	DNA metabolic process	7	329	0.012
GO:0006519	BP	Cellular amino acid and derivative metabolic process	7	345	0.017
GO:0009308	BP	Amine metabolic process	7	367	0.027
GO:0016829	MF	Lyase activity	6	272	0.018
GO:0004386	MF	Helicase activity	6	277	0.02
GO:0017111	MF	Nucleoside-triphosphatase activity	13	872	0.022
GO:0016462	MF	Pyrophosphatase activity	13	893	0.027
GO:0016818	MF	Hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	13	919	0.035
GO:0016817	MF	Hydrolase activity, acting on acid anhydrides	13	944	0.045

BP, biological process; MF, molecular function.

roles in DNA homologous recombination repair after UV-B-induced DNA damage (Culligan & Hays, 2000; Wu et al., 2003). A recent study based on qingke growing in alpine habitats also found rapid adaptive evolution of genes that are likely resistant to strong UV-B radiation (Zeng et al., 2020). It is highly likely that genes with similar functions were subject to strong selection pressure and therefore showed significant signatures of selection and elevated genetic differentiation between *P. rotundifolia* relative to its sister species, *P. davidiana*, although further functional experiments are strongly needed.

## Acknowledgements

This work was supported by the Second Tibetan Plateau Scientific Expedition and Research (STEP) program (grant number 2019QZKK0502) and the National Natural Science Foundation of China (grant numbers 41571054 and 31590821).

## Conflict of Interest

None declared.

## References

- Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, Boughman JW, Brelsford A, Buerkle CA, Buggs R, Butlin RK, Dieckmann U, Eroukhmanoff F, Grill A, Cahan SH, Hermansen JS, Hewitt G, Hudson AG, Jiggins C, Jones J, Keller B, Marczewski T, Mallet J, Martinez-Rodriguez P, Moest M, Mullen S, Nichols R, Nolte AW, Parisod C, Pfennig K, Rice AM, Ritchie MG, Seifert B, Smadja CM, Stelkens R, Szymura JM, Vainola R, Wolf JBW, Zinner D. 2013. Hybridization and speciation. *Journal of Evolutionary Biology* 26: 229–246.
- Aguirre-Liguori JA, Gaut BS, Jaramillo-Correa JP, Tenaillon MI, Montes-Hernández S, García-Oliva F, Hearne SJ, Eguiarte LE. 2019. Divergence with gene flow is driven by local adaptation to temperature and soil phosphorus concentration in teosinte subspecies (*Zea mays parviglumis* and *Zea mays mexicana*). *Molecular Ecology* 28: 2814–2830.
- Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19: 1655–1664.
- Berrill JP, Dagley CM, Coppeto SA, Gross SE. 2017. Curtailing succession: Removing conifers enhances understory light and growth of young aspen in mixed stands around Lake Tahoe, California and Nevada, USA. *Forest Ecology and Management* 400: 511–522.
- Bliss LC. 1962. Adaptations of arctic and alpine plants to environmental conditions. *Arctic* 15: 117–144.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.
- Britt AB. 1999. Molecular genetics of DNA repair in higher plants. *Trends in Plant Science* 4: 20–25.
- Browning BL, Browning SR. 2013. Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* 194: 459–471.
- Browning SR, Browning BL. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American Journal of Human Genetics* 81: 1084–1097.
- Chaturvedi S, Lucas LK, Buerkle CA, Fordyce JA, Forister ML, Nice CC, Gompert Z. 2020. Recent hybrids recapitulate ancient hybrid outcomes. *Nature Communications* 11: 2179.
- Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* 6: 80–92.
- Clarkson CS, Weetman D, Essandoh J, Yawson AE, Maslen G, Manske M, Field SG, Webster M, Antao T, MacInnis B. 2014. Adaptive introgression between *Anopheles* sibling species eliminates a major genomic island but not reproductive isolation. *Nature Communications* 5: 4248.

- Culligan KM, Hays JB. 2000. *Arabidopsis* MutS homologs—atmsh2, atmsh3, atmsh6, and a novel atmsh7—form three distinct protein heterodimers with different specificities for mismatched DNA. *The Plant Cell* 12: 991.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R. 2011. The variant call format and VCFtools. *Bioinformatics* 27: 2156–2158.
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernysky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* 43: 491–498.
- Edenius L, Ericsson G, Kempe G, Bergström R, Danell K. 2011. The effects of changing land use and browsing on aspen abundance and regeneration: A 50-year perspective from Sweden. *Journal of Applied Ecology* 48: 301–309.
- Excoffier L, Foll M. 2011. Fastsimcoal: A continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* 27: 1332–1334.
- Fan L, Zheng H, Milne RI, Zhang L, Mao K. 2018. Strong population bottleneck and repeated demographic expansions of *Populus adenopoda* (Salicaceae) in subtropical China. *Annals of Botany* 665–679.
- Fang ZF, Zhao SD, Skvortsov AK. 1999. Salicaceae. In: Wu Z-Y, Raven PH, Hong D-Y eds. *Flora of China*. Beijing: Science Press; St. Louis: Missouri Botanical Garden Press. 4: 139–274.
- Gao F, Ming C, Hu W, Li H. 2016. New software for the fast estimation of population recombination rates (FastEPFR) in the genomic era. *G3* 6: 1563–1571.
- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N. 2012. Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Research* 40: D1178–D1186.
- Guo X, Hu Q, Hao G, Wang X, Zhang D, Ma T, Liu J. 2018. The genomes of two *Eutrema* species provide insight into plant adaptation to high altitudes. *DNA Research* 25: 307–315.
- Han F, Lamichhaney S, Grant BR, Grant PR, Andersson L, Webster MT. 2017. Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. *Genome Research* 27: 1004–1015.
- Hoskin CJ, Higgie M, McDonald KR, Moritz C. 2005. Reinforcement drives rapid allopatric speciation. *Nature* 437: 1353–1356.
- Hou Z, Li A, Zhang J. 2020. Genetic architecture, demographic history, and genomic differentiation of *Populus davidiana* revealed by whole-genome resequencing. *Evolutionary Applications*, 1–15. <https://doi.org/10.1111/eva.13046>
- Hou Z, Wang Z, Ye Z, Du S, Liu S, Zhang J. 2018. Phylogeographic analyses of a widely distributed *Populus davidiana*: Further evidence for the existence of glacial refugia of cool-temperate deciduous trees in northern East Asia. *Ecology and Evolution* 8: 13014–13026.
- Jia H, Liu G, Li J, Zhang J, Sun P, Zhao S, Zhou X, Lu M, Hu J. 2020a. Genome resequencing reveals demographic history and genetic architecture of seed salinity tolerance in *Populus euphratica*. *Journal of Experimental Botany* 71: 4308–4320.
- Jia KH, Zhao W, Maier PA, Hu XG, Jin Y, Zhou SS, Jiao SQ, El-Kassaby YA, Wang T, Wang XR, Mao JF. 2020b. Landscape genomics predicts climate change-related genetic offset for the widespread *Platycladus orientalis* (Cupressaceae). *Evolutionary Applications* 13: 665–676.
- Karrenberg S, Liu X, Hallander E, Favre A, Herforth-Rahme J, Widmer A. 2019. Ecological divergence plays an important role in strong but complex reproductive isolation in *Silene*. *Evolution* 73: 245–261.
- Körner C. 2003. Alpine plant life: Functional plant ecology of high mountain ecosystems. Verlag Berlin, Heidelberg: Springer Science & Business Media.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760.
- Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475: 493.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078–2079.
- Li SC, Xu L, Guo YX, Qian WH, Zhang GQ, Li C. 2007. Change of annual precipitation over Qinghai-Xizang Plateau and sub-regions in recent 34 years. *Journal of Desert Research* 27: 307–314.
- Liang Q, Xu X, Mao K, Wang M, Wang K, Xi Z, Liu J. 2018. Shifts in plant distributions in response to climate warming in a biodiversity hotspot, the Hengduan Mountains. *Journal of Biogeography* 45: 1334–1344.
- Liu JQ, Duan YW, Hao G, Ge XJ, Sun H. 2014. Evolutionary history and underlying adaptation of alpine plants on the Qinghai-Tibet Plateau. *Journal of Systematics Evolution* 52: 241–249.
- Ma T, Wang K, Hu Q, Xi Z, Wan D, Wang Q, Feng J, Jiang D, Ahani H, Abbott RJ. 2018. Ancient polymorphisms and divergence hitchhiking contribute to genomic islands of divergence within a poplar species complex. *Proceedings of the National Academy of Sciences USA* 115: E236–E243.
- Ma Y, Wang J, Hu Q, Li J, Sun Y, Zhang L, Abbott RJ, Liu J, Mao K. 2019. Ancient introgression drives adaptation to cooler and drier mountain habitats in a cypress species complex. *Communications Biology* 2: 1–12.
- Manel S, Gugerli F, Thuiller W, Alvarez N, Legendre P, Holderegger R, Gielly L, Taberlet P, IntraBioDiv C. 2012. Broad-scale adaptive genetic variation in alpine plants is driven by temperature and precipitation. *Molecular Ecology* 21: 3729–3738.
- Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. 2010. Robust relationship inference in genome-wide association studies. *Bioinformatics* 26: 2867–2873.
- Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, Blaxter M, Manica A, Mallet J, Jiggins CD. 2013. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research* 23: 1817–1828.
- Mayr E. 1963. *Animal species and evolution*. Cambridge: Belknap Press.
- Norsang G, Kocbach L, Stamnes J, Tsoja W, Pincuo N. 2011. Spatial distribution and temporal variation of solar UV radiation over the Tibetan Plateau. *Applied Physics Research* 3: 37.
- Nosil P. 2008. Speciation with gene flow could be common. *Molecular Ecology* 17: 2103–2106.
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* 38: 904–909.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. 2007. PLINK: A tool set for whole-genome association and population-based

- linkage analyses. *American Journal of Human Genetics* 81: 559–575.
- Qu Y, Chen C, Xiong Y, She H, Zhang YE, Cheng Y, DuBay S, Li D, Ericson PGP, Hao Y, Wang H, Zhao H, Song G, Zhang H, Yang T, Zhang C, Liang L, Wu T, Zhao J, Gao Q, Zhai W, Lei F. 2019. Rapid phenotypic evolution with shallow genomic differentiation during early stages of high elevation adaptation in Eurasian Tree Sparrows. *National Science Review* 7: 113–127.
- Ren G, Mateo RG, Liu J, Suchan T, Alvarez N, Guisan A, Conti E, Salamin 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* 213: 1500–1512.
- Riesch R, Muschick M, Lindtke D, Villoutreix R, Comeault AA, Farkas TE, Lucek K, Hellen E, Soria-Carrasco V, Dennis SR, de Carvalho CF, Safran RJ, Sandoval CP, Feder J, Gries R, Crespi BJ, Gries G, Gompert Z, Nosil P. 2017. Transitions between phases of genomic differentiation during stick-insect speciation. *Nature Ecology & Evolution* 1: 0082.
- Rogers PC, Pinno BD, Šebesta J, Albrechtsen BR, Li G, Ivanova N, Kusbach A, Kuuluvainen T, Landhäusser SM, Liu H, Myking T, Pulkkinen P, Wen Z, Kulakowski D. 2020. A global view of aspen: Conservation science for widespread keystone systems. *Global Ecology and Conservation* 21: e00828.
- Ru DF, Sun YS, Wang DL, Chen Y, Wang TJ, Hu QJ, Abbott RJ, Liu JQ. 2018. Population genomic analysis reveals that homoploid hybrid speciation can be a lengthy process. *Molecular Ecology* 27: 4875–4887.
- Savolainen O, Lascoux M, Merilä J. 2013. Ecological genomics of local adaptation. *Nature Reviews Genetics* 14: 807–820.
- Shang H, Hess J, Pickup M, Field DL, Ingvarsson PK, Liu J, Lexer C. 2020. Evolution of strong reproductive isolation in plants: Broad-scale patterns and lessons from a perennial model group. *Philosophical Transactions of the Royal Society B: Biological Sciences* 375: 20190544.
- Streb P, Josse EM, Gallouet E, Baptist F, Kuntz M, Cornic G. 2005. Evidence for alternative electron sinks to photosynthetic carbon assimilation in the high mountain plant species *Ranunculus glacialis*. *Plant, Cell & Environment* 28: 1123–1135.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. Mega6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* 30: 2725–2729.
- Tarailo-Graovac M, Chen N. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics* Chapter 4: Unit 4.10.
- Taylor SA, Larson EL. 2019. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology & Evolution* 3: 170–177.
- Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, Xu W, Su Z. 2017. agriGO v2.0: A GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Research* 45: W122–W129.
- Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroove S, Dejardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehrling J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjarvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leple JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouze P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313: 1596–1604.
- Via S. 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367: 451–460.
- Wang GD, Zhang BL, Zhou WW, Li YX, Jin JQ, Shao Y, Yang HC, Liu YH, Yan F, Chen HM. 2018. Selection and environmental adaptation along a path to speciation in the Tibetan frog *Nanorana parkeri*. *Proceedings of the National Academy of Sciences USA* 115: E5056–E5065.
- Wang J, Street NR, Park EJ, Liu J, Ingvarsson PK. 2020a. Evidence for widespread selection in shaping the genomic landscape during speciation of *Populus*. *Molecular Ecology* 29: 1120–1136.
- Wang J, Street NR, Scofield DG, Ingvarsson PK. 2016. Variation in linked selection and recombination drive genomic divergence during allopatric speciation of European and American aspens. *Molecular Biology and Evolution* 33: 1754–1767.
- Wang M, Zhang L, Zhang Z, Li M, Wang D, Zhang X, Xi Z, Keefover-Ring K, Smart LB, DiFazio SP, Olson MS, Yin T, Liu JQ, Ma T. 2020b. Phylogenomics of the genus *Populus* reveals extensive interspecific gene flow and balancing selection. *New Phytologist* 225: 1370–1382.
- Wang TJ, Ru DF, Zhang D, Hu QJ. 2019. Analyses of genome-scale variation reveal divergence of two *Sinallaria* species (Brassicaceae) with continuous but limited gene flow. *Journal of Systematics and Evolution* 57: 268–277.
- Wu SY, Culligan K, Lamers M, Hays J. 2003. Dissimilar mismatch-recognition spectra of *Arabidopsis* DNA-mismatch-repair proteins msh2-msh6 (mut $\alpha$ ) and msh2-msh7 (mut $\gamma$ ). *Nucleic Acids Research* 31: 6027–6034.
- Xu J, Song X, Ruhsam M, Liu T, Li J, Neaves LE, Miao J, Xie S, Meng Q, Mao K. 2019. Distinctiveness, speciation and demographic history of the rare endemic conifer *Juniperus erectopatens* in the eastern Qinghai-Tibet Plateau. *Conservation Genetics* 20: 1289–1301.
- Zeng X, Yuan H, Dong X, Peng M, Jing X, Xu Q, Tang T, Wang Y, Zha S, Gao M. 2020. Genome-wide dissection of co-selected UV-B responsive pathways in the UV-B adaptation of qingke. *Molecular Plant* 13: 112–127.
- Zhang L, Xi Z, Wang M, Guo X, Ma T. 2018. Plastome phylogeny and lineage diversification of Salicaceae with focus on poplars and willows. *Ecology and Evolution* 8: 7817–7823.
- Zhang M, Suren H, Holliday JA. 2019a. Phenotypic and genomic local adaptation across latitude and altitude in *Populus trichocarpa*. *Genome Biology and Evolution* 11: 2256–2272.
- Zhang T, Qiao Q, Novikova PY, Wang Q, Yue J, Guan Y, Ming S, Liu T, De J, Liu Y. 2019b. Genome of *Crucihimalaya himalaica*, a close relative of *Arabidopsis*, shows ecological adaptation to high altitude. *Proceedings of the National Academy of Sciences USA* 116: 7137–7146.
- Zheng B, Xu Q, Shen Y. 2002. The relationship between climate change and Quaternary glacial cycles on the Qinghai-Tibetan Plateau: Review and speculation. *Quaternary International* 97–98: 93–101.

- Zheng H, Fan L, Milne RI, Zhang L, Wang Y, Mao K. 2017. Species delimitation and lineage separation history of a species complex of aspens in China. *Frontiers in Plant Science* 8: 375.
- Zheng H, Fan L, Wang T, Zhang L, Ma T, Mao K. 2016. The complete chloroplast genome of *Populus rotundifolia* (Salicaceae). *Conservation Genetics Resources* 8: 399–401.

## Supplementary Material

The following supplementary material is available online for this article at <http://onlinelibrary.wiley.com/doi/10.1111/jse.12665/supinfo>:

**Fig. S1.** Statistical summary of climate differentiation between *Populus davidiana* and *P. rotundifolia*.

**Fig. S2.** Speciation scenarios for *Populus davidiana* (Pda) and *P. rotundifolia* (Pro).

**Fig. S3.** Bar diagram of values of heterozygosity of each individual.

**Fig. S4.** Bar diagram of homozygosity (ROH) for each individual.

**Fig. S5.** Genetic structure analysis for 61 individuals using ADMIXTURE with  $K = 2-5$ .

**Fig. S6.** Single nucleotide polymorphism frequency spectrums for (A) *Populus davidiana* and (B) *P. rotundifolia*.

**Table S1.** Overview of sample information and sequencing statistics.

**Table S2.** Cross-validation value for  $K$  from 1 to 5 in ADMIXTURE analysis.

**Table S3.** Parameters of four candidate models inferred from *fastsimcoal2*.

**Table S4.** Amino acid change table for 245 positive selection genes.