Natural selection and intrinsic barriers play important roles in speciation of two closely related Populus (Salicaceae) species

Yang Tian[1], Shuyu Liu[1], Pär Ingvarsson[2], Dandan Zhao[1], Li Wang[1], Baoerjiang Abuduhamiti[1], Zhiqiang Wu[3], Jian-Guo ZHANG[1], and Zhaoshan Wang[1]

[1]Chinese Academy of Forestry Research Institute of Forestry
[2]Swedish University of Agricultural Sciences
[3]Chinese Academy of Agricultural Sciences Agricultural Genomes Institute at Shenzhen

February 5, 2021

Abstract

Despite the growing number of recent studies on genome-wide divergence during speciation, the genetic basis and mechanisms of genomic divergence are still incompletely understood. In most species, natural selection plays a key role in heterogeneous genomic divergence. Additionally, intrinsic barriers, such as chromosomal rearrangements or gene incompatibilities, can also cause genomic heterogeneity. Based on whole genome re-sequencing data from 27 Populus alba and 28 P. adenopoda individuals, we explored the reasons for heterogeneous genomic divergence of these two closely related species. The results showed that the two species diverged ~5-10 million years ago (Mya), when the Qinghai-Tibet Plateau reached a certain height and the inland climate of the Asian continent became arid, which is associated with the fact that the two species begin to diverge and eventually led to speciation. In highly differentiated regions, the absolute divergence (dxy) was significantly higher than genomic background, and relative and absolute divergence were highly correlated, which indicates that intrinsic barriers played an important role in maintaining genomic heterogeneous divergence. Additionally, $\vartheta\pi$ and shared polymorphisms decreased while fixed differences increased in highly differentiated regions, which are characteristics of natural selection. The above description indicates that the combination of intrinsic barriers and natural selection result in heterogeneous genomic divergence and reproductive isolation. We further found some genes that are related to reproduction may be involved in explaining the reproductive isolation of the two species.

Introduction

Understanding the origins of reproductive isolation is a major goal of modern speciation studies in plants because reproductive isolation is crucial for speciation (Widmer, Lexer, & Cozzolino, 2008; Seehausen et al., 2014; Ravinet et al., 2017; Stankowski et al., 2019). Reproductive isolation is a consequence of prezygotic and postzygotic barriers as well as their potentially complex interactions (Rieseberg & Willis, 2007; Widmer et al., 2008). Prezygotic isolation, due to reduced probabilities for meeting (spatial or temporal isolation), mating (sexual isolation) or successful fertilization (homogametic sperm preferences or sperm-egg incompatibilities), has been found to be a greater contributor to the total isolation between species than postzygotic isolation (Qvarnstrom & Bailey, 2009). Postzygotic isolation may be caused by inferior niche adaptation, reduced attractiveness of hybrids or reduced fertility and/or viability of hybrids resulting from intrinsic incompatibilities (Baack, Melo, Rieseberg, & Ortiz-Barrientos, 2015; Levin, 1978; Lynch & Force, 2000; MacNair & Christie, 1983; Orr & Turelli, 2001; Qvarnström & Bailey, 2008). Some degree of postzygotic isolation often occurs in recently diverged species and are followed by the slow establishment of stronger prezygotic isolation (Seehausen et al., 2014; Widmer et al., 2008).

The origins of reproductive isolation include multiple evolutionary forces, including divergent selection that

creates extrinsic reproductive isolation, genetic drift, gene incompatibilities that cause intrinsic reproductive isolation (Baack, Melo, Rieseberg, & Ortiz-Barrientos, 2015; Seehausen et al., 2014). Reproductive isolation is often considered to be an important by-product of genetic drift or differential adaptation when no gene flow is present (Noor, Grams, Bertucci, & Reiland, 2001; Nosil, Egan, & Funk, 2008). However, divergent selection in the face of gene flow is a common scenario that can increase the genomic differentiation of the related regions in organisms (Nosil, Funk, & Ortiz-barrientos, 2009). Positive and/or negative selection at individual loci can reduce the diversity in one or both daughter populations, thereby increasing the degree of differentiation at these loci (Geraldes, Basset, Smith, & Nachman, 2011; Rifkin, Castillo, Liao, & Rausher, 2019). Selective sweeps or background selection can also lead to great amount of differentiation in genomic regions with lower recombination rates (Nachman & Payseur, 2012). In addition, phenomena such as chromosomal rearrangements or gene incompatibilities can also contribute to reproductive isolation (Baack, Melo, Rieseberg, & Ortiz-Barrientos, 2015; Noor et al., 2001; Rieseberg, 2001; Schumer et al., 2014). Gene incompatibilities may lead to infertility or even mortality of hybrids between species and chromosomal rearrangements are known to play an important role in the origin of reproductive isolation because they sometimes disrupt meiosis in hybrids, resulting in inviability and/or infertility in hybrids (Hou, Friedrich, de Montigny, & Schacherer, 2014; Lynch & Force, 2000; Avelar, Perfeito, Gordo, & Ferreira, 2013).

A common interpretation of the small number of strongly differentiated regions (based on F_{ST}) identified during genome scanning is that they represent loci related to reproductive isolation or ecological specialization and strong selection prevent their introgression between species (Payseur & Rieseberg, 2016; Turner, Hahn, & Nuzhdin, 2005; Wu, 2001). Loci unrelated to reproductive isolation will experience homogenization from gene flow and will consequently exhibit low or no genetic differentiation (Ravinet et al., 2017). These differences in the action of natural selection is thought to explain a pattern of heterogeneous differentiation among sites in the genomes of closely related species (Cruickshank & Hahn, 2014). There are many factors leading to heterogeneous genomic divergence, including selection due to ecological causes (Funk et al., 2016) or genetic conflicts (Nosil et al., 2009; Presgraves, 2007), random effects of genetic drift (Funk et al., 2016), variable mutation rates (Lynch & Force, 2000; Baack, Melo, Rieseberg, & Ortiz-Barrientos, 2015) or the genomic distribution and effect size of selected genes and chromosome structure (Noor et al., 2001; Rieseberg, 2001; Avelar et al., 2013). And there are also intrinsic features of the genome, such as heterogeneous gene density, local diversity, and local recombination rates (Booker, Yeaman, & Whitlock, 2020; Noor & Bennett, 2009; Ravinet et al., 2017).

Therefore, it is important to analyse different driving factors to account for how heterogeneous genomic divergence is formed and what roles they play in the formation and maintenance of reproductive isolation and adaptive phenotypic differentiation (Jacobs, Hughes, Robinson, Adams, & Elmer, 2018). If there is no gene flow and natural selection between species that have recently diverged, the heterogeneous genomic differentiation is simply resulting from stochastic variation in coalescence times (Barton, 2006). However, if selection is present, regions that experience strong natural selection will exhibit larger differences than those under weaker selection and will thus share less ancestral variation (Cruickshank & Hahn, 2014). In the presence of gene flow, selection reduces the diversity of selected loci but also on sites linked to the selected loci thereby promoting genetic differentiation and resulting in an inflation of measures such as F_{ST} between species (Jakobsson, Edge, & Rosenberg, 2013; Walsh & Blows, 2009). Natural selection reduces effective gene flow near selected sites, however, selection does not affect the redistribution of alleles in daughter species. Other regions exhibit low differentiation, either because they maintain a high level of ancestral polymorphism or because the homogenizing effects of gene flow, thereby leading to heterogenous genetic differentiation across the genome or an organism (Turner & Hahn, 2010; Seehausen et al., 2014; Wang, Street, Scofield, & Ingvarsson, 2016).

Intrinsic barriers, such as chromosomal rearrangements or gene incompatibilities, can also cause heterogeneous genomic divergence because they inhibit the redistribution of alleles between daughter species. Such barriers can lead to the formation of highly differentiated regions by countering the homogenizing effects of gene flow (Feder & Nosil, 2009). At the same time, other areas of the genome are homogenized by gene flow, consequently showing lower levels of differentiation. This process results in a highly variable degree of gene-

tic differentiation throughout the genome (Nosil et al., 2009; Ortíz-Barrientos, Reiland, Hey, & Noor, 2002; Wolf & Ellegren, 2016). Chromosomal rearrangements are also known to reduce the fitness of heterozygotes via by affecting meiosis, which may eventually result in reproductive isolation (Baack, Melo, Rieseberg, & Ortiz-Barrientos, 2015; Feulner & De-Kayne, 2017).

In summary, the factors outlined above can explain heterogeneity in genomic differentiation, distinguishing them and deciding which are the main factors in a particular scenario is a challenging task. Of course, these assumptions are not mutually exclusive and detailed information about the process of speciation, such as the time of speciation, the time of differentiation, and the demographic history of the species pair, are required to understand the causes of heterogeneous differentiation across the genome (Nosil et al., 2008; Wang, Street, Scofield, & Ingvarsson, 2016).

Populus alba and *P. adenopoda* are two important tree species in the section *Populus* with large ecological and economic value (Wang et al., 2010). A phylogenetic tree constructed using a combination of cpDNA and nuclear DNA show that the two species are sister species (Wang et al., 2014), indicating that they have diverged recently. The former *P. alba* is widely distributed and cultivated across the North African deserts, European flood plain forests and in central Asian regions having a strong continental climate and severe winter frosts (Dickmann & Kuzovkina, 2014; Stölting et al., 2015). The latter *P. adenopoda* is endemic to China, where it often occurs on sunny slopes or along riversides and is common in the southwest, where it shows good natural regeneration (Fan, Zheng, Milne, Zhang, & Mao, 2018). Natural populations of *P. alba* often hybridise with populations of other closely related species such as *P. adenopoda* , have produced a large number of natural *P. tomentosa*hybrids, which have poor fertility and produce almost no offspring (Li, Huang, & Wang, 1997; Wang et al., 2011). Therefore, while the two species have diverged recently they nevertheless show evidence for very strong postzygotic isolation. And *P. alba* and *P. adenopoda*is a promising system for evaluating how different evolutionary forces contribute to the genomic patterns of divergence during speciation and how the accumulation of genetic differences leads to intrinsic barriers to reproduction (Palumbi, 1994).

In this study, we used genome-wide re-sequencing data from the two species to infer their divergence time and estimate their historical population dynamics. By examining population genomic statistics, such as genetic diversity and relative and absolute divergence, we expected to identify the factors leading to heterogeneous genomic differentiation between the two species and to find outlier regions and genes that may be associated with adaptation to novel environments during the speciation process.

Materials and Methods

Population samples, sequencing and mapping

27 individuals of *P. alba* and 28 individuals of *P. adenopoda* were sampled (Figure 1a and Table S1). The former is widely distributed in Central Asia and Europe, and the natural populations of*P. alba* in China are only distributed in the Irtsh River basin, and the latter is distributed in the subtropical region of China. The two populations are far away and do not overlap (Figure 1a). The sampling distance is more than 200 meters. We extracted genomic DNA from leaves of each individual using a DNA extraction kit (Aidlab, Beijing, China). Paired-end (PE) reads with an insert size of 150bp were constructed according to the Illumina library preparation protocol and sequenced on an Illumina HiSeq 2000 platform (Illumina, San Diego, CA). The sequencing coverage target for all samples was about 25× (see Table S1). By using Trimmomatic (Lohse et al., 2012) to truncate and filter the raw data, that is, remove the adapter and remove the bases whose leading and trailing bases are less than 20, and retain the minimum read length of 36, and you will get clean data. Then using BWA-MEM in bwa-0.7.15 (Li unpublished data, https://arxiv.org/abs/1303.3997v2) to set the default threshold to map clean data to *P. trichocarpa*reference genome (v3.0) (Tuskan et al., 2006), which is used as the reference genome because of its high-quality assembly and annotation as well as the high genomic synteny between members of sections*Populus* (*P. alba* and *P. adenopoda*) and Tacamahaca (*P. trichocarpa*) (Liu, Wang & Zeng, 2019; Pakull, Groppe, Meyer, Markussen, & Fladung, 2009; Tuskan et al., 2006). Next, SortSam in Picard can directly generate sorted bam files from the sam file. We then performe PCR duplication

removal on bam files using MarkDuplicates from the Picard toolkit (http://broadinstitute.github.io/picard/).

Filtering sites

We used several filtering steps to rule out errors that could be caused by paralogous or repeated DNA sequences before using the variant and genotype calls. First, We randomly selected 500,000 sites to calculate the distribution map of reads after obtaining the coverage of reads per site with SAMtools (Figure S3), and we removed sites with a very low (<13 reads for each sample per species) or high (>100 reads for each sample per species) read coverage after examining the empirical distribution of reads. Second, since zero-mapped mass fractions were assigned to reads that could have been mapped to multiple genomic locations, we removed sites containing more than 20 such reads in all samples of each species. Third, we removed sites that overlapped with known repetitive elements identified using RepeatMasker (Chen, 2004; Tarailograovac & Chen, 2009). After these filtering steps, 48.6% of the sites in the genome were used for downstream analysis.

SNP and genotype calling

In this study, we used two complementary bioinformatics methods for SNP and genotype calling. First, the genetic statistics of populations that depended on the inferred site-frequency spectrum (SFS) were estimated directly from the genotype likelihoods rather than by calling the genotype (Nielsen, Korneliussen, Albrechtsen, Li, & Wang, 2012) in ANGSD (Korneliussen, Albrechtsen, & Nielsen, 2014). We considered only reads with a minimum mapping quality of 30 and bases with a minimum quality score of 20. For all the filtered sites in these two species, we defined the allele that was identical to that found in the *P. trichocarpa* reference genome as the ancestral allele. Based on the SAMtools (Li et al., 2009) genotype likelihood model for all sites, we used -doSaf to calculate the site allele frequency likelihood and then used -realSFS for maximum likelihood estimation of the expanded SFS using the expectation maximization (EM) algorithm (Kim et al., 2011). Then, we calculated several population genetic statistics based on the global SFS, such as Fay & Wu's H; the proportions of shared, private and fixed polymorphisms; and population structure parameters. Second, multi-sample SNP and genotype calls were performed using HaplotypeCaller and GenotypeGVCFs in GATK (Danecek et al., 2011) . We also ran a number of filtering steps to decrease false positives from SNP and genotype calls. For example, removing SNPs in regions that did not pass all previous filtering criteria, then removing SNPs with more than two alleles in both species. Furthermore, if the quality scores of genotypes (GQ) were less than 10, MQ <40.0, QD <2.0 and DP <8.0, they were designated as missing genotypes, and then the SNPs with more than two missing genotypes in each species were filtered out.

Population structure

The NGSadmix (Skotte et al., 2013) was part of the package ANGSD (Korneliussen et al., 2014) and it was used to infer population genetic structures by selecting only sites with less than 10% missing data. The program inferred individual ancestry based on genotype likelihoods and examines the genetic relationships between individuals. Based on the output file from ANGSD, it was possible to infer individual admixture proportions and estimate the frequency of different ancestral populations from genotype likelihoods. The SAMTools model (Li et al., 2009) was used to estimate the genotype likelihoods in ANGSD, then generated a Beagle file for a subset of the genome, which was identified as a variable using the likelihood ratio test (P -value $<10^{-6}$) (Kim et al., 2011). The number of genetic clusters K was predetermined and varied from 2 to 4 and the maximum iteration of the EM algorithm was set to 10,000.

Similarly, the sample allele frequency likelihoods generated in ANGSD were used to perform principal component analysis (PCA) to visualize inter-individual genetic relationships included possible sequencing errors and all uncertainties in the genotype calls (Fumagalli, Vieira, Linderoth, & Nielsen, 2014). The expected covariance matrix of each individual in the two species was calculated according to the genotypic posterior probability of all the filtered sites.

Demographic history

We used *fastSimcoal2* (ver 2.6.0.3-14.10.17) (Excoffier et al., 2013) to infer the demographic history associated with the speciation of*P. alba* and *P. adenopoda* based on the coalescent simulation method. A

two-dimensional joint SFS (2D-SFS) was constructed using the posterior probability from ngsTools for the frequency of the sample allele. We used 100,000 joint simulations to estimate the expected 2D-SFS and log probabilities for a set of demographic parameters in each model. The global maximum likelihood estimates for each model were obtained by 50 independent runs. We compared the models based on the maximum likelihood of 50 independent runs using the Akaike information criterion (AIC) and Akaike's evidence weights. The model with the largest Akaike's weight value as the best model. The CI of the best model was obtained from 100 parametric bootstrap samples, with 50 independent runs for each sample. Then we used the PSMC method (Li & Durbin, 2011) to estimate historical changes in the population size of the two species. We used beagle.08Jun17.d8b.jar in Beagle (Browning & Browning, 2007) to phase and impute all of the sites within each species before analysis. The scaled time and population size were converted to actual time and size using a 15-year generation time and 2.5 x 10^{-9} mutations per nucleotide per year (Koch, Haubold, & Mitchell-Olds, 2000).

Genome-wide patterns of differentiation

We divided the genome into 39,406 non-overlapping 10 Kbp windows for studying the pattern of genomic differentiation between the species. After all filtering steps detailed above (see Materials and Methods), the windows less than 1000 bases left were removed to include a window in the downstream analysis. We also removed windows where the number of variable sites were less than 10. F_{ST} was calculated using the software VCFtools (v0.1.13) (Danecek et al., 2011) to estimate the degree of genetic differentiation between species at each site.

transformed F_{ST} values to find outlier regions

VCFtools (v0.1.13) (Danecek et al., 2011) was used to calculate the relative measure of divergence F_{ST} and then we utilized the Origin 8 (OriginLab, Massachusetts, USA) to perform Z-transformation F_{ST} (Axelsson et al., 2013). We chose to set the thresholds at $Z(F_{ST}) > 1.96$ and $Z(F_{ST}) < -1.96$ as high and low differentiated regions because it corresponds to roughly 2.5% of the data. We compared two outlier windows (highly differentiated and lowly differentiated regions) with the rest of the genome by analysing additional population genetic statistics of the two species. First, we calculated the values of $\vartheta\pi$, Tajima's D (Tajima, 1989) and Fay & Wu's H (Fay & Wu, 2000) according to the sample allele frequency likelihood of the non-overlapping 10 Kbp windows in ANGSD (Korneliussen et al., 2014). Furthermore, we estimated and compared LD levels. We used VCFtools (Danecek et al., 2011) to calculate the correlation coefficients (r^2) between SNPs with a distance greater than 1 Kbp to evaluate the level of LD in each 10 Kbp window. Finally, we used the ngsStat (Fumagalli et al., 2014) program to calculate several other genetic differentiation parameters: (1) the proportion of fixed differences caused by the fixation of derived alleles in *P. alba* and *P. adenopoda* , using *P. trichocarpa* as an outgroup; (2) the proportion of shared polymorphisms among species at all segregating sites; (3) the absolute measure of divergence (d_{xy}) based on the posterior probability of the sample allele frequency for each locus and then averaged for each 10 Kbp non-overlapping window; and (4) relative node depth (RND) (Feder et al. 2005), calculated by dividing the d_{xy} between*P. alba* and *P. adenopoda* by the d_{xy}between the two poplars and *P. trichocarpa* . For this process, we used one-sided Wilcoxon ranked-sum tests to detect significant differences between values in outlier windows and the genome-wide mean of all population genetic statistics.

Gene Ontology (GO) term enrichment

To determine if any functional gene classes were over-representesd in the outlier regions with natural selection and intrinsic barriers, we performed functional enrichment analysis of GO terms using Fisher's exact test (https://www.omicshare.com/tools/Home/Soft/gogsea). This analysis excluded GO groups with fewer than two genes and further corrected the P -values from Fisher's exact test using the Benjamini-Hochberg false discovery rate (Benjamini & Hochberg, 1995). GO terms with a modified P -value <0.05 were considered significantly enriched.

Results

A total of 27 *P. alba* and 28 *P. adenopoda* individuals were selected for whole genome re-sequencing. After removing reads containing adaptors and then trimming and filtering, the sequenced reads from the two species were mapped to the *P. trichocarpa* reference genome (v3.0) (Tuskan et al., 2006) with a moderately high mapping rate (84.3% on average, Table S1). The mean coverage of reads that were uniquely mapped per site was 23.9 and 24.8 in *P. alba* and *P. adenopoda* , respectively (Table S1). VCFtools (Danecek et al., 2011)was used to perform a series of filtering steps on the VCF files prior to data analysis. A total of 22,525,740 and 21,217,984 high-quality SNPs were obtained from the 27 *P. alba* and 28 *P. adenopoda*samples after filtering criteria (see Materials and Methods) respectively.

Population structure and Demographic histories

We used NGSadmix (Skotte et al. 2013) in ANGSD (Korneliussen et al., 2014) to infer admixture proportions for the two species based on genotype likelihoods. The analyses were run with the number of genetic clusters (K) varying from 2 to 4. When K=2 (Figure 1b), all samples were divided into two species-specific populations. In the case of K = 3 and K=4 no further structures were found (Figure 1b). This result was verified by principal component analysis (PCA) that show a clear separation of the two species and little evidence for further subdivision within species (Figure 1c). The first principal component (PC1) explained 59.13% of total genetic variance (Tracy–Widom test,P =1.49e-15), while the second one was 27.91% (Tracy–Widom test, P = 3.48e-10). Of the total polymorphisms of these two species, the fixed difference between *P. alba* and *P. adenopoda* accounted for 7.8%, the shared polymorphism between them accounted for 12.9%, and the remaining polymorphic sites were private in the two species (Figure 1d).

We used *fastsimcoal2.6* (Excoffier, Dupanloup, Huerta-Sánchez, Sousa, & Foll, 2013), which is a continuous-time coalescence simulator assuming arbitrarily complex evolutionary scenarios (Excoffier & Foll, 2011), to infer the evolutionary history of *P. alba* and *P. adenopoda* . 25 different divergence models were assessed (Figure S1; Table S2). All models began with the subdivision of an ancestral population into two derived groups and then differed in the following ways: (I) divergence time, (II) the occurrence and duration of gene flow following population divergence, and (III) effective population size before and after differentiation. The most appropriate model was an isolation-migration-isolation model, where the separation of the two species involved three phases (Figure 2a). The exact parameter estimates were provided for this model, including the divergence time, levels of gene flow and effective population sizes together with their associated 95% confidence intervals (CIs). Under the best fitting model (Table S10), *P. alba* and *P. adenopoda* diverged from an ancestral population approximately 9.54 million years ago (Mya) (bootstrap range [BR]: 5-10 Mya), and this differentiation initially occurred in allopatry as there was no evidence for gene flow during the early stages of differentiation. From 1.15 Mya to 0.34 Mya there was evidence for secondary contact and weak asymmetric gene flow between the two incipient species, after which the two species returned to an allopatric state without gene flow. The population size of the two isolated species showed stepwise changes, with *P. alba* experiencing a population decline during in the final stages. The migration rate of *P. alba*to *P. adenopoda* per generation was 1.28×10^{-6}(7.49×10^{-7}-2.33×10^{-6}), and that from*P. adenopoda* to *P. alba* was 9.69×10^{-6}(7.45×10^{-6}-1.40×10^{-5}). This result was not surprising because the geographical distance between the two species was large and their distributions were very scattered. The estimates of contemporary effective population size (*Ne*) of*P. alba* ($N_{e\text{-}P.\ alba}$) and *P. adenopoda* ($N_{e\text{-}P.\ adenopoda}$) were 19,551 (BR:18624.5-23509.0) and 31,333 (BR:30851.5-34287.8), respectively.

We also used the pairwise sequentially Markovian coalescent (PSMC) model (Li & Durbin, 2011) to reconstruct historical effective population size changes in the two species with a focus on diploid consensus sequences. Estimates of the effective population sizes of *P. alba* and*P. adenopoda* based on the PSMC model at the beginning of population divergence approximately 10 Mya were similar to those of their ancestral population obtained with *fastsimcoal2.6* . The two species experienced similar degrees of population expansion following divergence, possibly indicating successful adaptation to new niches (Figure 2b). The population size of *P. alba* then began to decline but rebounded with a second population expansion from 1.15-0.34 Mya. In contrast, the population of *P. adenopoda* has undergone considerable and continuous decline (Figure 2b).

Genome-wide patterns of differentiation and identification of outlier regions

We used F_{ST}, a relative measure of divergence (Charlesworth, 1998), to explore inter-specific genomic differentiation patterns in non-overlapping 10-kilobase (Kbp) windows (Figure 3). As shown in Figure 3, genetic differentiation varied widely across the genome with an average F_{ST} value of 0.542 and many windows showed substantially higher genetic differentiation between species. We identified 415 and 983 outlier windows which showed significantly ($P < 0.05$) high and lowF_{ST} values, respectively (Figure 4a). We found that both highly differentiated and lowly differentiated regions were randomly distributed throughout the genome (Figure 3) after checking the genome distribution, physical size, and overlaps of these outlier windows, and the size of these regions seemed to be quite small, and most of them occurred on a physical scale less than 10Kbp (Figure S2). .

Since F_{ST} is based on a comparison of intra- and inter-population diversity, any reduction in the former will increaseF_{ST} (Nachman Michael & Payseur Bret, 2012). We used three other methods to quantify and compare inter-specific genetic differentiation between the two classes of outlier windows and the rest of the genome. The first was d_{xy}, an absolute measure of divergence, which is the average number of pairwise differences between the two populations (Cruickshank & Hahn, 2014). The second was relative node depth (RND), which is d_{xy} divided by d_{xy} relative an outgroup species, in this case *P. trichocarpa* . This ratio corrects for possible variation in the mutation rate among genomic regions. The third was the proportion of ancestral polymorphisms shared by the two species. We conducted a Mann-Whitney U test and found that d_{xy} and RND were significantly higher (Figure 4b; Table S3) and that the proportion of shared ancestral polymorphisms was significantly lower in the highly differentiated regions of the two species compared to the genomic background (Figure 4b; Table S3). These regions also contained a greater proportion of fixed differences, caused by the fixation of derived alleles in both species relative to the genome-wide average (Figure 4c; Table S3). Both species showed positive Tajima's D values and negative Fay & Wu's H values on average. The values of Tajima's D and Fay & Wu's H test showed no significant difference when comparing highly differentiated regions to the genomic background in the two species. However, when we divided the genome into three subsets based on d_{xy} values (highly differentiated regions, top 1000 windows; middle and lowly differentiated regions, bottom 1000 windows), the d_{xy} values in the high and low parts showed significantly lower and higher value than the background in both the two species, respectively (Figure 4c; Table S3).

In contrast to the pattern found in highly differentiated regions, polymorphism in the lowly differentiated regions (based onF_{ST}) was higher. The value of Tajima's D test and Fay & Wu's H test were significantly higher for lowly differentiated regions than background regions. Compared with the genomic background, the proportion of polymorphisms shared between species was higher in lowly differentiated regions and the proportion of fixed differences was extremely low (Figure 4b, c; Table S3). We also investigated the correlation between relative and absolute divergence between *P. alba* and *P. adenopoda* and we found a positive correlation between them (Spearman's $\rho = 0.623$, P -value < 0.001).

Gene Ontology (GO) functional analysis of regions with high or lowF_{ST} values was performed using the *P. trichocarpa* genome annotation. We assessed whether specific GO terms were significantly overexpressed using gene ontology (GO) assignments of these candidate genes. 293 and 599 genes were identified in the outlier windows showing high differentiation or low genetic differentiation, respectively (Table S4 and S5). We identified 41 GO categories showing significant enrichment, with the most prominent group (7 terms) associated with the term "reproduction " ($P < 0.05$) (Table S6). Genes related to reproduction and adaptation in highly differentiated regions between *P. alba* and *P. adenopoda*are listed on Tables S7 and S8.

Discussion

1 Demography

We took advantage of genome-wide patterns of genetic diversity and differentiation to explore the evolutionary history of the two closely related species *P. alba* and *P. adenopoda* to study how the genomic heterogeneity in genetic diversity and differentiation is affected by various evolutionary forces.

Demographic models showed that *P. alba* and *P. adenopoda*diverged approximately 9.5 Mya (5 Mya-10

Mya), during the late Miocene. The climate of Asia has been strongly linked to the surface uplift of the Tibetan Plateau (Li et al., 2014; Li et al., 2018; Li & Chetelat, 2014; Zhang, Fengquan, & Jianmin, 2000; Zhisheng, Kutzbach, Prell, & Porter, 2001). The Asian winter monsoon has appeared in some form since about 13.1 Mya (Fan, Song, Dettman, Fang, & Xu, 2006). Rising vapour from the Indian Ocean cools, condenses and precipitates before it crosses the Tibetan Plateau, which results in drier air advancing to the northern plateau (Fang et al., 2019; Favre et al., 2015; Ge, 2006; Li et al., 2018). The combination of mountains and monsoons makes the climate in north-western China drier and the aridity reached its maximum approximately 9.6-8.0 Mya (Chen et al., 2019; Fan, Dettman, Song, Fang, & Garzione, 2007; Mertz-Kraus, Brachert, Jochum, Reuter, & Stoll, 2009; Zhisheng et al., 2001). From 5.3 Mya the climate has gradually became drier and/or cooler (Fan et al., 2007). Furthermore, the uplift of the Pamir Plateau and the central plateau of Anatolia, coupled with the collision between the Pamir Plateau and the Tianshan Mountains, further blocked the transport of water vapour from the winter westerly belt to the east ˜5 Mya (Li et al., 2018; Meijers et al., 2018; Shen et al., 2018; Thompson, Burbank, Li, Chen, & Bookhagen, 2015). Approximately 2-3 Mya, the climate was still very dry and cold because north-western China received only the water vapour in the westerly belt at that time (Caves et al., 2015; Li et al., 2017; Zhuang, Brandon, Pagani, & Krishnan, 2014). Approximately 1.80-1.10 Mya, the paleotemperature was approximately 7.5-13 °C (Meijers et al., 2018), which was not sufficient to cause contact between *P. alba* and*P. adenopoda* . This was evident in demographic models due to the lack of gene flow in the early stages of divergence of the two species (from 9.54 Mya to 1.15 Mya).

The model detects a small amount of asymmetric gene flow between the two species ˜ 1.15-0.34 Mya. This time period was close to the appearance of the Mindel-Riss Interglacial (Jing & Liu, 1999; Wright, 1926), which also included several cold and warm periods (Cheng, 1996; Lindner, 1981; Peiyuan, 1989). During this period, the paleotemperature increased significantly and the annual average temperature reached 28 °C (Cheng, 1996; Peiyuan, 1989). A form of subtropical forest or forest grassland environment was present during this time period and the climate was much warmer and wetter than in modern times. Based on the results of the PSMC simulation, the population of *P. alba* experienced exponential size changes at this stage. From ˜0.34 Mya to the present, gene flow between *P. alba* and *P. adenopoda* has been interrupted again and this period corresponds to the Riss Glacial (Jing & Liu, 1999; Kukla, 2005).

2 Heterogeneous differentiation across the genome

Relative divergence measures, such as F_{ST}, are based on a comparison of within-population diversity to between-population diversity (Charlesworth, 1998; Cruickshank & Hahn, 2014). Absolute divergence is usually measured as d_{xy}, the average number of pairwise differences between alleles sampled from two populations (Nachman & Payseur, 2012), and variation in d_{xy} (Ma et al., 2018) can be caused by variation in levels of ancestral polymorphism or substitution rates (Cruickshank & Hahn, 2014). A more comprehensive understanding of heterogeneous genomic differentiation requires the use of both relative (influenced by within-population variation) and absolute (not influenced by within-population variation) measures of sequence divergence (Noor & Bennett, 2009; Ravinet et al., 2017; Wolf & Ellegren, 2016). Any type of selection that reduces levels of linked neutral diversity and reduced intra-population/species diversity will lead to increased estimates of relative divergence (such as F_{ST}) (Charlesworth, 1998; Charlesworth, Nordborg, & Charlesworth, 1997; Cruickshank & Hahn, 2014). However, the increased F_{ST} seen in such regions does not always indicate a reduction in actual gene flow. In other words, there is a redistribution of alleles in these regions. Therefore, the value of d_{xy} in these regions does not necessarily increase. In contrast, if intrinsic barriers are the main driving force of heterogeneous differentiation across the genome, both absolute and relative divergence measures should be high because the redistribution of alleles in these regions is prevented while the remainder of the genome is homogenized by gene flow.

Most studies have implicated natural selection as the primary cause for genomic heterogeneous differentiation and even for speciation (Burri, Nater, Kawakami, Mugal, & Ellegren, 2015; Cruickshank & Hahn, 2014; Nosil, Vines, & Funk, 2005). Selection affects population differentiation by reducing polymorphism levels at sites under selection (Seehausen et al., 2014). The living environment of *P. alba* and*P. adenopoda* is very

different. Natural selection is inevitable in the early stage of speciation, but neither Tajima's D nor Fay & Wu's H has detected a difference in either of the two species. This may because the differentiation time is long enough for them to reach a new balance. However, the nucleotide diversity ($\vartheta\pi$) in highly differentiated regions is significantly reduced, proportion of shared polymorphisms is also decreased, while proportion of fixed difference is increased, which shows evidence that natural selection works. The size of differentiated regions is related to the intensity of selection and the stage of species divergence and the highly differentiated region may represent genes associated with reproductive isolation (Wu, 2001).

However, intrinsic barriers, such as chromosomal rearrangements and genetic incompatibilities can also lead to heterogeneous genomic differentiation (Faria & Navarro, 2010; Navarro, Betran, Barbadilla, & Ruiz, 1997; Kulmuni et al., 2020; Rieseberg, 2001). Intrinsic barriers inhibits allele exchange in such regions and in linked surrounding regions eventually leads to elevated genetic differentiation. If such highly differentiated regions contain genes associated with reproductive isolation, speciation may eventually occur (Ravinet et al., 2017; Wu, 2001). For example, the inversions of*D. pseudoobscura* and *D. persimilis* may have contributed to their divergence (Noor et al., 2001). Rieseberg (2001) suggested that chromosomal rearrangements can reduce gene flow between species by the effects of individual sterility or other isolating factors and potentially contribute to speciation (Rieseberg, 2001). Gene incompatibilities between two species of budding yeast has been found, the complete sterility of the F2 generation was a reproductive-isolation mechanism of the hybrid due to gene incompatibilities (Louis, 2009).

In this study, we Z-transformed the F_{ST}distributions to identify regions showing high, medium and low genetic differentiation across the genome. We evaluated multiple population genetic parameters to understand the mechanisms generating the heterogeneous genomic differentiation seen between the two species. Overall the values of Tajima's D in the two species were positive and Fay & Wu's H were negative, suggesting that recent population contractions have taken place (Fay & Wu, 2000; Holliday, Yuen, Ritland, & Aitken, 2010; Tajima, 1989), consistent with the PSMC results (Figure 2b).

The high correlation between relative and absolute divergence highlights the significant effects of intrinsic barriers in generating the heterogeneous differentiation landscape between the two species. Furthermore, when we divided the whole genome into three parts by d_{xy} values we found that the high part showed significantly lower correlation coefficients (r^2) value than the genomic background, further supporting the conclusion that intrinsic barriers play an important role in shaping the genome-wide divergence of the two species (Ellegren et al., 2012; Geraldes et al., 2011). That may because when the foreign DNA fragments were introduced by introgression, the observed value of correlation coefficients (r^2) increased in related regions, but not for the regions of intrinsic barriers because the foreign DNA fragment cannot be integrated into the genome. Compared to the genomic background, highly differentiated regions also contained an excess of derived fixed differences and a smaller proportion of shared polymorphisms.

3 Reproductive isolation and speciation

Postzygotic isolation can be caused by intrinsic and extrinsic factors, leading to phenomena such as hybrid inviability and sterility and ecological and behavioural sterility, respectively (Widmer et al., 2008). The natural hybrid of *P. alba* and *P. adenopoda* is *P. tomentosa* , which is a widely distributed species in China (Li et al., 1997; Wang et al., 2014) with strong adaptability and poor fertility (Wang et al., 2019). A recent study using more than 200 individuals of *P. tomentosa* found no progeny resulting from backcrossing or selfing (Wang et al., 2019). Furthermore, some studies have found that *P. tomentosa* exhibits unequal numbers of univalents and a decreased degree of synapsis during meiosis (Kang, 2001). A larger percentage of univalents occurs at diakinesis and metaphase I and at anaphase I and telophase I lagging chromosomes are frequently observed (Kang, Zhu, & Zhang, 1999). All of the above results confirmed that *P. alba* and *P. adenopoda* are subject to postzygotic isolation and imply that intrinsic barriers have played an important role in establishing and maintaining reproductive isolation between the two species.

There are several reasons for postzygotic intrinsic barriers, such as chromosomal rearrangements or gene incompatibilities (King & Templeton, 1993; Noor et al., 2001). Structural changes in the genome, including

deletions, insertions, duplications, inversions and translocations, alter the genome organization of individuals and may also contribute to reproductive isolation (Wolf & Ellegren, 2016). Genomic incompatibilities might result from chromosomal rearrangements, which lead to mis-segregation during meiosis in hybrids, or from epistatic interactions that act as loss-of-function alleles in hybrid backgrounds (Lynch & Force, 2000). These factors are often accompanied by a decrease in hybrid inviability, sterility and/or even mortality (Barbash, Siino, Tarone, & Roote, 2003; Allen Orr, 1989; Orr & Coyne, 1989; Schumer et al., 2014). To distinguish to what extent chromosomal rearrangements and gene incompatibilities contribute to postzygotic isolation, high-quality genomes for the two species must be assembled in the future.

The *P. alba* populations we collected are distributed in cold temperate regions (Stölting et al., 2015). The temperature of seed germination and branch germination in early spring is lower than the sprouting season of *P. adenopoda* , and the living environment of*P. alba* is relatively dry. While *P. adenopoda* lives in the subtropical regions of China (Fan et al., 2018) and are generally distributed along the river. The environmental conditions such as temperature, light, and air humidity when they bloom are also significantly different. In order to identify gene ontology terms that were enriched in these gene sets, we performed gene ontology (GO) analysis in highly differentiated regions. We identified 41 categories that showing significant enrichment and the most prominent group (7 terms) was associated with the word "reproduction". The seven genes belonging to this category include*CDC* 2,*CUL* 1,*emb* 2746,*AFO* ,*HK* 2,*TFL* 2 and *EDA* 8, several of which are known to affect cell division, pollen rejection pathways, embryogenesis, flowers/spike into seedlings and reproductive habits, plant germ cell development during meiosis, transition from photoperiod regulation to reproductive growth and the early development of plant embryos and pollen tubes (Table S9). These findings support the hypothesis that highly differentiated regions contain an excess of reproduction-related genes. The gene ontology analysis also identified three genes involved in seed germination and a gene for drought resistance. *RCAR* 3 and *GCR* 1 regulated seed germination and early seedling development while *CPK* 23 is involved in drought- and salt stress-induced calcium signaling cascades (Colucci, Apone, Alyeshmerni, Chalmers, & Chrispeels, 2002; Lim, Luan, & Lee, 2014; Ma & Wu, 2007).

Conclusion

Our study found that the divergence time of *P. alba* and *P. adenopoda* was about 9.54 Mya, the uplift of the Tibetan Plateau and the Tianshan mountains led to drastic climate change in the northwest of China, which triggered the genetic differentiation of the two species. Multiple population genetic parameters were assessed and indicated that natural selection and intrinsic barriers together cause the heterogeneous genomic differentiation. We also identified some genes related to reproductive isolation and adaptation, which may therefore be signatures of differential adaptation and growing environments in the two species.

Acknowledgments

We thank Dong Wang, Aiguo Duan, Wenhao Shao, Sirong Yi, Zhonghui Shi, Bin Zhang for some sample collection. We thank Prof. Jing Wang in Sichuan University, Associate Professor. Fumin Zhang, Dr. Zhe Cai and Chunyan Jing in the State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing, China, for their valuable suggestion on data analysis. This work was supported by China National GeneBank (CNGB). Financial support for this research was provided by the National Natural Science Foundation of China (No. 31770702).

Figures and Tables

Figure 1. Sample collection and population structure analysis of 27 *P. alba* and 28 *P. adenopoda*. (a) 28 *P. adenopoda* individuals (red) were collected in Anhui, Chongqing, Hunan and Tianmushan of China and 27 *P. alba* individuals (blue) were collected from the Altay area, Xinjiang. (b) Population genetic structure in the samples based on an analysis using NGSadmix in ANGSD based on genotype likelihoods. The y-axis represents the number of clusters, and the x-axis shows the name of each individual. (c) Results from a principal component analysis (PCA) on the genetic covariance matrix for all individuals of *P. alba* (blue circles) and *P. adenopoda* (red circles). (d) The pie chart shows the proportion of fixed, shared and private polymorphisms in the two species.

Figure 2. Demographic history of *P. alba* and *P. adenopoda*. (a) The most appropriate model inferred by*fastsimcoal2.6* . The ancestral population is colored gray while*P. alba* and the *P. adenopoda* are colored red and blue, respectively. The widths represents the relative effective population sizes (N e). Double-headed arrows represents the per generation gene flow between *P. alba* and *P. adenopoda* . All inferred demographic parameters are presented in Table 1. (b) Pairwise Sequentially Markovian coalescent (PSMC) estimates of the effective population size (N e) changes for *P. alba* (bold blue line) and *P. adenopoda* (bold red line) based on the inference from four phased haplotypes in each species, and performed 100 bootstrap replicates for four individuals (light-colored line respectively). Assume that the neutral mutation rate of each generation (μ) = 3.75×10^{-8}, and the generation time (g) = 15 years, to calculate the time scale on the x-axis.

Figure 3. Genome-wide divergence. Chromosomal distribution of pairwise genetic divergence (F_{ST}) between*P. alba* and *P. adenopoda* in 10 Kb sliding windows. Windows with significantly high or low genetic differentiation are shown in red and green, respectively.

Figure 4. Identifying candidate outlier windows. (a) Distribution of genetic differentiation (F_{ST}) between *P. alba* and *P. adenopoda* . The dotted line represents the threshold based on Z-scores to determine windows with high (red) or low (green) genetic differentiation ($P < 0.05$). (b) d_{xy} (absolute measure of divergence), relative node depth (RND) and interspecific shared polymorphisms were compared between the genomic background and regions with significant high and low genetic differentiation. (c) Comparison of population genetic statistics between regions with significantly high (red) or low (green) genetic differentiation and the genomic background (blue) in*P. alba* and *P. adenopoda* . Statistics include Tajima's D, Fay &Wu's H, the proportion of fixed differences caused by derived alleles, nucleotide diversity ($\vartheta\pi$), and values of $d_{xy}$and F_{ST} calculated using ther^2, respectively. Asterisks indicate significant differences between outliers and background genomic regions based on Mann-Whitney U tests (* P -value < 0.05; **P -value $<$ 1e-4; ***P -value <2.2e-16).

Table 1. Demographic parameter estimates for the best-supported model in Figure 2a.

Point estimation 95% CI[a]	Point estimation 95% CI[a]	Point estimation 95% CI[a]	Point estimation 95% CI[a]
Parameters		**Lower bound**	**Upper bound**
$N_{e\text{-}ANC}$	11506	3170.675	80404.47
$N_{e\text{-}POP0}$	19551	18624.45	23508.97
$N_{e\text{-}POP1}$	31333	30851.53	34287.77
MIG01	1.28×10^{-6}	7.49×10^{-7}	2.33×10^{-6}
MIG10	9.59×10^{-6}	7.45×10^{-6}	1.40×10^{-5}
TDIV01	22786	21118.95	27847.42
TDIV02	76468	50890.35	98329.9
TDIV03	635956	336770.3	668025.3

The parameters are defined in Figure 2a. $N_{e\text{-}POP0}$,$N_{e\text{-}POP1}$ and $N_{e\text{-}ANC}$represent the effective population size of *P. alba* , *P. adenopoda* and their ancestral population respectively. MIG01 represents the migration rate of each generation from *P. alba* to *P. adenopoda* , and that from *P. adenopoda* to *P. alba* is MIG10. TDIV01, TDIV02 and TDIV03 represents the estimated divergence time between the two species obtained from *fastsimcoal2.6* .

aParameter bootstrap estimation obtained by performing parameter estimation from 100 simulated data sets based on the total maximum composite likelihood estimates displayed in the point estimation column, per likelihood is estimated from 100,000 simulations.

References

Axelsson, E., Ratnakumar, A., Arendt, M.-L., Maqbool, K., Webster, M. T., Perloski, M., . . . Lindblad-

Toh, K. (2013). The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature, 495* (7441), 360-364. doi:10.1038/nature11837

Baack, E., Melo, M. C., Rieseberg, L. H., & Ortiz-Barrientos, D. (2015). The origins of reproductive isolation in plants. *New Phytologist, 207* (4), 968-984. doi:10.1111/nph.13424

Barbash, D. A., Siino, D. F., Tarone, A. M., & Roote, J. (2003). A rapidly evolving MYB-related protein causes species isolation in Drosophila. *Proceedings of the National Academy of Sciences, 100* (9), 5302. doi:10.1073/pnas.0836927100

Barton, N. H. (2006). Evolutionary Biology: How Did the Human Species Form? *Current Biology, 16* (16), R647-R650. https://doi.org/10.1016/j.cub.2006.07.032

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological), 57* (1), 289-300. doi:10.1111/j.2517-6161.1995.tb02031.x

Booker, T. R., Yeaman, S., & Whitlock, M. C. (2020). Variation in recombination rate affects detection of outliers in genome scans under neutrality. *Molecular Ecology, 29* (22), 4274-4279. https://doi.org/10.1111/mec.15501

Burri, R., Nater, A., Kawakami, T., Mugal, C. F., & Ellegren, H. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. *Genome research ,25* (11), 1656-1665. doi: 10.1101/gr.196485.115

Caves, J. K., Winnick, M. J., Graham, S. A., Sjostrom, D. J., Mulch, A., & Chamberlain, C. P. (2015). Role of the westerlies in Central Asia climate over the Cenozoic. *Earth and Planetary Science Letters, 428* , 33-43. https://doi.org/10.1016/j.epsl.2015.07.023

Charlesworth, B. (1998). Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution, 15* (5), 538-543. doi:10.1093/oxfordjournals.molbev.a025953

Charlesworth, B., Nordborg, M., & Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research, 70* (2), 155-174. doi:10.1017/S0016672397002954

Chen, C., Bai, Y., Fang, X., Guo, H., Meng, Q., Zhang, W., . . . Murodov, A. (2019). A Late Miocene Terrestrial Temperature History for the Northeastern Tibetan Plateau's Period of Tectonic Expansion. *Geophysical Research Letters, 46* (14), 8375-8386. doi:10.1029/2019GL082805

Chen, N. (2004). Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. *Current protocols in human genetics, 25* (1).

Cheng, Z. (1996). Comparative Study on the Sedimentary Environment of the Fourth Period in Lushan, Huangshan and Tianmu Mountain Areas. *Geographic science, 016* (1), 37-45.

Colucci, G., Apone, F., Alyeshmerni, N., Chalmers, D., & Chrispeels, M. J. (2002). GCR1, the putative Arabidopsis G protein-coupled receptor gene is cell cycle-regulated, and its overexpression abolishes seed dormancy and shortens time to flowering. *Proceedings of the National Academy of Sciences, 99* (7), 4736-4741. doi:10.1073/pnas.072087699

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology, 23* (13), 3133-3157.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Sherry, S. T. (2011). The variant call format and VCFtools. *Bioinformatics, 27* (15), 2156-2158.

Depristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J., Hartl, C., . . . Hanna, M. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature*

Genetics, 43 (5), 491-498.

Dickmann, D., & Kuzovkina, Y. (2014). Poplars and Willows in the World. In (pp. 8-91).

Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., . . . Wolf, J. B. W. (2012). The genomic landscape of species divergence in Ficedula flycatchers. *Nature, 491* , 756. doi:10.1038/nature11584

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust Demographic Inference from Genomic and SNP Data.*PLOS Genetics, 9* (10), e1003905. doi:10.1371/journal.pgen.1003905

Excoffier, L., & Foll, M. (2011). fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics, 27* (9), 1332-1334. doi:10.1093/bioinformatics/btr124

Fan, L., Zheng, H., Milne, R. I., Zhang, L., & Mao, K. (2018). Strong population bottleneck and repeated demographic expansions of Populus adenopoda (Salicaceae) in subtropical China. *Annals of Botany, 121* (4), 665-679. doi:10.1093/aob/mcx198

Fan, M., Dettman, D. L., Song, C., Fang, X., & Garzione, C. N. (2007). Climatic variation in the Linxia basin, NE Tibetan Plateau, from 13.1 to 4.3 Ma: The stable isotope record. *Palaeogeography, Palaeoclimatology, Palaeoecology, 247* (3), 313-328. https://doi.org/10.1016/j.palaeo.2006.11.001

Fan, M., Song, C., Dettman, D. L., Fang, X., & Xu, X. (2006). Intensification of the Asian winter monsoon after 7.4 Ma: Grain-size evidence from the Linxia Basin, northeastern Tibetan Plateau, 13.1 Ma to 4.3 Ma. *Earth and Planetary Science Letters, 248* (1), 186-197. https://doi.org/10.1016/j.epsl.2006.05.025

Fang, X., Fang, Y., Zan, J., Zhang, W., Song, C., Appel, E., . . . Zhang, T. (2019). Cenozoic magnetostratigraphy of the Xining Basin, NE Tibetan Plateau, and its constraints on paleontological, sedimentological and tectonomorphological evolution. *Earth-Science Reviews, 190* , 460-485. https://doi.org/10.1016/j.earscirev.2019.01.021

Faria, R., & Navarro, A. (2010). Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends in Ecology and Evolution, 25* (11), 660-669.

Favre, A., Päckert, M., Pauls, S. U., Jähnig, S. C., Uhl, D., Michalak, I., & Muellner-Riehl, A. N. (2015). The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas.*Biological Reviews, 90* (1), 236-253. doi:10.1111/brv.12107

Fay, J. C., & Wu, C. (2000). Hitchhiking Under Positive Darwinian Selection. *Genetics, 155* (3), 1405-1413.

Feder, J. L., & Nosil, P. (2009). Chromosomal inversions and species differences: when are genes affecting adaptive divergence and reproductive isolation expected to reside within inversions?*Evolution, 63* (12), 3061-3075. doi:10.1111/j.1558-5646.2009.00786.x

Fumagalli, M., Vieira, F. G., Linderoth, T., & Nielsen, R. (2014). ngsTools: methods for population genetics analyses from next-generation sequencing data. *Bioinformatics, 30* (10), 1486-1487. doi:10.1093/bioinformatics/btu041

Funk, W. C., Lovich, R. E., Hohenlohe, P. A., Hofman, C. A., Morrison, S. A., Sillett, T. S., . . . Andelt, W. F. (2016). Adaptive divergence despite strong genetic drift: genomic analysis of the evolutionary mechanisms causing genetic differentiation in the island fox (Urocyon littoralis). *Molecular Ecology, 25* (10), 2176-2194. doi:10.1111/mec.13605

Ge, X. (2006). Multi-stage uplifts of the Qinghai-Tibet plateau and their environmental effects. *Earth Sci. Frontiers, 13* (6), 118-130.

Geraldes, A., Basset, P., Smith, K. L., & Nachman, M. W. (2011). Higher differentiation among subspecies of the house mouse (Mus musculus) in genomic regions with low recombination. *Molecular Ecology, 20* (22), 4722-4736. doi:10.1111/j.1365-294X.2011.05285.x

Guo, X., Chen, F., Gao, F., Li, L., Liu, K., You, L., . . . Xu, X. (2020). CNSA: a data repository for archiving omics data.*bioRxiv* , 2020.2004.2007.030833. doi:10.1101/2020.04.07.030833

Holliday, J. A., Yuen, M., Ritland, K., & Aitken, S. N. (2010). Postglacial history of a widespread conifer produces inverse clines in selective neutrality tests. *Molecular Ecology, 19* (18), 3857-3864.

Hou, J., Friedrich, A., de Montigny, J., & Schacherer, J. (2014). Chromosomal Rearrangements as a Major Mechanism in the Onset of Reproductive Isolation in Saccharomyces cerevisiae. *Current Biology, 24* (10), 1153-1159. https://doi.org/10.1016/j.cub.2014.03.063

Jacobs, A., Hughes, M. R., Robinson, P. C., Adams, C. E., & Elmer, K. R. (2018). The Genetic Architecture Underlying the Evolution of a Rare Piscivorous Life History Form in Brown Trout after Secondary Contact and Strong Introgression. *Genes, 9* (6), 280.

Jakobsson, M., Edge, M. D., & Rosenberg, N. A. (2013). The Relationship Between FST and the Frequency of the Most Frequent Allele.*Genetics, 193* (2), 515. doi:10.1534/genetics.112.144758

Jing, C., & Liu, H. (1999). On the glacial and interglacial stages in Quaternary of China. *J. Cheng Du Univ. Tech, 26* , 97-100.

Kang, X. (2001). Study on mechanism of pollen abortion in chinese white poplar (*Populus* tomentosa carr.). *Scientia Silvae Sinicae* .

Kang, X., Zhu, Z., & Zhang, Z. (1999). Cytogenetic studies on the origin of Chinese white poplar. *Journal of Beijing Forestry University, 21* (6), 6-10.

Kim, S. Y., Lohmueller, K. E., Albrechtsen, A., Li, Y., Korneliussen, T., Tian, G., . . . Nielsen, R. (2011). Estimation of allele frequency and association mapping using next-generation sequencing data. BMC Bioinformatics, 12(1), 231. doi:10.1186/1471-2105-12-231

King, M., & Templeton, A. R. (1993). Species Evolution: The Role of Chromosome Change. *44* (4), 191-202.

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics, 15* (1), 356. doi:10.1186/s12859-014-0356-4

Kulmuni, J., Nouhaud, P., Pluckrose, L., Satokangas, I., Dhaygude, K., & Butlin, R. K. (2020). Instability of natural selection at candidate barrier loci underlying speciation in wood ants. *Molecular Ecology* , *29* (20), 3988-3999. https://doi.org/10.1111/mec.15606

Kukla, G. (2005). Saalian supercycle, Mindel/Riss interglacial and Milankovitch's dating. *Quaternary Science Reviews, 24* (14), 1573-1583. https://doi.org/10.1016/j.quascirev.2004.08.023

Levin, D. A. (1978). The Origin of Isolating Mechanisms in Flowering Plants. In M. K. Hecht, W. C. Steere, & B. Wallace (Eds.),*Evolutionary Biology* (pp. 185-317). Boston, MA: Springer US.

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature, 475* (7357), 493-496. doi:10.1038/nature10231

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics, 25* (16), 2078-2079. doi:10.1093/bioinformatics/btp352

Li, J., Fang, X., Song, C., Pan, B., Ma, Y., & Yan, M. (2014). Late Miocene–Quaternary rapid stepwise uplift of the NE Tibetan Plateau and its effects on climatic and environmental changes. *Quaternary Research, 81* (3), 400-423. doi:10.1016/j.yqres.2014.01.002

Li, J. X., Yue, L. P., Roberts, A. P., Hirt, A. M., Pan, F., Guo, L., . . . Liu, Q. S. (2018). Global cooling and enhanced Eocene Asian mid-latitude interior aridity. *Nature Communications, 9* (1), 3026. doi:10.1038/s41467-018-05415-x

Li, K., Huang, M., & Wang, M. (1997). Study on origin of *Populus tomemtosa* Carr. *Acta Phytotaxonomica Sinica, 35* (1), 24-31.

Li, W., & Chetelat, R. T. (2014). The Role of a Pollen-Expressed Cullin1 Protein in Gametophytic Self-Incompatibility in Solanum. *Genetics, 196* (2), 439-442. https://doi.org/10.1534/genetics.113.158279

Li, X., Hao, Q., Wei, M., Andreev, A. A., Wang, J., Tian, Y., . . . Shi, W. (2017). Phased uplift of the northeastern Tibetan Plateau inferred from a pollen record from Yinchuan Basin, northwestern China. *Scientific Reports, 7* (1), 18023. doi:10.1038/s41598-017-16915-z

Lim, C. W., Luan, S., & Lee, S. C. (2014). A Prominent Role for RCAR3-Mediated ABA Signaling in Response to Pseudomonas syringae pv. tomato DC3000 Infection in Arabidopsis. *Plant and Cell Physiology, 55* (10), 1691-1703. doi:10.1093/pcp/pcu100

Lindner, L. (1981). Organogenic deposits of the Mazovian Interglacial (Mindel II /Riss I) in the middle Vistula basin, compared to coeval European localities. *Acta Geologica Polonica, 31* , 111-126.

Liu, Y.-J., Wang, X.-R., & Zeng, Q.-Y. (2019). De novo assembly of white poplar genome and genetic diversity of white poplar population in Irtysh River basin in China. *Science China Life Sciences, 62* (5), 609-618.

Louis, E. J. (2009). Origins of reproductive isolation. *Nature ,457* (7229), 549-550. https://doi:10.1038/457549a

Lynch, M., & Force, A. G. (2000). The Origin of Interspecific Genomic Incompatibility via Gene Duplication. *The American Naturalist, 156* (6), 590-605. doi:10.1086/316992

Ma, S.-Y., & Wu, W.-H. (2007). AtCPK23 functions in Arabidopsis responses to drought and salt stresses. *Plant Molecular Biology, 65* (4), 511-518. doi:10.1007/s11103-007-9187-2

Ma, T., Wang, K., Hu, Q., Xi, Z., Wan, D., Wang, Q., . . . Liu, J. (2018). Ancient polymorphisms and divergence hitchhiking contribute to genomic islands of divergence within a poplar species complex. *Proceedings of the National Academy of Sciences, 115* (2), E236. doi:10.1073/pnas.1713288114

MacNair, M. R., & Christie, P. (1983). Reproductive isolation as a pleiotropic effect of copper tolerance in Mimulus guttatus? *Heredity, 50* (3), 295-302. doi:10.1038/hdy.1983.31

Meijers, M. J. M., Brocard, G. Y., Cosca, M. A., Lüdecke, T., Teyssier, C., Whitney, D. L., & Mulch, A. (2018). Rapid late Miocene surface uplift of the Central Anatolian Plateau margin. *Earth and Planetary Science Letters, 497* , 29-41. https://doi.org/10.1016/j.epsl.2018.05.040

Mertz-Kraus, R., Brachert, T. C., Jochum, K. P., Reuter, M., & Stoll, B. (2009). LA-ICP-MS analyses on coral growth increments reveal heavy winter rain in the Eastern Mediterranean at 9 Ma. *Palaeogeography, Palaeoclimatology, Palaeoecology, 273* (1), 25-40. https://doi.org/10.1016/j.palaeo.2008.11.015

Nachman Michael, W., & Payseur Bret, A. (2012). Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367* (1587), 409-421. doi:10.1098/rstb.2011.0249

Navarro, A., Betran, E., Barbadilla, A., & Ruiz, A. (1997). Recombination and gene flux caused by gene conversion and crossing over in inversion heterokaryotypes. *Genetics, 146* (2), 695-709. doi: 10.1017/S0016672397002784

Navarro, A., & Ruiz, A. (1997). On the Fertility Effects of Pericentric Inversions. *Genetics, 147* (2), 931-933.

Nielsen, R., Korneliussen, T., Albrechtsen, A., Li, Y., & Wang, J. (2012). SNP Calling, Genotype Calling, and Sample Allele Frequency Estimation from New-Generation Sequencing Data. *PLOS ONE, 7* , e37558. doi:10.1371/journal.pone.0037558

Noor, M. A. F., Grams, K. L., Bertucci, L. A., & Reiland, J. (2001). Chromosomal inversions and the reproductive isolation of species.*Proceedings of the National Academy of Sciences, 98* (21), 12084. doi:10.1073/pnas.221274498

Noor, M. A. F., & Bennett, S. M. (2009). Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity* , *103* (6), 439-444. doi:10.1038/hdy.2009.151

Nosil, P., Egan, S. P., & Funk, D. J. (2008). Heterogeneous genomic differentiation between walking-stick ecotypes: "isolation by adaptation" and multiple roles for divergent selection.*Evolution, 62* (2), 316-336. doi:10.1111/j.1558-5646.2007.00299.x

Nosil, P., Funk, D. j., & Ortiz-barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology, 18* (3), 375-402. doi:10.1111/j.1365-294X.2008.03946.x

Nosil, P., Vines, T. H., & Funk, D. J. (2005). Reproductive isolation caused by natural selection against immigrants from divergent habitatats. *Evolution, 59* (4), 705-719. doi:10.1111/j.0014-3820.2005.tb01747.x

Orr, H. A. (1989). Genetics of sterility in hybrids between two subspecies of drosophila. *Evolution, 43* (1), 180-189. doi:10.1111/j.1558-5646.1989.tb04216.x

Orr, H. A., & Coyne, J. A. (1989). The Genetics of Postzygotic Isolation in the Drosophila Virilis Group. *Genetics, 121* (3), 527-537.

Orr, H. A., & Turelli, M. (2001). The evolution of postzygotic isolation: accumulating dobzhansky-muller incompatibilities.*Evolution, 55* (6), 1085-1094. doi:10.1111/j.0014-3820.2001.tb00628.x

Ortíz-Barrientos, D., Reiland, J., Hey, J., & Noor, M. A. F. (2002). Recombination and the divergence of hybridizing species. In W. J. Etges & M. A. F. Noor (Eds.), *Genetics of Mate Choice: From Sexual Selection to Sexual Isolation* (pp. 167-178). Dordrecht: Springer Netherlands.

Pakull, B., Groppe, K., Meyer, M., Markussen, T., & Fladung, M. (2009). Genetic linkage mapping in aspen (Populus tremula L. and Populus tremuloides Michx.). *Tree Genetics & Genomes, 5* (3), 505-515. doi:10.1007/s11295-009-0204-2

Palumbi, & R, S. (1994). Genetic Divergence, Reproductive Isolation, and Marine Speciation. *Annu Rev Ecol Syst, 25* (1), 547-572.

Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology, 25* (11), 2337-2360. doi:10.1111/mec.13557

Peiyuan, H. (1989). Discussion on Paleotemperature Change from the Ratio of High-priced Iron to Low-priced Iron in the Quaternary Strata of Lushan Mountain. *Published by the Institute of Geomechanics, Chinese Academy of Geological Sciences* (2), 61-70.

Presgraves, D. C. (2007). Does genetic conflict drive rapid molecular evolution of nuclear transport genes in Drosophila? *BioEssays, 29* (4), 386-391. doi:10.1002/bies.20555

Qvarnstrom, A., & Bailey, R. I. (2009). Speciation through evolution of sex-linked genes. *Heredity, 102* (1), 4-15. https://doi.org/10.1038/hdy.2008.93

Qvarnström, A., & Bailey, R. I. (2008). Speciation through evolution of sex-linked genes. *Heredity, 102* , 4. doi:10.1038/hdy.2008.93

Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., . . . Westram, A. M. (2017). Interpreting the genomic landscape of speciation: a road map for finding intrinsic barriers.*J. Evol Biol, 30* . https://doi.org/10.1111/jeb.13047

Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation.*Trends in Ecology and Evolution, 16* (7), 351-358. https://doi.org/10.1016/S0169-5347(01)02187-5

Rieseberg, L. H., & Willis, J. H. (2007). Plant speciation.*Science, 317* (5840), 910-914. doi:10.1126/science.1137729

Rifkin, J. L., Castillo, A. S., Liao, I. T., & Rausher, M. D. (2019). Gene flow, divergent selection and resistance to introgression in two species of morning glories (Ipomoea). *Molecular Ecology, 28* (7), 1709-1729. doi:10.1111/mec.14945

Schumer, M., Cui, R., Powell, D. L., Dresner, R., Rosenthal, G. G., & Andolfatto, P. (2014). High-resolution mapping reveals hundreds of genetic incompatibilities in hybridizing fish species. *eLife, 3* . https://doi.org/10.7554/eLife.02535

Shen, X., Wan, S., Colin, C., Tada, R., Shi, X., Pei, W., . . . Li, A. (2018). Increased seasonality and aridity drove the C4 plant expansion in Central Asia since the Miocene–Pliocene boundary. *Earth and Planetary Science Letters, 502* , 74-83. https://doi.org/10.1016/j.epsl.2018.08.056

Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2013). Estimating Individual Admixture Proportions from Next Generation Sequencing Data.*Genetics, 195* (3), 693. doi:10.1534/genetics.113.154138

Stölting, K. N., Paris, M., Meier, C., Heinze, B., Castiglione, S., Bartha, D., & Lexer, C. (2015). Genome-wide patterns of differentiation and spatially varying selection between postglacial recolonization lineages of Populus alba (Salicaceae), a widespread forest tree.*New Phytologist, 207* (3), 723-734. doi:10.1111/nph.13392

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics, 123* (3), 585.

Tarailograovac, M., & Chen, N. (2009). UNIT 4.10 Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. *Current protocols in human genetics* .

Teresa Avelar, A., Perfeito, L., Gordo, I., & Godinho Ferreira, M. (2013). Genome architecture is a selectable trait that can be maintained by antagonistic pleiotropy. *Nature Communications, 4* , 2235. doi:10.1038/ncomms3235

Thompson, J. A., Burbank, D. W., Li, T., Chen, J., & Bookhagen, B. (2015). Late Miocene northward propagation of the northeast Pamir thrust system, northwest China. *Tectonics, 34* (3), 510-534. doi:10.1002/2014TC003690

Turner, T. L., & Hahn, M. W. (2010). Genomic islands of speciation or genomic islands and speciation? *Molecular Ecology, 19* (5), 848-850. doi:10.1111/j.1365-294X.2010.04532.x

Tuskan, G. A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., . . . Rokhsar, D. (2006). The genome of black cottonwood,*Populus trichocarpa* (Torr. & Gray). *Science, 313* (5793), 1596. doi:10.1126/science.1128691

Walsh, B., & Blows, M. W. (2009). Abundant Genetic Variation + Strong Selection = Multivariate Genetic Constraints: A Geometric View of Adaptation. *Annual Review of Ecology, Evolution, and Systematics, 40* (1), 41-59. doi:10.1146/annurev.ecolsys.110308.120232

Wang, D., Wang, Z., Kang, X., & Zhang, J. (2019). Genetic analysis of admixture and hybrid patterns of Populus hopeiensis and P. tomentosa.*Scientific Reports, 9* (1), 4821. doi:10.1038/s41598-019-41320-z

Wang, J., Street, N. R., Scofield, D. G., & Ingvarsson, P. K. (2016). Variation in Linked Selection and Recombination Drive Genomic Divergence during Allopatric Speciation of European and American Aspens.*Molecular Biology and Evolution, 33* (7), 1754-1767. doi:10.1093/molbev/msw051

Wang, Y., Sun, X., Tan, B., Zhang, B., Xu, L.-a., Huang, M., & Wang, M. (2010). A genetic linkage map of Populus adenopoda Maxim. × P. alba L. hybrid based on SSR and SRAP markers. *Euphytica, 173* (2), 193-205. doi:10.1007/s10681-009-0085-3

Wang, Y., Zhang, B., Sun, X., Tan, B., Xu, L.-a., Huang, M., & Wang, M. (2011). Comparative genome mapping among *Populus adenopoda* ,*P. alba, P. deltoides* , P. *euramericana* and *P. trichocarpa* . *Genes & Genetic Systems, 86* (4), 257-268. doi:10.1266/ggs.86.257

Wang, Z., Du, S., Dayanandan, S., Wang, D., Zeng, Y., & Zhang, J. (2014). Phylogeny Reconstruction and Hybrid Analysis of Populus (Salicaceae) Based on Nucleotide Sequences of Multiple Single-Copy Nuclear Genes and Plastid Fragments. *PLOS ONE, 9* (8), e103645. doi:10.1371/journal.pone.0103645

Widmer, A., Lexer, C., & Cozzolino, S. (2008). Evolution of reproductive isolation in plants. *Heredity, 102* , 31. doi:10.1038/hdy.2008.69

Wolf, J. B. W., & Ellegren, H. (2016). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics, 18* , 87. doi:10.1038/nrg.2016.133

Wright, W. B. (1926). The German Outwash Terraces and the Astronomical Theory of the Ice Age. *Nature, 117* (2934), 113-114. doi:10.1038/117113a0

Wu, C.-I. (2001). The genic view of the process of speciation.*Journal of Evolutionary Biology, 14* (6), 851-865. doi:10.1046/j.1420-9101.2001.00335.x

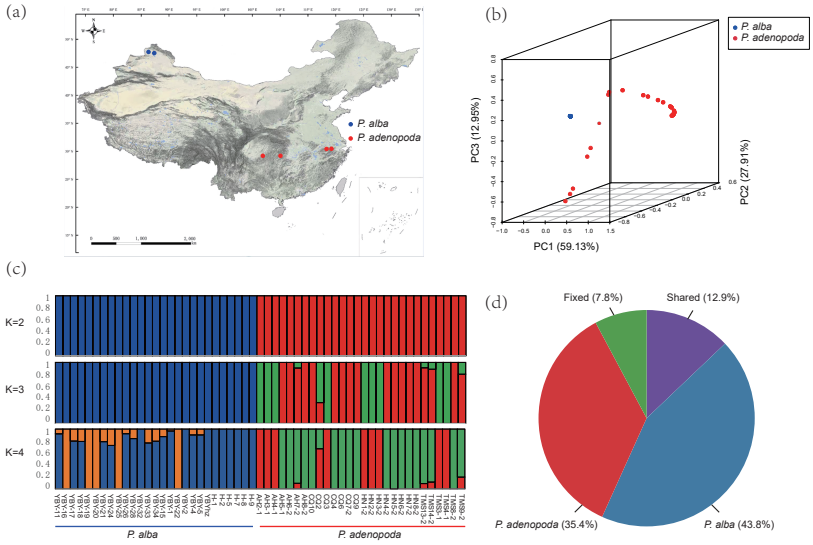
Zhang, D., Fengquan, L., & Jianmin, B. (2000). Eco-environmental effects of the Qinghai-Tibet Plateau uplift during the Quaternary in China. *Environmental Geology, 39* (12), 1352-1358. doi:10.1007/s002540000174

Zhisheng, A., Kutzbach, J. E., Prell, W. L., & Porter, S. C. (2001). Evolution of Asian monsoons and phased uplift of the Himalaya–Tibetan plateau since Late Miocene times. *Nature, 411* (6833), 62-66. doi:10.1038/35075035

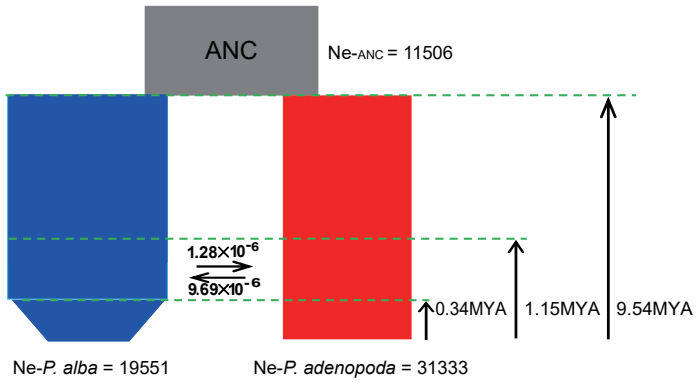
Zhuang, G., Brandon, M. T., Pagani, M., & Krishnan, S. (2014). Leaf wax stable isotopes from Northern Tibetan Plateau: Implications for uplift and climate since 15 Ma. *Earth and Planetary Science Letters, 390* , 186-198. https://doi.org/10.1016/j.epsl.2014.01.003

Hosted file

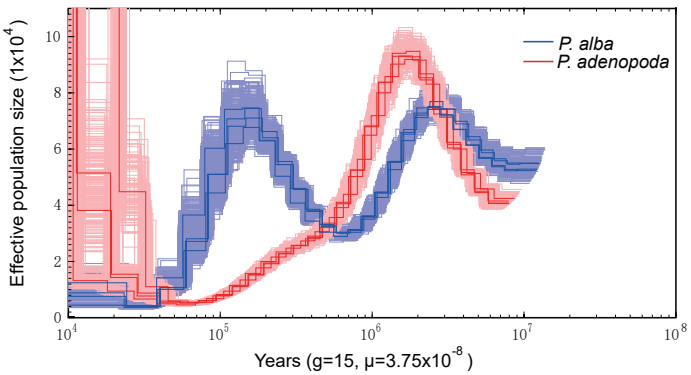
`tracked_manuscript.pdf` available at https://authorea.com/users/362208/articles/507429-natural-selection-and-intrinsic-barriers-play-important-roles-in-speciation-of-two-closely-related-populus-salicaceae-species

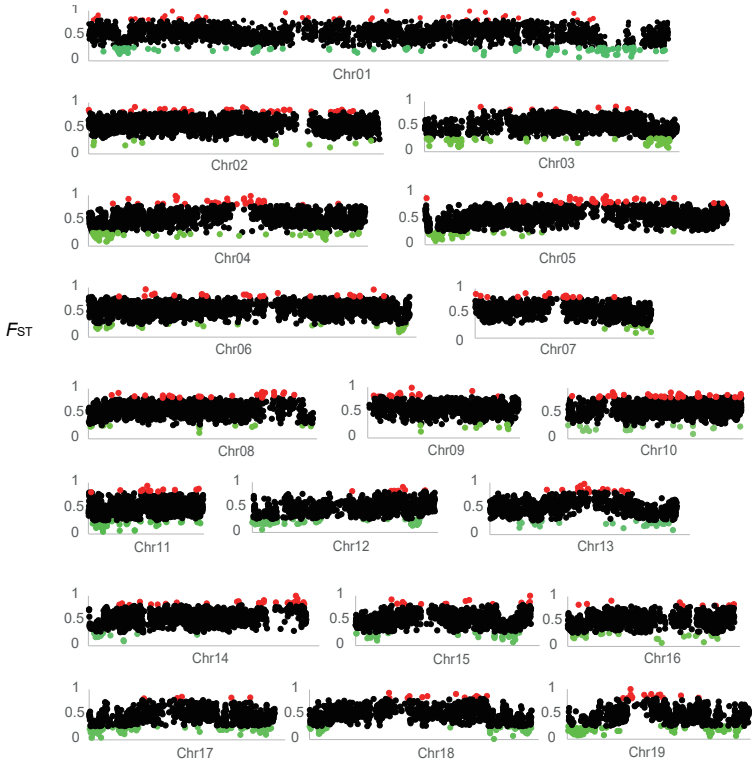


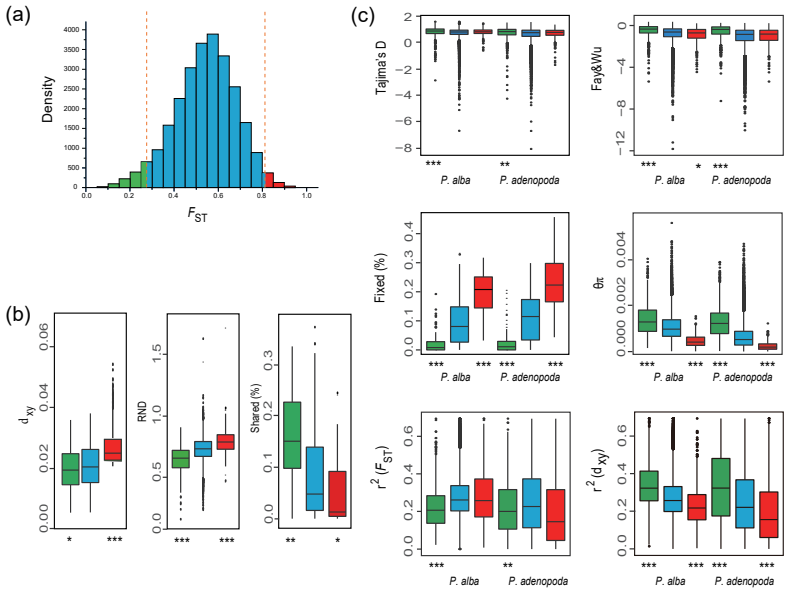
(a)



(b)







Hosted file

Figure S1.pdf available at https://authorea.com/users/362208/articles/507429-natural-selection-and-intrinsic-barriers-play-important-roles-in-speciation-of-two-closely-related-populus-salicaceae-species

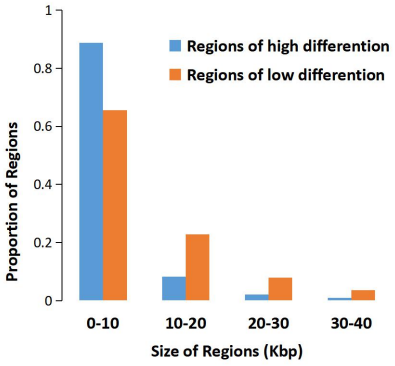


Figure S2. The physical size distribution between *P. alba* and *P. adenopoda* showing extremely high (blue bars)and low (orange bars) genetic differentiation.

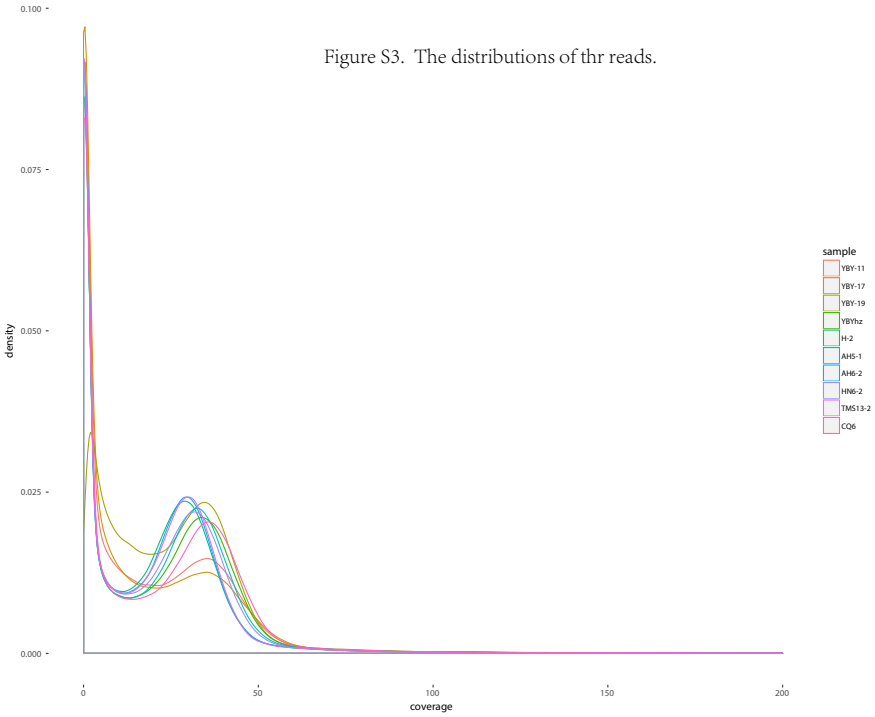


Figure S3. The distributions of thr reads.

Hosted file

`Table S1.pdf` available at https://authorea.com/users/362208/articles/507429-natural-selection-and-intrinsic-barriers-play-important-roles-in-speciation-of-two-closely-related-populus-salicaceae-species

Table S2. Relative likelihood of the different models shown in Figure S1

model	Max(log10(Lhoodi)[a]	No. Of parameters(d)	AICi[b]	Δi[b]	Model normalized relative likelihood(wi)b
Model1	-85025483.6	4	391556830.1	681380.8686	~0
Mode2	-85103413.95	5	391915714.6	1040271.393	~0
Model3	-85056421.02	6	391699306.2	823858.9528	~0
Model4	-85033399.28	5	391593285.2	717843.9221	~0
Model5	-85019567.11	6	391529587.7	654144.4252	~0
Model6	-85048908.67	7	391664712.6	789263.3025	~0
Model7	-84982673.75	8	391359691.5	484250.2237	~0
Model8	-85023913.29	7	391549604.6	674153.3238	~0
Model10	-85043369.01	6	391639199.5	763756.2255	~0
Model11	-85022818.26	7	391544561.8	669112.5243	~0
Model12	-85025552.15	6	391557149.8	681706.553	~0
Model13	-85025440.78	7	391556638.9	681185.6752	~0
Model14	-85061928.28	8	391724672.1	849230.8223	~0
Model15	-84982165.92	9	391357354.8	481897.5801	~0
Model16	-85040874.24	8	391627714.6	752265.385	~0
Model17	-128561324.8	7	592046794	201171346.8	~0
Model18	-84929383.02	8	391114278.6	238827.3427	~0
Model19	-84974058.69	9	391320019.7	444566.4062	~0
Model20	-84967887.1	8	391291596.4	416147.1839	~0
Model21	-84962919.61	9	391268722.3	393273.0471	~0
Model22	-84877520.15	9	390875443.3	0	1
Model23	-84892761.84	12	390945639.8	70184.57637	~0
Model24	-84900737.75	12	390982370.3	106912.9993	~0
Model25	-84934178.07	12	391136368.6	260909.364	~0

[a] Based on the best likelihood among the 50 independent runs for each model (Figure S1).

[b] The calculation of AICi, Δ_i and w_i are according to the methods shown in Excoffier *et al.* (2013).

Table S3. A comparison of the regions showing extreme genetic differentiation in *P. alba* and *P. adenopoda* with other genomic regions is presented (give the mean± standard deviation).

Parameters	Species	Regions displaying high differentiation	Regions displaying low differentiation	Background
θπ	*P .alba*	0.0005(±0.0003)***	0.0014(±0.0007)***	0.0011(±0.0006)
	P .adenopoda	0.0002(±0.0002)***	0.0013(±0.0007)***	0.0007(±0.0005)
Tajima's D	*P .alba*	1.2889(±0.5060)	1.3644(±0.6763)***	1.1893(±0.6189)
	P .adenopoda	1.1161(±0.6088)	1.2098(±0.8052)**	1.0416(±0.7827)
Fay & Wu's H	*P .alba*	-0.5939(±0.3420)*	-0.3731(±0.3552)***	-0.5392(±0.3493)
	P .adenopoda	-0.7136(±0.4016)	-0.4640(±0.4417)***	-0.7501(±0.4315)
$r^2(F_{ST})$	*P .alba*	0.3323(±0.2043)	0.2568(±0.1531)***	0.3289(±0.1550)
	P .adenopoda	0.2354(±0.2368)	0.2634(±0.1985)**	0.3105(±0.2376)
r^2(Dxy)	*P .alba*	0.2674(±0.1470)***	0.4213(±0.1776)***	0.3216(±0.1523)
	P .adenopoda	0.2350(±0.2104)***	0.4249(±0.2706)***	0.3048(±0.2341)
Fixed(%)	*P .alba*	0.2146(±0.0887)***	0.0258(±0.0389)***	0.0999(±0.0851)
	P .adenopoda	0.2535(±0.1122)***	0.0296(±0.0497)***	0.1211(±0.0904)
Shared(%)		0.0517(±0.0680)*	0.1695(±0.1009)**	0.0908(±0.0965)
F_{ST}		0.8594(±0.0396)***	0.2070(±0.0496)***	0.5496(±0.1194)
Dxy		0.0280(±0.0092)***	0.0212(±0.0083)*	0.0220(±0.0085）
RND		1.2226(±0.2831)***	0.9213(±0.2040)***	1.0906(±0.1958）

By Mann-Whitney U test, there were significant differences between the genome background and regions showing extreme genetic differentiation (* *P*-value < 0.05; ** *P*-value < 1e-4; ***P-value <2.2e-16).

Hosted file

Table S4.pdf available at https://authorea.com/users/362208/articles/507429-natural-selection-and-intrinsic-barriers-play-important-roles-in-speciation-of-two-closely-related-populus-salicaceae-species

Hosted file

Table S5.pdf available at https://authorea.com/users/362208/articles/507429-natural-selection-and-intrinsic-barriers-play-important-roles-in-speciation-of-two-closely-related-populus-salicaceae-species

Hosted file

Table S6.pdf available at https://authorea.com/users/362208/articles/507429-natural-selection-and-intrinsic-barriers-play-important-roles-in-speciation-of-two-closely-related-populus-salicaceae-species

Hosted file

Table S7.pdf available at https://authorea.com/users/362208/articles/507429-natural-selection-and-intrinsic-barriers-play-important-roles-in-speciation-of-two-closely-related-populus-salicaceae-species

Hosted file

Table S8.pdf available at https://authorea.com/users/362208/articles/507429-natural-selection-and-intrinsic-barriers-play-important-roles-in-speciation-of-two-closely-related-populus-salicaceae-species

Hosted file

Table S9.pdf available at https://authorea.com/users/362208/articles/507429-natural-selection-and-intrinsic-barriers-play-important-roles-in-speciation-of-two-closely-related-populus-salicaceae-species

Hosted file

Table S10.txt available at https://authorea.com/users/362208/articles/507429-natural-selection-and-intrinsic-barriers-play-important-roles-in-speciation-of-two-closely-related-populus-salicaceae-species