doi:10.1111/evo.13

# Reticulate evolution within a spruce (*Picea*) species complex revealed by population genomic analysis

**Yongshuai Sun,[1,2,3] Richard J. Abbott,[4] Zhiqiang Lu,[2] Kangshan Mao,[1] Lei Zhang,[1] Xiaojuan Wang,[1] Dafu Ru,[1] and Jianquan Liu[1,5,6]**

[1]*Key Laboratory for Bio-resource and Eco-environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu 610065, P. R. China*

[2]*CAS Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla 666303, P. R. China*

[3]*E-mail: sunyongshuai@xtbg.ac.cn*

[4]*School of Biology, University of St Andrews, St Andrews, Fife KY16 9TH, United Kingdom*

[5]*State Key Laboratory of Grassland Agro-Ecosystem, Institute of Innovation Ecology & College of Life Science, Lanzhou University, Lanzhou 730000, Gansu, P. R. China*

[6]*E-mail: liujq@nwipb.cas.cn*

Received November 9, 2017
Accepted October 5, 2018

**The role of reticulation in the rapid diversification of organisms is attracting greater attention in evolutionary biology. Evidence of genetic exchange between diverging taxa is reported frequently, although most studies fail to show how hybridization and introgression contribute to the adaptation and differentiation of introgressed taxa. Here, we report a population genomics approach to test the role of hybridization and introgression in the evolution of the *Picea likiangensis* species complex, which comprises four taxa occurring in the biodiversity hotspot of the Hengduan–Himalayan mountains. Based on 84,793 SNPs detected in transcriptomes of 82 trees collected from 35 localities, we identified 18 hybrids (including backcrosses) distributed within the range boundaries of the four taxa. Coalescent simulations, for each pair of taxa and for all taxa taken together, rejected several tree-like divergence models and supported instead a reticulate evolution model with secondary contacts occurring during Pleistocene glacial cycles after initial divergence in the late Pliocene. Significant gene flow occurred among some taxa after secondary contact according to an analysis based on modified ABBA–BABA statistics that accommodated a rapid diversification scenario. A novel finding was that introgression between certain taxa can contribute to increasing divergence (and possibly reproductive isolation) between those taxa and other taxa within a complex at some loci. These results illuminate the reticulate nature of evolution within the *P. likiangensis* complex and highlight the value of population genomic data in detecting the effects of introgression in the rapid diversification of related taxa.**

**KEY WORDS:** coalescent analyses, genetic divergence, introgression, *Picea*, population genomics, reticulate evolution.

Clarifying evolutionary processes that initiate and maintain species diversity is a central aim of natural biodiversity studies (Givnish and Sytsma 2000; Andrew et al. 2013; Grant and Grant 2017). Because hybridization can generate genetic variation acted on by selection at higher rates than mutation, it has been proposed that it may often play an important role in evolutionary divergence, adaptation, and speciation (Seehausen 2004; Mallet 2007; Rieseberg and Willis 2007; Abbott et al. 2013; Arnold and Kunte 2017). In the last 20 years or so, considerable evidence of gene exchange between diverging taxa has accumulated (Harrison



© 2018 The Author(s). *Evolution* © 2018 The Society for the Study of Evolution.

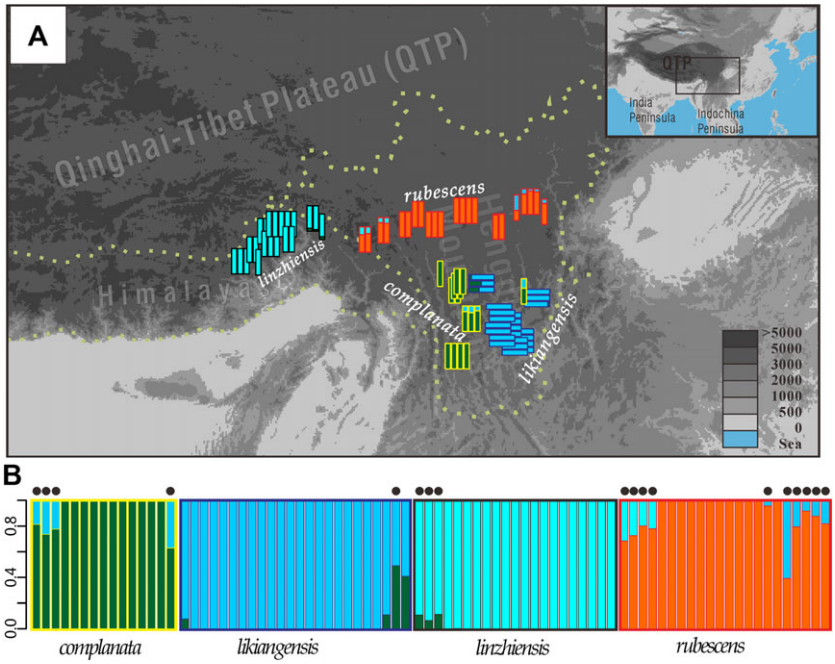


**Figure 1.** **(A) The distribution of the sampled 82 trees within the *Picea likiangensis* species complex. Each tree is represented by a bar indicating its inferred ancestry. The dotted lines indicate boundaries of the Himalayan mountains, Hengduan mountains and Qinghai-Tibet Plateau (QTP). (B) Assignment of individuals to different genetic groups (*K*) when $K = 4$ as inferred by ADMIXTURE. Each individual is represented by a vertical bar displaying the individual's probability (*Q*-value) belonging to one or another cluster (genetic group). Individuals with *Q*-values < 0.99 (bootstrap test, $P < 0.05$) are indicated by solid circles and considered to be hybrids.**

and Larson 2014; Abbott 2017; Suarez-Gonzalez et al. 2018a), leading to the belief that a reticulate pattern of evolution (particularly apparent for closely related taxa) is a more reasonable evolutionary paradigm than a tree-like pattern within many groups of organisms (Linder and Rieseberg 2004; Arnold 2015). However, to date, the contribution of hybridization and introgression to the diversification of extant taxa has not been verified in detail, largely because appropriate methods of analysis and genome-scale datasets were unavailable until recently (Abbott et al. 2016; Lexer et al. 2016; Payseur and Rieseberg 2016; Pease et al. 2016; Leroy et al. 2017; Suarez-Gonzalez et al. 2018b). Thus, no empirical study has yet indicated how introgression across the genome may have altered the genetic composition of recipient taxa and increased their divergence within a species complex during the course of its evolution.

In the present study, we used the *Picea likiangensis* species complex of spruce taxa as a system for testing if evolutionary divergence within this complex followed a reticulate or tree-like pattern, and second to clarify the role of introgression in promoting differentiation between diverging taxa. Traditionally, the *Picea likiangensis* species complex has been thought to comprise three varieties of the species (*rubescens*, *likiangensis*, and *linzhiensis*) that exhibit reduced density of stomatal lines on the abaxial surfaces of leaves relative to most other *Picea* species (Fu et al. 1999). Recent studies, however, indicate that *P. brachytyla* var. *complanata*, which has no stomatal lines on its abaxial leaf surfaces, should also be included in this complex (Ru et al. 2016). These four taxa occupy sites that differ in environmental conditions and have distinct geographical ranges (Fig. 1), spanning the Hengduan–Himalayan biodiversity hotspot in Asia (Fu et al. 1999; Li et al. 2010; Li et al. 2013; Lockwood et al. 2013; Wang et al. 2017). Analyses of molecular genetic variation have shown that the taxa are genetically differentiated (Li et al. 2013; Ru et al. 2016), however their evolutionary relationships remain incompletely resolved. Nuclear phylogenies have short branch lengths and vary in topology according to the nuclear genes analyzed (Li et al. 2013; Ru et al. 2016), indicating that the four taxa may have diversified rapidly and that hybridization and gene flow occurred during diversification (Li et al. 2013; Lockwood et al. 2013; Ran et al. 2015; Ru et al. 2016).

Here, we examine these possibilities further through population genomic analysis. First, we test whether the *P. likiangensis* species complex originated by means of rapid diversification involving hybridization among the four taxa or alternatively in a tree-like scenario of evolutionary divergence in which hybridization played no substantial role during diversification. Second, having shown that reticulation was important in the evolution of the *P. likiangensis* species complex, we examine how introgression contributed to the divergence of taxa within the complex.

**Table 1.** Locations of the 35 populations sampled within the *Picea likiangensis* species complex plus the collection site of one *P. breweriana* tree sampled. Numbers in brackets indicate samples previously examined by Ru et al. 2016.

| Population | | Collection site | Latitude | Longitude | Altitude (m) | *n* |
|---|---|---|---|---|---|---|
| *P. breweriana* (Bre) | | Common garden, Gansu | N35°55′59″ | E104°9′0.3 | 1774 | **1** |
| *P. likiangensis* var. *rubescens* (RUB) | | | | | | **22** (5) |
| 1 | MSZ-01 | Airport, Kangding, Sichuan | N30°7′14.1″ | E101°45′22″ | 4221 | 1 |
| 2 | MSZ-02 | Tacheng, Kangding, Sichuan | N30°17′41.9″ | E101°36′34″ | 3978.51 | 3 (1) |
| 3 | MSZ-03 | Tagongsi, Sichuan | N30°16′27.7″ | E101°31′19.2″ | 3589.66 | 1 (1) |
| 4 | MSZ-04 | Jianziwan mountain, Sichuan | N29°59′58.5″ | E100°52′17″ | 4178.94 | 2 (1) |
| 5 | MSZ-05 | Heni, Sichuan | N30°17′13″ | E99°31′9.2″ | 4289.50 | 4 (1) |
| 6 | MSZ-06 | Zongla mountain, Sichuan | N29°43′47.4″ | E98°37′47.1″ | 4026.58 | 3 (1) |
| 7 | MSZ-07 | Rumei, Tibet | N29°36′40.6″ | E98°9′25″ | 4104.20 | 2 |
| 8 | MSZ-08 | Zuogong, Tibet | N29°40′59.6″ | E97°55′54.9″ | 4122.95 | 2 |
| 9 | MSZ-09 | Ranwu, Tibet | N29°33′20.8″ | E96°46′38.4″ | 4186.15 | 2 |
| 10 | MSZ-35 | Ranwu lake, Tibet | N29°29′24.6″ | E96°40′20.2″ | 3920.11 | 2 |
| var. *linzhiensis* (LIN) | | | | | | **21** (5) |
| 11 | MSZ-15 | Milin, Linzhi, Tibet | N29°11′5.1″ | E93°58′42.8″ | 2988.12 | 3 (1) |
| 12 | MSZ-25 | Milin, Linzhi, Tibet | N29°27′48.2″ | E94°37′3.9″ | 2913.14 | 2 (1) |
| 13 | MSZ-26 | Qiangna, Linzhi, Tibet | N29°27′52.2″ | E94°35′36.6″ | 2919.63 | 1 (1) |
| 14 | MSZ-29 | Nixi, Linzhi, Tibet | N29°45′20.8″ | E94°15′36.1″ | 3042.92 | 2 |
| 15 | MSZ-30 | Sejila mountain, Linzhi, Tibet | N29°34′10.6″ | E94°33′28.7″ | 3421.91 | 3 (1) |
| 16 | MSZ-31 | Sejila mountain, Linzhi, Tibet | N29°40′26.1″ | E94°43′12.1″ | 3663.20 | 5 (1) |
| 17 | MSZ-32 | Lulang, Tibet | N29°49′32.2″ | E94°44′23.2″ | 3125.35 | 2 |
| 18 | MSZ-33 | Bomi, Tibet | N29°53′25.8″ | E95°31′23.8″ | 2698.53 | 1 |
| 19 | MSZ-34 | Bomi, Tibet | N29°49′25.8″ | E95°42′41.5″ | 3262.82 | 2 |
| var. *likiangensis* (LIK) | | | | | | **24** (5) |
| 20 | MSZ-41 | Pudacuo, Yunnan | N27°41′35.1″ | E100°0′54.6″ | 3208.26 | 3 (2) |
| 21 | MSZ-42 | Pudacuo, Yunnan | N27°34′8.4″ | E100°1′25.8″ | 3025.85 | 1 |
| 22 | MSZ-44 | Daju mountain, Yunnan | N27°13′12.5″ | E100°15′58.1″ | 2993.41 | 2 (1) |
| 23 | MSZ-45 | Daju mountain, Yunnan | N27°12′13.5″ | E100°16′33.9″ | 3132.56 | 2 |
| 24 | MSZ-46 | Daju mountain, Yunnan | N27°11′54.1″ | E100°16′43.9″ | 3260.41 | 2 |
| 25 | MSZ-47 | Yulongxue mountain, Yunnan | N27°7′53.8″ | E100°13′58.9″ | 2947.51 | **4** (1) |
| 26 | MSZ-48 | Yulongxue mountain, Yunnan | N27°8′31.7″ | E100°14′0.6″ | 3197.45 | 3 |
| 27 | MSZ-50 | Yuhu village, Yunnan | N27°1′30.3″ | E100°12′32.3″ | 2845.13 | 4 (1) |
| 28 | MYs | Mianya south, Sichuan | N27°31′9″ | E101°21′34″ | 3300 | 3 |
| *P. brachytyla* var. *complanata* (COM) | | | | | | **15** (5) |
| 29 | MSZ-38 | Deqin, Yunnan | N28°24′25.4″ | E98°59′16.6″ | 3882.86 | 1 |
| 30 | MSZ-40 | Napahai, Yunnan | N27°55′51″ | E99°36′59.3″ | 3511.80 | 2 |
| 31 | MY | Mianya, Sichuan | N27°33′49″ | E101°21′34″ | 3511 | 1 (1) |
| 32 | NPH | Napahai, Yunnan | N27°55′40″ | E99°36′51.2″ | 3506 | 2 (1) |
| 33 | XGLL | Xianggelila, Yunnan | N27°48′ | E99°39′ | 3329 | 2 (1) |
| 34 | XSQ | Xinshengqiao, Yunnan | N26°27.43′ | E99°18.83′ | 2925 | 4 (1) |
| 35 | Zhong | Xiaozhongdian, Yunnan | N27°27.399′ | E99°53.588′ | 3128 | 3 (1) |

# Methods

## PLANT MATERIAL AND RNA SEQUENCING

We collected young leaves from 62 trees throughout the range of the *P. likiangensis* species complex (Table 1). Trees from each population were spaced at least 100 m apart. We also sampled leaves from one *P. breweriana* tree used as an outgroup. We extracted RNA from fresh mature leaf needles collected from first year branches of each selected tree and stored it in liquid nitrogen in the field. Libraries were constructed individually using a NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA). Subsequently, standard RNA-seq procedures (Wang et al. 2009; Jiang et al. 2011) and an Illumina HiSeq 2500 platform were used to generate paired-end raw reads. We deposited the 62 novel transcriptomes in BioSample. For analysis, we added to these 20 transcriptomes of *P. brachytyla* var. *complanata*, and *P. likiangensis vars. rubescens*, *likiangensis*, and *linzhiensis*

available from a previous study (Ru et al. 2016). Thus a total of 83 transcriptomes were used in analyses with average number of raw bases obtained per individual > 6 Gb (Table S1).

## READ MAPPING AND VARIANT CALLING

By controlling the Phred score, sequence length, and percentage of ambiguous bases, raw reads were filtered using fastq_quality_trimmer (parameters -v -t 20 -l 30), fastq_quality_filter (-v -q 20 -p 90), and fastq_masker (-q 20 -v) elements of the FASTX Toolkit (from http://hannonlab.cshl.edu/fastx_toolkit/). Raw reads for each sample were mapped to the reference transcriptome of *P. abies* (Nystedt et al. 2013) using BWA-MEM ver. 0.7.10 (Li and Durbin 2009). Possible fungal transcripts in the *P. abies* transcriptome were deleted (Delhomme et al. 2015), and only the longest transcript of each gene was retained. Reads were sorted in BAM format using SAMTOOLS ver. 0.1.19 (Li et al. 2009) with duplicate reads marked by Picard tools (version 1.106) and excluded from further analysis. Genotypes and single-nucleotide polymorphisms (SNPs) were called using READS2SNPS v2.0 (Tsagkogeorga et al. 2012; Gayral et al. 2013). The minimum read depth was set to 10, minimum mapping quality to 20, minimum base quality to 20, and the number of threads to 10. When the posterior probability of the best-supported homozygote or heterozygote was below 0.95, the base was treated as missing. After deleting sites with missing bases in all 83 individuals, 4743 loci without missing bases remained (Table 2). Tests of Hardy–Weinberg equilibrium implemented in vcftools0.1.12b (Danecek et al. 2011) were used on these loci to filter out polymorphic sites that deviated significantly ($P < 0.05$). A total of 3646 loci with length > 200 bps and at least one polymorphic site was retained for further analysis (Table 2).

## ANALYSES OF NUCLEOTIDE DIVERSITY AND CLUSTERING

To examine nucleotide diversity and skew of the site frequency spectrum, we computed the number of segregating sites ($S$), $\pi$ per site (Nei 1987), Tajima's $D$ (Tajima 1989), and Fay and Wu's $H$ (Fay and Wu 2000) for each taxon in turn. We used the maximum frequency of derived mutation (MFDM) method to test the neutrality of the frequency spectrum for each locus (Li 2011) with homozygous bases in the transcriptome of *P. breweriana* used to determine ancestral states.

Individual-based clustering analysis was conducted using the maximum-likelihood approach implemented in ADMIXTURE ver. 1.23 (Alexander and Lange 2011). The data file was converted using VCFTOOLS ver. 0.1.12b and PLINK ver. 1.07 (Purcell et al. 2007; Danecek et al. 2011) and cross-validation was used to explore convergence and determine the optimum number of clusters ($K$, from 1 to 10, with the optimum $K$ value indicated by lowest cross-validation error). One hundred datasets were generated using random sampling with replacement to calculate the confidence bounds of $Q$-values for each individual tree.

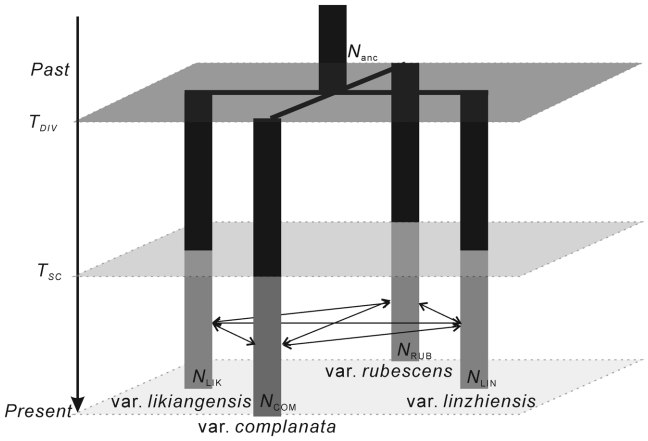


**Figure 2. The most likely reticulate evolutionary model involving a period of secondary contact among the four taxa comprising the *Picea likiangensis* species complex. The top layer (parallelogram) indicates the divergence time of the four taxa at $T_{DIV}$, when the four taxa diverged from an ancestral population with size $N_{anc}$. The middle layer indicates the time ($T_{SC}$) that secondary contact was initiated among taxa. Arrows indicate genetic exchange among taxa from initial secondary contact ($T_{SC}$) to the present. For each pair of taxa, two free migration parameters in different directions are assumed, indicating 12 parameters in the migration matrix (Text S1). Current population sizes of the four taxa are represented by $N_{COM}$, $N_{LIK}$, $N_{LIN}$, and $N_{RUB}$.**

## TESTING RETICULATE EVOLUTION AND TREE-LIKE DIVERGENCE MODELS

To determine the evolutionary relationships among all four taxa, we used FASTSIMCOAL2 (Excoffier et al. 2013) to compare different evolutionary models based on the site frequency spectrum (SFS) and coalescent simulations. We performed 200,000 simulations to calculate the expected SFS using the infinite site model (–I) and 200 ECM cycles to estimate parameters with the highest likelihood for each possible divergence model and the reticulation model that allowed gene flow among taxa after secondary contact (Fig. 2 and Text S1). The reticulate evolutionary model included 19 parameters (Text S1): the effective population size for each taxon and their ancestral population, 12 migration rate parameters for each pair of taxa, time of initial division ($T_{DIV}$), and time of secondary contact ($T_{SC}$). We also computed the relative $T_{SC}$ as $T_{SC}/T_{DIV}$. The stopping criterion in parameter optimization was set to $10^{-5}$. We collapsed all SFS entries less than five into a single category (–C5 of fastsimcoal2) with number of monomorphic sites considered. The mutation rate was set to $4 \times 10^{-8}$ substitutions per site per generation (De La Torre et al. 2017), and 50 years per generation was assumed (Li et al. 2010). AIC

**Table 2.** Summary statistics obtained from an analysis of sampled transcriptomes.

| | All samples | var. *complanata* | var. *likiangensis* | var. *linzhiensis* | var. *rubescens* |
|---|---|---|---|---|---|
| **Full data set**[1] | | | | | |
| **Number of (No.) loci** | 20743 | 16178 | 16558 | 17086 | 16775 |
| **No. alignment positions** | 16297461 | 13873969 | 14040232 | 14375147 | 13990046 |
| **No. SNPs** | 480258 | 207857 | 217561 | 192168 | 284698 |
| **Reduced data set**[2] | | | | | |
| **No. loci** | 4743 | 5640 | 5574 | 6307 | 6092 |
| **No. alignment positions** | 3438344 | 4217541 | 4819161 | 6017771 | 5782898 |
| **No. SNPs** | 93245 | 59158 | 65631 | 60640 | 111561 |
| **Reduced data set**[3] | | | | | |
| **No. loci** | 3646 | 3646 | 3646 | 3646 | 3646 |
| **No. alignment positions** | 3325106 | 3325106 | 3325106 | 3325106 | 3325106 |
| **No. SNPs** | 84793 | 37549 | 38258 | 27722 | 55893 |
| **$\pi \pm$ SD (×100)** | 0.256 ± 0.196 | 0.237 ± 0.193 | 0.219 ± 0.187 | 0.185 ± 0.176 | 0.264 ± 0.197 |
| **Tajima's $D \pm$ SD** | −1.211 ± 0.712 | −0.598 ± 0.910 | −0.514 ± 0.948 | −0.204 ± 1.033 | −1.054 ± 0.759 |
| **Fay and Wu 's $H \pm$ SD** | −0.091 ± 0.780 | −0.030 ± 0.978 | −0.137 ± 1.070 | −0.211 ± 1.192 | 0.012 ± 0.790 |

[1]All loci after deleting contigs in the reference transcriptome without mapped reads.

[2]Only sites without missing data.

[3]84793 SNPs retained across 3646 loci that met the following criteria: Hardy–Weinberg Equilibrium test ($P > 0.05$), length of locus > 200 bps, at least one segregating site per locus, no missing sites among four varieties.

values for all models were computed and used to determine the best model. We estimated parameters in this best model and the SDs by setting variable "C" parameters from 1 to 20.

## TESTING INTROGRESSION PATTERNS BETWEEN EACH PAIR OF TAXA

Although gene exchange had been reported among the four taxa of the *P. likiangensis* species complex, the frequency and timing of such introgression remain unknown. We carried out a statistical comparison of four models of introgression during speciation (Fig. 3) for each of the six possible pairs of taxa. The four models included (i) complete isolation (CI), (ii) continuous migration (CM), (iii) introgression during primary contact (PC), and (iv) introgression following secondary contact (SC).

Each of the four models included the division of an ancestral population with population size $N_A$ into two daughter populations with population sizes $N_1$ and $N_2$ at time $T$. In the CI model, hybridization between the two daughter populations was prohibited. In the CM and PC models gene flow occurred after division, continuing to the present time in the CM model, but stopping at time $T$pc in the PC model. In the SC model, secondary gene flow occurred from $T$sc to the present time following an intial allopatric phase.

For each pair of taxa, we computed the likelihoods of the observed SFS underlying the four models using FASTSIMCOAL2 (Excoffier et al. 2013). For each of the four models, we performed 200,000 simulations to calculate the expected SFS using the infinite site model, and estimated parameters with the highest likelihood in 300 ECM cycles. The stopping criterion in parameter optimization was set to $10^{-5}$. The number of monomorphic sites was included. Mutation rate was again set to $4 \times 10^{-8}$ substitutions per site per generation (De La Torre et al. 2017), and 50 years per generation was assumed (Li et al. 2010). The "C" parameter in FASTSIMCOAL2 (the minimum size of entry of the observed and simulated SFS) was set from 1 to 20. The four models were ranked according to the Akaike Informartion Criterion (AIC) values for each pair of taxa and each "C" value, and documented estimated values of parameters of the best-fit model. SDs were determined from estimates of different C parameter settings.

## TESTING GENE FLOW USING A MODIFIED ABBA–BABA TEST

The ABBA–BABA test provides a powerful means for detecting intertaxon gene flow by examining excess allele sharing between two taxa ($L_1$ and $L_2$), utilizing a sister taxon ($L_0$) that is more closely related to one of them. A high level of gene flow between $L_1$ and $L_2$ relative to between $L_0$ and $L_2$ would lead to a positive ABBA–BABA statistic, given that derived alleles in $L_2$ matched derived alleles in $L_0$ and $L_1$ with equal probability (Durand et al. 2011; Martin et al. 2015). However, the original formulae of the test assume a phylogeny with dichotomous topology, which was rejected here. Thus, we extended the test to accommodate a radiation divergence scenario with four taxa and an outgroup-species (here *P. breweriana*) used to determine the ancestral nucleotide state. We assumed that SNPs were genotyped in *P. brachytyla* var. *complanata* ($L_1$), and *P. likiangensis* vars.

*likiangensis* ($L_2$), *linzhiensis* ($L_3$), and *rubescens* ($L_4$) and denote the observed frequency of SNP *i* in taxon $L_j$ as $p_{ij}$. Because we used only homozygous bases in *P. breweriana*'s transcriptome, the observed frequency of SNP *i* in *P. breweriana* would be zero and was reduced in the following expressions.

by the jackknife algorithm, and DIJ values on the basis of each pseudo-replicate were computed using custom R scripts. The R scripts used for this are available online (Script S1 and S2 in Supporting Information). According to the calculated DIJ values

$$D12 = \frac{\sum_{i=1}^{n}\{(1-p_{i1})(1-p_{i2})p_{i3}p_{i4} - [(1-p_{i1})p_{i2}(1-p_{i3})p_{i4} + (1-p_{i1})p_{i2}p_{i3}(1-p_{i4}) + p_{i1}(1-p_{i2})(1-p_{i3})p_{i4} + p_{i1}(1-p_{i2})p_{i3}(1-p_{i4}) + p_{i1}p_{i2}(1-p_{i3})(1-p_{i4})]/5\}}{\sum_{i=1}^{n}\{(1-p_{i1})(1-p_{i2})p_{i3}p_{i4} + [(1-p_{i1})p_{i2}(1-p_{i3})p_{i4} + (1-p_{i1})p_{i2}p_{i3}(1-p_{i4}) + p_{i1}(1-p_{i2})(1-p_{i3})p_{i4} + p_{i1}(1-p_{i2})p_{i3}(1-p_{i4}) + p_{i1}p_{i2}(1-p_{i3})(1-p_{i4})]/5\}}$$

$$D13 = \frac{\sum_{i=1}^{n}\{(1-p_{i1})p_{i2}(1-p_{i3})p_{i4} - [(1-p_{i1})(1-p_{i2})p_{i3}p_{i4} + (1-p_{i1})p_{i2}p_{i3}(1-p_{i4}) + p_{i1}(1-p_{i2})(1-p_{i3})p_{i4} + p_{i1}(1-p_{i2})p_{i3}(1-p_{i4}) + p_{i1}p_{i2}(1-p_{i3})(1-p_{i4})]/5\}}{\sum_{i=1}^{n}\{(1-p_{i1})p_{i2}(1-p_{i3})p_{i4} + [(1-p_{i1})(1-p_{i2})p_{i3}p_{i4} + (1-p_{i1})p_{i2}p_{i3}(1-p_{i4}) + p_{i1}(1-p_{i2})(1-p_{i3})p_{i4} + p_{i1}(1-p_{i2})p_{i3}(1-p_{i4}) + p_{i1}p_{i2}(1-p_{i3})(1-p_{i4})]/5\}}$$

$$D14 = \frac{\sum_{i=1}^{n}\{(1-p_{i1})p_{i2}p_{i3}(1-p_{i4}) - [(1-p_{i1})p_{i2}(1-p_{i3})p_{i4} + (1-p_{i1})(1-p_{i2})p_{i3}p_{i4} + p_{i1}(1-p_{i2})(1-p_{i3})p_{i4} + p_{i1}(1-p_{i2})p_{i3}(1-p_{i4}) + p_{i1}p_{i2}(1-p_{i3})(1-p_{i4})]/5\}}{\sum_{i=1}^{n}\{(1-p_{i1})p_{i2}p_{i3}(1-p_{i4}) + [(1-p_{i1})p_{i2}(1-p_{i3})p_{i4} + (1-p_{i1})(1-p_{i2})p_{i3}p_{i4} + p_{i1}(1-p_{i2})(1-p_{i3})p_{i4} + p_{i1}(1-p_{i2})p_{i3}(1-p_{i4}) + p_{i1}p_{i2}(1-p_{i3})(1-p_{i4})]/5\}}$$

$$D23 = \frac{\sum_{i=1}^{n}\{p_{i1}(1-p_{i2})(1-p_{i3})p_{i4} - [(1-p_{i1})p_{i2}(1-p_{i3})p_{i4} + (1-p_{i1})p_{i2}p_{i3}(1-p_{i4}) + (1-p_{i1})(1-p_{i2})p_{i3}p_{i4} + p_{i1}(1-p_{i2})p_{i3}(1-p_{i4}) + p_{i1}p_{i2}(1-p_{i3})(1-p_{i4})]/5\}}{\sum_{i=1}^{n}\{p_{i1}(1-p_{i2})(1-p_{i3})p_{i4} + [(1-p_{i1})p_{i2}(1-p_{i3})p_{i4} + (1-p_{i1})p_{i2}p_{i3}(1-p_{i4}) + (1-p_{i1})(1-p_{i2})p_{i3}p_{i4} + p_{i1}(1-p_{i2})p_{i3}(1-p_{i4}) + p_{i1}p_{i2}(1-p_{i3})(1-p_{i4})]/5\}}$$

$$D24 = \frac{\sum_{i=1}^{n}\{p_{i1}(1-p_{i2})p_{i3}(1-p_{i4}) - [(1-p_{i1})p_{i2}(1-p_{i3})p_{i4} + (1-p_{i1})p_{i2}p_{i3}(1-p_{i4}) + p_{i1}(1-p_{i2})(1-p_{i3})p_{i4} + (1-p_{i1})(1-p_{i2})p_{i3}p_{i4} + p_{i1}p_{i2}(1-p_{i3})(1-p_{i4})]/5\}}{\sum_{i=1}^{n}\{p_{i1}(1-p_{i2})p_{i3}(1-p_{i4}) + [(1-p_{i1})p_{i2}(1-p_{i3})p_{i4} + (1-p_{i1})p_{i2}p_{i3}(1-p_{i4}) + p_{i1}(1-p_{i2})(1-p_{i3})p_{i4} + (1-p_{i1})(1-p_{i2})p_{i3}p_{i4} + p_{i1}p_{i2}(1-p_{i3})(1-p_{i4})]/5\}}$$

$$D34 = \frac{\sum_{i=1}^{n}\{p_{i1}p_{i2}(1-p_{i3})(1-p_{i4}) - [(1-p_{i1})p_{i2}(1-p_{i3})p_{i4} + (1-p_{i1})p_{i2}p_{i3}(1-p_{i4}) + p_{i1}(1-p_{i2})(1-p_{i3})p_{i4} + p_{i1}(1-p_{i2})p_{i3}(1-p_{i4}) + (1-p_{i1})(1-p_{i2})p_{i3}p_{i4}]/5\}}{\sum_{i=1}^{n}\{p_{i1}p_{i2}(1-p_{i3})(1-p_{i4}) + [(1-p_{i1})p_{i2}(1-p_{i3})p_{i4} + (1-p_{i1})p_{i2}p_{i3}(1-p_{i4}) + p_{i1}(1-p_{i2})(1-p_{i3})p_{i4} + p_{i1}(1-p_{i2})p_{i3}(1-p_{i4}) + (1-p_{i1})(1-p_{i2})p_{i3}p_{i4}]/5\}}$$

The positive statistic DIJ represents excess allele sharing between taxa $L_I$ and $L_J$, indicating intertaxon gene flow. We tested the significance of DIJ statistics for each pair of taxa employing a jackknife bootstrap test using density distributions. In this test, 1000 pseudo-replicates from the original dataset were generated

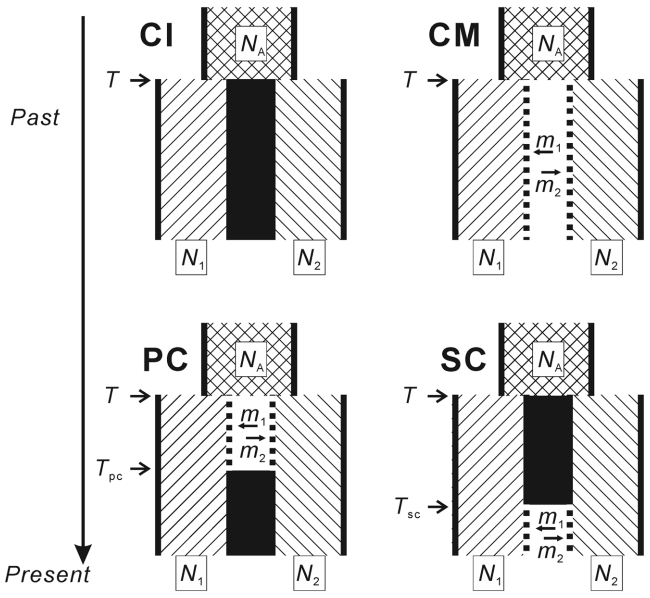


**Figure 3. Four models (evolutionary scenarios) used to test temporal patterns of divergence and migration: Complete isolation (CI), Continuous migration (CM), Primary contact (PC), and Secondary contact (SC). *T* is the divergence time in all models. $T_{PC}$ is the time when the two populations stopped exchanging migrants. $T_{SC}$ is the time when the two populations began exchanging migrants. $N_A$, $N_1$, and $N_2$ are the effective population sizes (for haploid genome) of the ancestral population and two daughter populations, respectively. The migration rates per generation per individual are denoted by $m_1$ and $m_2$.**

and density distributions, loci with positive DIJ values higher than the 97.5% quantiles were identified as introgressed between taxa $L_I$ and $L_J$. Having detected loci showing signals of introgression in this way, we computed population differentiation ($\Phi_{ST}$) between each pair of taxa using PopGenome (Excoffier et al. 1992; Pfeifer et al. 2014), at each of these loci, and then compared values of $\Phi_{ST}$ between different taxon pairs. Such comparisons were used to examine whether introgression between two taxa at a particular locus resulted in higher divergence of the introgressed taxon from a third taxon of the complex at this locus than at other loci. BLAST2GO (Conesa et al. 2005) was used to annotate the possible functions of putatively introgressed genes, while FLOR-ID (Bouché et al. 2016) was employed to examine the orthology of some of these genes to flowering-time genes in *Arabidopsis thaliana*.

# Results

### VARIANT CALLING, GENETIC DIVERSITY, AND CLUSTERING

RNA-seq produced an average of 53.49 million (M) raw reads with an average of 4.73 gillion (G) clean bases for each sampled individual of the *P. likiangensis* species complex, and 90.68 M raw reads with 6.48 G clean bases for the one sample of *P. breweriana* (Table S1). Variant calling analysis identified 480,258 SNPs with a total length of 16.30 Mb in all samples of the *P. likiangensis* species complex. After excluding loci indicative of paralogous signals or unequal expression bias, and sites not present in at least one of the 82 individuals of the species complex, 4743 loci with a total length of 11.9 M bases were left (Table 2). After filtering out polymorphic sites that deviated from Hardy–Weinberg equilibrium and loci < 200 bps in length, 3646 loci containing 84,793

SNPs remained for analysis. Final numbers of SNPs identified for each variety are listed in Table 2.

Estimates of nucleotide diversity were similar in all four taxa (Table 2). Tajima's *D* values were negative in all taxa, while Fay and Wu's *H* was negative in var. *complanata*, var. *likiangensis*, and var. *linzhiensis*, but positive in var. *rubescens*. However, MFDM tests showed no signal of positive selection in any taxon, in accordance with the large SDs calculated for both *D* and *H* (Table 2).

Genetic clustering of the 82 individuals using ADMIXTURE showed that the cross-validation error was lowest when the number of genetic clusters (*K*) was 4. With $K = 4$, most individuals of a taxon were assigned to a particular genetic cluster representing that taxon (Fig. 1), thus showing consistency with the morphologically based classification of taxa (Fu et al. 1999). However, 18 individuals (22% of the total) showed significant signals of admixture (*Q*-values $< 0.99$; Bootstrap test, $P < 0.05$; Fig. 1) and therefore were taken to be hybrids. Five of these were hybrids between var. *complanata* and var. *likiangensis*, three were hybrids between var. *complanata* and var. *linzhiensis*, six were hybrids between var. *likiangensis* and var. *rubescens*, and four were hybrids between var. *linzhiensis* and var. *rubescens*. No hybrid was detected between var. *complanata* and var. *rubescens* or between var. *likiangensis* and var. *linzhiensis* among the individuals surveyed.

## TESTING RETICULATE EVOLUTION AND TREE-LIKE DIVERGENCE MODELS

For all four taxa, a comparison between reticulate and tree-like models of divergence showed that the AIC value of the reticulate model (Fig. 2) was the lowest (Table 3), suggesting that secondary contacts occurred after the four taxa diverged from the common ancestor of the *P. likiangensis* species complex. This initial divergence time ($T_{DIV}$) was dated to 2.75 ($\pm$ 0.03) million years ago (Mya), during the late Pliocene period, while the time of initial secondary contact between taxa ($T_{SC}$) was dated to 0.77 Mya, i.e., during the mid-Pleistocene, with relative $T_{SC}$ (the relative timing of secondary contacts to divergence time, $T_{SC}/T_{DIV}$) estimated to be 0.28 ($\pm$ 0.3) (Table 4). The maximum effective population size (*N*e) of var. *rubescens* (45,939 $\pm$ 15,840) and minimum effective population size in var. *linzhiensis* (10,685 $\pm$ 104) were consistent with previous estimations for these taxa (Li et al. 2013). The *N*e of the ancestral population was estimated to be 50,163 ($\pm$ 6106) (Table 4).

In the reticulate model, the estimated immigration rate into any taxon from the three other taxa combined was higher than one individual per generation during the period of secondary contact (Table 4b). In addition, our results indicate highly asymmetric gene flow between some taxon pairs. For example, the mean migration rate from var. *likiangensis* to var. *complanata* was higher than in the reverse direction. Furthermore, gene flow from vars. *likiangensis* and *linzhiensis* into other taxa was greater than that from var. *complanata* and var. *rubescens* into other taxa, although corresponding SDs were very large (Table 4a).

**Table 3. AIC values for the reticulate model and all possible tree-like models for the evolution of the four taxa.**

| Model | Topology | AIC |
|---|---|---|
| Reticulate | see figure 2 | **3403620** |
| Polytomous | (COM,LIK,LIN,RUB) | 3485790 |
| AB_C_D | (((COM,LIK),LIN),RUB) | 3590390 |
| AB_CD | ((COM,LIK),(LIN,RUB)) | 3723828 |
| AB_D_C | (((COM,LIK),RUB),LIN) | 3546428 |
| AC_B_D | (((COM,LIN),LIK),RUB) | 3580112 |
| AC_BD | ((COM,LIN),(LIK,RUB)) | 3773072 |
| AC_D_B | (((COM,LIN),RUB),LIK) | 3602820 |
| AD_B_C | (((COM,RUB),LIK),LIN) | 3560660 |
| AD_BC | ((COM,RUB),(LIK,LIN)) | 3763128 |
| AD_C_B | (((COM,RUB),LIN),LIK) | 3675782 |
| BC_A_D | (((LIK,LIN),COM),RUB) | 3588372 |
| BC_AD | ((LIK,LIN),(COM,RUB)) | 3614766 |
| BC_D_A | (((LIK,LIN),RUB),COM) | 3581342 |
| BD_A_C | (((LIK,RUB),COM),LIN) | 3585060 |
| BD_AC | ((LIK,RUB),(COM,LIN)) | 3617176 |
| BD_C_A | (((LIK,RUB),LIN),COM) | 3600046 |
| CD_A_B | (((LIN,RUB),COM),LIK) | 3611890 |
| CD_AB | ((LIN,RUB),(COM,LIK)) | 3592800 |
| CD_B_A | (((LIN,RUB),LIK),COM) | 3587710 |

**Note: model AB_CD indicates that the divergence time between varieties A and B was more recent than divergence between C and D; model AB_C_D indicates that the first divergence event in coalescent time occurred between A and B, as shown in the topology. Bold type indicates the lowest AIC value. Abbreviations are: COM = var. *complanata*, LIK = var. *likiangensis*, LIN = var. *linzhiensis*, RUB = var. *rubescens*.**

## INTROGRESSION PATTERNS BETWEEN EACH PAIR OF TAXA

For each of the six pairs of taxa, we tested for patterns of introgression using coalescent simulations. The Secondary Contact (SC) model (Fig. 3) had lower AICs than the Complete Isolation (CI), Continuous Migration (CM), and Primary Contact (PC) models for each taxon-pair and each "C" parameter value (Table S2), indicating again that a period of secondary contact occurred between all four taxa during the history of the *P. likiangensis* species complex. Although, in the SC model the estimated divergence time between var. *linzhiensis* and var. *likiangensis* was smaller than those between var. *linzhiensis* and either var. *complanata* or var. *rubescens*, more recent divergence times were evident between var. *likiangensis* and both var. *complanata* and var. *rubescens*, indicating that times of initial divergence between each pair of taxa were similar and unordered (Table S3).

**Table 4a.** Estimated parameters of the reticulate evolutionary model illustrated in Figure 2. A generation time of 50 years (y) per generation (g) was assumed.

| Parameter | $N_{COM}$ | $N_{LIK}$ | $N_{LIN}$ | $N_{RUB}$ | $N_{anc}$ | Relative $T_{sc}$ | $T_{sc}$ (g) | $T_{DIV}$ (g) | $T_{sc}$ (y) | $T_{DIV}$ (y) |
|---|---|---|---|---|---|---|---|---|---|---|
| Mean | 28,036 | 17,881 | 10,685 | 45,939 | 50,163 | 0.28 | 15,364.50 | 55,091.05 | 768,225 | 2,754,553 |
| S.D. | 3,564 | 645 | 104 | 15,840 | 6,106 | 0.03 | 1,902.30 | 647.28 | 95,115 | 32,364 |

**Table 4b.** Mean migration rates per million generations and corresponding SDs for each pair of varieties in Figure 2.

| Migration rate matrix | Mean | | | | SD | | | |
|---|---|---|---|---|---|---|---|---|
| | From COM | From LIK | From RUB | From LIN | From COM | From LIK | From RUB | From LIN |
| $M_{COM}$ | – | 225.42 | 0.05 | 105.04 | – | 201.88 | 0.04 | 135.18 |
| $M_{LIK}$ | 0.74 | – | 0.33 | 84.66 | 0.71 | – | 0.43 | 132.75 |
| $M_{RUB}$ | 0.07 | 130.32 | – | 64.70 | 0.13 | 88.65 | – | 101.33 |
| $M_{LIN}$ | 33.41 | 159.57 | 0.51 | – | 102.57 | 129.41 | 0.73 | – |

### INTER-TAXON GENE FLOW AND TAXON DIFFERENTIATION

Analysis by means of modified ABBA–BABA statistics and jackknife tests of signicance showed that gene flow between vars. *complanata* and *likiangensis* (D12 = 0.288) and between vars. *linzhiensis* and *rubescens* (D34 = 0.098) was significant (Fig. 4). By comparing the density distribution and DIJ values of all 3646 loci, we identified a total of 3314 genes that showed signals of introgression between each of six taxon pairs. Following this, we examined the frequency-dynamics of such introgressed alleles by comparing taxon differentiation ($\Phi_{ST}$) per locus between each pair of taxa in turn (Fig. 4). When $\Phi_{ST}$ between vars. *linzhiensis* and *complanata* was compared with $\Phi_{ST}$ between vars. *linzhiensis* and *likiangensis*, it was evident that 32 genes, showing signals of introgression between vars. *complanata* and *likiangensis*, were located in the highest 5% tail of taxon differentiation between vars. *complanata* and *linzhiensis* and also in the highest 5% tail of taxon differentiation between vars. *likiangensis* and *linzhiensis*. This indicates that these 32 alleles occur at high frequencies in both vars. *complanata* and *likiangensis* as a result of introgression between these two taxa. Moreover, because of this, both taxa are highly differentiated from var. *linzhiensis*, which may contain the same alleles at either low frequency or not at all. Thus, it is apparent that in this case introgression most likely resulted in an increase in divergence of the introgressed pair (var. *complanata* and var. *likiangensis*) from var. *linzhiensis*. Similarly, significant gene flow between var. *linzhiensis* and var. *rubescens* likely resulted in this taxon pair becoming highly differentiated from var. *complanata* at 24 putatively introgressed loci between var. *linzhiensis* and var. *rubescens.* We annotated the functions of 53 of these putatively introgressed genes (Table S4) and found that three were highly orthologous to certain genes involved in flowering time in *Arabidopsis* (identified by FLOR-ID with *E*-value > 5), suggesting that they might, in turn, be involved in controlling flowering time differences and therefore prezygotic reproductive isolation among the four taxa comprising the *P. likiangensis* complex (Table S4).

## Discussion

Our population genomic analysis combined with tests of different evolutionary scenarios using coalescent simulations indicated that the *Picea likiangensis* complex, comprising three varieties of *P. likiangensis* (vars. *likiangensis, linzhiensis*, and *rubescens*) and one variey of *P. brachytyla* (var. *complanata*), originated rapidly at the end of the Pliocene and beginning of the Pleistocene. Following an initial allopatric phase, our analyses indicate that the four taxa came into secondary contact during the mid-Pleistocene and that their divergence, thereafter, has been influenced by hybridization and gene exchange (see Leroy et al. 2017, for a similar example of allopatric divergence followed by secondary contact among oaks). Signatures of high levels of gene flow were evident between vars. *rubescens* and *linzhiensis* and between vars. *complanata* and *likiangensis* that would explain the incongruent phylogenetic relationships resolved among the four taxa in previous studies (Li et al. 2013; Ru et al. 2016). Importantly, our analyses indicated that some introgressed alleles reached high frequencies in certain taxon pairs (vars. *complanata* and *likiangensis*, and vars. *rubescens* and *linzhiensis*, respectively) causing both taxa of such pairs to show increased divergence when compared with other taxa of the complex in which the same alleles were at low frequency or absent. To our knowledge, this is the first study to show that hybridization during secondary contacts can

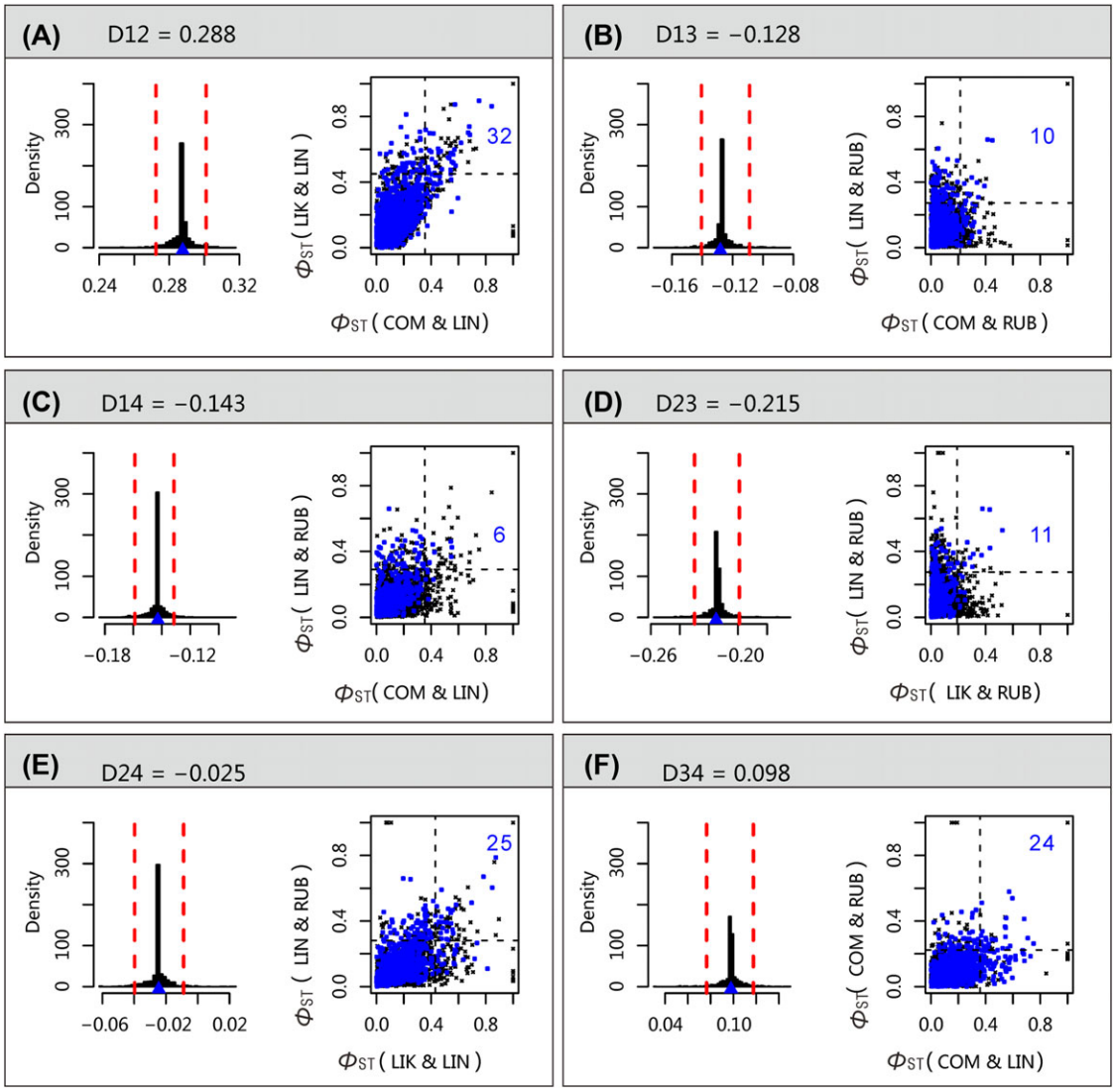


**Figure 4.** Tests of gene flow among each pair of taxa and population differentiation at putatively introgressed loci. Abbreviations for the four taxa: 1, var. *complanata*; 2, var. *likiangensis*; 3, var. *linzhiensis*; 4, var. *rubescens*. Positive DIJ represents gene flow between a pair of taxa I and J (I = 1, 2, 3; J = 2, 3, 4). For each pair of taxa, 1000 jackknife replicates were used to generate each density distribution, where the broken lines indicate the 95% confidence intervals and solid triangles denote observed value for each pair (shown on left side of each panel, A–F). For each pair of taxa, we also computed the $\Phi_{ST}$ per locus and then plotted values for a particular taxon pair against another pair where one taxon was common to both pairs while the other two taxa involved were different (shown on right side of each panel, A–F). This was done for both putative introgressed loci (blue squares) and non-introgressed loci (black crosses). Thus, in panel A the plot is shown of values of $\Phi_{ST}$ per locus for var. *likiangensis* verus var. *linzhiensis* against $\Phi_{ST}$ values for var. *complanata* versus var. *linzhiensis*. Broken lines within each plot represent the 95% $\Phi_{ST}$-quantiles for all loci. Blue integers represent the number of putative introgressed loci for which $\Phi_{ST}$ is higher than 95% $\Phi_{ST}$-quantiles in two dimensions.

promote differentiation among multiple related taxa as a result of introgression occurring between particular taxon pairs within a complex.

## RAPID DIVERSIFICATION AND RETICULATION

Rapid diversification giving rise to an evolutionary radiation has traditionally been inferred from a phylogenetic pattern where the number of DNA substitutions accumulated among taxon sequences is small causing branching to be compressed and nodes often poorly supported (Richardson et al. 2001). However, such poorly resolved or unresolved phylogenies can also arise from hybridization and introgression (Szöllősi et al. 2015). Indeed, evidence for hybridization among taxa has been detected in some groups where evolutionary radiations have occurred (Comes and Abbott 2001; Seehausen 2004; Wan et al. 2014; Lamichhaney et al. 2015). Therefore, it is necessary to consider the possible

effects of hybridization and gene flow when investigating the diversification of groups of closely related taxa.

Using the *P. likiangensis* species complex as a case study, we tested which evolutionary model best explained its origin and evolution. We concluded that following an initial phase of rapid divergence in allopatry, the evolution of the four taxa within this complex was influenced by hybridization and gene flow based on the following evidence. First, approximately 22% of the 82 trees examined in the complex were identified as hybrids (or backcross descendents) according to ADMIXTURE analysis. These putative hybrids occurred in regions where the distributions of different taxa were adjacent or overlapped, indicating that different pairs of taxa hybridize when in contact. Second, a comparison of possible evolutionary relationships among all four taxa showed that a reticulate model involving initial divergent radiation followed by secondary contact and gene exchange among taxa was best supported (Fig. 2, Table 3). Third, coalescent simulations for each pair of taxa consistently suggested that models allowing gene flow between taxa were better supported than a model assuming complete isolation (CI, Fig. 3). Furthermore, a model assuming secondary contact (SC, Fig. 3) following an initial period of allopatric divergence was better supported (Table S2), than those assuming continual migration (CM, Fig. 3) or initial divergence with gene flow followed by complete isolation (Primary Contact, PC, in Fig. 3).

Analyses using modified ABBA–BABA statistics showed significant signatures of gene flow had occurred between vars. *rubescens* and *linzhiensis* and between vars. *complanata* and *likiangensis*, respectively (Fig. 4). Generally, it is assumed that gene flow will lead to a reduction of population differentiation between related taxa in areas of overlap and gene exchange (Abbott et al. 2013; Arnold 2015). However, it is feasible that such gene flow could increase the divergence of a hybridizing taxon pair from other taxa in a complex at loci where introgression has occurred. Our results showed this was the case for 32 genes introgressed between vars. *complanata* and *likiangensis*. For these genes this pair of taxa showed significantly increased divergence from var. *linzhiensis*. Similarly, for 24 genes introgressed between vars. *rubescens* and *linzhiensis*, these taxa showed significantly increased divergence from var. *likiangensis* (Fig. 4). Because our coalescent simulations indicated that the major direction of introgression was from var. *likiangensis* to var. *complanata* in this pair of taxa, and from var. *linzhiensis* into var. *rubescens* in this taxon pair (Table 4), it seems that introgression from var. *likiangensis* will have increased the divergence of var. *complanat*a from var. *linzhiensis*, while introgression from var. *linzhiensis* has elevated divergence of var. *rubescens* from var. *likiangensis*. Although the functions of these introgessed genes within the complex require detailed analysis, we found that three of them were orthologous to those involved in the control of flowering time in *A. thaliana* (Table S4). It is possible, therefore, they are similarly involved in controlling differences in flowering time within the *P. likiangensis* complex and consequently might play a partial role in the evolution of prezygotic reproductive isolation between taxa. In summary, our results are consistent with the hypothesis that gene flow occurring among taxa comprising the *P. likiangensis* complex will have greatly influenced their pattern of divergence.

## EVOLUTION OF THE *P. likiangensis* SPECIES COMPLEX

We dated the origin and initial divergence of taxa within the *P. likiangensis* complex to the late Pliocene, 2.75 (± 0.03) Mya, which is broadly consistent with a previous estimate of 2.05–5.2 Mya based on population variation for 13 nuclear genes (Li et al. 2013). According to palynological records, *Picea* pollen was widespread in the Hengduan and Himalayan mountain regions from the Pliocene into the Pleistocene (Lü et al. 2001; Jiang et al. 2010; Xu et al. 2010). Therefore, the ancestor of the *P. likiangensis* complex might have been widely distributed in this region before undergoing rapid divergence and speciation over a relatively short timescale triggered by climatic shifts and possible geographical changes occurring during the Pliocene–Pleistocene period (Spicer et al. 2003; Clark et al. 2005; Wang et al. 2012; Wang et al. 2014a; Wang et al. 2014b). According to our analysis, this divergence most likely occurred initially under conditions of spatial population isolation and therefore in the absence of gene flow as assumed in the SC model (Figs. 2 and 3). Under such conditions of allopatry, isolated populations may have diverged rapidly in response to selection for adaptation to different ecological niches and the effects of genetic drift (Li et al. 2013; Wang et al. 2017).

Initial secondary contact among the four taxa was dated to approximately 0.77 (± 0.10) Mya (Table 4), coinciding with the start of the major glacial cycles during the Pleistocene (Elderfield et al. 2012; Martínez-Botí et al. 2015). The relative distributions of the different taxa are likely to have been markedly affected by Pleistocene climatic oscillations causing taxa to come into secondary contact at different times allowing hybridization and gene exchange to occur (Arnold 2016). Previous studies of the *P. likiangensis* complex indicated that changes in the distributions of taxa most likely occurred in response to climate change during recent glacial cycles (Li et al. 2013; Sun et al. 2015). Such cycles, beginning in the mid-Pleistocene, are likely therefore to have played a critical role in shaping the reticulate evolution of the *P. likiangensis* species complex.

Our coalescent estimations suggest that gene exchange among the four taxa continued over a lengthy period, approximately 0.77 million years, representing 28% of the period after the taxa first diverged from each other (Table 4). Theory predicts that some alleles would be fixed or increase to high frequencies within different taxa over such a timescale by random drift

and/or other evolutionary drivers (Durrett 2008). Pairwise comparisons between taxa showed divergence among taxa was not elevated by the introgression of many alleles, with introgression restricted to areas of parapatry (Figs. 1 and 4). Nonetheless, some introgressed alleles showed high levels of divergence among certain taxon pairs (Fig. 4), indicating that introgression of these alleles had contributed to divergence within the complex (see above). Although the MFDM test showed no obvious signal of selection for each introgressed gene, a recent investigation proposed that uncertainty in the inference of derived states from one outgroup may lead to low accuracy in detecting such signals (Keightley et al. 2016). Consequently, detailed investigations of selection are required to determine whether or not introgressed alleles that contributed to divergence have been subject to selection.

Our refined understanding of reticulate relationships among the four taxa comprising the *P. likiangensis* complex helps shed new light on the origin of variation in density of leaf stomatal lines within the complex, which might be critical for adaptation to past and future changes in environmental moisture level (Beerling and Kelly 1997; Chen et al. 2017). Whereas vars. *rubescens*, *likiangensis*, and *linzhiensis* of *P. likiangensis* show a reduced density of stomatal lines on the abaxial surfaces of leaves, they are completely absent from leaves of *P. brachytyla* var. *complanata*. Ru et al. (2016) previously suggested that adaptive introgression may have caused the trait to be absent in var. *complanata* (Ru et al. 2016), however the secondary contact model supported by the present population genomic analysis suggests that variation in density of stomatal lines on the abaxial surfaces of leaves could have been present among ancestral populations that gave rise to the four taxa. Thus, the loss of stomatal lines from leaves in extant populations of var. *complanata* could have arisen at an early stage in the formation of the four taxa and before they came into secondary contact approximately 0.77 Mya.

In summary, our study has shown that a population genomic analysis combined with coalescent simulations provides a powerful means of statistically detecting signals of reticulate evolution and distinguishing between this type of evolution and a tree-like pattern within closely related groups of taxa. In particular, our analysis has provided a deeper understanding of the reticulate relationships among the four taxa comprising the *P. likiangensis* complex, and how reticulate evolution within the complex may have been affected by climatic oscillations and geological changes that occurred in the Hengduan and Himalayan regions during the Pleistocene.

## AUTHOR CONTRIBUTIONS

J.L. and Y.S. designed the study. Y.S., K.M., L.Z., and Z.L. collected materials and performed the experiments. Y.S. proposed the modified statistics and wrote the scripts. Y.S., X.W. ,and D.R. performed the analyses. J.L., R.A., and Y.S. wrote the manuscript.

## ACKNOWLEDGMENTS

The authors thank staff at the Public Technology Service Centre, XTBG-CAS, for their assistance in using the HPC Platform. The authors thank Maria Servedio, Carole Smadja, Stuart Baird, and anonymous referees for their excellent suggestions. This work was supported by grants from National Key Research and Development Program (2017YFC0505203), National Natural Science Foundation of China (grant numbers 31590821, 31670665, 91731301), National Key Project for Basic Research (2014CB954100), "1000 Youth Talents Plan" of Yunnan Province and CAS "Light of West China" Program. The authors declare no conflict of interest.

## DATA ARCHIVING

All new RNA-seq data sets used in this study have been submitted to the NCBI SRA under BioProject Accession no. PRJNA392950 and PRJNA378930. The doi for our data is https://doi.org/10.5061/dryad.gb6c3b0.

## LITERATURE CITED

Abbott, R. J. 2017. Plant speciation across environmental gradients and the occurrence and nature of hybrid zones. J. Syst. Evol. 55:238–258.

Abbott, R. J., D. Albach, S. Ansell, J. Arntzen, S. Baird, N. Bierne, J. Boughman, A. Brelsford, C. Buerkle, and R. Buggs. 2013. Hybridization and speciation. J. Evol. Biol. 26:229–246.

Abbott, R. J., N. H. Barton, and J. M. Good. 2016. Genomics of hybridization and its evolutionary consequences. Mol. Ecol. 25:2325–2332.

Alexander, D. H. and K. Lange. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics 12:246.

Andrew, R. L., L. Bernatchez, A. Bonin, C. A. Buerkle, B. C. Carstens, B. C. Emerson, D. Garant, T. Giraud, N. C. Kane, and S. M. Rogers. 2013. A road map for molecular ecology. Mol. Ecol. 22:2605–2626.

Arnold, M. L. 2015. Divergence with genetic exchange. Oxford Univ. Press, Oxford, U.K.

———. 2016. Anderson's and Stebbins' prophecy comes true: genetic exchange in fluctuating environments. Syst. Bot. 41:4–16.

Arnold, M. L. and K. Kunte. 2017. Adaptive genetic exchange: a tangled history of admixture and evolutionary innovation. Trends Ecol. Evol. 32:601–611.

Beerling, D. and C. Kelly. 1997. Stomatal density responses of temperate woodland plants over the past seven decades of $CO_2$ increase: a comparison of Salisbury (1927) with contemporary data. Am. J. Bot. 84:1572–1572.

Bouché, F., G. Lobet, P. Tocquin, and C. Périlleux. 2016. FLOR-ID: an interactive database of flowering-time gene networks in *Arabidopsis thaliana*. Nucleic Acids Res. 44:D1167–D1171.

Chen, Z.-H., G. Chen, F. Dai, Y. Wang, A. Hills, Y.-L. Ruan, G. Zhang, P. J. Franks, E. Nevo, and M. R. Blatt. 2017. Molecular evolution of grass stomata. Trends Plant Sci. 22:124–139.

Clark, M. K., M. House, L. Royden, K. Whipple, B. Burchfiel, X. Zhang, and W. Tang. 2005. Late Cenozoic uplift of southeastern Tibet. Geology 33:525–528.

Comes, H. P. and R. J. Abbott. 2001. Molecular phylogeography, reticulation and lineage sorting in the Mediterranean species complex of *Senecio* sect. *Senecio* (Asteraceae). Evolution 55:1943–1962.

Conesa, A., S. Götz, J. M. García-Gómez, J. Terol, M. Talón, and M. Robles. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676.

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

De La Torre, A. R., Z. Li, Y. Van de Peer, and P. K. Ingvarsson. 2017. Contrasting rates of molecular evolution and patterns of selection among gymnosperms and flowering plants. Mol. Biol. Evol. 34:1363–1377.

Delhomme, N., G. Sundström, N. Zamani, H. Lantz, Y.-C. Lin, T. R. Hvidsten, M. P. Höppner, P. Jern, Y. Van de Peer, and J. Lundeberg. 2015. Serendipitous meta-transcriptomics: the fungal community of Norway spruce (*Picea abies*). PloS One 10:e0139080.

Durand, E. Y., N. Patterson, D. Reich, and M. Slatkin. 2011. Testing for ancient admixture between closely related populations. Mol. Biol. Evol. 28:2239–2252.

Durrett, R. 2008. Probability models for DNA sequence evolution. Springer, New York.

Elderfield, H., P. Ferretti, M. Greaves, S. Crowhurst, I. McCave, D. Hodell, and A. Piotrowski. 2012. Evolution of ocean temperature and ice volume through the mid-Pleistocene climate transition. Science 337:704–709.

Excoffier, L., I. Dupanloup, E. Huerta-Sánchez, V. C. Sousa, and M. Foll. 2013. Robust demographic inference from genomic and SNP data. PLoS Genet. 9:e1003905.

Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491.

Fay, J. C. and C.-I. Wu. 2000. Hitchhiking under positive Darwinian selection. Genetics 155:1405–1413.

Fu, L. G., N. Li, and T. S. Elias. 1999. Pinaceae. Pp. 11–52 *in* Z. Wu, and P. H. Raven, eds. Flora of China. Science Press, Beijing.

Gayral, P., J. Melo-Ferreira, S. Glémin, N. Bierne, M. Carneiro, B. Nabholz, J. M. Lourenco, P. C. Alves, M. Ballenghien, and N. Faivre. 2013. Reference-free population genomics from next-generation transcriptome data and the vertebrate–invertebrate gap. PLoS Genet. 9:e1003457.

Givnish, T. J. and K. J. Sytsma. 2000. Molecular evolution and adaptive radiation. Cambridge Univ. Press, Cambridge, U.K.

Grant, B. R. and P. R. Grant. 2017. Watching speciation in action. Science 355:910–911.

Harrison, R. G. and E. L. Larson. 2014. Hybridization, introgression, and the nature of species boundaries. 105:795–809.

Jiang, L., F. Schlesinger, C. A. Davis, Y. Zhang, R. Li, M. Salit, T. R. Gingeras, and B. Oliver. 2011. Synthetic spike-in standards for RNA-seq experiments. Genome Res. 21:1543–1551.

Jiang, S., S. Xiang, and Y. Xu. 2010. Geological significances of Late Pliocene: early Pleistocene palynological assemblages in the Zanda Basin, Tibet (China). Geol. Sci. Technol. Inform. 29:21–31.

Keightley, P. D., J. L. Campos, T. R. Booker, and B. Charlesworth. 2016. Inferring the frequency spectrum of derived variants to quantify adaptive molecular evolution in protein-coding genes of *Drosophila melanogaster*. Genetics 203:975–984.

Lü, H., S. Wang, N. Wu, G. Tong, X. Yang, C. Sheng, S. Li, L. Zhu, and L. Wang. 2001. A new pollen record of the last 2.8 Ma from the Co Ngoin, central Tibetan Plateau. Sci. China D 44:292–300.

Lamichhaney, S., J. Berglund, M. S. Almén, K. Maqbool, M. Grabherr, A. Martinez-Barrio, M. Promerová, C.-J. Rubin, C. Wang, and N. Zamani. 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature 518:371–375.

Leroy, T., C. Roux, L. Villate, C. Bodénès, J. Romiguier, J. A. Paiva, C. Dossat, J. M. Aury, C. Plomion, and A. Kremer. 2017. Extensive recent secondary contacts between four European white oak species. New Phytol. 214:865–878.

Lexer, C., F. Marthaler, S. Humbert, T. Barbará, M. Harpe, E. Bossolini, M. Paris, G. Martinelli, and L. M. Versieux. 2016. Gene flow and diversification in a species complex of Alcantarea inselberg bromeliads. Bot. J. Linn. Soc. 181:505–520.

Li, H. 2011. A new test for detecting recent positive selection that is free from the confounding impacts of demography. Mol. Biol. Evol. 28:365–375.

Li, H. and R. Durbin. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079.

Li, L., R. J. Abbott, B. Liu, Y. Sun, L. Li, J. Zou, X. Wang, G. Miehe, and J. Liu. 2013. Pliocene intraspecific divergence and Plio-Pleistocene range expansions within *Picea likiangensis* (Lijiang spruce), a dominant forest tree of the Qinghai-Tibet Plateau. Mol. Ecol. 22:5237–5255.

Li, Y., M. Stocks, S. Hemmilä, T. Källman, H. Zhu, Y. Zhou, J. Chen, J. Liu, and M. Lascoux. 2010. Demographic histories of four spruce (*Picea*) species of the Qinghai-Tibetan Plateau and neighboring areas inferred from multiple nuclear loci. Mol. Biol. Evol. 27:1001–1014.

Linder, C. R. and L. H. Rieseberg. 2004. Reconstructing patterns of reticulate evolution in plants. Am. J. Bot. 91:1700–1708.

Lockwood, J. D., J. M. Aleksić, J. Zou, J. Wang, J. Liu, and S. S. Renner. 2013. A new phylogeny for the genus *Picea* from plastid, mitochondrial, and nuclear sequences. Mol. Phylogenet. Evol. 69:717–727.

Mallet, J. 2007. Hybrid speciation. Nature 446:279–283.

Martínez-Botí, M., G. L. Foster, T. Chalk, E. Rohling, P. Sexton, D. J. Lunt, R. Pancost, M. Badger, and D. Schmidt. 2015. Plio-Pleistocene climate sensitivity evaluated using high-resolution $CO_2$ records. Nature 518: 49–54.

Martin, S. H., J. W. Davey, and C. D. Jiggins. 2015. Evaluating the use of ABBA–BABA statistics to locate introgressed loci. Mol. Biol. Evol. 32:244–257.

Nei, M. 1987. Molecular evolutionary genetics. Columbia Univ. Press, New York.

Payseur, B. A. and L. H. Rieseberg. 2016. A genomic perspective on hybridization and speciation. Mol. Ecol. 25:2337–2360.

Pease, J. B., D. C. Haak, M. W. Hahn, and L. C. Moyle. 2016. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. PloS Biol. 14:e1002379.

Pfeifer, B., U. Wittelsbürger, S. E. R. Onsins, and M. J. Lercher. 2014. PopGenome: an efficient Swiss army knife for population genomic analyses in R. Mol. Biol. Evol. 31:1929–1936.

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

Ran, J.-H., T.-T. Shen, W.-J. Liu, P.-P. Wang, and X.-Q. Wang. 2015. Mitochondrial introgression and complex biogeographic history of the genus *Picea*. Mol. Phylogenet. Evol. 93:63–76.

Richardson, J. E., R. T. Pennington, T. D. Pennington, and P. M. Hollingsworth. 2001. Rapid diversification of a species-rich genus of neotropical rain forest trees. Science 293:2242–2245.

Rieseberg, L. H. and J. H. Willis. 2007. Plant Speciation. Science 317:910–914.

Ru, D., K. Mao, L. Zhang, X. Wang, Z. Lu, and Y. Sun. 2016. Genomic evidence for polyphyletic origins and inter-lineage gene flow within complex taxa: a case study of *Picea brachytyla* in the Qinghai-Tibet Plateau. Mol. Ecol. 25:2373–2386.

Seehausen, O. 2004. Hybridization and adaptive radiation. Trends Ecol. Evol. 19:198–207.

Spicer, R. A., N. B. Harris, M. Widdowson, and A. B. Herman. 2003. Constant elevation of southern Tibet over the past 15 million years. Nature 421:622–624.

Suarez-Gonzalez, A., C. Lexer, and Q. C. B. Cronk. 2018a. Adaptive introgression: a plant perspective. Biol. Lett. 14:2017068.

Suarez-Gonzalez, A., C. A. Hefer, C. Lexer, Q. Cronk, and C. J. Douglas. 2018b. Scale and direction of adaptive introgression between black cottonwood (*Populus trichocarpa*) and balsam poplar (*P. balsamifera*). Mol. Ecol. 27:1667–1680.

Sun, Y., L. Li, L. Li, J. Zou, and J. Liu. 2015. Distributional dynamics and interspecific gene flow in *Picea likiangensis* and *P. wilsonii* triggered by climate change on the Qinghai–Tibet Plateau. J. Biogeogr. 42:475–484.

Szöllősi, G. J., E. Tannier, V. Daubin, and B. Boussau. 2015. The inference of gene trees with species trees. Syst. Biol. 64:e42–e62.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595.

Tsagkogeorga, G., V. Cahais, and N. Galtier. 2012. The population genomics of a fast evolver: high levels of diversity, functional constraint, and molecular adaptation in the tunicate *Ciona intestinalis*. Genome Biol. Evol. 4:852–861.

Wan, D., Y. Sun, X. Zhang, X. Bai, J. Wang, A. Wang, and R. Milne. 2014. Multiple ITS copies reveal extensive hybridization within *Rheum* (Polygonaceae), a genus that has undergone rapid radiation. PloS One 9: e89769.

Wang, C., J. Dai, X. Zhao, Y. Li, S. A. Graham, D. He, B. Ran, and J. Meng. 2014a. Outward-growth of the Tibetan Plateau during the Cenozoic: a review. Tectonophysics 621:1–43.

Wang, E., E. Kirby, K. P. Furlong, M. Van Soest, G. Xu, X. Shi, P. J. Kamp, and K. Hodges. 2012. Two-phase growth of high topography in eastern Tibet during the Cenozoic. Nat. Geosci. 5:640–645.

Wang, G.-H., H. Li, H.-W. Zhao, and W.-K. Zhang. 2017. Detecting climatically driven phylogenetic and morphological divergence among spruce (*Picea*) species worldwide. Biogeosciences 14:2307–2319.

Wang, P., D. Scherler, J. Liu-Zeng, J. Mey, J.-P. Avouac, Y. Zhang, and D. Shi. 2014b. Tectonic control of Yarlung Tsangpo Gorge revealed by a buried canyon in Southern Tibet. Science 346:978–981.

Wang, Z., M. Gerstein, and M. Snyder. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10:57–63.

Xu, Y., K. Zhang, G. Wang, S. Xiang, S. Jiang, and F. Chen. 2010. Geological significance of Miocene-early pleistocene palynological zones in the Gyirong basin, southern Tibet. Earth Sci. J. China U. Geosci. 35:759–773.

Associate Editor: C. Smadja
Handling Editor: M. Servedio

## *Supporting Information*

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Table S1**. Numbers of raw reads, clean reads, mapped reads, bases of mapped reads, and percentages of reads mapped to the reference (mapping ratios) for each sample.

**Table S2**. AICs of the four models in Figure 3 for each of six variety-pairs and "C" parameter values from 1 to 20.

**Table S3**. Estimated divergence times in generations (T) of the SC model in Figure 3 for each of six variety-pairs.

**Table S4**. GO analysis annotated 53 putatively introgressed genes that located in the highest 5% tail of taxon differentiation for any of six pairs of varieties.

**Text S1**. List of parameters for the reticulate evolutionary model (Fig. 2).

**Script S1.** R functions used in calculating six D statistics.

**Script S2.** Custom R script: bootstrap test based on 'jack.knife' method.