

# Origin and speciation of *Picea schrenkiana* and *Picea smithiana* in the Center Asian Highlands and Himalayas

Lili Li · Yongshuai Sun · Jiabin Zou · Wei Yue ·  
Xi Wang · Jianquan Liu

Published online: 17 August 2014

© The Author(s) 2014. This article is published with open access at Springerlink.com

**Abstract** Elucidating the evolutionary history of current species diversity, especially trees with large effective population sizes and long generation times, is a complicated exercise confounded by gene flow and incomplete lineage sorting. In the present study, we aim to determine the origin and speciation of *Picea schrenkiana* and *Picea smithiana* using population genetic data from chloroplast (cp), mitochondrial (mt), and nuclear (nr) genomes. These two species occur in the Central Asian Highlands and Himalayas, respectively, where they are isolated from other Asian congeneric species by the Qinghai–Tibet Plateau (QTP) or adjacent deserts. Previous studies based on both morphological and molecular evidence suggest that they have contrasting phylogenetic relationships with *Picea likiangensis* or *Picea wilsonii* which are closely related and both located in the QTP. We examined genetic variation among 16 loci of three genomes from 30 populations of these four species. At both cpDNA loci and mtDNA loci, *P. schrenkiana* appeared to be closely related to *P. likiangensis*, although statistical support for this was weak. However, phylogenetic analyses and speciation tests based on the nuclear data from 11 loci provided evidence that *P. schrenkiana* and *P. smithiana* are sister species. These two species diverged around five million years ago (Mya) while the divergence between them and the *P. likiangensis*–*P. wilsonii* clade occurred about 18.4 Mya. We also detected gene flow accompanying these speciation events. Our results highlight the complex speciation histories of these alpine conifers due to interspecific gene flow and/or incomplete

lineage sorting, and the importance of the early QTP uplifts in promoting the origin of these important conifer species in the Asian highlands.

**Keywords** Gene flow · *Picea schrenkiana* · *P. smithiana* · Population genetic data · Speciation test

## Introduction

Geographical or other selective barriers to genetic exchange may give rise to new species by isolating previously interbreeding populations (Coyne and Orr 2004). Such geographical isolation can play an important role in dividing a common ancestor into two (rarely more) sister species in which different random alleles become fixed and adaptations to different local conditions are selected (Mayr 1954; Grant 1981; Coyne 1992; Orr and Presgraves 2000). In addition, gene flow and incomplete lineage sorting complicate speciation process, especially in tree species with large effective population sizes and long generation times (Gavrilets 2003; Levin 2003; Rieseberg and Willis 2007). Evidence increasingly suggests that population-genetic data, derived from sequencing multiple loci, is the best approach for clarifying the origin and speciation of such tree species (e.g., Chen et al. 2010; Zou et al. 2013; Sun et al. 2014). Such data allow for alternative speciation models and estimation of the extent of gene flow that has accompanied speciation (e.g., Hey and Nielsen 2004, 2007; Hey 2006, 2010a, b; Li et al. 2010). In addition, estimates of divergence timescales based on analysis of multilocus population genetic data in a coalescent framework provide good temporal hierarchies for understanding the roles of geological events in triggering speciation (Wakeley 2003; Takayama et al. 2013).

Numerous alpine plant species are endemic in the Qinghai–Tibet Plateau (QTP), the Himalayas, and the Asian

**Electronic supplementary material** The online version of this article (doi:10.1007/s11105-014-0774-5) contains supplementary material, which is available to authorized users.

L. Li · Y. Sun · J. Zou · W. Yue · X. Wang · J. Liu (✉)  
State Key Laboratory of Grassland Agro-Ecosystems, School of Life Sciences, Lanzhou University, Lanzhou 730000, China  
e-mail: liujq@lzu.edu.cn

highlands—regions which together are recognized as one of the world's most important alpine biodiversity hotspots (Wilson 1992; Myers et al. 2000). These endemic species may have originated during and/or after the extensive uplifts of the QTP and Himalayas. However, the date when both the QTP and Himalayas were uplifted to their current heights remains debatable, three extensive uplifts have been dated between 22 and 20, 15–8 and 5–3 Mya (e.g., Shi et al. 1998; An et al. 2001; Guo et al. 2002; Buslov et al. 2007; Wang et al. 2008). Previous studies have found that a few herbaceous genera diversified greatly during these three stages, especially the latter two (e.g., Liu et al. 2002, 2006; Wang et al. 2009). Studies designed to examine the origin and speciation of alpine trees, especially studies based on population genetic data, are rare (but see Mao et al. 2010 and Xu et al. 2010). In the present paper, we examine the phylogenetic relationships and speciation patterns of two Asian highland spruce species: *Picea schrenkiana* and *Picea smithiana*. The genus *Picea* diversified greatly in Asia with 24 out of the total of 35 spruce species occurring there (Fu et al. 1999; Farjón 2001). Most of them are distributed in cold and temperate mountainous regions, especially in the QTP and adjacent highlands. However, *P. schrenkiana* and *P. smithiana* occur with narrow distributions in the Central Asian Highlands (Tian Shan Mountains; Zhang and Tang 1989) and the Himalayas, respectively. These two species are isolated from each other by the western Kunlun Mountains, and both are isolated from spruces in the QTP by the Himalaya and Tarim depressions. All these geographic isolations seem to have developed before the Pliocene (5.3–2.5 Mya; Fig. 1). For example, the QTP–Himalayas uplift and Tarim depression occurred at the same as the first extensive uplift of the QTP around 22 Mya (Guo et al. 2002). The second extensive QTP uplift, which occurred between 15 and 8 Mya, was accompanied by further increases of the western Kunlun Mountains and the Himalayas, as well as the Tarim depressions (Wang et al. 1990, 2006, 2008; Abdrakhmatov et al. 1996; Shi et al. 1998; An et al. 2001; Bullen et al. 2001, 2003; Charreau et al. 2006, 2009; Buslov et al. 2007; Dupont–Nivet et al. 2008). It is likely that the origin and speciation of *P. schrenkiana* and *P. smithiana* are correlated with geographical isolations consequent upon these geological events.

These two species are morphologically distinct from each other, despite having similar quadrangular leaves: *P. schrenkiana* usually has thick leaves while leaves of *P. smithiana* are linear and more slender. These two species were placed together in the section *Picea* (Liu 1982; Farjón 2001) along with other species (e.g., *Picea wilsonii*, *Picea neveitchii*, *Picea crassifolia*, and *Picea abies*) with quadrangular leaves and stomatal lines present and almost equal in number on each surface (Fu et al. 1999). Those species represented by *Picea likiangensis* of Sect. *Casieta* (Liu 1982; Farjón 2001) have flattened leaves with stomatal lines mostly

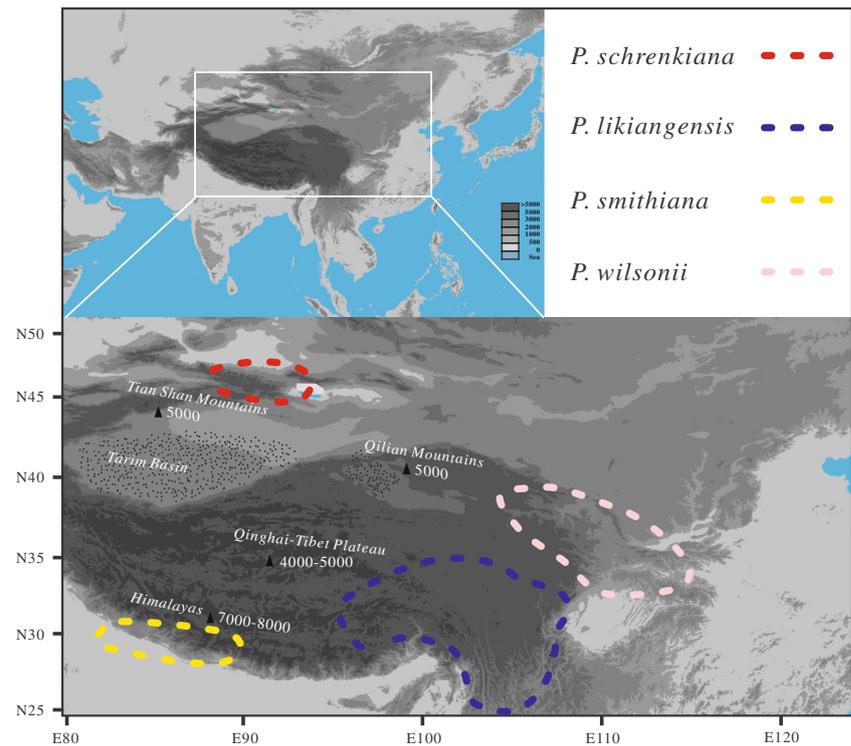
or only present on the adaxial surface. However, phylogenetic analyses based on chloroplast (cp) DNA sequence variations suggest that both *P. schrenkiana* and *P. smithiana* cluster together with *P. likiangensis* into a well-supported subclade, sister to that represented by *P. wilsonii*. The remaining species of Sect. *Picea* form a separate clade with distant relationships (Ran et al. 2006; Bouillé et al. 2011). In contrast, phylogenetic analyses based mainly on variations in the mitochondrial (mt) genome suggest that *P. schrenkiana* clusters into a subclade with *P. wilsonii* while *P. smithiana* should be placed at the basal position of the subclade containing *P. likiangensis* (Lockwood et al. 2013). Another recent phylogenetic analysis based on 11 concatenated nuclear gene sequences suggested that *P. schrenkiana* and *P. smithiana* clustered into a separate clade, sister to the other one comprising the other eight species (Sun et al. 2014). It should be noted that all other species, i.e., *Picea spinulosa*, *Picea farreri*, *Picea neveitchii*, *Picea morrisonicola*, *Picea purpurea*, and *Picea brachytyla* are closely related to *P. likiangensis* or *P. wilsonii*, or are derived from an original hybrid of these two key species (Liu 1982; Fu et al. 1999; Ran et al. 2006; Bouillé et al. 2011; Lockwood et al. 2013; Zou et al. 2013; Sun et al. 2014). We therefore chose *P. likiangensis* and *P. wilsonii* to represent the other species and further examined their phylogenetic relationships with *P. schrenkiana* and *P. smithiana* based on three genomes data. We sequenced three cpDNA loci, two mtDNA loci and 11 nuclear DNA loci from 30 populations comprising 330 individuals of four species. We used approximate Bayesian computation (ABC) methods (Wegmann et al. 2010) to test alternative speciation hypotheses between these four species, estimate the divergence timescales, and examine whether gene flow occurred during speciation events.

## Material and Methods

**Population Sampling** Populations of four species were sampled from across their major distributions and three varieties recognized under *P. likiangensis* were included (Fig. 1; Supplementary Table 1). The number of sampled individuals in each population was between 4 and 21, spaced  $\geq 100$  m apart. A total of 330 individuals were used in this study. Fresh leaf needles were dried and stored in the field using silica gel; seeds were stored at  $-20$  °C. The latitude, longitude, and altitude of each sampling location were measured using an Extrex GIS monitor (Germany) (Supplementary Table 1). One *Picea meyeri* individual was used as an outgroup for phylogenetic analyses.

**Sequencing and Phasing** We used either the modified cetyltrimethylammonium bromide (CTAB) procedure (Doyle and Doyle 1990) or a QIAGEN DNeasy Plant Mini Kit (QIAGEN, Valencia, CA, USA) to extract the total

**Fig. 1** Distributions of *P. schrenkiana*, *P. smithiana*, *P. likiangensis*, and *P. wilsonii*. *P. schrenkiana* distributed in Tian Shan Mountains and neighboring areas (TSMs), *P. smithiana* distributed in Himalaya Mountains, *P. likiangensis* and *P. wilsonii* distributed in Qinghai–Tibet plateau and adjacent regions (QTPs). Different colors represent different species. The mean altitudes of mountains are given in Arabic numerals. Extensive geographical barriers, including deserts, are indicated by the areas of black dots; high mountains between the TSMs and the QTPs are shown by small triangles



genomic DNA from needles. Eleven nuclear loci (*4CL*, *EBS*, *GI*, *MOO2*, *M007D1*, *Sb16*, *Sb29*, *Sb62*, *se1364*, *se1390*, and *xy1420*) were selected and sequenced following Li et al. (2010) using an ABI 3130xl or 3730xl Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) and an ABI Prism BigDye Terminator Cycle V3.1 Sequencing Kit. We amplified and sequenced three chloroplast loci (*trnL-trnF*, *trnS-trnG*, and *ndhK/C*; Taberlet et al. 1991; Hamilton 1999; Anderson et al. 2006) and two mitochondrial loci (*nad1* intron b/c and *nad5* intron1; Meng et al. 2007). Basecalling of nuclear genotypic sequences was done using Phred v0.020425.c (score >20) (Ewing et al. 1998) with CODONCODE ALIGNER software (CodonCode Corporation) and manually checked whether the SNPs consistent with the chromatogram peaks using the MEGA5 (Tamura et al. 2011). Sequences from cp and mt genomes were identified and checked using MEGA5. Finally, we used CLUSTAL W (Thompson et al. 1997) in MEGA5 to perform the alignment of sequence matrix at each locus.

Nuclear sequences with heterozygous sites were rephased and separated into two allele sequences by PHASE (Stephens et al. 2001, 2003) in the software package DnaSP V5 (Librado and Rozas 2009). A Hardy–Weinberg equilibrium was assumed and the phased sequences were inferred using a coalescent Bayesian method. Homozygous genotypes, and genotypes with single heterozygous sites, were set as known alleles to improve the performance of PHASE analyses. A general recombination model (Li and Stephens 2003) and 10 iterations of the final run were used and other optional parameters were

set as default. Runs were repeated twice with different seed numbers to ensure that results were robust. In this research, three nuclear fragments (*Sb16*, *Sb29*, and *Sb62*.) were cloned and 10 clones were sequenced.

All newly obtained sequences of *P. schrenkiana* and *P. smithiana* were submitted to GenBank (accession numbers KJ176997–KJ179244) and three genomes sequences of *P. likiangensis* and *P. wilsonii* were cited from Li et al. (2013) and Zou et al. (2013).

**Phylogenetic Analyses** For cpDNA, mtDNA, and nuclear DNA, we used NETWORK V4.2.1.1 (Bandelt et al. 1999) (available at <http://www.fluxus-engineering.com/>) to construct their genealogies. We further used a Bayesian coalescent-based method for species tree estimation based on three genomes data, as implemented in BEAST 1.7.2 package (Drummond and Rambaut 2007; available at [http://beast.bio.ed.ac.uk/Main\\_Page](http://beast.bio.ed.ac.uk/Main_Page)). The approach makes use of multilocus nuclear data, embedding individual gene trees in a multilocus species tree, allowing estimation of species genealogy (Hailer et al. 2012). Nonrecombining blocks of each locus were chosen using IMGIC software (Woerner et al. 2007) because recombination cannot be modeled in the coalescence approach in BEAST. Using a strict molecular clock and the models of sequence evolution indicated by jModeltest (Posada 2008), we performed three independent runs of  $2 \times 10^8$  generations and sampled parameters every 5,000 iterations, with a burn-in of 30 %. Convergence was checked in Tracer 1.4 (available at <http://beast.bio.ed.ac.uk/>)

Tracer). For both cpDNA and mtDNA, all indels were coded as “0” and “1”. Branch supports were also evaluated using the program MrBayes 3.1.2 (Huelsenbeck et al. 2001; Ronquist and Huelsenbeck 2003).

**Population Genetic Parameters and Neutrality Tests** We used DnaSP V5 (Librado and Rozas 2009) to calculate basic population genetic parameters: the number of segregating sites ( $S$ ), Watterson’s parameter ( $\theta_w$ , Watterson 1975), nucleotide diversity ( $\pi$ , Tajima 1983). For nuclear loci, we further showed the divergence between each pair of species using ARLEQUIN, version 3.0 (Excoffier et al. 2005) with significance tests based on 10,000 permutations.

Before estimating evolutionary history, we used a recently developed maximum frequency of derived mutation (MFDM) method (Li 2011) to test the neutrality of variation at each locus. The MFDM method is not affected by the impact of varying population size, such as expansion or shrinkage, which may produce similar signals to selection (Li 2011). Because the demographic dynamics cannot change the genealogies of a locus (Tajima 1983; Hein et al. 2004), an unbalanced topology shaped by recent selection could be captured and stand as evidence of non-neutrality. Simulation analysis and comparisons with several other methods showed higher power and a lower false-positive rate when detecting recent selection using the MFDM method (Li 2011). Here, we tested the neutrality per locus in each species with this method using a 5 % significance level. Ancestral states at variable sites were determined by comparison with two outgroup species (*Picea breweriana* and *P. meyeri*).

**Linkage Disequilibrium and Population Structure** Linkage disequilibrium (LD) was measured by  $r^2$ , the square of the correlation coefficient between each single nucleotide polymorphism (SNP) pair (Heuertz et al. 2006), using DnaSP V5 (Librado and Rozas 2009). The significance level of the statistical association between alleles at different sites was measured using Fisher’s exact test, and Bonferroni correction was used to correct for false positives.  $r^2=0$  means loci are in complete linkage equilibrium and  $r^2=1$  indicates loci are in complete linkage disequilibrium. However, the result showed that  $r^2$  values were very small, almost equal to zero, which indicated loci were nearly no linkage disequilibrium (Supplementary Table 2). Therefore, we used all SNPs of nuclear loci to identify the genetic structure and inter-relationships of four species at the population level by STRUCTURE V2.3.4 analysis (Pritchard et al. 2000; Hubisz et al. 2009). The most likely number of populations in the dataset ( $K$ ) was estimated by conducting 15 independent runs for each  $K$  value ranging from 1 to 6. Each run had a burn-in of 200,000 iterations and additional 500,000 iterations. Finally, we used Distruct v.1.1 (Rosenberg 2004) to draw the graphics. We

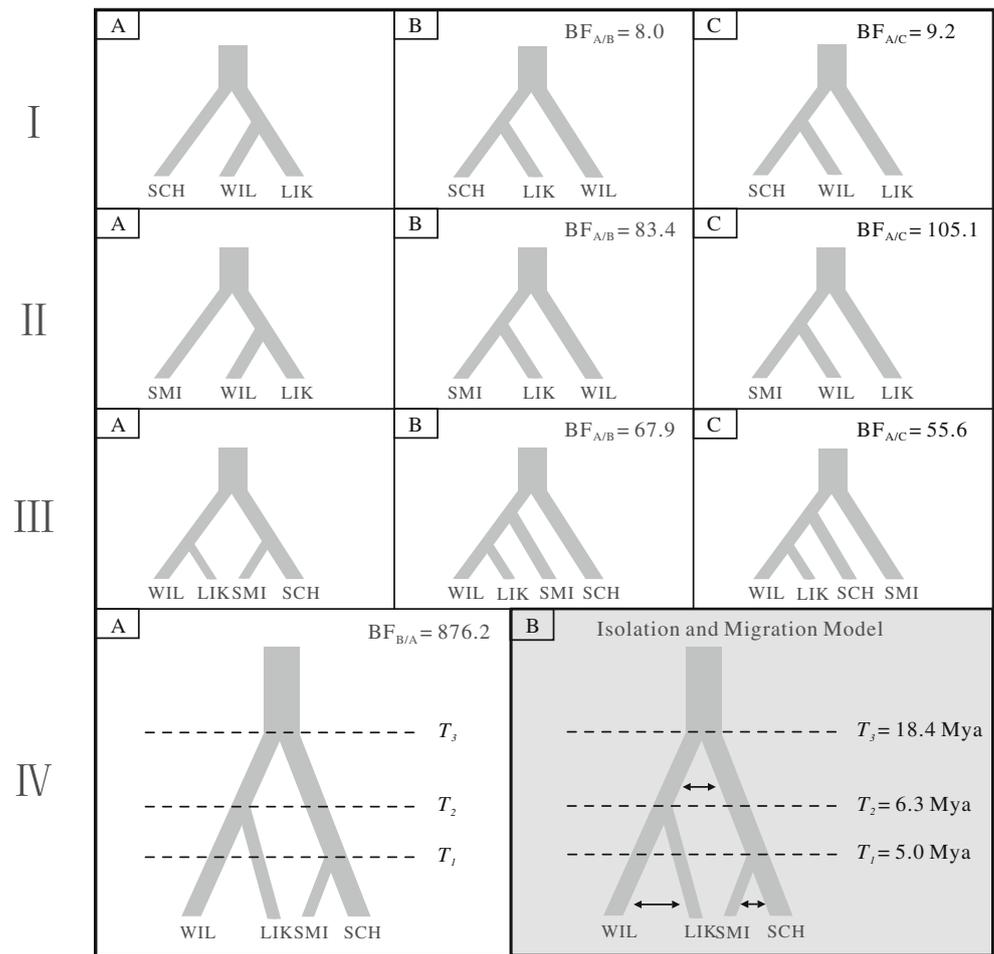
computed  $\Delta K$ , using the method in Evanno et al. (2005), to determine the most likely number of clusters.

**Tests for Speciational Bottlenecks and Gene Flow Using ABCtoolbox** We compared four speciation models with and without gene flow using Approximate Bayesian Computation (ABC) implemented in the ABCtoolbox software package (Wegmann et al. 2010) based on the 11 nuclear loci sequence data. We then used Bayesian factors (BF) to choose the more suitable of two compared models, if  $BF > 3$ , model A was better than model B or C (Fig. 2). We adopted four steps to test speciation orders between four species in order to reduce the number of the hypothesized models (together 64 models) and shorten modeling time (together more than 1 year). Firstly, we examined the interspecific relationships among *P. schrenkiana*, *P. likiangensis*, and *P. wilsonii* based on all three possible models. Secondly, we tested whether *P. smithiana* is closely related to *P. likiangensis* or *P. wilsonii* using a similar approach to the first step. Both tests from the first and second steps suggested that *P. likiangensis* and *P. wilsonii* always clustered together relative to *P. schrenkiana* or *P. smithiana* (Supplementary Table 3). Thirdly, we treated *P. likiangensis*–*P. wilsonii* as one clade, and examined its relationships with *P. schrenkiana* and *P. smithiana*. This test also suggested that *P. smithiana* and *P. schrenkiana* clustered together, diverging late from the *P. likiangensis*–*P. wilsonii* clade. Finally, we compared two different models with or without gene flow during these series of speciation events.

The current population sizes of *P. schrenkiana*, *P. smithiana*, *P. likiangensis*, *P. wilsonii*, and ancestor species were recorded as *NSCH*, *NSMI*, *NL*, *NW*, *N1* (ancestor of *P. likiangensis* and *P. wilsonii*), *N2* (ancestor of *P. schrenkiana* and *P. smithiana*), and *N3* (ancestor of all four species), respectively. The divergence time was recorded as *T1*, *T2*, and *T3*, representing the divergence time between *P. schrenkiana* and *P. smithiana*, *P. likiangensis* and *P. wilsonii*, and *P. likiangensis*–*P. wilsonii* and *P. schrenkiana*–*P. smithiana*, respectively. *Nm* was used to represent the migration numbers per generation.

We assumed uniform priors on the  $\log_{10}$  scale for all population sizes (for *N1* (4.5, 5.7), for *N2* (3.8, 5.0), for *N3* (3.8, 6.0), for *NL* (5.3, 5.5), *NW* (5.4, 5.6), and for *NSCH/NSMI* (4.5, 5.0)) and three timing parameters (for *T1* (5.1, 5.3), for *T2* (5.2, 5.5), and for *T3* (5.5, 5.8)). We used 34 statistics computed by ARLEQUIN v3 (Excoffier et al. 2005) to summarize the population genetic information of *P. schrenkiana* or *P. smithiana* together with *P. likiangensis* and *P. wilsonii*, and we used 34 statistics to estimate the speciation models of four species. Thus, for each species, we computed the number of segregating sites, the number of private segregating sites, Tajima’s  $D$  (Tajima 1989), Fu and Li’s  $D^*$  (Fu et al. 1999), and the number of pairwise

**Fig. 2** Models for origin of *P. schrenkiana* (SCH), *P. smithiana* (SMI), *P. likiangensis* (LIK) and *P. wilsonii* (WIL). If  $BF > 3$ , model A was better than model B/C or model B was better than model A.  $T_1$ ,  $T_2$ , and  $T_3$  indicates the divergence time between *P. schrenkiana* and *P. smithiana*, *P. likiangensis* and *P. wilsonii*, and *P. likiangensis*–*P. wilsonii* and *P. schrenkiana*–*P. smithiana*. The black arrows indicate the migrations between pairs of species



differences. For each pair of species, we computed  $F_{ST}$ , the average number of pairwise differences and the significance was tested by 1,000 permutations as implemented in ARLEQUIN version 3.0 (Excoffier et al. 2005). We also computed the total number of segregating sites over both species. To decrease the redundancy of statistics, we extracted partial least squares (PLS) components from the total of 34 summary statistics using the specific R script of ABCtoolbox (Wegmann et al. 2009). A total of 28 PLS components were used according to root mean square error plots and the conversion equations were inferred from the 10,000 samples simulated by a standard algorithm for each of the last speciation models. We used the likelihood-free ABC–MCMC (Markov chain Monte Carlo) methods (Wegmann et al. 2009) and the program fastsimcoal (Excoffier and Foll 2011) to simulate a total of 1,000,000 samples with a proposed range of  $\varphi=1$  and tolerance  $\delta=1\%$ . The 10,000 simulated samples used for inferring PLS components were used as the calibration in the MCMC sampling step in this study. We retained the 10,000 simulated samples that compared closest with the observed summary statistics of actual dataset following the PLS transformations and applied the regression adjustment

general linear model (GLM) to generate posterior distributions of all parameters in each model. Parameters were estimated based on a mean mutation rate of  $\mu=1.41 \times 10^{-8}$  per site per generation, with a conservative confidence interval between an extremely “slow” rate ( $1.11 \times 10^{-8}$  per site per generation) and an extremely “fast” rate ( $1.71 \times 10^{-8}$  per site per generation). Average generation time was set to 50 years according to previous studies for spruce species (Bousquet and Bouillé 2005; Chen et al. 2010) and the dated timescales need further confirmation from independent evidence. This generation time is about three to five times to the age at the first reproduction, but less than the maximum life expectancy of spruce species (Burns and Honkala 1990). It is likely that alpine spruces in our studied region may have longer generation times than those distributed in the low-altitude region. Even if these demographic parameters are underestimated, our population genetic estimates are of the same order of magnitude and largely consistent with those from other studies of spruce species (Bousquet and Bouillé 2005; Chen et al. 2010; Li et al. 2010). To detect gene flow after speciation, we further used Ima2 (Hey 2010a, 2010b) to estimate the interspecific gene flow after divergence based on the isolation-with-migration

(IM) model (Nielsen and Wakeley 2001; Hey and Nielsen 2004, 2007; Hey 2010a, 2010b).

**Demographic Expansion Analysis** To assess the demographic history of four spruce species, we used LAMARC v2.1.8 (Kuhner 2006; available at <http://evolution.genetics.washington.edu/lamarc/index.html>), with a coalescent simulation approach that considered the genealogical relationships among the nuclear haplotypes to estimate the exponential population growth rate parameter ( $g$ ) within the two species. The analysis was based on a Bayesian method, using Metropolis-coupled MCMC with replication of chains and adaptive heating to obtain better sampling of the parameter space. Large, positive values of the exponential growth parameter ( $g$ ) indicate positive population expansion, but negative values indicate population shrinkage while relatively small positive values ( $g=10$ ) may indicate little or no growth.

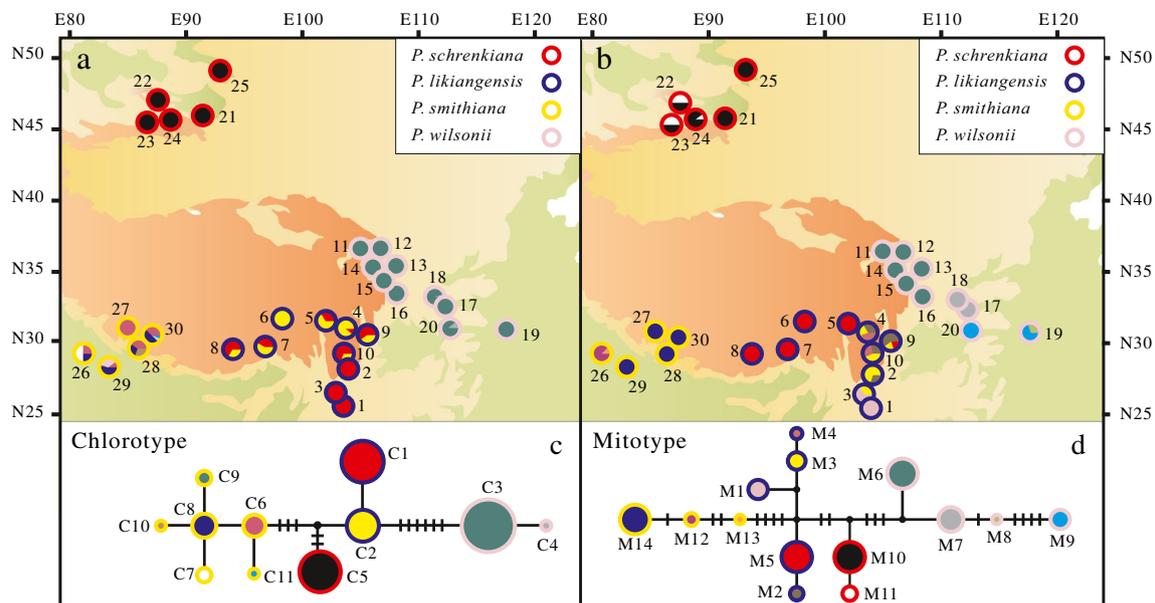
## Results

**Variation in Chloroplast and Mitochondrial Loci and Interspecific Relationships** A total of 11 chlorotypes were identified based on the concatenated data of three chloroplast loci (*trnL-trnF*, *trnS-trnG*, and *ndhK/C*) across all individuals of the four spruce species (Supplementary Table 4). Chlorotypes of four species were grouped into two distinct groups and no haplotype was shared between species (Fig. 3). Two haplotypes from *P. wilsonii* comprised a single distinct clade with

six steps detected. The other comprised six haplotypes (C6–C11) from *P. smithiana*, a single haplotype C5 from *P. schrenkiana*, and two (C2 and C1) from *P. likiangensis*. C5 appeared to be more closely related to C2 and C1 than to C6–C11 in the network. Phylogenetic analyses also identified two distinct clades, and close relationships between haplotypes from *P. schrenkiana* and *P. likiangensis* were not supported by the statistics (Fig. 4).

Fourteen mitotypes were identified across all samples from the examination of the sequence variation based on concatenated data of two mitochondrial loci (*nad5* intron 1 and *nad1* intron b/c; Supplementary Table 5). No mitotype was shared between any two species. Similarly, two distinct groups were identified according to both network and phylogenetic analyses. Four mitotypes M6 to M9 from *P. wilsonii* (Fig. 3) comprised one clade at five steps from the other three species. In the other clade, two mitotypes (M10 and M11) in *P. schrenkiana* appeared to be closer to those (M1 to M5) found in *P. likiangensis* although such a relationship was not well supported (Fig. 4).

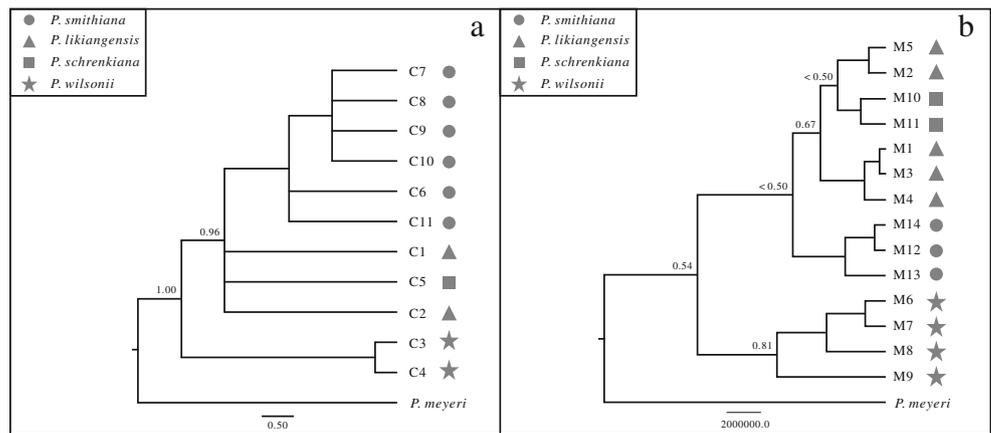
**Variation in Nuclear Loci and Interspecific Relationships** We sequenced 11 unlinked nuclear loci with a total concatenated length of 4,986 bp. We revealed 32, 38, 133, and 143 segregating sites in *P. schrenkiana*, *P. smithiana*, *P. likiangensis*, and *P. wilsonii*, respectively. *P. schrenkiana* and *P. smithiana* had a low level of silent nucleotide diversity,  $\pi_s$  (0.0012) and  $\pi_s$  (0.0016). A similar pattern was observed for the total nucleotide diversity (Table 1). *P. likiangensis* showed the highest nucleotide diversity in cytoplasmic DNA. MFDM



**Fig. 3** Distributions (a, b) and networks (c, d) of chlorotypes and mitotypes within four species: *P. schrenkiana*, *P. smithiana*, *P. likiangensis*, and *P. wilsonii*. The color of the circumference of the circles indicates each species. The proportion of each colored sector filling the

circles in a and b indicates the frequency of each chlorotype or mitotype in each species. Each haplotype in c and d is represented by a circle whose size is proportional to its frequency over all populations

**Fig. 4** Phylogenetic trees constructed for chlorotypes (a) and mitotypes (b) recovered from all sampled individuals among four spruce species. Numbers above nodes indicate statistical significance



tests revealed nearly no substantial evidence for selection acting on each locus across all four species (Table 2).

Eight haplotype genealogies constructed by NETWORK V4.2.1.1 are shown in Supplementary Fig. 1, and the remaining three are not shown because of too many haplotypes. *P. schrenkiana* was highly differentiated from the other four species (*se1390*, *se1364*, *Sb29*, and *MOO2*), while *P. smithiana* could be distinguished from the other species at only three loci (*se1390*, *GI*, and *4CL*). At the remaining loci, interspecific differentiations were low and indistinct. However, AMOVAs conducted on all nuclear data showed that divergence between each pair of species was highly significant ( $P < 0.001$ , Table 3) and the highest genetic divergence existed between *P. schrenkiana* and *P. smithiana* (Table 3) which may be caused by large geographic distance.

We then used a Bayesian clustering algorithm (Structure version 2.3.4) to construct the genetic structures of all examined populations and individuals. The  $\Delta K$  (Evanno et al. 2005) tests showed that the most likely number of groups

( $K$ ) for the entire dataset, using the total nuclear sequence data, was  $K=4$  and that the individuals of all four species clustered into four distinct clusters with high probability. In the clustering results for  $K=2$ , the first cluster contained exclusively individuals from *P. schrenkiana* and *P. smithiana*, while the second cluster contained *P. likiangensis* and *P. wilsonii*. When  $K=3$ , *P. schrenkiana* and *P. smithiana* still clustered into one group, but *P. likiangensis* and *P. wilsonii* clustered into two separate groups (Fig. 5, Supplementary Fig. 2). The phylogenetic tree based on the concatenated sequences of all nuclear loci indicated *P. schrenkiana* and *P. smithiana* clustered to one clade, and *P. likiangensis* and *P. wilsonii* formed the other with high-supporting values (Fig. 6). However, the individuals from *P. likiangensis* and *P. wilsonii* did not cluster into separate subclades, probably due to gene flow or intraspecific substructure.

*Tests of Speciation Models and Demographic Histories of Four Species* The simulations with two demographic models

**Table 1** Average nucleotide diversity across three different genomes in each of the four spruce species

Genomes	Species	Total					Nonsynonymous sites				Silent sites			
		<i>n</i>	<i>L</i>	<i>S</i>	$\theta_W$	$\pi$	<i>L</i>	<i>S</i>	$\theta_W$	$\pi$	<i>L</i>	<i>S</i>	$\theta_W$	$\pi$
nrDNAs	<i>P. schrenkiana</i>	60	4,986	32	0.0014	0.0012	2,444.37	14	0.0012	0.0012	2,532.63	18	0.0015	0.0012
	<i>P. smithiana</i>	98	4,986	38	0.0015	0.0017	2,442.88	17	0.0014	0.0018	2,534.12	20	0.0015	0.0016
	<i>P. likiangensis</i>	100	4,986	133	0.0053	0.0056	2,445.49	NA	NA	0.0042	2,531.51	NA	NA	0.00688
	<i>P. wilsonii</i>	72	4,986	143	0.006	0.0054	2,445.25	59	0.005	0.0044	2,531.75	86	0.007	0.0064
mtDNAs	<i>P. schrenkiana</i>	53	1,361	1	0.00016	0.00025	107.33	0	0	0	1,253.67	1	0.00018	0.00027
	<i>P. smithiana</i>	49	1,413	3	0.00048	0.00026	107.33	0	0	0	1,305.67	3	0.00052	0.00028
	<i>P. likiangensis</i>	91	1,367	5	0.00072	0.00101	107.33	0	0	0	1,259.67	5	0.00078	0.00109
	<i>P. wilsonii</i>	86	1,256	4	0.00063	0.00127	107.33	0	0	0	1,148.67	4	0.00069	0.00139
cpDNAs	<i>P. schrenkiana</i>	53	1,693	0	0	0	–	–	–	–	–	–	–	–
	<i>P. smithiana</i>	49	1,567	4	0.00057	0.00055	–	–	–	–	–	–	–	–
	<i>P. likiangensis</i>	91	1,699	9	0.00104	0.00039	–	–	–	–	–	–	–	–
	<i>P. wilsonii</i>	86	1,704	1	0.00012	0.00003	–	–	–	–	–	–	–	–

*n* sample size, *L* length in base pairs, *S* number of segregating sites,  $\theta_W$  Watterson’s parameter,  $\pi$  nucleotide diversity, NA failed to compute

**Table 2** *P* values of neutrality tests at each locus as measured by MFDM tests

Locus	<i>P. schrenkiana</i>	<i>P. smithiana</i>	<i>P. likiangensis</i>	<i>P. wilsonii</i>
4CL	NA	1.0000	0.6061	0.2254
EBS	1.0000	1.0000	1.0000	0.7887
GI	1.0000	1.0000	0.4444	1.0000
MOO2	1.0000	0.2062	0.0808	0.3380
M007D1	1.0000	1.0000	0.6869	1.0000
Sb16	1.0000	0.0512	0.7071	0.3099
Sb29	0.3389	1.0000	1.0000	1.0000
Sb62	0.3729	0.2062	0.7677	0.5634
sel364	1.0000	0.7835	NA	1.0000
sel390	0.2034	0.5361	0.6869	0.6761
xy1420	NA	NA	1.0000	1.0000
nad1 intron b/c	1.0000	1.0000	1.0000	1.0000
nad5 intron1	NA	NA	0.2222	0.4471
trnL-trnF	NA	0.04433*	1.0000	1.0000
trnS-trnG	NA	NA	0.0222*	NA
ndhK/C	NA	NA	0.0223	NA

NA failed to compute due to insufficient variation

\* $P < 0.05$ ; \*\* $P < 0.01$ ; and \*\*\* $P < 0.001$

of *P. schrenkiana*, *P. smithiana*, *P. likiangensis*, and *P. wilsonii* were obtained using approximate Bayesian computations (ABC–MCMC) methods and these simulated samples were used to calculate Bayes factors. All tests suggested the divergence between the four species to follow in the order: *P. schrenkiana*–*P. smithiana* versus *P. likiangensis*–*P. wilsonii*, then *P. likiangensis* versus *P. wilsonii*, and *P. schrenkiana* versus *P. smithiana*. In addition, the model that included gene flow between these speciation events (Model B; Fig. 2) provided a better fit to the observed nuclear data than did the model without gene flow (model B versus model A; BF=876.2). ABC analyses (Fig. 2) further estimated that the divergence between *P. likiangensis*–*P. wilsonii* and *P. schrenkiana*–*P. smithiana* occurred around 18.4 Mya (95 % HPDI, 11.7–22.4 Mya) (Supplementary Table 6; Supplementary Fig. 3). The divergence between *P. schrenkiana* and *P. smithiana* was estimated at 5.0 Mya (95 % HPDI, 3.98–7.27 Mya) while *P. likiangensis* and *P. wilsonii* diverged at 6.31 Mya (95 % HPDI, 5.01–

**Table 3**  $\Phi_{ST}$  values over all loci among four spruce species

	<i>P. schrenkiana</i>	<i>P. smithiana</i>	<i>P. likiangensis</i>	<i>P. wilsonii</i>
<i>P. schrenkiana</i>	–			
<i>P. smithiana</i>	0.6312***	–		
<i>P. likiangensis</i>	0.5923***	0.54811***	–	
<i>P. wilsonii</i>	0.5835***	0.5231***	0.2585***	–

\* $P < 0.05$ ; \*\* $P < 0.01$ ; and \*\*\* $P < 0.001$

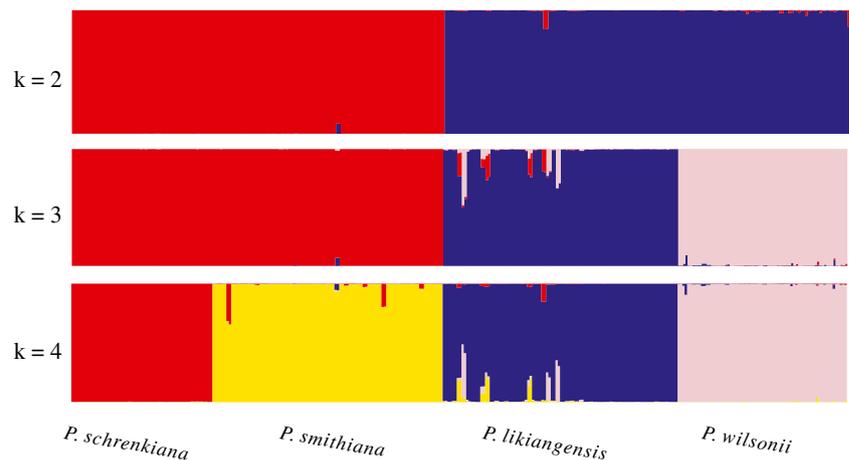
10.2 Mya). The results of ABC analyses revealed symmetric post-divergence gene flow between two ancestral clades and two pairs of the current species. We also examined gene flow between these four species using IMA2 program and our estimates (Table 4) were roughly consistent with the ABC results. In addition, large expansion population growth rate parameter ( $g$ ) values of *P. schrenkiana* (378.99), *P. likiangensis* (1205.29), and *P. wilsonii* (35.92) showed all these species may have experienced sudden and recent expansions. The small ( $g$ ) (0.0021) of *P. smithiana* indicated it to have had little or no growth.

## Discussion

In the present study, we combined population genetic analyses from chloroplast, mitochondrial, and nuclear genomes to examine the origin and speciation model of *P. schrenkiana* and *P. smithiana*, occurring in the Central Asian Highlands and the Himalayas. Although some remote populations were not sampled due to the difficulty to access, all studied populations can represent the core distributions of two species. The sampled individuals for some populations are fewer than expected because of the failure during gene amplifying and sequencing. Therefore, we focused our discussions and implications on the divergence and speciation between two species and their closely related congeners, rather than intraspecific divergences and diversity. In both chloroplast and mitochondrial loci, we found that *P. schrenkiana* is more closely related to *P. likiangensis* than to *P. wilsonii* or *P. smithiana*. However, nuclear datasets suggested that *P. schrenkiana* and *P. smithiana* are more closely related to each other than to either of the other two species. Modeling and testing of the speciation series of four species not only supported this conclusion but further suggested gene flow accompanying speciation. This inconsistency of phylogenetic relationships and speciation series of the four species between three genomes may derive from two different but interacting factors: (1) different rates of the incomplete lineage sorting; (2) interspecific gene flow accompanying speciation. Our further estimates of the speciation timescales between these four species suggest that their initial and further divergences seem to be correlated with the early QTP uplifts and the corresponding TSMs depressions.

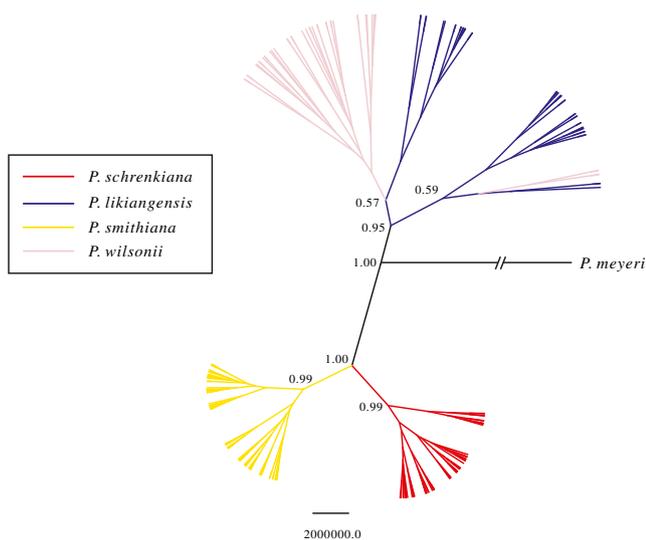
*Incomplete Lineage Sorting and Interspecific Gene Flow* Three previous studies, each mainly based on chloroplast, mitochondrial, or nuclear sequence variations, but from fewer individuals, discovered a close affinity of *P. schrenkiana* with *P. likiangensis* (Ran et al. 2006) or *P. wilsonii* (Lockwood et al. 2013) or *P. smithiana* (Sun et al. 2014). In the present study, our phylogenetic analyses of both chloroplast and mitochondrial sequence variations suggest that *P. schrenkiana* seems to be more closely related to *P. likiangensis* (Fig. 4) than to *P. smithiana*. The

**Fig. 5** Structure analysis of four species when  $K=2-4$  clusters are assumed. For each  $K$  value, results of the run with the highest value of LnPD were used. Variation among runs was limited



clade containing haplotypes from these three species comprise a distinct clade that differs from that comprising the *P. wilsonii* haplotypes. However, both phylogenetic and population genetic analyses based on nuclear genetic sequence data suggest that *P. schrenkiana* and *P. smithiana* cluster first, while *P. likiangensis* and *P. wilsonii* together form the other clade (Fig. 6). It should be noted that within the latter clade, individuals from two species did not cluster into separate subclades. This would have resulted from substructures of each species and interspecific gene flow or introgressions, which may have distorted phylogenetic analyses of population data. In fact, STRUCTURE analysis clearly supported these two distinct species although gene flow and introgression between them were detected for a few individuals (Fig. 5). The divergence series between four species were further confirmed by modeling and testing of the genetic population data from the nuclear loci (Figs. 2 and 6). The inconsistent interspecific relationships at different genomes may have been caused by the following two factors. First, In contrast to the biparental

inheritance exhibited by nuclear DNA (nrDNA), mtDNA is maternally transmitted through seeds, while cpDNA is paternally inherited via pollen in spruce species (Corriveau and Coleman 1988; Harris and Ingram 1991; Reboud and Zeyl 1994); therefore, the average coalescence time of two randomly picked chloroplast or mitochondrial alleles ( $N_e$  generation) is half of the nuclear DNA alleles ( $2N_e$  generation; Nei and Tajima 1981). Therefore, DNA variation would be lost more easily and species-specific mutations would be accumulated more quickly in cytoplasmic DNAs than in nuclear DNAs, especially in extremely small isolated populations. This may account for species-specific haplotypes in these two cytoplasmic DNAs. For *P. smithiana* which was isolated with other species by Himalayas, the small effective population size may have led to a high level of accumulation of species-specific mutations in the cytoplasmic DNA, which probably distorted the true interspecific relationships in the results of earlier studies (Ran et al. 2006;



**Fig. 6** The phylogenetic tree based on the concatenated nuclear data using the BEAST 1.7.2. The color of each line indicates each species, and the length of each line shows the genetic distance

**Table 4** Maximum-likelihood estimates (MLE) and the 95 % highest posterior density (HPD) Intervals of migration rate among pair-wise comparisons

		MLE	HPD95Lo	HPD95Hi
<i>P. schrenkiana</i> / <i>P. smithiana</i>	$m_1$	0.079	0	1.763
	$m_2$	0.001	0	1.615
<i>P. schrenkiana</i> / <i>P. likiangensis</i>	$m_1$	0.175	0	0.593
	$m_2$	0.001	0	0.427
<i>P. schrenkiana</i> / <i>P. wilsonii</i>	$m_1$	0.0645	0	0.3615
	$m_2$	0.0015	0	0.6345
<i>P. smithiana</i> / <i>P. likiangensis</i>	$m_1$	0.1375	0	0.8425
	$m_2$	0.0025	0	1.127
<i>P. smithiana</i> / <i>P. wilsonii</i>	$m_1$	0.1005	0	0.4995
	$m_2$	0.0045	0	0.6885
<i>P. likiangensis</i> / <i>P. wilsonii</i>	$m_1$	0.005	0	0.805
	$m_2$	0.445	0.055	1.325

$m_1$  is the migration rate from species 1 to species 2 forward in time, and  $m_2$  is the migration from species 2 to species 1

Bouillé et al. 2011; Zou et al. 2013). However, at the nuclear loci, incomplete lineage sorting still remains for most sampled loci. The additive mutations from multiple nuclear loci therefore recovered a different interspecific relationship from those based on mitochondrial and chloroplast sequence variations. Second, gene flow and interspecific introgressions, especially cytoplasmic haplotypes occur frequently in spruce (Li et al. 2010; Du et al. 2011; Zou et al. 2012). It is therefore likely that *P. schrenkiana* captured mitochondrial and chloroplast haplotypes from *P. likiangensis*. In addition, modeling of the nuclear population genetic data suggests that gene flow continued for a long time after the divergence. Gene flow at nuclear loci maybe occurred more frequently between *P. schrenkiana* and *P. smithiana* than between *P. wilsonii* and/or *P. likiangensis* (Supplementary Table 6). Therefore, it is also likely the regional nuclear gene flow might have grouped *P. schrenkiana* and *P. smithiana* together. All these lines of evidence suggested that speciation histories of these species are more complex than expected and that both incomplete lineage sorting and gene flow might have together resulted in different interspecific relationships discovered between cytoplasmic and nuclear genomes.

**Geological Isolation and Speciation History** According to our parameter estimates based on  $4 \times 10^6$  simulated samples and the current data pattern, two clades *P. schrenkiana*–*P. smithiana* and *P. likiangensis*–*P. wilsonii* diverged around 18.4 Mya. The further divergences between *P. likiangensis* and *P. wilsonii* and between *P. schrenkiana* and *P. smithiana* were estimated to occur at 6.31 and 5.0 Mya, respectively. It is interesting to note that the divergence timescales estimated here agree well with the geological evidence that the QTP (and Himalayas) were uplifted extensively during these stages. For example, the first large-scale QTP–Himalayas uplifts were dated to occur at 22 Mya, although the uplift of the QTP and Himalayas began about 50 Mya, which caused some peaks that were high enough to trigger Asian desertification (Guo et al. 2002). Between 10 and 8 Mya, the second uplift might have occurred, which further enhanced the aridity of the Asian interior and the onset of the Asian monsoon (An et al. 2001). This uplift may have occurred as early as 15 Mya, and in particular may have caused further increases of the western Kunlun Mountains, the Himalayas, and the vast depressions of the Tarim (Wang et al. 1990; Abdrakhmatov et al. 1996; Shi et al. 1998; An et al. 2001; Bullen et al. 2001, 2003; Charreau et al. 2006, 2009; Wang et al. 2006; Buslov et al. 2007; Dupont–Nivet et al. 2008; Wang et al. 2008). It is also likely that the first extensive uplift promoted the initial divergence between two clades, *P. likiangensis*–*P. wilsonii* and *P. schrenkiana*–*P. smithiana*. It is interesting to find this divergence was accompanied with gene flow according to the selected fitter model. The second large-scale uplift might have terminated this gene flow and also initiated further divergences between *P. likiangensis* and *P. wilsonii* and between *P. schrenkiana* and *P. smithiana*. The TSMs deserts formed at this stage (Sun

et al. 2009) and the further increase of the Himalayas should have restricted natural migration or dispersal between *P. schrenkiana* and *P. smithiana* (Fig. 1). Therefore, these two extensive uplifts of the QTP (–Himalayas) might have played an important role in the speciation histories of these four species.

In addition to previous studies, which have demonstrated that several herbal genera originated or diversified greatly during the QTP uplift stages (e.g., Liu et al. 2002, 2006; Wang et al. 2009), our population genetic analyses have indicated that trees with long generation times may also have speciated in response to these uplifts. Further analyses of more tree speciation events in the QTP and adjacent regions are needed to demonstrate this generality. Overall, these findings together highlight the complex speciation histories of the alpine plants, and the importance of the geological events in promoting plant diversification.

**Acknowledgments** The authors thank Dr. Kun Wang, Long Li, and Ting Xu for their helps in analyzing data. This research was supported by grants from National Key Project for Basic Research (2014CB954100), National Natural Science Foundation of China (30930072), Key Project of International Collaboration Program, Ministry of Science and Technology of China (2010DFB63500) and ‘111’ collaboration Program.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## References

- Abdrakhmatov KY, Aldazhanov SA, Hager BH, Hamburger MW, Herring TA, Kalabaev KB, Makarov VI, Molnar P, Panasyuk SV, Prilepin MT, Reilinger RE, Sadybakasov IS, Souter BJ, Trapeznikov YA, Tsurkov VY, Zubovich AV (1996) Relatively recent construction of the Tien Shan inferred from GPS measurements of present-day crustal deformation rates. *Nature* 384:450–453
- An ZS, Kutzbach JE, Prell WL, Port SC (2001) Evolution of Asian monsoons and phased uplift of the Himalayan–Tibetan plateau since late Miocene times. *Nature* 411:62–66
- Anderson LL, Hu FS, Nelson DM, Petit RJ, Paige KN (2006) Ice-age endurance: DNA evidence of a white spruce refugium in Alaska. *Proc Natl Acad Sci U S A* 103:12447–12450
- Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37–48
- Bouillé M, Senneville S, Bousquet JB (2011) Discordant mtDNA and cpDNA phylogenies indicate geographic speciation and reticulation as driving factors for the diversification of the genus *Picea*. *Tree Genet Genomes* 7:469–484
- Bousquet J, Bouillé M (2005) Trans-species shared polymorphism at orthologous nuclear gene loci among distant species in the conifer *Picea* (Pinaceae): implications for the long-term maintenance of genetic diversity in trees. *Am J Bot* 92:63–73
- Bullen ME, Burbank DW, Garver JI, Abdrakhmatov KY (2001) Late Cenozoic tectonic evolution of the northwestern Tien Shan: new age estimates for the initiation of mountain building. *Geol Soc Am Bull* 113:1544–1559

- Bullen ME, Burbank DW, Garver JI (2003) Building the northern Tien Shan: integrated thermal, structural, and topographic constraints. *J Geol* 111:149–165
- Burns RM, Honkala BH (1990) Silvics of North America. Vol 1 conifers. agriculture handbook. U.S. Department of Agriculture and Forestry Service, Washington, p 654
- Buslov MM, Grave JD, Bataleva EAV, Batalev VY (2007) Cenozoic tectonic and geodynamic evolution of the Kyrgyz Tien Shan mountains: a review of geological, thermochronological and geophysical data. *J Asian Earth Sci* 29:205–214
- Charreau J, Gilder S, Chen Y, Dominguez S, Avouac JP, Sen S, Jolivet M, Li YG, Wang WM (2006) Magnetostratigraphy of the Yaha Section, Tarim basin (China): 11 Ma acceleration in erosion and uplift of the Tian Shan mountains. *Geology* 34:181–184
- Charreau J, Chen Y, Gilder S, Barrier L, Dominguez S, Augier R, Sen S, Avouac JP, Gallaud A, Graveleau F, Wang QC (2009) Neogene uplift of the Tian Shan Mountains observed in the magnetic record of the Jingou River section (northwest China). *Tectonics* 28:1–22
- Chen J, Källman T, Gyllenstrand N, Lascoux M (2010) New insights on the speciation history and nucleotide diversity of three boreal spruce species and a Tertiary relict. *Heredity* 104:3–14
- Corriveau JL, Coleman AW (1988) Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *Am J Bot* 75:1443–1458
- Coyne JA (1992) Genetics and speciation. *Nature* 355:511–515
- Coyne JA, Orr HA (2004) Speciation. Sinauer, Sunderland, MA
- Doyle JJ, Doyle JL (1990) Isolation of plant DNA from plant tissue. *Focus* 12:13–15
- Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214
- Du FK, Peng XL, Liu JQ, Lascoux M, Hu FS, Petit RJ (2011) direction and extent of organelle DNA introgression between two spruce species in the Qinghai–Tibetan Plateau. *New Phytol* 192:1024–1033
- Dupont–Nivet G, Hoorn C, Konert M (2008) Tibetan uplift prior to the eocene–oligocene climate transition: evidence from pollen analysis of the Xining basin transition. *Geology* 6:987–990
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol* 14:2611–2620
- Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. accuracy assessment. *Genome Res* 8:175–185
- Excoffier L, Foll M (2011) fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* 27:1332–1334
- Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evol Bioinforma* 1:47–50
- Farjón A (2001) World checklist and bibliography of conifers. Royal Botanic Gardens, Kew
- Fu L, Li N, Mill RR (1999) *Picea*. In: Wu, ZY, Raven, PH (eds.). Flora of China. Science Press, Beijing, pp 25–32
- Gavrilets S (2003) Perspective: models of speciation: what have we learned in 40 years? *Evolution* 57:2197–2215
- Grant V (1981) Plant speciation. Columbia University Press, New York
- Guo ZT, Ruddiman WF, Hao QZ, Wu HB, Qiao YS, Zhu RX, Peng SZ, Wei JJ, Yuan BY, Liu TS (2002) Onset of Asian desertification by 22 Mya ago inferred from loess deposits in China. *Nature* 416:159–163
- Hailer F, Kutschera VE, Hallström BM, Klassert SR, Leonard JA, Arnason U, Janke A (2012) Nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage. *Science* 336:44–347
- Hamilton MB (1999) Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Mol Ecol* 8: 521–523
- Harris SA, Ingram R (1991) Chloroplast DNA and biosystematics, the effects of intraspecific diversity and plastid transmission. *Taxon* 40: 393–412
- Hein J, Schierup MH, Wiuf C (2004) Gene genealogies, variation and evolution, a primer in coalescent theory. Oxford University Press, Oxford
- Heuertz M, Paoli ED, Kallman T, Larsson H, Jurman M, Morgante M, Lascoux M, Gyllenstrand N (2006) Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce (*Picea abies* (L.) Karst). *Genetics* 174:2095–2105
- Hey J (2006) Recent advances in assessing gene flow between diverging populations and species. *Curr Opin Genet Dev* 16:592–596
- Hey J (2010a) Isolation with migration models for more than two populations. *Mol Biol Evol* 27:905–920
- Hey J (2010b) The divergence of chimpanzee species and subspecies as revealed in multipopulation isolation-with-migration analyses. *Mol Biol Evol* 27:921–933
- Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* 167:747–760
- Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proc Natl Acad Sci U S A* 104:2785–2790
- Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour* 9:1322–1332
- Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294:2310–2314
- Kuhner MK (2006) LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics* 22:768–770
- Levin D (2003) The ecological transition in speciation. *New Phytol* 161: 91–96
- Li HP (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 N, Stephens M (2003) Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* 165:2213–2233
- Li Y, Stocks M, Hemmälä S, Källman T, Zhu HT, Zhou YF, Chen J, Liu JQ, Lascoux M (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
- Li L, Abbott RJ, Liu BB, Sun YS, Li LL, Zou JB, Wang X, Miede G, Liu JQ (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
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452
- Liu TS (1982) A new proposal for the classification of the genus *Picea*. *Acta Phytol Geobot* 33:227–244
- Liu JQ, Gao TG, Chen ZD, Lu AM (2002) Molecular phylogeny and biogeography of the Qinghai–Tibet Plateau endemic *Nannoglottis* (Asteraceae). *Mol Phylogenet Evol* 23:307–325
- Liu JQ, Wang YJ, Wang AL, Hideaki O, Abbott RJ (2006) Radiation and diversification within the *Ligularia–Cremanthodium–Parasenecio* complex (Asteraceae) triggered by uplift of the Qinghai–Tibetan Plateau. *Mol Phylogenet Evol* 38:31–49
- Lockwood JD, Jelena MA, Zou JB, Wang J, Liu JQ, Susanne SR (2013) A new phylogeny for the genus *Picea* from plastid, mitochondrial, and nuclear sequences. *Mol Phylogenet Evol* 60:717–727

- Mao KS, Hao G, Liu JQ, Adams RP, Milne RI (2010) Diversification and biogeography of *Juniperus* (Cupressaceae): variable diversification rates and multiple intercontinental dispersals. *New Phytol* 188:254–272
- Mayr E (1954) Change of genetic environment and evolution. Columbia University Press, New York
- Meng L, Yang R, Abbott RJ, Miede G, Hu TH, Liu JQ (2007) Mitochondrial and chloroplast phylogeography of *Picea crassifolia* Kom. (Pinaceae) in the Qinghai–Tibetan Plateau and adjacent highlands. *Mol Ecol* 16:4128–4137
- Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature* 403:853–858
- Nei M, Tajima F (1981) Genetic drift and estimation of effective population size. *Genetics* 98:625–640
- Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* 158:885–896
- Orr HA, Presgraves DC (2000) Speciation by postzygotic isolation, forces, genes and molecules. *Bioessays* 22:1085–1094
- Posada D (2008) JModelTest: phylogenetic model averaging. *Mol Biol Evol* 25:1253
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Ran JH, Wei XX, Wang XQ (2006) Molecular phylogeny and biogeography of *Picea* (Pinaceae): implication for phylogeographical studies using cytoplasmic haplotypes. *Mol Phylogenet Evol* 41:405–419
- Reboud X, Zeyl C (1994) Organelle inheritance in plants. *Heredity* 72:132–140
- Rieseberg LH, Willis JH (2007) Plant speciation. *Science* 317:910–914
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574
- Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. *Mol Ecol Notes* 4:137–138
- Shi YF, Li JJ, Li BY (1998) Uplift and environmental changes of Qinghai–Xizang Plateau in the Late Cenozoic. *Guangdong Science & Technology Press, Guangzhou*, p 154
- Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 73:1162–1169
- Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68:978–989
- Sun J, Zhang Z, Zhang L (2009) New evidence on the age of the Taklimakan desert. *Geology* 37:159–162
- Sun YS, Abbott JR, Li LL, Zou JB, Liu JQ (2014) Evolutionary history of purple cone spruce (*Picea Purpurea*) in the Qinghai–Tibet Plateau: homoploid hybrid origin and Pleistocene expansion. *Mol Ecol* 23(2):343–359. doi:10.1111/Mec.12599
- Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol Biol* 17:1105–1109
- Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105:437–460
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595
- Takayama K, Sun BY, Stuessy TF (2013) Anagenetic speciation in Ullung Island, Korea: genetic diversity and structure in the island endemic species, *Acer takesimensis* (Sapindaceae). *J Plant Res* 126:323–333
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731–2739
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882
- Wakeley J (2003) Inferences about the structure and history of populations: coalescents and intraspecific phylogeography. In: Singh R, Veuille M (eds) *The evolution of population biology*. Cambridge University Press, New York
- Wang D, Sun X, Zhao Y (1990) Late cretaceous to tertiary palynofloras in Xinjiang and Qinghai, china. *Rev Palaeobot Palynol* 65:95–104
- Wang Y, Deng T, Biasatti D (2006) Ancient diets indicate significant uplift of southern Tibet after ca. 7 Ma. *Geology* 34:309–312
- Wang CS, Zhao XX, Liu ZF, Lippert PC, Graham SA, Coe RS, Yi HS, Zhu LD, Liu S, Li YL (2008) Constraints on the early uplift history of the Tibetan Plateau. *Proc Natl Acad Sci U S A* 105:4987–4992
- Wang LY, Abbott RJ, Zheng W, Chen P, Wang YJ, Liu JQ (2009) History and evolution of alpine plants endemic to the Qinghai–Tibet Plateau: *Aconitum gymnantrum* (Ranunculaceae). *Mol Ecol* 18:709–721
- Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 7:256–276
- Wegmann D, Leuenberger C, Excoffier L (2009) Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics* 182:1207–1218
- Wegmann D, Leuenberger C, Neuenchwander S, Excoffier L (2010) ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinforma* 11:116–122
- Wilson EO (1992) *The diversity of life*. Harvard University Press, Cambridge
- Woerner AE, Cox MP, Hammer MF (2007) Recombination-filtered genomic datasets by information maximization. *Bioinformatics* 23:1851–1853
- Xu TT, Abbott R, Milne R, Mao KS, Du FK, Wu GL, Ciren ZX, Miede G, Liu JQ (2010) Phylogeography and allopatric divergence of cypress species (*Cupressus* L.) in the Qinghai–Tibetan Plateau and adjacent regions. *BMC Evol Biol* 10:194–203
- Zhang YS, Tang GC (1989) *Picea schrenkiana* forest. In: Editorial Committee of Xinjiang Forest (ed) *Xinjiang Forest*. Xinjiang People Press, Urumqi, pp 121–149
- Zou JB, Peng XL, Li L, Liu JQ, Miede G, Opgenoorth L (2012) Molecular phylogeography and evolutionary history of *Picea likiangensis* in the Qinghai–Tibetan Plateau inferred from mitochondrial and chloroplast DNA sequence variation. *J Syst Evol* 50:341–350
- Zou JB, Sun YS, Li L, Wang GN, Yue W, Lu ZQ, Wang Q, Liu JQ (2013) Population genetic evidence for speciation pattern and gene flow between *Picea wilsonii*, *P. morrisonicola* and *P. neoveitchii*. *Ann Bot* 12:1829–1844