

ARTICLE



Backcrossing to different parents produced two distinct hybrid species

Donglei Wang^{1,5}, Yongshuai Sun^{2,5}, Weixiao Lei^{3,5}, Hui Zhu³, Ji Wang¹, Hao Bi¹, Shuo Feng⁴, Jianquan Liu^{1,3} and Dafu Ru^{1,3}

© The Author(s), under exclusive licence to The Genetics Society 2023

Repeated homoploid hybrid speciation (HHS) events with the same parental species have rarely been reported. In this study, we used population transcriptome data to test paraphyly and HHS events in the conifer *Picea brachytyla*. Our analyses revealed non-sister relationships for two lineages of *P. brachytyla*, with the southern lineage being placed within the re-circumscribed *P. likiangensis* species complex (PLSC) and *P. brachytyla sensu stricto* (s.s.) consisted solely of the northern lineage, forming a distinct clade that is paratactic to both the PLSC and *P. wilsonii*. Our phylogenetic and coalescent analyses suggested that *P. brachytyla* s.s. arose from HHS between the ancestor of the PLSC before its diversification and *P. wilsonii* through an intermediate hybrid lineage at an early stage and backcrossing to the ancestral PLSC. Additionally, *P. purpurea* shares the same parents and an extinct lineage with *P. brachytyla* s.s. but backcrossing to the other parent, *P. wilsonii* at a later stage. We reveal the first case that backcrossing to different parents of the same extinct hybrid lineage produced two different hybrid species. Our results highlight the existence of more reticulate evolution during species diversification in the spruce genus and more complex homoploid hybrid events than previously identified.

Heredity (2023) 131:145–155; <https://doi.org/10.1038/s41437-023-00630-9>

INTRODUCTION

Homoploid hybrid speciation (HHS) is a fascinating phenomenon in which the hybridization of two distinct species generates a new species without any change in ploidy level (Rieseberg et al. 2003; Nolte and Tautz 2010). Although HHS was once thought to be rare, recent studies indicate that it may be more common than previously thought (e.g. Rieseberg 1997; Gross and Rieseberg 2005; Mavárez et al. 2006; Hermansen et al. 2011; Nieto Feliner et al. 2017; Lamichhaney et al. 2018; Sun et al. 2020; Wang et al. 2021). However, only a few cases fulfill the rigorous criterion set forth by Schumer et al. (2014), which requires the presence of hybridization-derived reproductive isolation (RI) between the hybrid and parent species. Theoretical models suggest that such RI may arise through inheriting parental genetic barriers, such as genetic incompatibilities or extrinsic barrier loci (Schumer et al. 2015; Brennan et al. 2019; Sun et al. 2020; Wang et al. 2021). However, direct evidence of this mechanism can be challenging to obtain using traditional approaches like tree-based, egegetic/association mapping, or other methods. Revealing more candidate HHS cases is therefore highly informative in advancing our understanding of reticulate species diversification even when direct evidence of RI caused by hybridization is absent (Nieto Feliner et al. 2017).

Here we investigate the potential homoploid hybrid origin of *Picea brachytyla sensu stricto* (s.s.), a diploid species within the genus *Picea* (Wright 1955). Traditionally, *P. brachytyla* (Franch.) E. Pritz. is described as comprising two varieties in China: a southern

lineage in the Yunnan province and a northern lineage in the Sichuan and Chongqing provinces adjacent to the Qinghai-Tibet Plateau (QTP) (Ru et al. 2016; Lyu et al. 2020). Previous studies proposed that the southern lineage (*P. brachytyla*-southern lineage) shares a recent common ancestor with three varieties of *P. likiangensis* (*likiangensis*, *rubescens*, and *linzhiensis*) within the *P. likiangensis* species complex (PLSC), which underwent diversification through a radiation with gene flow (Ru et al. 2016; Sun et al. 2018; Shen et al. 2019). However, little is known about the evolutionary history of the northern lineage, which includes the population where the *P. brachytyla* s.s. type specimen was collected, although a hybrid origin has been proposed (Shen et al. 2019).

Phylogenetic analyses of *Picea* genus have suggested that *P. brachytyla* s.s. is more closely related to either the PLSC or *P. wilsonii* based on various nuclear loci and chloroplast DNA (Ran et al. 2015; Shao et al. 2019; Shen et al. 2019). Further genetic analyses supported this inference in part, but also found that *P. brachytyla* s.s. shared some genomic information with both the PLSC and *P. wilsonii* (Lyu et al. 2020). In addition, *P. brachytyla* s.s. exhibits intermediate morphological traits of both *P. likiangensis* (i.e., quadrangular leaves and 4–5 stomatal lines on each of four surfaces) and *P. wilsonii* (i.e., tightly arranged seed scales before maturity) (Fu et al. 1999; Lyu et al. 2020). These conflicting phylogenetic relationships and intermediate morphological traits suggest that *P. brachytyla* s.s. may have originated from hybridization between two other species (or species complexes)

¹Key Laboratory for Bio-resource and Eco-environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu, PR China. ²CAS Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla 666303 Yunnan, China. ³State Key Laboratory of Grassland Agro-ecosystem, College of Ecology, Lanzhou University, Lanzhou 730000, China. ⁴State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University, Xi'ning, Qinghai, China. ⁵These authors contributed equally: Donglei Wang, Yongshuai Sun, Weixiao Lei. Associate editor: Dario Grattapaglia. email: liujq@nwpb.ac.cn; rudf@lzu.edu.cn

Received: 29 September 2022 Revised: 23 May 2023 Accepted: 23 May 2023

Published online: 1 June 2023

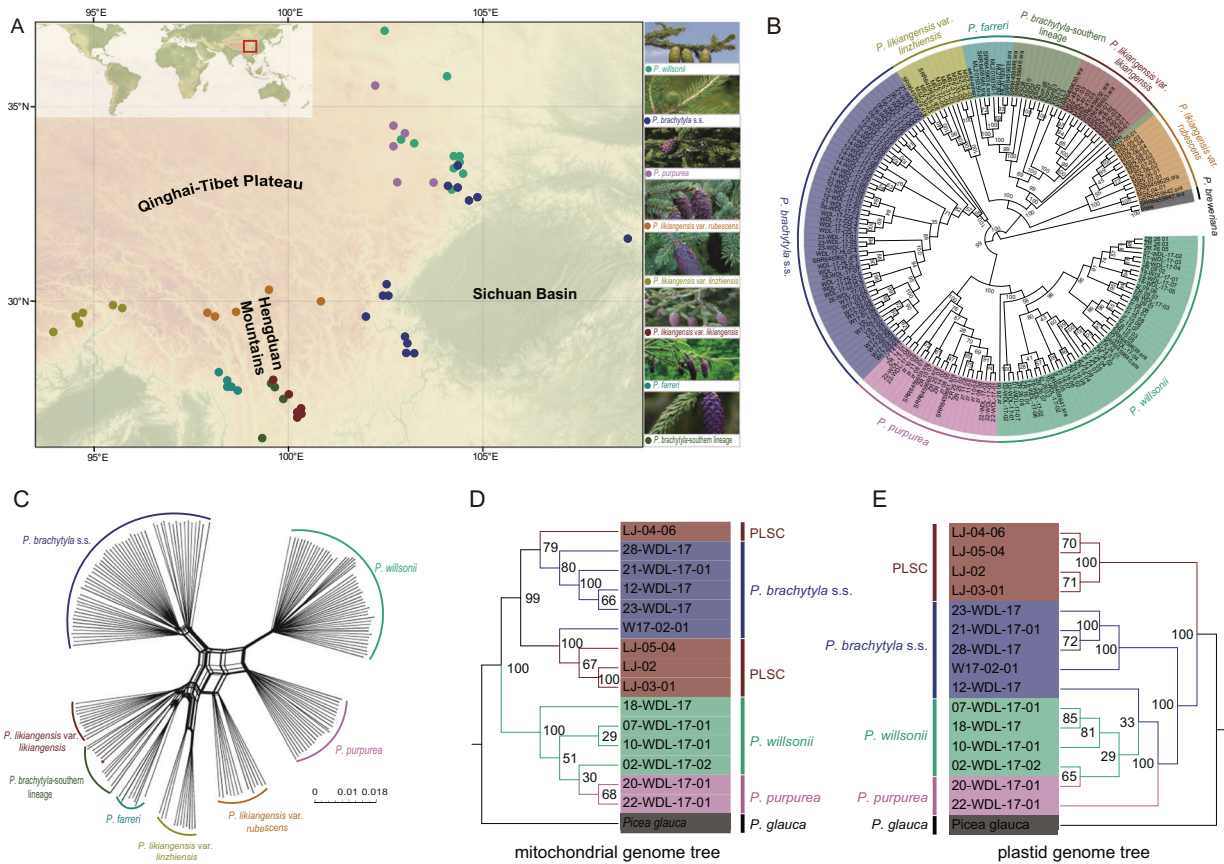


Fig. 1 Geographic distribution of samples and genetic relationships. **A** Distributions of *P. wilsonii*, *P. purpurea*, *P. farreri* and *P. likiangensis* var. *rubescens*, *P. likiangensis* var. *linzhiensis*, *P. likiangensis* var. *likiangensis*, *P. brachytyla* sensu stricto (s.s.) and *P. brachytyla*-southern lineage sampled for transcriptome sequencing for this study. Different colors indicate the taxon collected as indicated. **B** Maximum likelihood tree based on all 186 transcriptome sequences with *P. breweriana* as an outgroup. **C** Phylogenetic network based on the W-RNA-seq dataset. **D** Mitochondrial genome tree of four taxonomic groups with *P. glauca* as an outgroup. **E** Plastid genome tree of four groups with *P. glauca* as an outgroup. PLSC: *P. likiangensis* species complex, red: *P. likiangensis* var. *likiangensis*, green: *P. wilsonii*, pink: *P. purpurea*, blue: *P. brachytyla* s.s., olive: *P. likiangensis* var. *linzhiensis*, cyan: *P. farreri*, linen: *P. likiangensis* var. *rubescens*, celadongreen: *P. brachytyla*-southern lineage.

(Shao et al. 2019; Shen et al. 2019; Lyu et al. 2020). Geographically, *Picea brachytyla* s.s. is found in low-altitude humid valleys, while the PLSC is found on high-altitude mountains and *P. wilsonii* is found in northern regions of the QTP with low altitudes and relatively dry conditions (Fu et al. 1999; Lyu et al. 2020). It is worth noting that *P. purpurea*, another high-altitude species, was shown to originate from hybridization between *P. likiangensis* and *P. wilsonii* (Sun et al. 2014; Ru et al. 2018), indicating that hybridization between two parental lineages may produce more than one diploid hybrid species, similar to HHS in sunflowers (Rieseberg 1997). In addition, *P. likiangensis* var. *rubescens* was found to share substantially nuclear ancestry with *P. purpurea* based on population genetic structure analyses (Ru et al. 2018), large-scale shared genetic variations from introgression (Sun et al. 2018), and conflicting phylogenetic relationships based on DNA sequences (Shao et al. 2019; Shen et al. 2019). When modeling the reticulate evolution involving the PLSC and *P. wilsonii*, it may be better to remove var. *rubescens* from the PLSC, while incorporating the *P. brachytyla*-southern lineage into the PLSC. Overall, these findings suggest that the evolutionary history of hybridization and relationships of these lineages and their ancestors are more complex than expected and warrants additional examination.

In this study, we employed transcriptomics to better understand the origin of *P. brachytyla* s.s. Using novel and published transcriptomes from *P. brachytyla* and related species, we aimed

to address the following questions: (1) Is there support for polyphyly of *P. brachytyla* and how should the PLSC best be defined? (2) Did *P. brachytyla* s.s. originate through HHS between *P. wilsonii* and the common ancestor of the PLSC, or bifurcate from one lineage with further introgression from the other? (3) If it originated through HHS, did *P. brachytyla* s.s. originate earlier or later than *P. purpurea* which is derived from hybridization between the same parents?

MATERIAL AND METHODS

Material and RNA sequencing

We collected fresh, mature leaf needles from first-year branches of 78 individuals (Table S1) representing *P. brachytyla* and related taxa, including *P. farreri*, *P. wilsonii*, and *P. purpurea*. To avoid the confounding effects of hybridization and introgression, we excluded forest stands where two or more spruce species co-occurred, and sampled individuals spaced at least 500 m apart across their core distributions. We also excluded areas where previous literature indicated that two taxa may have experienced high levels of gene flow in the recent past. Transcriptomes were sequenced using the methods described by Ru et al. (2018), and the resulting 150 bp paired-end raw reads were deposited in BioSample (average number of raw bases > 6 Gb; Table S2). Additionally, we included previously published transcriptomes obtained from leaf needle samples of 108 individuals, comprising three varieties of *P. likiangensis* (vars. *likiangensis*, *linzhiensis* and *rubescens*), *P. wilsonii* and *P. purpurea* (Ru et al. 2016, 2018; Feng et al. 2019;

Shao et al. 2019; Shen et al. 2019). In total, we used 186 transcriptomes for our analyses (Fig. 1A and Tables S1 and S2).

Read trimming, mapping and individual variant calling

We performed quality control on the raw data using Trimmomatic ver. 0.38 (Bolger et al. 2014) to trim Poly-N, and low quality bases and discarded reads with <36 bases with the following parameters “ILLUMINACLIP:adapter.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:36”. The resulting high-quality reads were then aligned to a reference transcriptome from *P. abies* (Nystedt et al. 2013) using the “mem” module in the Burrows-Wheeler Aligner (BWA) software version 0.7.10 (Li and Durbin 2019) with default parameters. Before variant calling, PCR duplicates were marked and removed using MarkDuplicates in PICARD ver. 1.129 (<http://broadinstitute.github.io/picard/>), and local realignment was performed around each indel using Genome Analysis Toolkit (GATK) ver. 3.8 (Danecek et al. 2011). Single nucleotide polymorphisms (SNPs) were extracted using “mpileup” in SAMTOOLS ver. 1.8 (Li et al. 2009) based on uniquely mapped reads for all individuals, with a minimum base quality (-Q) and mapping quality (-q) set to 20 and 30, respectively. To obtain high quality variants, a custom Perl script was used to filter our raw SNPs with the following criteria: (1) SNPs located within a 5-bp window of an indel; (2) SNPs with a Phred quality score <20; (3) SNPs with >20% missing bases within each species. Additionally, bases with a depth of coverage (DP) < 10 were set as missing for each individual (Chapman et al. 2013; Li et al. 2014). Finally variant sites with minimum allele frequency <0.01 (to ensure that at least 4 alleles were found in our sample set) were filtered using VCFTOOLS ver. 0.1.14 (Danecek et al. 2011) to reduce the false discovery rate. For convenience, we refer to the resulting nuclear transcriptomic sequences as the N-RNA-seq dataset.

Population differentiation, phylogenetic tree and network reconstruction

The population differentiation index F_{ST} (Weir and Cockerham 1984) between populations was calculated using VCFTOOLS (Danecek et al. 2011) on the N-RNA-seq dataset. The negative values were reassigned to zero when calculating the mean genome-wide F_{ST} . In addition, we calculated the d_{XY} value per locus (Foote et al. 2016), which is the average number of nucleotide substitutions, using a custom Perl script. For the loci with $F_{ST} = 1$ (both fixed in *P. wilsonii* v.s. PLSC and PLSC v.s. *P. brachytyla* s.s., and/or both fixed in *P. wilsonii* v.s. PLSC and *P. wilsonii* v.s. *P. brachytyla* s.s.), we performed phylogenetic tree reconstruction using RAxML ver.8.1.20 (Stamatakis 2014) with the GTRGAMMA model and 1000 bootstrap replications, due to the small dataset size. To identify functional categories associated with environmental adaptation, we aligned the loci with $F_{ST} = 1$ to the *Arabidopsis thaliana* proteome using BLASTX.

Given the size of the N-RNA-seq dataset, we constructed a maximum-likelihood (ML) tree using RAxML ver.8.1.20 (Stamatakis 2014) with the GTRCAT model and 200 bootstrap replications performed to assess the branch reliability. An unrooted phylogenetic network analysis was performed with SplitTree4 ver. 4.14.6 (Huson 1998; Huson and Bryant 2005), which can also account for more realistic models such as gene recombination, hybridization, and duplication events. The program is suitable when there is evidence of hybridization events between species. We excluded the outgroup individuals from the N-RNA-seq dataset and designated the resulting dataset for this analysis as the W-RNA-seq dataset. For all phylogenetic analyses, unless otherwise noted, we utilized *P. breweriana* as the outgroup and visualized the resulting trees using FigTree ver. 1.4.0 (<http://tree.bio.ed.ac.uk/software/figtree/>), except in cases where this was unfeasible.

Plastid and mitochondrial genome trees reconstruction

To infer a history of hybrid speciation, we investigated evolutionary relationships from the plastid and mitochondrial genomes with paternal or maternal inheritances. We selected needle leaves from 15 trees (consisting of *P. likiangensis*, *P. purpurea*, *P. wilsonii*, *P. brachytyla* s.s.) and obtained resequencing data using an Illumina HiSeq 2500 (Table S3). After cleaning the raw data using Trimmomatic ver. 0.38 (Bolger et al. 2014) as mentioned above, high-quality reads were used for plastid and mitochondrial genome assembly with NOVOPlasty ver. 2.7 (Dierckxsens et al. 2017) and MITObim ver. 1.8 (Hahn et al. 2013), respectively, using default parameters. Target CDS sequences for constructing phylogenetic trees were identified from draft genomes using BLAST ver. 2.2.30+ (Camacho et al. 2009) with *P. glauca* plastid and mitochondrial CDS sequences as reference, and only

complete CDS sequences were retained for subsequent analyses. The complete CDS sequences from 16 individuals (with *P. glauca* as outgroup) were aligned using MAFFT ver. 7.313 (Katoh and Standley 2013) and then trimmed using TrimAl v1.2 (Capella-Gutiérrez et al. 2009). The final alignments from the plastid and mitochondrial CDS sequences were concatenated separately and analyzed using RAxML's rapid bootstrap algorithm with the GTRGAMMA model and 1000 bootstrap replicates to obtain the best-scoring ML trees.

Population structure and ADMIXTURE analyses

We thinned the W-RNA-seq dataset based on linkage disequilibrium (LD) using PLINK ver.1.07 (Purcell et al. 2007; Danecek et al. 2011) with the parameters “--indep-pairwise 50 5 0.2” to reduce bias. This resulted in a set of 208,633 SNPs for population structure analyses. We performed a principal component analysis (PCA) using the ‘smartpca’ program from the EIGENSOFT package ver. 6.0.1 (Price et al. 2006), and eigenvectors were generated with the R function ‘region’. A Tracy–Widom test was performed in R to determine the significance level of the eigenvectors. We used ADMIXTURE ver. 1.23 (Alexander and Lange 2011) to perform an unsupervised ancestry component analysis, with the K value (number of assumed ancestral components) ranging from 2 to 10. For each K , we performed 200 bootstrap replicates to calculate cross-validation error (CV). The optimal K value was indicated by the lowest CV values among the numbers assumed.

Species-level transcriptome assembly

To gain deeper insights into the history of hybrid speciation, we created an additional dataset. The high-quality reads obtained from the samples were used to assemble species-level transcriptomes for several species. For *P. brachytyla* s.s., 14 individuals were randomly selected from each of the populations examined (marked with a dark green color in Table S2) and Trinity ver. 2.6.6 (Grabherr et al. 2011) with default parameters was used to assemble the transcriptome based on pooled libraries to reduce the gene loss caused by random variation in expression. For the PLSC, one individual from each of the sampled populations of two varieties of *P. likiangensis* (var. *linzhiensis* and var. *likiangensis*), the *P. brachytyla* southern lineage in Yunnan, and *P. farreri* (marked with a light green color in Table S2) were selected. *P. likiangensis* var. *rubescens* was excluded from the PLSC because our preliminary ADMIXTURE analyses suggested that this variety shared many species-specific nuclear elements from another hybrid species, *P. purpurea*, due to a second hybridization event. The species-level transcriptomes for *P. wilsonii* and *P. breweriana* were obtained from a previous study (Ru et al. 2018).

To obtain a high-quality transcriptome, several analyses were performed on the assembled transcriptomes. First, CD-HIT ver. 4.6.1 (Huang et al. 2010) was used to retain a set of non-redundant, representative sequences for the assembled with a threshold value of 0.95. Next, coding and peptide sequences in the open reading frame (ORF) were predicted using TransDecoder ver. 2.0.1 (Haas et al. 2013) following the instructions described in the relevant wiki (<https://github.com/TransDecoder/TransDecoder/wiki>). This involved extracting the long open reading frames, identifying ORFs with homology to known proteins via Blast or Pfam (Finn et al. 2016) searches, and predicting the likely coding regions. In addition, the high-quality transcriptome generated from the above steps was further processed to remove bacterial contaminants using BLAST ver. 2.2.30+ (Camacho et al. 2009) and the longest transcripts were extracted with a custom Perl script. The completeness of assembled transcriptome was assessed for gene completeness using BUSCO ver. 4 (Simão et al. 2015) with the embryophyta database (<https://busco.ezlab.org/>).

Phylogenetic analyses of the orthologous sequences of four species or species complex

To identify orthologous genes for phylogenetic analyses, we used OrthoMCL (Li et al. 2003) to delineate gene families and cluster all genes into paralogous and orthologous groups based on species-level transcriptomes of the four species/species complex. We then generated a 1:1:1:1 orthologous gene dataset for *P. brachytyla* s.s., *P. wilsonii*, PLSC, and *P. breweriana* (outgroup) with a custom Perl script.

The amino acid sequences for each ortholog group (OG) were aligned using MAFFT ver. 7.313 (Katoh and Standley 2013), and poorly aligned regions were excluded using TrimAl ver. 1.2 with “-fasta -gappout -columbering”. The protein-coding nucleotide sequences for each OG were then aligned based on the corresponding amino acid alignments

using PAL2NAL ver. 14 (Suyama et al. 2006) to ensure the correct reading frames.

We constructed phylogenies for each gene using RAxML's rapid bootstrap algorithm under the GTRGAMMA model with 1000 bootstrap replicates to find the best-scoring ML tree. We restricted this analysis to those groups satisfying the following criteria: sequence length >300 bp with the '-' character excluded. Gene trees with <70% bootstrap support were excluded from further analysis. A custom R script was used to count the number of resulting phylogenies showing different topologies.

We used 3305 orthologous gene trees with >70% bootstrap support to infer interspecific relationships with PhyloNet ver. 3.6.1 (Than et al. 2008; Yu et al. 2014). Rooted trees were converted into the required input format with a custom Perl script. For the analysis in PhyloNet, we calculated maximum likelihood (using the command InferNetwork_MPL) in a coalescent framework, taken into account both incomplete lineage sorting and gene flow, and allowing 0, 1, and 2 reticulations in 100 runs to return the best network.

We applied a synonymous substitution based (K_s -based) method to estimate divergence between species pairs. K_s values for each species pair were calculated using the ML method implemented in codeml of the PAML package (Yang 1997) under the F3 × 4 model (Goldman and Yang 1994). We discarded all pairs with a K_s value of less than 0.001 from the time estimation, as this threshold would include transcript isoforms as well as recent tandem duplications.

Testing HHS using coalescent simulations based on population genomic data

To examine the evolutionary relationships among PLSC, *P. brachytyla* s.s., and *P. wilsonii*, we used fastsimcoal2 ver. 2.6.0.3 (Excoffier et al. 2013) to compare predefined demographic models using coalescent simulations based on the site frequency spectrum of all sampled individuals of these species. When constructing two-dimensional joint site frequency spectra (2D-SFS) for each pair of species, we included only the four-fold Degenerate Synonymous Sites (4DTV) with ngsTools (Fumagalli et al. 2014). As we did not have information about the ancestral state, we treated the transcriptome of *P. abies* as both the reference and the ancestral state. Subsequently, we folded all the 2D-SFSs with the 'fold' function implemented in dadi ver. 1.7.0 (Gutenkunst et al. 2009). In total, we compared 16 different evolutionary models (Fig. S5), of which 11 models (1–11) represented dichotomous or radiative topologies with or without gene flow after divergence, three (models 12–14) represented HHS via a single hybridization event with or without migration/size-change after divergence, and two (models 15–16) represented models of hybrid speciation involving an extinct intermediate hybrid lineage in the origin of *P. brachytyla* s.s.

For each model, we performed 100,000 coalescent simulations to estimate the expected 2D-SFS and computed log-likelihoods based on simulated and observed 2D-SFS matrixes. Global maximum likelihood estimates for each model were obtained from 50 independent runs, with 30–50 conditional maximization algorithm cycles. The relative fit of each of the different demographic models to the data was evaluated using the Akaike Information Criterion (AIC), and the model with the minimum AIC value was considered optimal. We assumed a mutation rate of 4.01×10^{-8} per site per generation (μ) and a generation time (T) of 50 years (Li et al. 2010; De La Torre et al. 2017). A parametric bootstrapping approach was used to construct 95% confidence intervals with 50 independent runs for each bootstrap.

We used the reduced PLSC (with var. *rubescens* excluded) to examine the HHS origin of *P. brachytyla* s.s. because var. *rubescens* contains numerous introgressions from *P. purpurea*, which is assumed to have originated from the same parents (Ru et al. 2018). This introgression, and/or the likely hybrid origin through *P. purpurea*, may complicate the modeling results.

We further examined the diploid hybrid origin of *P. purpurea* from the reduced PLSC and *P. wilsonii*. We examined four alternative speciation models for the origin of *P. purpurea* because we had tested multiple models and all models suggested that this species originated through HHS (Ru et al. 2018) (Fig. S6). We tested which of four HHS models fit the reduced PLSC and combined the origins of *P. brachytyla* s.s. and *P. purpurea* to outline the evolutionary relationships among the reduced PLSC, *P. wilsonii*, *P. brachytyla* s.s., and *P. purpurea*.

RESULTS

Sampling, sequencing and single nucleotide polymorphism (SNP) calling

After quality control, we retained an average of 46.37 million (M) reads (50.50 M raw reads) with 6.18 gigabases (G) of clean data per individual (Table S2). The species-level assemblies for *P. brachytyla* s.s. and PLSC produced, respectively, 222,203 and 313,706 transcripts with N50 values of 549 and 667 after redundancy reduction and ORF prediction. Quality metrics (i.e., numbers of total assembled bases, total Trinity transcripts, genes, average contig length, contig N50, and percent GC) are similar to those of previously sequenced transcriptomes of *P. likiangensis*, *P. purpurea*, and *P. wilsonii* (Ru et al. 2018) (Table S4). All the assembled transcriptomes had >80% BUSCO completeness (Table 1), but they have more contigs with lower N50 values than the transcriptome of *P. abies* (Table S4). Therefore, we mapped quality-filtered reads to the revised transcriptome of *P. abies* as in our previous work (Ru et al. 2016) and called SNPs for each individual. The average mapping rate for all individuals was 56.6%, with an average coverage of the reference transcriptome assembly being 73.9% and a 48.5× average effective depth (Table S2).

Our strict filtering criteria for SNP-calling using SAMTools resulted in 10,237 contigs with 339,165 SNPs. The number of SNPs varied among the different species, with *P. brachytyla* s.s. (180,329), *P. likiangensis* (including all varieties, 160,394) and *P. wilsonii* (160,224) having the highest number of SNPs, *P. farreri* (99,459) had the lowest number of SNPs, followed by *P. brachytyla*-southern lineage (116,558), and *P. purpurea* (140,250) (Table S5).

We also looked at SNP sharing between species and found that *P. brachytyla* s.s. and *P. wilsonii* shared more SNPs with each other (114,657) and with *P. likiangensis* (119,636) than *P. likiangensis* and *P. wilsonii* shared with each other (9553) (Fig. S1A). Among *P. likiangensis*, *P. brachytyla* s.s., and *P. wilsonii*, 31,319 SNPs were specific to *P. brachytyla* s.s., 35,314 SNPs to *P. wilsonii*, and 30,505 SNPs to *P. likiangensis* (Fig. S1A).

In terms of species-specific SNPs among *P. likiangensis*, *P. farreri* and *P. brachytyla*-southern lineage, *P. likiangensis* had the highest number (41,626), followed by *P. brachytyla*-southern lineage (3898) and *P. farreri* (2545) (Fig. S1B). We also found that *P. farreri* and *P. brachytyla*-southern lineage shared a large portion of SNPs (88.51%; 88,028 of 99,459), but *P. farreri* still shared more with *P. likiangensis* (94.6%; 94,136; Fig. S1B).

Nucleotide diversity, interspecific differentiation and phylogenetic analyses

The nucleotide diversities (π) were similar among *P. likiangensis*, *P. brachytyla* s.s., *P. wilsonii*, *P. brachytyla*-southern lineage, *P. farreri*,

Table 1. BUSCO results for assembly completeness of four spruce transcriptomes.

Species	Classification	BUSCO results
PLSC	Spermatophytes	C: 84.4% S: 47.1%, D: 37.3%, F: 2.7%, M: 12.9%, n: 1440
<i>P. brachytyla</i> s.s.	Spermatophytes	C: 81.2% S: 71.7%, D: 9.5%, F: 3.0%, M: 15.8%, n: 1440
<i>P. wilsonii</i>	Spermatophytes	C: 86.1% S: 15.5%, D: 70.6%, F: 2.3%, M: 11.6%, n: 1440
<i>P. breweriana</i>	Spermatophytes	C: 85.1% S: 10.1%, D: 75.0%, F: 2.4%, M: 12.5%, n: 1440

PLSC *P. likiangensis* species complex.

and *P. purpurea* (Table S6). The mean global differentiation (F_{ST}) between each pair of the six taxa was greater than 0.05, except for *P. farreri* and *P. likiangensis* (0.043 ± 0.052) (Table S7). The F_{ST} value between *P. likiangensis* and *P. wilsonii* (0.116 ± 0.093) was higher than that between *P. brachytyla* s.s. and either *P. likiangensis* (0.089 ± 0.082) or *P. wilsonii* (0.105 ± 0.092 ; Table S7 and Figs. S1B & S2D). The absolute genetic divergence estimated as d_{XY} was higher between *P. likiangensis* and *P. wilsonii* than between *P. brachytyla* s.s. and either *P. likiangensis* or *P. wilsonii* (Table S7 and Figs. S2A & S2C). Both F_{ST} and d_{XY} indicated that *P. wilsonii* and *P. farreri* have the greatest divergence among all comparisons, followed by comparisons between *P. wilsonii* and *P. brachytyla*-southern lineage. The *P. brachytyla*-southern lineage and *P. farreri* exhibit low indications of divergence (F_{ST} : 0.064 ± 0.080 ; d_{XY} : 0.0099 ± 0.013 ; Table S7 and Figs. S2A & S2B). These patterns suggested a close relationship among *P. likiangensis*, *P. farreri*, and *P. brachytyla*-southern lineage, and we tentatively treated them as the *P. likiangensis* species complex (PLSC) for analytical tractability.

Using the N-RNA-seq dataset and *P. breweriana* as an outgroup, we inferred the genealogy for 186 individuals and identified seven lineage-specific clades. Notably, *P. brachytyla* demonstrated two distinct clusters: *P. brachytyla*-southern lineage and *P. brachytyla* s.s. (Fig. 1B). Additionally, *P. brachytyla*-southern lineage was found to be non-monophyletic. Further analysis revealed that *P. brachytyla*-southern lineage was closely related to var. *likiangensis* and *P. farreri*, and together they were related to var. *linzhiensis* and var. *rubescens*. These lineages comprise a monophyletic PLSC clade that is paratactic to *P. brachytyla* s.s., *P. purpurea*, and *P. wilsonii*. *P. purpurea* was more closely related to *P. wilsonii* than to the others, and *P. brachytyla* s.s. was sister to the *P. purpurea* – *P. wilsonii* clade (Fig. 1B).

Our analysis based on W-RNA-seq dataset, yielded a well resolved phylogenetic network that exhibited highly consistency with the ML phylogenetic tree. *P. brachytyla* was also found to be comprised of two groups: the *P. brachytyla*-southern lineage and *P. brachytyla* s.s. The former was clustered closely with the monophyletic PLSC clade, while the latter formed a cluster that is clearly distinct from other clades (Fig. 1C). These findings align with previously published results (Lookwood et al. 2013; Sun et al. 2014; Ru et al. 2018; Lyu et al. 2020). It was clear that the ancestry of *P. brachytyla* is polyphyletic (Fig. 1B & 1C).

Furthermore, we discovered that *P. brachytyla* s.s. possessed two fixed genes (comp42862_c0_seq1, comp96487_c6_seq1) that were derived from *P. wilsonii*. Interestingly, phylogenetic trees constructed using these two genes showed that PLSC formed a monophyletic clade while *P. brachytyla* s.s. and *P. wilsonii* clustered together (Figs. S4A and S4B). Importantly, we found no evidence of genes fixed from PLSC in *P. brachytyla* s.s. Notably, the genes comp42862_c0_seq1 and comp96487_c6_seq1 are involved in stomatal regulation and response to water deprivation, respectively (Table S8).

Phylogenetic discordance base on the plastid and mitochondrial genome

Previous studies have demonstrated that the species within the PLSC formed a monophyletic group in both the mitochondrial, plastid sequences as well as nuclear sequences, suggesting that *P. likiangensis* can represent PLSC (Ran et al. 2006; Bouille et al. 2011; Lockwood et al. 2013; Ran et al. 2015; Feng et al. 2019; Lyu et al. 2020). Based on the results of the maximum likelihood phylogenetic analyses of the 16 mitochondrial genomes two well resolved clades were identified: (1) the PLSC and *P. brachytyla* s.s. and (2) *P. wilsonii* and *P. purpurea* (Fig. 1D). The plastid genome phylogeny was similar to the nuclear genomic ML tree with resolved two strongly supported clades: (1) the PLSC and (2) *P. brachytyla* s.s., *P. wilsonii*, and *P. purpurea* together (Fig. 1E). Notably, the phylogenetic relationships of *P. brachytyla* s.s. were discordant between the maternally inherited mitochondrial genome tree and the paternally

inherited plastid genome tree. The topological inconsistency between the two phylogenies supports the hypothesis of a hybrid origin of *P. brachytyla* s.s. Additionally, the placement of *P. purpurea* also showed subtle differences between the mitochondrial and plastid genome phylogenies.

Population structure and ADMIXTURE analyses

The results of ADMIXTURE and PCA clustering analyses were largely consistent with the N-RNA-seq phylogenetic trees, with the exception of var. *rubescens*. At $K = 2-4$, four taxa of the PLSC, namely var. *likiangensis*, var. *linzhiensis*, *P. farreri*, and *P. brachytyla*-southern lineage, shared the same genetic composition (Fig. 2A). However, var. *rubescens* showed genetic ancestry shared with the reduced PLSC, *P. wilsonii*, and *P. purpurea*. Additionally, *P. brachytyla* s.s. exhibited mixed ancestry with the reduced PLSC and *P. wilsonii* at $K = 2$, but formed a separate cluster at $K = 3$. At $K = 4$, *P. purpurea* formed a distinct cluster, although it exhibited mixed genetic ancestry with both the reduced PLSC and *P. wilsonii* at lower values of K . The ADMIXTURE analyses suggested that if both *P. brachytyla* s.s. and *P. purpurea* originated from the same parents, composed of the reduced PLSC and *P. wilsonii*, then *P. purpurea* originated later than *P. brachytyla* s.s. Within the reduced PLSC, var. *linzhiensis* separated from the other three early while *P. farreri* contained a mixture of genetic elements from var. *linzhiensis* and var. *likiangensis* + *P. brachytyla*-southern lineage. Both var. *likiangensis* + *P. brachytyla*-southern lineage belonged to the same genetic pool without clear separation at $K = 5-6$ (best $K = 5$, with lowest CV error = 0.162), although var. *rubescens* separated as an independent cluster (Fig. 2A). Further ADMIXTURE analyses of the PLSC individuals produced similar results (Fig. S3).

In the PCA analysis, the first three components (PC1, PC2 and PC3) explained 12.43, 9.10 and 5.11% of the total variance respectively (Fig. 2B and Table S9). The analysis distinguished four distinct clusters: *P. brachytyla* s.s., *P. wilsonii*, *P. purpurea*, and the PLSC. Within the PLSC, all individuals of var. *rubescens*/*linzhiensis* formed a separate cluster from the other four, while var. *likiangensis*, *P. brachytyla*-southern lineage and *P. farreri* clustered together. These results support the conclusion that *P. likiangensis* can represent the PLSC, as demonstrated in previous studies based on mitochondrial, plastid, and nuclear sequences.

PhyloNet test of HHS

After identifying 6471 orthologous gene groups across four taxa, filtering was performed and 6226 of these were used to generate gene trees. Out of the 3305 gene trees with $\geq 70\%$ bootstrap support for all branches, 1129 (34.16%) clustered the reduced PLSC with *P. brachytyla* s.s. (topo1), 959 (29.02%) showed *P. brachytyla* s.s. as an isolated clade with *P. wilsonii* and the reduced PLSC clustered together (topo2), and 1217 (36.82%) clustered *P. wilsonii* and *P. brachytyla* s.s. together (topo3; Fig. 2C). The resulting phylogenetic network inferred with an assumption of one hybridization event (Fig. 2D) indicated a hybrid origin for *P. brachytyla* s.s. The contributions from the reduced PLSC (~80%) were greater than those from *P. wilsonii* (~20%; Fig. 2D).

Relative divergence time and K_s test of HHS

Timing of divergence between *P. brachytyla* s.s. and its putative parents, *P. wilsonii* and the PLSC, was examined using a K_s -based method. This method allowed us to determine if the hybrid species diverged from its parents at approximately the same time. Our results showed the K_s value between *P. brachytyla* s.s. and PLSC was similar to the that between *P. brachytyla* s.s. and *P. wilsonii*, indicating a close divergence time between these three species. Furthermore, the K_s value between the PLSC and *P. wilsonii* was greater than those between *P. brachytyla* s.s. and its parents, suggesting that these two parental species diverged from each other earlier than they diverged from *P. brachytyla* s.s. (Fig. 2E).

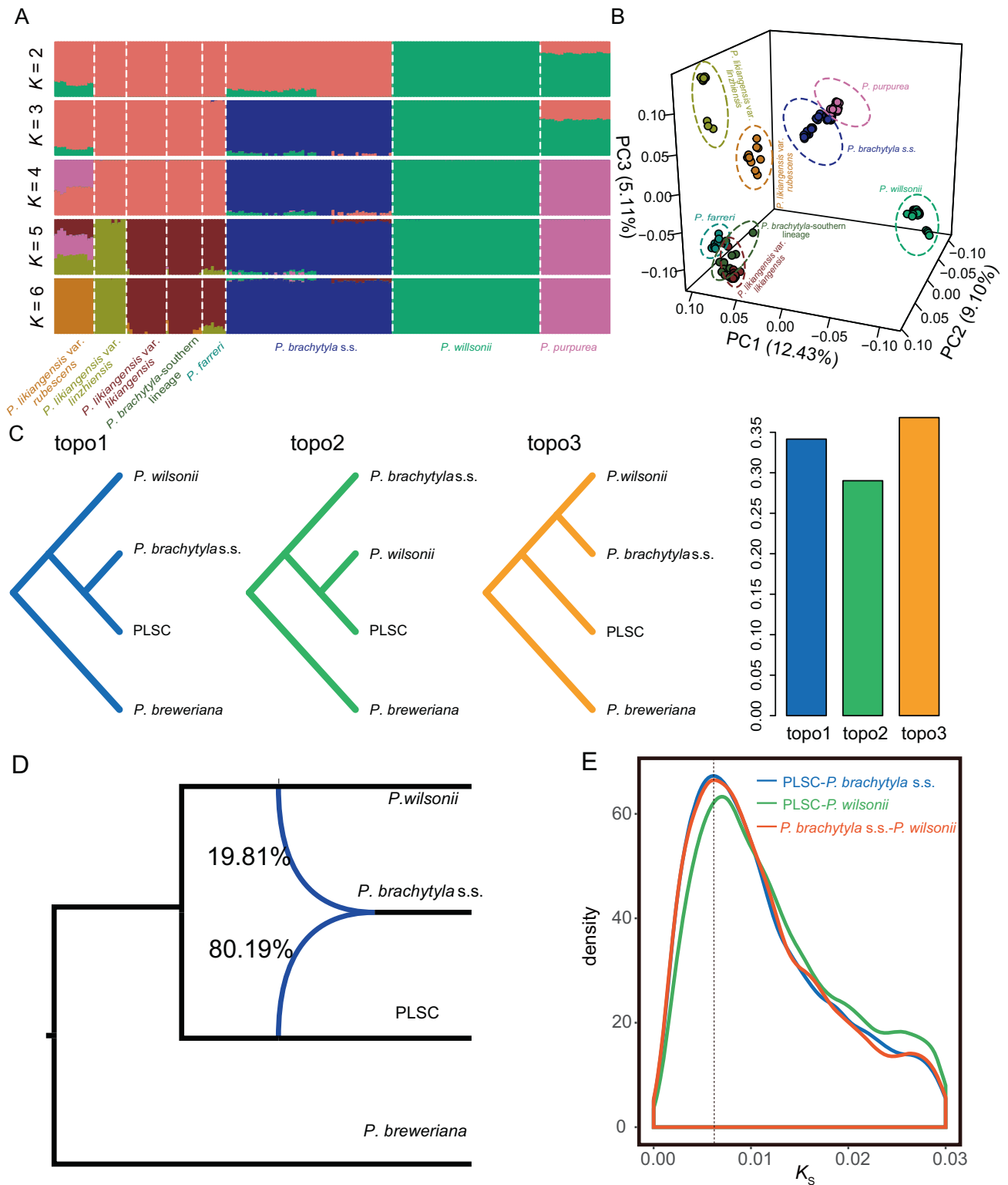


Fig. 2 Genetic clustering using SNPs and HHS testing with orthologous genes. **A** Bar plots indicative of assignment probabilities from ADMIXTURE analysis of 184 transcriptomes assuming numbers of clusters (K) 2 to 6. **B** Principal component analysis (PCA) plots showing the first three principal components. **C** Relationships between *P. likiangensis* species complex (PLSC), *P. brachytyla* s.s., and *P. wilsonii* (using *P. breweriana* as outgroup) according to ML analysis of 3305 orthologous gene sequences identified by OrthoMCL as having a ratio of 1:1:1. And the proportion of each topology is based on the ortholog groups. **D** ML-bootstrap network for 3305 orthologous gene trees generated by PhyloNet after runs allowing 0, 1, or 2 reticulations. Reticulations are shown in blue with inheritance probabilities. Note: Only trees with branch bootstrap values > 70% were analyzed. **E** K_s age distributions for 1:1 orthologs between the PLSC and *P. brachytyla* s.s. and *P. wilsonii*. PLSC: *P. likiangensis* species complex.

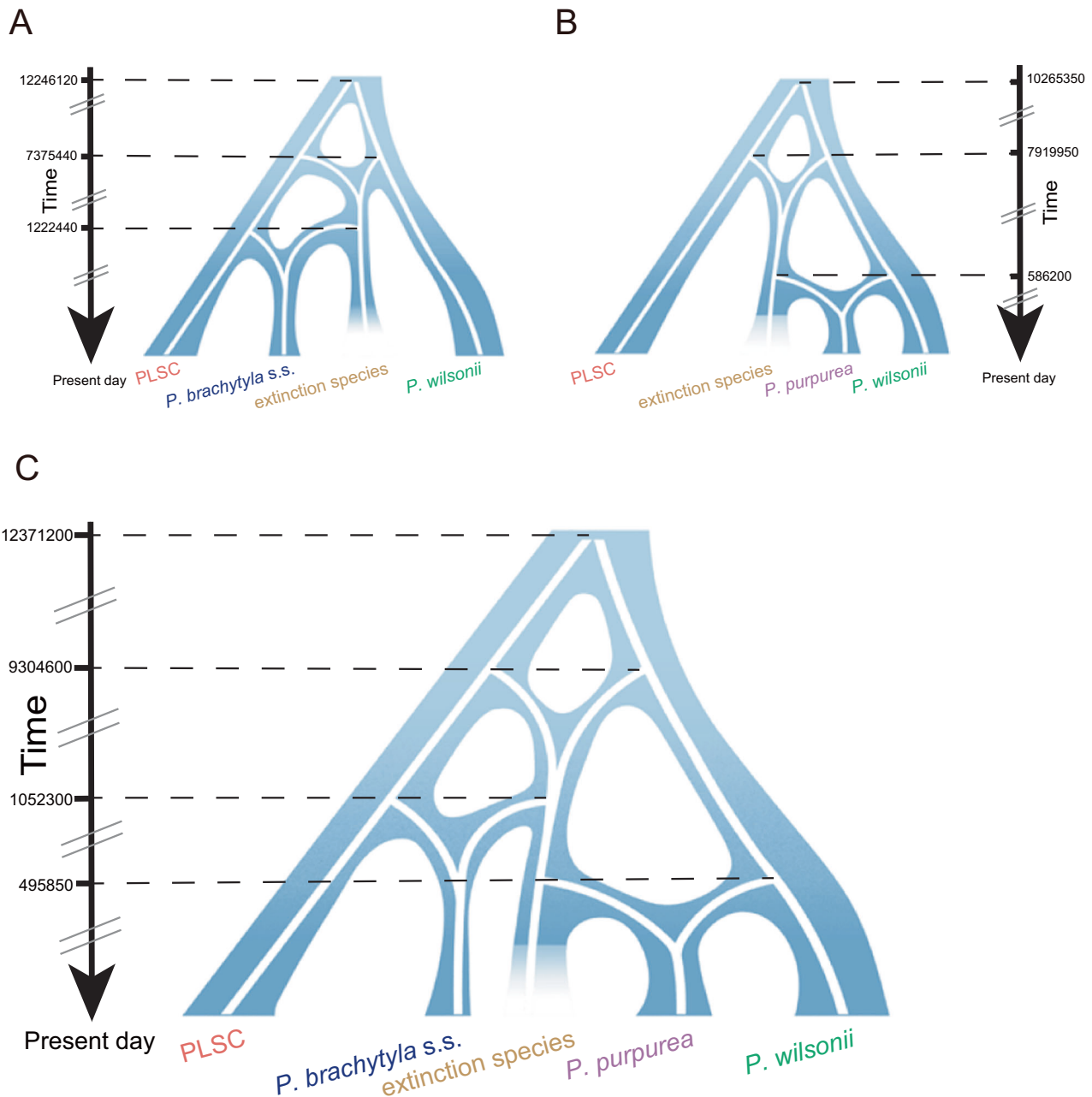


Fig. 3 Inference of demographic histories using fastsimcoal2. **A** Simplified graphical summary of the best-fitting demographic model inferred by fastsimcoal2 for the reduced PLSC, *P. brachytyla* s.s., and *P. wilsonii*. **B** Simplified graphical summary of the best-fitting demographic model inferred by fastsimcoal2 for the reduced PLSC, *P. purpurea*, and *P. wilsonii*. **C** Simplified graphical summary of the best-fitting demographic model inferred by fastsimcoal2 for the reduced PLSC, *P. brachytyla* s.s., *P. purpurea*, and *P. wilsonii*. The percentages indicate nuclear genomic compositions contributed from parents to the hybrid offspring. The right-hand axis indicates the timescale in years before the present. PLSC: *P. likiangensis* species complex.

Coalescent analysis of alternative speciation patterns

The best-fitting model for the origin of *P. brachytyla* s.s. was determined using the AIC value, with model15 (involving hybridization and backcrossing) having the lowest value (Table S10). This indicates that *P. brachytyla* s.s. originated through backcrossing between the reduced PLSC and an extinct hybrid lineage which was initially formed through hybridization between PLSC and *P. wilsonii* (Fig. 3A). This is also supported by estimated effective population size of each lineage, divergence of the two assumed parents, and the timescale of the homoploid hybridization events (Table 2 and Fig. 3A). Similarly, the AIC value for the origin of *P. purpurea* suggested backcrossing between *P. wilsonii*

and an extinct hybrid lineage initially formed through hybridization between the reduced PLSC and *P. wilsonii* (model4; Fig. 3B and S6 and Table S11). The estimated effective population sizes for both parents and the timescale of the origin of the extinct hybrid lineage were similar to those for an HHS origin of *P. brachytyla* s.s. (Table 2 and S12).

We therefore assumed that two homoploid hybrid species, *P. brachytyla* s.s. and *P. purpurea*, originated through hybridization and subsequent backcrossing to different parents (PLSC and *P. wilsonii*), rather than through bifurcation followed by introgression. Specifically, our analysis suggests that both species share a common extinct intermediate hybrid lineage formed by

Table 2. Inferred demographic parameters of the best-fitting demographic model in Fig. S4.

Parameters	Point estimation	95% CI Lower bound	95% CI Upper bound
N_{e-PLSC}	16614	11,301	22,342
$N_{e-P. wilsonii}$	32321	21,375	36,121
$N_{e-P. brachytyla s.s.}$	18111	13,031	21,685
$N_{e-ghost}$	78107	10,442	1,72,368
$m^1_{PLSC \rightarrow P. wilsonii}$	2.48e-06	1.29e-08	9.83e-04
$m^1_{P. wilsonii \rightarrow PLSC}$	3.37e-06	2.22e-08	9.25e-04
$m^2_{PLSC \rightarrow P. wilsonii}$	2.63e-03	6.85e-04	4.84e-02
$m^2_{P. wilsonii \rightarrow PLSC}$	1.36e-07	6.99e-09	2.58e-02
$m_{PLSC \rightarrow P. brachytyla s.s.}$	4.84e-05	2.65e-05	9.42e-05
$m_{P. brachytyla s.s. \rightarrow PLSC}$	6.09e-05	1.59e-05	1.31e-04
$m_{P. wilsonii \rightarrow P. brachytyla s.s.}$	7.70e-05	3.22e-05	1.21e-04
$m_{P. brachytyla s.s. \rightarrow P. wilsonii}$	1.27e-05	9.32e-07	2.87e-05
$m_{P. wilsonii \rightarrow PLSC}$	2.94e-05	7.73e-06	2.31e-02
$m_{PLSC \rightarrow P. wilsonii}$	1.42e-05	2.34e-07	6.82e-05
T_{ADM1}	1222440	1005800	5478200
T_{ADM2}	7375440	5817900	13122100
T_{DIV}	12246120	10157600	16086500

N_{e-PLSC} , $N_{e-P. brachytyla s.s.}$, $N_{e-P. wilsonii}$, $N_{e-ghost}$ indicate the effective population sizes of the reduced *P. likiangensis* species complex (PLSC), *P. brachytyla s.s.*, *P. wilsonii*, the ghost intermediate lineage and ancestral population respectively. $m^1_{PLSC \rightarrow P. wilsonii}$, $m^1_{P. wilsonii \rightarrow PLSC}$, $m^2_{PLSC \rightarrow P. wilsonii}$, $m^2_{P. wilsonii \rightarrow PLSC}$ indicate migration per generation before and after hybridization between PLSC and *P. wilsonii*; $m_{PLSC \rightarrow ghost}$: migration per generation from PLSC to the ghost lineage and $m_{ghost \rightarrow PLSC}$ migration per generation from ghost lineage to PLSC. $m_{PLSC \rightarrow P. brachytyla s.s.}$, $m_{P. brachytyla s.s. \rightarrow PLSC}$, $m_{P. wilsonii \rightarrow P. brachytyla s.s.}$ and $m_{P. brachytyla s.s. \rightarrow P. wilsonii}$ indicate, respectively, migration per generation between *P. brachytyla s.s.* and PLSC or *P. wilsonii* in both directions. T_{ADM1} indicates time (years) of backcrossing of the ghost lineage to *P. wilsonii* that gave rise to *P. brachytyla s.s.*, while T_{ADM2} indicates time (years) of formation of the ghost lineage between PLSC and *P. wilsonii*. T_{DIV} indicates the estimated divergence time (years) between PLSC and *P. wilsonii* obtained from fastsimcoal2.

hybridization between the reduced PLSC and *P. wilsonii*. Our estimates suggest that the extinct hybrid lineage originated ~9.3 Ma (95%HPDI: 5.5–12.3 Ma), while *P. brachytyla s.s.* and *P. purpurea* originated more recently, ~1.1 Ma (95%HPDI: 0.8–5.2 Ma), and ~0.50 Ma (95%HPDI: 0.4–3.4 Ma), respectively, through further backcrossing to the reduced PLSC or *P. wilsonii* (Fig. 3C and Table S13). We also estimated the current effective population sizes (N_e) of the reduced PLSC, *P. wilsonii*, and *P. brachytyla s.s.*, which were estimated to be 15,016 (95%HPDI: 11,106–51,868), 27,194 (95%HPDI: 18,162–89,603), and 16,340 (95%HPDI: 11,700–57,543), respectively. In addition, gene flow from *P. brachytyla s.s.* to PLSC was estimated to be greater than the opposite direction, and gene flow from *P. brachytyla s.s.* to *P. wilsonii* estimated to be less than the opposite direction. Both values were greater than that between PLSC and *P. wilsonii* (Table 2 and S13).

The estimated timescale of the extinct hybrid lineage was slightly earlier (9.3 Ma) compared to the previous estimations of 7.3 Ma or 7.9 Ma when only one hybrid species was involved. However, the estimated timescale of hybrid origin for both *P. brachytyla s.s.* (1 Ma) and *P. purpurea* (0.49 Ma) were later than the previous estimates of 1.2 Ma and 0.58 Ma, respectively, when only one hybrid species was involved. This may be a consequence of the changes in the N_e of the four lineages when all of them were involved in the estimations.

DISCUSSION

Our analyses confirmed the polyphyly of the previously circumscribed *P. brachytyla* (Ru et al. 2016; Lyu et al. 2020), and identified two non-sister lineages: the *P. brachytyla*-southern lineage and *P. brachytyla s.s.* We found evidence for a hybrid origin of *P. brachytyla s.s.* through an extinct intermediate hybrid lineage which backcrossed to one parent in the PLSC. Interestingly, we found that the same parents may have given rise to another homoploid hybrid species, *P. purpurea*, through the same extinct hybrid lineage which backcrossed to the other parent, *P. wilsonii*. To our knowledge, this is the first case illustrating that backcrossing to two parents produced two different homoploid species. Our findings recovered the high HHS complexity in *Picea*.

Polyphyly of *P. brachytyla*

Our population genomic data provides strong support for the polyphyly of the previously circumscribed *P. brachytyla*, consistent with previous studies based on a limited number of individuals (Lockwood et al. 2013; Zou et al. 2016; Ru et al. 2016; Shao et al. 2019; Shen et al. 2019) and genetic loci (Lyu et al. 2020). Phylogenetic (i.e., trees and networks) and population genetic analyses (i.e., PCA and ADMIXTURE) identified two groups: the *P. brachytyla*-southern lineage being closely related to var. *likiangensis* and *P. farreri* of the PLSC and the northern lineage (*P. brachytyla s.s.*) being independent of all other taxa (Figs. 1B, 1C, 2A, & 2B). In particular, the *P. brachytyla*-southern lineage, *P. likiangensis*, and *P. farreri* were embedded within the PLSC clade. Admixture analyses further revealed that *P. brachytyla s.s.* likely originated earlier than the *P. brachytyla*-southern lineage. Interestingly, we found that *P. brachytyla s.s.* and the *P. brachytyla*-southern lineage have similar leaf traits, including, for example, a lack of a stomatal line in the abaxial surface; however, they differ from each other in cone colors. Therefore, the polyphyly of *P. brachytyla* inferred here based on more samples supported our previous findings that leaf traits might have experienced parallel adaptation in response to the climatic stress (Ru et al. 2016).

Species distinctness and homoploid hybrid origin of *P. brachytyla s.s.*

The reconstructed phylogenetic tree and network based on the consensus sequences revealed that the samples of *P. brachytyla s.s.* clustered together into a distinct group, despite possible reticulation and intra-lineage divergence (Fig. 1B and C). Our population genomic results showed that the genetic structure of *P. brachytyla s.s.* was distinct from that of *P. wilsonii* and the PLSC (Fig. 2A and B). A detailed investigation of SNPs found 31,319 variants unique to the genomes of *P. brachytyla s.s.* (Fig. S1). These results supported the distinctness of *P. brachytyla s.s.* as an evolutionary lineage. Additionally, an independent niche modeling study of spruce species suggested that *P. brachytyla s.s.* occupies a unique ecological space (Wang et al. 2017). Thus, *P. brachytyla s.s.* is genetically and ecologically delimited from both *P. wilsonii* and the PLSC.

The present results showed that *P. brachytyla s.s.* had accumulated some private mutations and that hybridization between the PLSC and *P. wilsonii* had contributed the fundamental genetic components to *P. brachytyla s.s.* These facts satisfy two primary criteria of the HHS model (Schumer et al. 2014). To clarify the hybrid origin of *P. brachytyla s.s.*, we performed multiple analyses including coalescent testing and phylogenetic tree/network analyses.

Firstly, direct evidence for the hybrid origin includes the discovery that the mitochondrial genome of *P. brachytyla s.s.* seemed to be sourced from the PLSC, while the plastid genome of *P. brachytyla s.s.* seemed to be inherited from *P. wilsonii* (Figs. 1D and 1E). This inconsistency in inheritance patterns has also been previously reported based on the sequencing of a few loci

(Lockwood et al. 2013; Ran et al. 2015; Zou et al. 2016; Shao et al. 2019; Shen et al. 2019; Lyu et al. 2020).

Secondly, analyses of genetic structure similarly suggested that both the PLSC and *P. wilsonii* contributed genetic information to *P. brachytyla* s.s., but with the former making a greater contribution (80.19%) than the latter (19.81%) (Figs. 2A, 2B, and 2D). *P. wilsonii* contributed little genomic information into *P. brachytyla* s.s., but *P. brachytyla* s.s. had fixed genetic variants from *P. wilsonii* at two genes (comp42862_c0_seq1, comp96487_c6_seq1; Figs. S4A and S4B and Table S8), supporting the hybrid origin of *P. brachytyla* s.s. substantially. This is further supported by our coalescent analyses (Fig. 3A).

It is noteworthy that the two identified genes may have helped *P. brachytyla* s.s. to adapt to water scarcity, which is a critical limiting factor for plant growth and productivity, especially in arid or semi-arid environments (Meinzer et al. 2006; Battipaglia et al. 2013). Studies have shown that the shortage of water and increased drought stress can lead to tree dieback (Anderegg et al. 2015; Hember et al. 2017; Choat et al. 2018). This adaptation may have allowed *P. brachytyla* s.s. to occupy new niches successfully.

Finally, the hybrid origin of *P. brachytyla* s.s. is confirmed by its genetic divergence from its two putative parents. K_s analysis based on shared orthologous genes revealed a similar divergence time between *P. brachytyla* s.s. and either *P. likiangensis* or *P. wilsonii*, but more recent than that between the PLSC and *P. wilsonii* (Fig. 2E).

Interestingly, when we modeled an alternative speciation model with an extinct lineage, *P. brachytyla* s.s. seemed to be formed in two steps with the first intermediate hybrid lineage being formed between the PLSC and *P. wilsonii* followed by backcrossing of this extinct intermediate lineage with the PLSC (Fig. 3A and Table 2). This two-step evolutionary model is very similar to that of *P. purpurea* (Ru et al. 2018).

The same extinct parental lineage for *P. purpurea* and *P. brachytyla* s.s.

In a previous study of *P. purpurea* (Ru et al. 2018), we used three varieties of *P. likiangensis* to represent the PLSC (Sun et al. 2014; Ru et al. 2018). Recent studies proposed that both *P. farneri* and *P. brachytyla*-southern lineage should be included in the PLSC (Sun et al. 2018; Lyu et al. 2020). In addition, var. *rubescens* shared some common genetic information with *P. purpurea*, according to our population structure analyses and previous studies (Sun et al. 2018; Shao et al. 2019; Shen et al. 2019). These results may alter the conclusion of hybrid origin of *P. purpurea*, and also may complicate the modeling of speciation hypotheses. We therefore removed var. *rubescens* from the PLSC in present analyses and used the reduced PLSC and *P. wilsonii* to model HHS versus an alternative, bifurcating, origin for *P. purpurea*. Similar to previous findings (Ru et al. 2018), *P. purpurea* originated through HHS between the reduced PLSC and *P. wilsonii* by a two-step process that formed an intermediate hybrid lineage, which further backcrossed with *P. wilsonii* to produce *P. purpurea* (Fig. 3B, S6 and Table S11). Thus both *P. purpurea* and *P. brachytyla* s.s. originated by HHS from the same parents, PLSC and *P. wilsonii*. However, *P. purpurea* originated later than *P. brachytyla* s.s., accumulating fewer species-specific mutations and retaining more parental ancestry. The mosaic of parental ancestry in *P. purpurea* was always obvious in the ADMIXTURE result based on population genomic data (Fig. 2A). However, for *P. brachytyla* s.s., more species-specific mutations blurred the evidence of a mixed ancestry, although inconsistent phylogenies (Fig. 1B, C, D and E), PhyloNet suggestions (Fig. 2D) and coalescent tests (Fig. 3A) together support an HHS origin for this species.

In addition, we assumed that the two homoploid hybrid species originated through the same extinct intermediate hybrid lineage but that it backcrossed with different parents for two reasons. First, this is the most parsimonious hypothesis. Although it is likely

that two ancient hybrid lineages may have originated from hybridization between the PLSC and *P. wilsonii*, it is less likely that both of them became extinct. Second, the time of origin of the extinct hybrid lineage was estimated to be similar (7.4 or 7.9 Ma; Fig. 3A, B; Tables 3 and S11) when only two parents and one assumed hybrid species, *P. brachytyla* s.s. or *P. purpurea*, was involved. When all four taxa were involved into one complex model, the origin of the extinct hybrid lineage was estimated to be earlier, around 9.3 Ma (Fig. 3C), while the two hybrid species, *P. purpurea* and *P. brachytyla* s.s., originated 1 Ma and 0.49 Ma, respectively (Fig. 3A and B; Table S13). These divergence events occurred in the late Miocene and the Pleistocene on the Qinghai-Tibet Plateau (QTP) and adjacent regions, where extensive geological activities and climatic oscillations had occurred (Deng and Ding 2015; Mulch and Chamberlain 2006). During these oscillations, numerous new species originated and interspecific hybridizations occurred (Liu et al. 2013; Du et al. 2017; Ma et al. 2019). These extensive changes may therefore have facilitated the secondary contacts of diverged *Picea* taxa, resulting in inter-lineage hybridizations and backcrosses, and leading to the origin of the hybrid lineage. It remains unknown how the intermediate hybrid lineage was extinguished. It is likely that environmental changes or adaptive advantages of the newly formed hybrid species resulted in the extinction by hybridization-derived lineage fusion or replacement of the intermediate hybrid lineage (Ru et al. 2018).

To our knowledge, this is the second reported case in which two parental taxa produced more than one homoploid hybrid species. In sunflowers, three homoploid hybrid species occurring in different extreme habitats were found to be derived from hybridization between two sunflower species (Rieseberg et al. 1997, 2003; Todesco et al. 2020). In our case, the genetic components of two hybrid spruce taxa (*P. purpurea* and *P. brachytyla* s.s.) were contributed by the PLSC and *P. wilsonii*. However, we found the signal of one extinct hybrid lineage which produced *P. purpurea* and *P. brachytyla* s.s. through backcrossing with *P. wilsonii* and the PLSC, respectively (Fig. 3). This finding supports the hypothesis that in the spruce genus, reticulate species diversification through hybridization rather than non-bifurcating divergence seems to be more frequent than previously thought (Feng et al. 2019).

DATA AVAILABILITY

The sequencing data have been deposited in National Genomics Data Center under the BioProject ID: PRJCA003239.

REFERENCES

- Alexander DH, Lange K (2011) Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinforma* 12:246. <https://doi.org/10.1186/1471-2105-12-246>
- Anderegg WRL, Flint A, Huang C-Y, Flint L, Berry JA, Davis FW et al. (2015) Tree mortality predicted from drought-induced vascular damage. *Nat Geosci* 8:367–371
- Battipaglia G, DeMicco V, Brand WA, Saurer M, Aronne G, Linke P et al. (2013) Drought impact on water use efficiency and intra-annual density fluctuations in *Erica arborea* on Elba (Italy). *Plant Cell Environ* 37:382–391
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bouille M, Senneville S, Bousquet J (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
- Brennan AC, Hiscock SJ, Abbott RJ (2019) Completing the hybridization triangle, the inheritance of genetic incompatibilities during homoploid hybrid speciation in ragworts (*Senecio*). *AoB Plants* 11:ply078. <https://doi.org/10.1093/aobpla/ply078>
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K et al. (2009) BLAST+: architecture and applications. *BMC Bioinforma* 10:421. <https://doi.org/10.1186/1471-2105-10-421>

- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>
- Chapman MA, Hiscock SJ, Filatov DA (2013) Genomic divergence during speciation driven by adaptation to altitude. *Mol Biol Evol* 30:2553–2567. <https://doi.org/10.1093/molbev/mst168>
- Choat B, Brodribb TJ, Brodersen CR, Duursma RA, López R, Medlyn BE (2018) Triggers of tree mortality under drought. *Nature* 558:531–539
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA et al. (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- De La Torre AR, Li Z, Van D, Peer Y, Ingvarsson PK (2017) Contrasting rates of molecular evolution and patterns of selection among gymnosperms and flowering plants. *Mol Biol Evol* 34:1363–1377. <https://doi.org/10.1093/molbev/msx069>
- Deng T, Ding L (2015) Paleoaltimetry reconstructions of the Tibetan Plateau, progress and contradictions. *Natl Sci Rev* 2:417–437. <https://doi.org/10.1093/nsr/nwv062>
- Dierckxsens N, Mardulyn P, Smits G (2017) NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res* 45:e18. <https://doi.org/10.1093/nar/gkw955>
- Du F, Hou M, Wang W, Mao K, Hampe A (2017) Phylogeography of *Quercus aquifolioides* provides novel insights into the Neogene history of a major global hotspot of plant diversity in south-west China. *J Biogeogr* 44:294–307. <https://doi.org/10.1111/jbi.12836>
- Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M (2013) Robust demographic inference from genomic and SNP data. *PLoS Genet* 9:e1003905. <https://doi.org/10.1371/journal.pgen.1003905>
- Feng S, Ru D, Sun Y, Mao K, Milne R, Liu J (2019) Trans-lineage polymorphism and nonbifurcating diversification of the genus *Picea*. *N Phytologist* 222:576–587. <https://doi.org/10.1111/nph.15590>
- Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL et al. (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 44:D279–285. <https://doi.org/10.1093/nar/gkv1344>
- Foote AD, Vijay N, Avila-Arcos MC, Baird RW, Durban JW, Fumagalli M et al. (2016) Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. *Nat Commun* 7:11693. <https://doi.org/10.1038/ncomms11693>
- Fu L, Li N, Mills R (1999) Pinaceae, flora of China. China, Science Press, Beijing
- Fumagalli M, Vieira FG, Linderöth T, Nielsen R (2014) ngsTools: methods for population genetics analyses from next-generation sequencing data. *Bioinformatics* 30:1486–1487. <https://doi.org/10.1093/bioinformatics/btu041>
- Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol* 11:725–736. <https://doi.org/10.1093/oxfordjournals.molbev.a040153>
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29:644–652. <https://doi.org/10.1038/nbt.1883>
- Gross BL, Rieseberg LH (2005) The ecological genetics of homoploid hybrid speciation. *J Heredity* 96:241–252. <https://doi.org/10.1093/jhered/esi026>
- Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet* 5:e1000695. <https://doi.org/10.1371/journal.pgen.1000695>
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J et al. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8:1494–1512. <https://doi.org/10.1038/nprot.2013.084>
- Hahn C, Bachmann L, Chevreaux B (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Res* 41:e129. <https://doi.org/10.1093/nar/gkt371>
- Hember RA, Kurz WA, Coops NC (2017) Relationships between the individual-tree mortality and water-balance variables indicate positive trends in water stress-induced tree mortality across North America. *Glob Change Biol* 23:1691–1710
- Hermansen JS, Saether SA, Elgvin TO, Borge T, Hjelte E, Saetre GP (2011) Hybrid speciation in sparrows I: phenotypic intermediacy, genetic admixture and barriers to gene flow. *Mol Ecol* 20:3812–3822. <https://doi.org/10.1111/j.1365-294X.2011.05183.x>
- Huang Y, Niu B, Gao Y, Fu L, Li W (2010) CD-HIT suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26:680. <https://doi.org/10.1093/bioinformatics/btq003>
- Huson DH (1998) Splitstree, analyzing and visualizing evolutionary data. *Bioinformatics* 14:68–73. <https://doi.org/10.1093/bioinformatics/14.1.68>
- Huson DH, Bryant D (2005) Estimating phylogenetic trees and networks using SplitsTree 4. Manuscript in preparation, software retrieved from www.splitsite.org
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>
- Lamichanay S, Han F, Webster MT, Andersson L, Grant BR, Grant PR (2018) Rapid hybrid speciation in Darwin's finches. *Science* 359:224–228. <https://doi.org/10.1126/science.aao4593>
- Li H, Durbin R (2019) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N et al. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li L, Stoeckert JC, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 13:2178–2189. <https://doi.org/10.1101/gr.1224503>
- Li M, Tian S, Yeung C, Meng X, Tang Q, Niu L et al. (2014) Whole-genome sequencing of Berkshire (European native pig) provides insights into its origin and domestication. *Sci Rep* 4:4678. <https://doi.org/10.1038/srep04678>
- Li Y, Stocks M, Hemmilla S, Kallman T, Zhu H, Zhou Y et al. (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. <https://doi.org/10.1093/molbev/msp301>
- Liu J, Molloy M, Provan J, Gao L, Poudel RC, Li D (2013) Geological and ecological factors drive cryptic speciation of yews in a biodiversity hotspot. *N Phytologist* 199:1093–1108. <https://doi.org/10.1111/nph.12336>
- Lockwood JD, Aleksic JM, Zou J, Wang J, Liu J, Renner SS (2013) A new phylogeny for the genus *Picea* from plastid, mitochondrial, and nuclear sequences. *Mol Phylogenetics Evol* 69:717–727. <https://doi.org/10.1016/j.ympev.2013.07.004>
- Lyu L, Wang D, Li L, Zhu Y, Jiang D, Liu J et al. (2020) Polyphyly and species delimitation of *Picea brachytyla* (Pinaceae) based on population genetic data. *J Syst Evol* 59:515–523. <https://doi.org/10.1111/jse.12584>
- Ma Y, Wang J, Hu Q, Li J, Sun Y, Zhang L et al. (2019) Ancient introgression drives adaptation to cooler and drier mountain habitats in a cypress species complex. *Commun Biol* 2:213. <https://doi.org/10.1038/s42003-019-0445-z>
- Mavárez J, Salazar CA, Bermingham E, Salcedo C, Jiggins CD, Linares M (2006) Speciation by hybridization in *Heliconius* butterflies. *Nature* 441:868–871. <https://doi.org/10.1038/nature04738>
- Meinzer FC, Brooks JR, Gartner BL, Warren JM, Wodruff DR, Bible K et al. (2006) Dynamics of water transport and storage in conifers studied with deuterium and heat tracing techniques. *Plant, Cell Environ* 29:105–114
- Mulch A, Chamberlain CP (2006) Earth science, the rise and growth of Tibet. *Nature* 439:670–671. <https://doi.org/10.1038/439670a>
- Nieto Feliner G, Álvarez I, Fuentes-Aguilar J, Heuertz M, Marques I, Moharrek F et al. (2017) Is homoploid hybrid speciation that rare? An empiricist's view. *Heredity* 118:513–516. <https://doi.org/10.1038/hdy.2017.7>
- Nolte AW, Tautz D (2010) Understanding the onset of hybrid speciation. *Trends Genet* 26:54–58. <https://doi.org/10.1016/j.tig.2009.12.001>
- Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG et al. (2013) The Norway spruce genome sequence and conifer genome evolution. *Nature* 497:579–584. <https://doi.org/10.1038/nature12211>
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38:904–909. <https://doi.org/10.1038/ng1847>
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575. <https://doi.org/10.1086/519795>
- Ran J, Shen T, Liu W, Wang P, Wang X (2015) Mitochondrial introgression and complex biogeographic history of the genus *Picea*. *Mol Phylogenet Evol* 93:63–76. <https://doi.org/10.1016/j.ympev.2015.07.020>
- Ran J, Wei X, Wang X (2006) Molecular phylogeny and biogeography of *Picea* (Pinaceae): implications for phylogeographical studies using cytoplasmic haplotypes. *Mol Phylogenet Evol* 41:405–419
- Rieseberg LH (1997) Hybrid origins of plant species. *Annu Rev Ecol Syst* 28:359–389. <https://doi.org/10.1146/annurev.ecolsys.28.1.359>
- Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T et al. (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301:1211–1216. <https://doi.org/10.1126/science.1086949>
- Ru D, Sun Y, Wang D, Chen Y, Wang T, Hu Q et al. (2018) Population genomic analysis reveals that homoploid hybrid speciation can be a lengthy process. *Mol Ecol* 27:4875–4887. <https://doi.org/10.1111/mec.14909>
- Ru D, Ma K, Zhang L, Wang X, Lu Z, Sun Y (2016) Genomic evidence for polyphyletic origins and interlineage gene flow within complex taxa: a case study of *Picea brachytyla* in the Qinghai-Tibet Plateau. *Mol Ecol* 25:2373–2386. <https://doi.org/10.1111/mec.13656>

- Schumer M, Cui R, Rosenthal GG, Andolfatto P (2015) Reproductive isolation of hybrid populations driven by genetic incompatibilities. *PLoS Genet* 11:e1005041. <https://doi.org/10.1371/journal.pgen.1005041>
- Schumer M, Rosenthal GG, Andolfatto P (2014) How common is homoploid hybrid speciation? *Evolution* 68:1553–1560. <https://doi.org/10.1111/evo.12399>
- Shao C, Shen T, Jin W, Mao H, Ran J, Wang X (2019) Phylotranscriptomics resolves interspecific relationships and indicates multiple historical out-of-North America dispersals through the Bering Land Bridge for the genus *Picea* (Pinaceae). *Mol Phylogenet Evol* 141:106610. <https://doi.org/10.1016/j.ympev.2019.106610>
- Shen T, Ran J, Wang X (2019) Phylogenomics disentangles the evolutionary history of spruces (*Picea*) in the Qinghai-Tibetan Plateau: implications for the design of population genetic studies and species delimitation of conifers. *Mol Phylogenet Evol* 141:106612. <https://doi.org/10.1016/j.ympev.2019.106612>
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatic* 31:3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Stamatakis A (2014) RAXML version 8, a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Sun Y, Abbott RJ, Li L, Li L, Zou J, Liu J (2014) Evolutionary history of purple cone spruce (*Picea purpurea*) in the Qinghai-Tibet Plateau: homoploid hybrid origin and Pleistocene expansion. *Mol Ecol* 23:343–359. <https://doi.org/10.1111/mec.12599>
- Sun Y, Abbott RJ, Lu Z, Mao K, Zhang L, Wang X et al. (2018) Reticulate evolution within a spruce (*Picea*) species complex revealed by population genomic analysis. *Evolution* 72:2669–2681. <https://doi.org/10.1111/evo.13624>
- Sun Y, Lu Z, Zhu X, Ma H (2020) Genomic basis of homoploid hybrid speciation within chestnut trees. *Nat Commun* 11:3375. <https://doi.org/10.1038/s41467-020-17111-w>
- Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* 34:W609–612. <https://doi.org/10.1093/nar/gkl315>
- Than C, Ruths D, Nakhleh L (2008) PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinforma* 9:322. <https://doi.org/10.1186/1471-2105-9-322>
- Todesco M, Owens GL, Bercovich N, L  gar   J-S, Soudi S, Burge DO et al. (2020) Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature* 584:602–607
- Wang G, Li H, Zhao H, Zhang W (2017) Detecting climatically driven phylogenetic and morphological divergence among spruce (*Picea*) species worldwide. *Bio-geosciences* 14:2307–2319. <https://doi.org/10.5194/bg-14-2307-2017>
- Wang Z, Jiang Y, Bi H, Lu Z, Ma Y, Yang X et al. (2021) Hybrid speciation via inheritance of alternate alleles of parental isolating genes. *Mol Plant* 14:208–222. <https://doi.org/10.1016/j.molp.2020.11.008>
- Weir BS, Cockerham CC (1984) Estimating f-statistics for the analysis of population-structure. *Evolution* 38:1358–1370. <https://doi.org/10.1111/j.1558-5646.1984.tb05657.x>
- Wright JW (1955) Species crossability in spruce in relation to distribution and taxonomy. *For Sci* 1:319–349. <https://doi.org/10.1093/forestscience/1.4.319>
- Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13:555–556. <https://doi.org/10.1093/bioinformatics/13.5.555>

- Yu Y, Dong J, Liu KJ, Nakhleh L (2014) Maximum likelihood inference of reticulate evolutionary histories. *Proc Natl Acad Sci USA* 111:16448–16453. <https://doi.org/10.1073/pnas.1407950111>
- Zou J, Yue W, Li L, Wang X, Lu J, Duan B et al. (2016) DNA barcoding of recently diversified tree species: a case study on spruces based on 20 DNA fragments from three different genomes. *Trees* 30:959–969. <https://doi.org/10.1007/s00468-015-1337-6>

ACKNOWLEDGEMENTS

We would like to thank the HighEdit company for assistance with English language editing of this manuscript. This work was supported by grants from the National Natural Science Foundation of China (grant numbers 32001085, 31971392, 31590821, 31670665, 91731301, 31960319), the National Key Research and Development Program (2017YFC0505203), the Fundamental Research Funds for the Central Universities (Grant No. lzujbky-2020-34, lzujbky-2020-ct02), the Strategic Priority Research Program of Chinese Academy of Sciences (Grant No. XDB01000000), and the Project of Qinghai Science & Technology Department (2020-ZJ-944Q).

AUTHOR CONTRIBUTIONS

J.L. and D.R. planned and designed the research. D.W., Y.S., W.L. and D.R. conducted fieldwork, performed experiments and analysed data etc. D. R., J.L., Y.S. and D.W. wrote the manuscript; and all authors revised and approved the final manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41437-023-00630-9>.

Correspondence and requests for materials should be addressed to Jianquan Liu or Dafu Ru.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.