Divergence amid recurring gene flow: complex demographic processes during speiation are the growing expectation for forest trees

Constance E. Bolte[1,4], Trevor M. Faske[2], Christopher J. Friedline[3], and Andrew J. Eckert[3]

[1]Integrative Life Sciences Doctoral Program, Virginia Commonwealth University, Richmond, VA, USA
[2]Ecology, Evolution & Conservation Biology Doctoral Program, University of Nevada, Reno, NV, USA
[3]Department of Biology, Virginia Commonwealth University, Richmond, VA, USA
[4]Corresponding author; boltece@vcu.edu

Abstract

Long-lived species of trees, especially conifers, often display weak patterns of reproductive isolation, but clear patterns of local adaptation and phenotypic divergence. Discovering the evolutionary history of these patterns is paramount to a generalized understanding of speciation for long-lived plants. We focus on two closely related yet phenotypically divergent pine species, *Pinus pungens* and *P. rigida*, that co-exist along high elevation ridgelines of the southern Appalachian Mountains. Based on genome-wide RADseq data, patterns of population structure for each species were uncorrelated to geography and the environment. Signals of admixture, however, were present range-wide. When combined with information from contemporary and historical species distribution models, these patterns are consistent with a complex evolutionary history of speciation. This was confirmed using inferences based on the multidimensional site-frequency spectrum, where demographic modeling inferred recurring gene flow since divergence (9.3 – 15.4 million years ago) and population size reductions that align in timing with the last interglacial period (~120 – 140 thousand years ago). This suggests that phenotypic and genomic divergence, including the evolution of divergent phenological schedules leading to partial reproductive isolation, as previously

documented for these two species, can happen rapidly, even between long-lived species of pines.

Keywords: conifer speciation, *Pinus pungens, Pinus rigida*, reproductive isolation, population genetics, species distributions

Statements and Declarations

The authors have no financial or proprietary interests in any material discussed in this article. Field sampling permits at National Parks and State Parks were obtained prior to collecting leaf tissue: ACAD-2017-SCI-007, BLRI-2013-SCI-0027, GRSM-2017-SCI-2028, ZZ-RCP-112514, and SFRA-1725.

Acknowledgements

This research was funded by Virginia Commonwealth University (VCU) Department of Biology, VCU Integrative Life Sciences, and National Science Foundation (NSF) awards to Andrew J. Eckert (NSF-EF-1442486) and Christopher J. Friedline (NSF-NPGI-PRFB-1306622). We thank Mitra Menon and Rebecca Piri for their assistance with field sampling, Brandon Lind for providing computational support, and the undergraduate researchers who helped with the DNA extraction protocol: Kaylyn Carver, Casey Harless, and Leslie Ranson. Demographic inference requires large computational resources, and we thank the VCU Center for High Performance Computing for proving those.

Introduction

The process of speciation has been characterized as a continuum of divergence underpinned with the expectation that reproductive isolation strengthens over time leading to increased genomic conflict between species (Seehausen et al. 2014). While the term continuum suggests linear directionality, it is better thought of as a multivariate trajectory that is nontemporal, allowing stalls and even breakdown of reproductive barriers in the overall progression toward complete reproductive isolation (Cannon and Petit 2020; Kulmuni et al. 2020). Indeed, speciation can occur with or without ongoing gene flow and demographic processes such as expansions, contractions, isolation, and introgression leave detectable genetic patterns within and among populations of species that affect the evolution of reproductive isolation (Nosil 2012; e.g., Gao et al. 2012). Divergence histories with gene flow are an emerging pattern for species of forest trees with reproductive isolation often developing through prezygotic isolating mechanisms and reinforced by environmental adaptation (Abbott 2017; Cavender-Bares 2019). Together, these two processes can facilitate the development of genomic incompatibilities over time (Baack et al. 2015).

Climate and geography are well-established drivers of demographic processes and patterns (Hewitt 2001). For the past 2.6 million years, Quaternary climate has oscillated between glacial and interglacial periods causing changes in species distributions, but the significance of these changes and their influence on population differentiation has varied by region and taxa (Hewitt 2004; Lascoux et al. 2004). In North America, the effects of

Quaternary climate on tree species distributions and patterns of genetic diversity have been profound but more drastic for species native to northern (i.e., previously glaciated) and eastern regions. For instance, the geographical distribution of white oak (*Quercus alba* L.), a native tree species to eastern North America, experienced greater shifts since the last interglacial period, approximately 120 thousand years ago (kya), compared to the distributional shifts of valley oak (*Quercus lobata* Née) in California (Gugger et al. 2013). For the latter, distributional, and hence niche, stability was correlated with higher levels of genetic diversity.

Given the climate instability of eastern North America since the last interglacial period (LIG; ~120 kya), a host of phylogeographic studies have reported genetic diversity estimates for taxa of this region and the genetic structuring of populations due to geographic barriers such as the Appalachian Mountains and Mississippi River (Soltis et al. 2006) as well as postglacial expansion (e.g., Gougherty et al. 2020). The vast majority of tree taxa in these studies, however, were angiosperms, with the divergence history of only one closely related pair of conifer species native to this region, *Picea mariana* (Mill.) Britton, Sterns, & Poggenb. and *P. rubens* Sarg., being fully characterized (Perron et al. 2000; Lafontaine et al. 2015). The relative differences in geographical distributions and genetic diversities across *P. mariana* and *P. rubens,* as well as models of demographic inference, suggest a progenitor-derivative species relationship that initiated approximately 110 kya through population contractions and geographical isolation. Despite this history, these two species actively hybridize today. In general, speciation among conifer lineages remains an enigmatic process (Bolte and Eckert 2020), largely

because there is a mismatch between species-level taxonomy and the existence of reproductive isolation, so that hybridization among species is common both naturally as well as artificially (Critchfield 1986). The ability to hybridize, moreover, is idiosyncratic, with examples ranging from well-developed incompatibilities among populations within species (e.g., *P. muricata* D. Don; Critchfield 1967) to the almost complete lack of incompatibilities among diverged and geographically distant species (*P. wallichiana* A. B. Jacks. from central Asia and *P. monticola* Douglas ex D. Don from western North America; Wright 1959). Thus, the tempo and mode for the evolution of reproductive isolation for conifers remains largely unexplained despite decades of research into patterns of natural hybridization, crossing rates, and the mechanisms behind documented incompatibilities (McWilliam 1959; Kriebel 1972; Hagman 1975; Critchfield 1986; Vasilyeva and Goroshkevich 2018).

The key to understanding the evolution of reproductive isolation, and hence a more developed explanation of the process of speciation for conifers, is the role of demography and gene flow during the divergence among lineages. Analytical approaches have been developed to infer past demographic processes from population genomic data, which can now easily be generated even for conifers (Parchman et al. 2018). While many studies have used demographic inference methods to describe the phylogeographic history of a single species (e.g., Gugger et al. 2013; Li et al. 2013; Bagley et al. 2020; Ju et al. 2019; Park and Donoghue 2019; Capblancq et al. 2020; Yang et al. 2020; Labiszak et al. 2021), some of these established methods have also be used to infer divergence histories between two or three species (e.g., Zou et al. 2013; Christe et al. 2017; Kim et al. 2018;

Menon et al. 2018). Single species inferences have found that the last glacial maximum (LGM; ~22,000 years ago) affected distributional shifts and intraspecific gene flow dynamics, while multispecies studies have focused almost solely on how these climatic oscillations drove periods of increased and decreased interspecific gene flow which contributed to the formation of environmentally dependent hybrid zones, ancient periodical introgression, or adaptive divergence in the development of reproductive isolation.

The number of potential divergence histories underlying even a modest number of species is vast. The preemptive formation of a hypothesis from historical species distribution modeling (SDM), however, can aid in defining a more realistic set of models from which to make inference, as well as to examine the impact of the climate change on genetic diversity and demographic processes (Carstens and Richards, 2007). For example, Lima et al. (2017) modeled distributional changes for *Eugenia dysenterica* DC. between the LGM and today which led to a hypothesis that range stability was more likely than range expansion or contraction in this South American region. Their SDM informed hypothesis was supported by range-wide, *E. dysenterica* genetic data. Likewise, SDMs across several time points allows for estimation of habitat suitability change (i.e., a proxy for contraction or expansion) and distributional overlap of multiple species (i.e., potential gene flow). With these quantified changes, testable hypotheses emerge, lending to more deliberate investigations of speciation through justified parameter selection (Richards et al. 2007). Of course, there are inherent limitations associated with SDMs and interpreting historical distributions should be done cautiously but using SDMs to complement

demographic inference is now common in the field of phylogeography (Hickerson et al. 2010; Gavin et al. 2014; Peterson and Anamza 2015). Indeed, where a species occurs is determined by its traits and thus genetics. Ikeda et al. (2017) found that SDM predictions under future climate scenarios improved with acknowledgement of local adaptation in *Populus fremontii* S. Watson (i.e., three identified genetic clusters across the full species distributional range were modeled independently).

Here, we focus on two closely related, yet phenotypically diverged, pine species, Table Mountain pine (*Pinus pungens* Lamb.) and pitch pine (*Pinus rigida* Mill.). Recent estimates from multiple time-calibrated phylogenies have placed the time of divergence in the range of 1.5 to 17.4 million years ago (mya; Hernandez-Leon et al. 2013; Saladin et al. 2017; Gernandt et al. 2018; Jin et al. 2021), with these studies either placing them as sister species (e.g., Hernandez-Leon et al. 2013; Saladin et al. 2017) or as part of a clade with *P. serotina* Michx. as the sister to *P. rigida* (e.g., Gernandt et al. 2018; Jin et al. 2021). Changes in climate, fire regime, and geographic distributions have likely influenced species divergence (Keeley 2012). This is plausible given that *P. pungens* populations are restricted to high elevations of the Appalachian Mountains, while the much larger distribution of *P. rigida* ranges from Georgia into portions of eastern Canada. It is particularly interesting that these recently diverged species are found in sympatry, yet hybridization has rarely been observed in the field (Zobel 1969), although they can be reciprocally crossed to yield viable offspring (Critchfield 1963). An ecological study of three sympatric *P. pungens* and *P. rigida* populations indicated that the timing of pollen release was separated by approximately four weeks, enough to sustain partial

reproductive isolation at these sites (Zobel 1969), which is a common contributor to pre-zygotic isolation among conifer species (Dorman and Barber 1956; Critchfield 1963). It was also noted that while *P. pungens* was most densely populated on arid, rocky, steep southwestern slopes, *P. rigida* was less confined to these areas (Zobel 1969), thus suggesting environmental adaptation may also be important in the divergence of these two closely related species.

We hypothesized that *P. pungens* and *P. rigida* experienced speciation with gene flow followed by population contraction and isolation (i.e., different refugia) initiated during the LGM. This period of isolation could potentially explain the striking divergence in trait values and phenological schedules between these two species. This hypothesis was assessed and revised using distributional overlap in habitat suitability over the last 120,000 years and tested using a multidimensional, folded site frequency spectrum from 2168 genome-wide, unlinked single nucleotide polymorphisms (SNPs) across 300 trees. The SDM-informed hypothesis was supported. Divergence occurred amid ongoing symmetrical gene flow for the past 9.3 – 15.4 million years, depending on generation time, and both species experienced major contraction in effective population size during the LIG. The best supported model also included a reduction in gene flow since the LIG which we sought to explain using population genetic analyses and environmental associations. Prezygotic reproductive isolation due to differing phenological schedules appears to be the primary mechanism involved in the maintenance of species boundaries and an emerging pattern in conifer speciation.

Methods

Sampling

Range-wide samples of needle tissue were obtained from 14 populations of *Pinus pungens* and 19 populations of *Pinus rigida* (Fig. 1). Each population consisted of 4-12 trees with each sampled tree distanced by approximately 50 m from the next to avoid potential kinship (Table 1). Needle tissue was dried using silica beads, then approximately 10 mg of tissue was cut and lysed for DNA extraction.

DNA sequence data

Genomic DNA was extracted from all 300 sampled trees using DNeasy Plant Kits (Qiagen) following the manufacturer's protocol. Four ddRADseq libraries (Peterson et al. 2012), each containing up to 96 multiplexed samples, were prepared using the procedure from Parchman et al. (2012). EcoRI and MseI restriction enzymes were used to digest all four libraries before performing ligation of adaptors and barcodes. After PCR, agarose gel electrophoresis was used to separate then select DNA fragments between 300-500 bp in length. The pooled DNA was isolated using a QIAquick Gel Extraction Kit (Qiagen). Single-end sequencing was conducted on Illumina HiSeq 4000 platform by Novogene Corporation (Sacramento, CA). Raw fastq files were demultiplexed using GBSX (Herten et al. 2015) version 1.2, allowing two mismatches (-mb 2). The dDocent bioinformatics pipeline (Puritz et al. 2014) was used to generate a reference assembly and call variants.

The reference assembly optimized using shell scripts and documentation within dDocent (cutoffs: individual = 6, coverage = 6; clustering similarity: -c 0.92), utilizing cd-hit-est (Fu et al. 2012) for assembly. The initial variant calling produced 87,548 single nucleotide polymorphisms (SNPs) that was further filtered using *vcftools*, (Danecek et al. 2011), version 0.1.15. In final, we retained only biallelic SNPs with sequencing data for at least 50% of the samples, minor allele frequency (MAF) > 0.01, summed depth across samples > 100 and < 10000, and alternate allele call quality ε 50. To account for linkage disequilibrium, which if not properly acknowledged can lead to erroneous inferences of demographic history (Gutenkunst et al. 2009), we thinned the dataset to one SNP per contig (--thin 100). Additionally, stringent filtering steps to were taken to minimize the potential misassembly of paralogous genomic regions. Removing loci with excessive coverage and retaining only loci with two alleles present, as above, should ameliorate the influence of misassembled paralogous loci in our data (Hapke and Thiele 2016; McKinney et al. 2018). Lastly, we retained loci with F_{IS} > –0.5, as misassembly to paralogous genomic regions can lead to abnormal heterozygosity (Hohenlohe et al. 2013; McKinney et al. 2017). The reduced 2168 SNP dataset was used in all analyses.

Population structure and genetic diversity

Patterns of genetic diversity and structure within and between *P. pungens* and *P. rigida* were assessed using a suite of standard methods. Overall patterns of genetic structure were investigated using principal component analysis (PCA), as employed in the prcomp function of the *stats* version 4.0.4 package, and population graphs using the *popgraph* version 1.5.2 package (Dyer and Nason 2004) in R version 3.6.2 (R Development Core

Team, 2021). Genetic diversity within each species was examined using multilocus estimates of observed and expected heterozygosity (H_o and H_e) for each population using a custom R script. An individual-based assignment test was conducted using *fastSTRUCTURE* (Raj et al. 2014) with cluster assignments ranging from $K = 2$ to $K = 7$. Ten replicate runs of each cluster assignment were conducted. The cluster assignment with the highest log-likelihood value was determined to be the best fit. Individual admixture assignments were then aligned and averaged across the 10 runs using the *pophelper* version 1.2.0 (Francis 2017) package in R. Third, multilocus, hierarchical fixation indices (*F*-statistics) were defined by nesting trees into populations and populations into species, with F_{CT} describing differentiation between species and F_{SC} describing population differentiation within species (Yang 1998). *F*-statistics and associated confidence intervals (95% CIs) from bootstrap resampling ($n = 100$ replicates) were calculated in the *hierfstat* version 0.5-7 package (Goudet and Jombart 2020) in R.

To assess influences on within species genetic structure, Mantel tests (Mantel 1967) were used to examine Isolation-by-Distance (IBD; Wright 1943) and Isolation-by-Environment (IBE; Wang and Bradburd 2014). In these analyses, the Mantel correlation coefficient (r) was calculated between linearized, pairwise F_{ST}, estimated with the method of Weir and Cockerham (1984) using the *hierfstat* package in R, and either geographical (IBD) or environmental (IBE) distances. For geographical distances, latitude, and longitude records for each tree in a population were averaged to obtain one representative coordinate per population. Geographic distances among populations were then calculated using the Vincenty (ellipsoid) method within the *geosphere* version 1.5-10

package (Hijmans 2019) in R. Environmental distances were calculated as Euclidean distances using extracted raster values associated with the mean population coordinates from 19 bioclimatic variables, downloaded from WorldClim at 30 arc second resolution (version 2.1; Fick and Hijmans 2017). Values associated with the mean population coordinates for were extracted using the *raster* version 2.5-7 R package. Environmental data were centered and scaled prior to estimation of distances. Additionally, we used a Mantel test to assess correlation between population-based environmental distances and population-based geographic distances.

Associations between genetic structure and environment

To test the multivariate relationships among genotype, climate, and geography, redundancy analysis (RDA) was conducted using the *vegan* version 2.5-7 package (Oksanen et al. 2020) in R version 4.0.4 (R Core Development Team, 2021). Genotype data were coded as counts of the minor allele for each sample (i.e., 0,1, or 2 copies) and then standardized following Patterson et al. (2006). Climate raster data (i.e., 19 bioclimatic variables at 30 arc second resolutions), as well as elevational raster data from WorldClim, were extracted, as mentioned above, from geographic coordinates for each sampled tree and then tested for correlation using Pearson's correlation coefficient (r). Five bioclimatic variables that were not highly correlated ($r < |0.75|$) and known to influence diversification in the genus *Pinus* were retained for analysis: Bio 2 (mean diurnal range), Bio 10 (maximum temperature of the warmest quarter), and Bio 11 (minimum temperature of the coldest quarter), Bio 15 (precipitation seasonality), and Bio 17 (precipitation of the driest quarter). The full explanatory data set included these five

bioclimatic variables, latitude, longitude, and elevation. The multivariate relationship between genetic variation, climate, and geography was then evaluated through RDA. Statistical significance of the RDA model ($\alpha = 0.05$), as well as each axis within the model, was assessed using a permutation-based analysis of variance (ANOVA) procedure with 999 permutations (Legendre and Legendre 2012). The influence of predictor variables, as well as their confounded effects, in RDA were quantified using variance partitioning as employed in the *varpart* function of *vegan* package in R.

Species distribution modeling

To help formulate a testable hypothesis in the inference of demography from genomic data, species distribution modeling (SDM) was performed for each species to identify areas of suitable habitat under current climate conditions and across three historical time periods (see Richards et al. 2007). These temporal inferences were then used to help identify plausible demographic responses. For example, if overlap in modeled habitat suitability changed over time, the hypothesis for demographic inference would include changes in gene flow parameters over time. If the amount of suitable habitat changed over time, the hypothesis would also include changes in effective population size to allow for potential expansions or contractions. This in effect helps to constrain the possible parameter space for exploration.

Occurrence records for *P. pungens* were downloaded from GBIF (www.gbif.org) and combined with known occurrences published by Jetton et al. (2015). For *P. rigida*, all

occurrence records were downloaded from GBIF. Records were examined for presence within or close to the known geographical range of each species (Little 1971). Records far outside the known geographic range were pruned. The remaining locations were then thinned to one occurrence per 10 km to reduce the effects of sampling bias using the *spThin* version 0.1.0.1 package (Aiello-Lammens et al. 2015) in R. The resulting occurrence dataset included 84 records for *P. pungens* and 252 records for *P. rigida* (Online Resource 2). All subsequent analyses were performed in R version 3.6.2 (R Development Core Team, 2021).

The same bioclimatic variables (Bio2, Bio10, Bio11, Bio15, Bio17) selected for RDA were used in species distribution modeling but were downloaded from WorldClim version 1.4 (Hijmans et al. 2005) at 2.5 arc minute resolution. The change in resolution from above was because paleo-climate data in 30 arc second resolution were not available for the LGM. Paleoclimate raster data for the LGM (~21,000 years ago) and Holocene (HOL, ~6000 years ago) were based on three General Circulation Models (GCMs; CCSM4, MIROC-ESM, and MPI-ESM). Ensembles were built by averaging the grid cell values across the three GCMs for each time period, which were then used to predict species distributions and habitat suitability in the past. For comparative purposes, SDMs were also produced from each individual GCM for each time period since the LGM. Paleoclimate data for the LIG (~120,000 years ago; Otto-Bliesner et al. 2006) were only available at 30 arc second resolution and required downscaling to 2.5 arc minute resolution using the aggregate function (fact=5) of the *raster* package. Only one GCM is available for the LIG; therefore, no ensemble was built.

Raster layers were cropped to the same extent using the *raster* package to include the most northern and eastern extent of *P. rigida*, and the most western and southern extent of *P. pungens.* Species distribution models (SDMs) were built using MAXENT version 3.4.1 (Phillips et al. 2017) and all possible features and parameter combinations were evaluated using the *ENMeval* version 2.0.0 R package (Kass et al. 2021). Metadata about model fitting and evaluation are available within Online Resource 2.

The selected features used in predictive modeling were those associated with the best-fit model as determined using AIC. Raw raster predictions were standardized to have the sum of all grid cells equal the value of one using the *raster.standardize* function in the *ENMTools* version 1.0.5 (Warren et al. 2021) R package. Standardized predictions were then transformed to a cumulative raster prediction with habitat suitability scaled from 0 to 1, allowing for quantitative SDM comparisons across species and time. Next, SDM cumulative raster predictions were converted into coordinate points using the *sf* version 0.9-7 R package to calculate the number of points with habitat suitability values greater than 0.5 (i.e., moderate to high suitability areas). Population size expansion or contraction was hypothesized if the number of points increased or decreased over time, respectively. Overlap (i.e., shared points across species) in SDM predictions for each time period was measured using the *inner_join* function in the *dplyr* version 1.0.5 R package. The extent of modeled species distributional overlap was also quantified using the *raster.overlap* function in *ENMTools,* thus providing measures for Schoener's *D* (1968) and Warren's *I* (Warren et al. 2008).

353

354 ***Demographic modeling***

355 Demographic modeling was conducted using Diffusion Approximation for Demographic
356 Inference ($\partial\alpha\partial i$ v.2.0.5; Gutenkunst et al. 2009). A model of pure divergence (SI; strict
357 isolation) was compared against eleven other demographic models representing different
358 potential divergence scenarios with or without gene flow and effective population size
359 changes (Online Resource 4, Fig. S4). Based on SDM predictions across four time points,
360 we hypothesized that a model that allowed changes in effective population size and rate
361 of gene flow before the LIG would best fit the genetic data. Ten replicate runs of each
362 model were performed in $\partial\alpha\partial i$ with a 200 x 220 x 240 grid space and the nonlinear
363 Broyden-Fletcher-Goldfarb-Shannon (BFGS) optimization routine. Model selection was
364 conducted using Akaike information criterion (AIC; Akaike 1974). The best replicate run
365 (highest log composite likelihood) for each model was then used to calculate ΔAIC
366 ($AIC_{model\ i} - AIC_{best\ model}$) scores (Burnham and Anderson 2002). From the best supported
367 model, upper and lower 95% confidence intervals (CIs) for all parameters were obtained
368 using the Fisher Information Matrix (FIM)-based uncertainty analysis. Unscaled
369 parameter estimates and their 95% CIs were obtained using a per lineage substitution
370 rate of 7.28×10^{10} substitutions/site/year rate for *Pinaceae* (De La Torre et al. 2017) and
371 two possible values for generation time, 15 and 25 years, for comparative purposes.

372

373

374

375

Results

Population structure and genetic diversity

Principal component analysis (PCA) showed clear separation at the species level along PC1 which explained 4.232% of the variation across the 2168 SNP x 300 tree data set (Fig. 2a). Of the 2168 SNPs analyzed, 380 and 196 SNPs were fixed in *P. pungens* and *P. rigida*, respectively. Lack of population clustering within each species was observed when the PCA was labeled by population (Online Resource 4, Fig. S1). Using hierarchical *F*-statistics, the estimate of differentiation between species (F_{CT}) was 0.117 (95% CI: 0.099 – 0.136) and similarly to that among all sampled populations (F_{ST} = 0.123, 95% CI: 0.106 – 0.143), thus highlighting structure is largely due to differences between species. Differentiation among populations within species was consequently much lower (F_{SC} = 0.007 (95% CI: 0.0055-0.0088) whether analyzed jointly (F_{SC}) or separately (see Table 2). In the analysis of structure, K = 2 had the highest log-likelihood values (Fig. 2b). Admixture in small proportions (assigning to the other species by 2-10%) was observed in 41 out of the 300 samples (13.67% of samples) across both species. There were 16 trees with ancestry coefficients higher than 10% assignment to the other species: four *P. rigida* samples (2.29% of sampled *P. rigida*) and twelve *P. pungens* samples (9.60% of sampled *P. pungens*). Admixture proportions were moderately correlated to latitude (Pearson's r = -0.414), longitude (Pearson's r = -0.291), and elevation (Pearson's r = 0.445). All three correlative relationships were significant ($p < 0.001$).

Pairwise F_{ST} estimates for *P. pungens* ranged from 0 to 0.0457, while a similar but narrower range of values (0 – 0.0257) was noted for *P. rigida*. The highest pairwise F_{ST} value across both species was between two *P. pungens* populations located in Virginia, PU_DT and PU_BB (Table 1). Interestingly, PU_DT in general had higher pairwise F_{ST} values (0.0146 – 0.0457) compared to all the other sampled *P. pungens* populations (Online Resource 1). For *P. rigida*, the RI_SH population located in Ohio had higher pairwise F_{ST} values for 16 out of the 18 comparisons (0.0123 – 0.0257). The two populations that had low pairwise F_{ST} values with RI_SH were geographically nearby: RI_OH located in Ohio (pairwise F_{ST} = 0, distance: 90.1 km) and RI_KY located in Kentucky (pairwise F_{ST} = 0.0089, distance: 107.7 km). The highest pairwise F_{ST} value among *P. rigida* populations was between RI_SH and RI_HH, which are geographically distant from one another. From the Mantel tests for IBD and IBE, Pearson correlations were low (Table 2). The correlation with geographical distances was highest for *P. rigida* (Mantel r = 0.176, p = 0.055). From the Mantel test, Pearson correlation between geographic distance and environmental distance was high for both *P. rigida* (r = 0.611, p = 0.001) and *P. pungens* (r = 0.893, p = 0.001). Observed heterozygosity of *P. pungens* (H_o = 0.127 ± 0.015 SD), averaged across SNPs and populations, was higher than the average expected heterozygosity (H_e = 0.118 ± 0.008 SD), both of which were higher than the almost equal values for *P. rigida* (H_o = 0.102 ± 0.009 SD; H_e = 0.104 ± 0.005 SD; Table 2). Heterozygosity estimates for each population are listed in Table 1. Across both species, observed heterozygosity was mildly associated with geography and elevation. For *P. rigida*, the highest correlation was with elevation (r = 0.253), followed by correlation with longitude (r = 0.113). Observed heterozygosity in *P. pungens* had a negative

correlative relationship with elevation (r = -0.168) and positive correlative relationship with longitude (r = 0.175). Correlations between latitude and heterozygosity were low in both species (r = 0.008 for *P. rigida*; r = 0.08 for *P. pungens*).

Population graphs provide a visualization of the relative genetic connectivity across and among populations, as well as the relative genetic diversity within each population. Edges in population graphs represent unique associations in allele frequencies between populations. Collinearity across measures of genetic connectivity between populations, or across species labels, leads to edges being dropped from the population graph. Consistent with the PCA, fastSTRUCTURE, and *F*-statistics results, two disjunct groups of populations were apparent in the population graph, with each group largely corresponding to species labels (Fig. 3). The exception was the PU_DT population from *P. pungens,* which is located in Virginia, that showed genetic connectivity with two geographically distant populations from *P. rigida*, located in Ohio (RI_SH) and Maine (RI_ME).

Associations between genetic structure and environment

The combined effects of climate and geography explained 4.16% (r^2), or rather 1.52% (adj. r^2), of the genetic variance across 2168 SNPs and 300 sampled trees. The first RDA axis accounted for the bulk of the explanatory variance (42.3%, Fig. 4) and was the only RDA axis with a p-value ($p < 0.001$) less than commonly accepted thresholds of significance. Average elevation associated with *P. pungens* samples was 724.68 m (±

224.17 SD), while average elevation across *P. rigida* samples was lower (399.69 m, ± 292.26 SD). The average for Bio15 (precipitation seasonality) was 11.33 (± 1.83 SD) for *P. pungens*, and higher for *P. rigida* (14.23 ± 3.97 SD). Precipitation seasonality is the coefficient of variation calculated from monthly mean precipitation values across the span of one year. Considering the standard deviations around the mean, overlap in values for elevation and precipitation seasonality provide some context to present day overlap in species distributions along the southern Appalachian Mountains. Comparisons of predictor loadings across both RDA axes show latitude, longitude, and Bio11 (mean temperature of the coldest quarter) as important to explaining the variance both within (RDA 2, 9.77%) and across species (RDA1).

Partitioning the effects of each predictor set revealed that climate independently (i.e., conditioned on geography) accounted for 31.93% of the explained variance. Geography independently (i.e., conditioned on climate) accounted for 34.10% of the explained variance. The confounded effect, due to the correlations inherent to the chosen geographic and climatic predictor variables, was 33.97%.

Species distribution modeling

Past geographical distributions during the LIG, LGM, and HOL were predicted using MAXENT to form testable hypotheses within the demographic inference framework of $\partial\alpha\partial i$, v.2.0.5. Because population structure within each of the focal species was not observed from our genetic data (i.e., no genetic clusters were identified), we produced SDMs using occurrence records across the full distributional range of each species. The

best fit SDM for *P. pungens* used a linear and quadratic feature class with a 1.0 regularization multiplier, while the SDM for *P. rigida* used a linear, quadratic, and hinge feature class with a regularization multiplier of 3.0. Best-fit models were those with the lowest AIC when evaluated against other models with varying feature class and regularization multiplier settings. The AUC associated with the training data of the *P. pungens* and *P. rigida* SDMs was 0.929 and 0.912, respectively. Metadata, data inputs, outputs, and statistical results for model evaluation are available in Online Resource 2. The climatic variables with the highest permutation importance were Bio11 (mean temperature of the coldest quarter) and Bio15 (precipitation seasonality) which contributed 41.1% and 39.7% to the *P. pungens* SDM and 19.5% and 62.4% to the *P. rigida* SDM. Of the five climate variables included in the RDA, Bio15 and Bio11 had the highest loadings along RDA axis 1, helping to explain differences across species. The tandem reporting of Bio15 and Bio11 importance to both genetic differentiation and species distributions could be indicative that these climatic variables were drivers in the divergence of these two species.

Current SDMs indicate a larger area of suitable habitat for *P. rigida* (11,128 grid cells had > 0.5 habitat suitability) compared to *P. pungens* (6,632 grid cells) with the highest overlap (14.07% shared cells) across all four time points (Fig. 5). According to the SDM predictions, the areas of moderate to high suitability shifted, contracted, and expanded over time for both species, with overlapping areas of suitable habitat exhibiting some of these fluctuations as well. SDM predictions for HOL indicate the lowest overlap (8.25% of grid cells with > 0.5 habitat suitability), while LGM predictions indicate the highest

overlap (16.33%). Contrastingly, calculations of overlap from full distributional predictions were the lowest (Schoener's *D* = 0.170) for LGM followed by the LIG (Schoener's *D* = 0.288). The highest full distributional overlap was associated with the current SDM (Schoener's *D* = 0.612).

The ensemble prediction for *P. pungens* during the LGM shows multiple potential refugial areas. One of these areas may have been glaciated (Fig. 5). Likewise, multiple refugia were predicted for *P. rigida* during the LGM, but these areas were below the mapped glacial extent (~18 kya; Dyke 2003). Interspecific gene flow during the LGM may thus have been possible just south of the glacial extent as well as in a disjunct refugia farther south than where either species currently occurs. Modeled distributions of *P. pungens* and *P. rigida* during the HOL distributions were proximal to each other, with high habitat suitability west of and along the Appalachian Mountains. These distributions may have promoted both intraspecific and interspecific gene flow to occur ~6 kya. The above quantifications are based on the SDM predictions from ensembled GCMs for the HOL and LGM time periods. SDM predictions associated with each independent GCM (CCSM4, MIROC-ESM, and MPI-ESM) are available in Online Resource 4 (Fig. S2, Fig. S3).

Demographic modeling

Results from the SDMs and fastSTRUCTURE analysis informed the hypothesis that divergence occurred with ongoing gene flow given the overlap in suitable habitat across the four time points. Gene flow was hypothesized to be symmetrical given the rather even

distribution in admixture *Q*-scores across both species (Fig. 2). We further hypothesized effective population sizes to have been stable over the past 120,000 years as potentially impactful expansion or contraction in habitat suitability was not observed. The best replicate run (highest composite log-likelihood) for each of the twelve modeled divergence scenarios, their associated parameter outputs, and ΔAIC ($AIC_{model\ i} - AIC_{best\ model}$) are summarized in Online Resource 3. A model that allowed changes in both effective population size and rate of symmetrical gene flow across two time periods (PSCMIGCs) best fit the 2168 SNP data set (Table 2) and had small, normally distributed residuals (Fig. S5). This model was 20.84 AIC units better than the second best-fit model (PSCMIGs; Table 3), which had allowed change in population size estimates across two time intervals but only one symmetrical gene flow parameter was inferred, representing a constant rate across both time periods.

Assuming a generation time of 15 years, the unscaled parameters from the best-fit model estimated an initial divergence time at 9.27 mya (95% CI: 7.60 – 10.94). With a 25-year generation time, initial divergence was estimated to be 15.44 mya (95% CI: 12.64 – 18.24). The first time interval during divergence (T_1) lasted 98.7% of the total divergence time with symmetrical gene flow (M_i) occurring at a rate of 48.6 (95% CI: 33.1 – 64.1) migrants per generation (Fig. 6). Parameters associated with effective population size were independent of generation time. The effective size of the ancestral population (*Nref*) was 203,431 (95% CI: 176,575 – 230,287; Fig. 6) prior to divergence. For most of the divergence history, *P. pungens* had an effective population size of N_{P1} = 5,767,720 (95% CI: 791,500 – 10,743,940) while *P. rigida* had a relatively smaller, but still large, effective

size of N_{R1} = 4,272,253 (95% CI: 1,207,070 – 7,337,436). The second time interval (T_2) during divergence was estimated to have begun between 118,897 (15 year generation time; 95% CI: 111,400 – 126,394) and 198,115 (25 year generation time; 95% CI: 185,622 – 210,608) years ago when effective population sizes decreased instantaneously to 19,408 (95% CI: 18,160 – 20,656) for *P. pungens* (N_{P2}) and 22,151 (95% CI: 20,710 – 23,592) for *P. rigida* (N_{R2}). During this time interval the relative rate of symmetrical gene flow dropped from 48.6 to 38.4 (95% CI: 35.7 – 41.1) migrants per generation.

Discussion

To understand the divergence history and development of reproductive isolation between *P. pungens* and *P. rigida*, we combined inferences of population structure, admixture, environmental associations, and historical species distribution modeling to form a testable hypothesis within a demographic inference framework. Our hypothesis was supported. Since the LIG, effective population sizes for both species did not change, and interspecific gene flow continued to occur. The best-fit demographic model using 2168 SNPs as summarized using the multidimensional site frequency spectrum also indicated that at the onset of the LIG there was a large reduction in effective population size which coincided with a reduction in gene flow. Even though gene flow occurred throughout the history of these two species, distinctive life histories and phenotypes have developed. The morphology and distributional differences across *P. pungens* and *P. rigida* suggest

558 divergence was driven by fire and climate with phenological differences contributing to
559 the maintenance of species boundaries through partial, prezygotic reproductive isolation.
560
561
562 ***Climate drives divergence***
563 Adaptation to seasonality among temperate species was influenced by Quaternary
564 climate (Dobzhansky 1950; Savolainen et al. 2004; Jump and Penuelas 2005; Williams
565 and Jackson 2007; Bonebrake and Mastrandea 2010). Phenological traits have been
566 linked to seasonal variation within species of North America (Jump and Penuelas 2005),
567 and differences in seasonality requirements for *P. pungens* and *P. rigida* likely explain the
568 observed phenological trait differences in seed size, reproductive age, timing of pollen
569 release, and rates of seedling establishment across these two species (Zobel 1969; Della-
570 Bianca 1990; Ledig et al. 2015). Using niche and trait data, the phylogenetic inference of
571 Jin et al. (2021) identified precipitation seasonality (Bio15) and elevation as drivers of
572 diversification in eastern North American pines. The results of our RDA corroborate these
573 conclusions, as clear separation of species-level genetic differences was explained by
574 Bio15 and elevation along axis 1 (Fig. 4). Quaternary climate began 2.6 mya and in terms
575 of the total divergence time inferred for these two species (9.3 – 15.4 mya) adaptations
576 to seasonality occurred rather recently (~16 – 25% of total divergence time). Based on
577 our data, *P. pungens* has lower and narrower niche requirements in terms of Bio15, which
578 helps explain its patchy distribution along the southern Appalachian Mountains. In
579 contrast, populations of *P. rigida* may have evolved a response to increased precipitation

seasonality during the Quaternary period, explaining its larger distribution and latitudinal expansion into northeastern North America.

Divergence between *P. pungens* and *P. rigida* began during the Miocene between 9.3 and 15.4 mya, which is consistent with divergence times previously estimated for these two pine species (Gernandt et al. 2018; Jin et al. 2021). The evolution of fire-related traits in pines has been linked to the mid-Miocene period (Jin et al. 2021) suggesting fire as an initiator of differentiation across ancestral populations. Distributions of our focal species are locally divergent across slope aspects in the Appalachian Mountains, with *P. pungens* primarily distributed on southwestern slopes and *P. rigida* primarily distributed on southeastern slopes (Zobel 1969). Due to higher fire frequency and intensity on western slopes, *P. pungens* has evolved strategies that confer population persistence such as high cone serotiny and fast seedling development. Although some northern *P. rigida* populations exhibit serotiny, the populations found along the southern Appalachian Mountains, and proximal to *P. pungens,* have nonserotinous cones and other traits consistent with enduring fire (e.g., epicormics; Zobel 1969) as opposed to relying on it (Jin et al. 2021). With these factors in mind and the correlative evidence between fire intensity and level of serotiny presented across populations of other pine species (*P. halepensis* and *P. pinaster*; Hernandez-Serrano et al. 2013), we suspect genomic regions involved in the complex, polygenic trait of serotiny (Parchman et al. 2012; Budde et al. 2014) may have contributed to the rapid development of prezygotic reproductive isolation between our focal species.

603 ***Reproductive isolation can evolve rapidly during conifer speciation***

604 While *P. pungens* and *P. rigida* can be found on the same mountain and even established
605 within a few meters of each other, mountains are heterogeneous, complex landscapes
606 offering opportunity for niche evolution along multiple axes of biotic and abiotic influence
607 for parental species and hybrids alike. The distances to disperse into novel environments
608 are relatively short in these heterogenous landscapes thus suggesting diversification
609 could be more rapid as environmental complexity increases (Bolte and Eckert 2020).
610 Mountains have rain shadow regions characterized by drought and thus more active fire
611 regimes (Parisien and Moritz 2009). A host of adaptive traits in trees are associated with
612 fire frequency and intensity (Pausas and Schwilk 2012). Among those, the genetic basis
613 of serotiny is characterized as being polygenic with large effect loci in *P. contorta* Dougl.
614 (Parchman et al. 2012) and in *P. pinaster* Aiton (Budde et al. 2014). Such genetic
615 architectures, even in complex demographic histories such as the one described here,
616 can evolve relatively rapidly to produce adaptive responses to shifting optima (e.g., Stetter
617 et al. 2018; reviewed for forest trees by Lind et al. 2018), so that it is not unreasonable to
618 expect divergence in fitness-related traits such as serotiny to also contribute to niche
619 divergence and reproductive isolation. Considering large effect loci associated with
620 serotiny were also associated with either water stress response, winter temperature, cell
621 differentiation, or root, shoot, and flower development (Budde et al. 2014), serotiny may
622 be a trait that contributes to widely distributed genomic islands of divergence thus
623 explaining the development of ecologically based reproductive isolation between *P.*
624 *pungens* and *P. rigida* amid recurring gene flow (Nosil and Feder 2012). Given that our
625 focal species are reciprocally crossable to yield viable offspring (Critchfield 1963), it is

likely that postzygotic ecological processes, such as selection for divergent fire-related and climatic niches limits hybrid viability in natural stands. Indeed, hybrids are rarely identified in sympatric stands (Zobel 1969; Brown 2021). Thus, it appears that niche divergence is associated with divergence in reproductive phenologies during speciation for our focal taxa. However, whether niche divergence reinforces reproductive isolation based on pollen release timing or divergent pollen release timing is an outcome of niche divergence itself remains an open question.

The rate of symmetrical gene flow in our best-fit demographic model was reduced by approximately 10 migrants per generation providing evidence that prezygotic reproductive isolation may have strengthened under Quaternary climatic conditions. This reduction reflects a scenario of reduced effective population sizes, reduced rates of migration (*m*), or both. The rate of gene flow associated with a given time interval should not be interpreted as constant though. Sousa et al. (2011) found that posterior distributions for the timing of gene flow parameters in demographic inference were highly variable across the simulations they performed making pulses of gene flow (i.e., a gene flow event occurring within a time frame of no active gene flow), as probable as constant, ongoing gene flow. While acknowledging this blurs interpretation of parameter estimates for gene flow, a history with recurring gene flow events fits the narrative of prezygotic isolation being labile especially when geographical distributions or reproductive phenology are the factors involved. Indeed, observations of hybridization occurring between once-prezygotically isolated species have been made and suggests phenological barriers such as timing of pollen release and flowering may not be

649 permanently established and can shift towards synchrony in warming climates (Vallejo-
650 Marín and Hiscock 2016).
651
652 ***Climate instability reduces genetic diversity***
653 Conifers often have high genetic diversity and low population differentiation because of
654 outcrossing, wind-dispersion, and introgression (Petit and Hampe 2006). *Pinus pungens*
655 and *P. rigida* both have modest levels of genetic diversity within and across the
656 populations we sampled, and no detectable within-species population structure given our
657 genome-wide data. Our best fit model inferred a drastic effective population size reduction
658 (*P. pungens,* ~99.7%; *P. rigida*, ~99.5%) between 118-198 kya. Since then, climate has
659 continued to oscillate between glacial and interglacial periods for geologic time intervals
660 too short for species with long generation times and low migratory potential to sufficiently
661 track causing a mismatch between the breadth of a species' climatic niche and where
662 populations are established (Svenning et al. 2015). This dynamic affects population
663 persistence, reduces genetic variation within populations, and thus to some degree limits
664 the potential for local adaptation in climatically unstable regions and may explain the lack
665 of IBD and IBE across the populations of our focal species. Our SDM predictions showed
666 substantial shifts in habitat suitability since the LIG providing evidence of high climate
667 instability in temperate eastern North America during the Quaternary period.
668
669 From a theoretical standpoint, we anticipated the patchy, mountain top distribution of *P.*
670 *pungens* to be characterized by strong patterns of population differentiation. Lack of
671 structure in *P. pungens* could be attributed to long distant dispersal or a recent move up

672 in elevation with genomes still housing elements of historical panmixia. Indeed, suitable
673 habitat predictions during the HOL, just 6000 years ago, were rather contiguously
674 distributed (Fig. 5) and may have allowed an increase in intraspecific gene flow. For *P.*
675 *rigida* some structure differentiating the northern populations from those along the
676 southern Appalachian Mountains was expected from an empirical standpoint because
677 previously reported trait values in a common garden study led to identification of three
678 latitudinally arranged genetic groupings (Ledig et al. 2015). Although structure analysis
679 did not support groupings within *P. rigida*, our estimates for isolation-by-distance (IBD)
680 yielded a correlation of 0.177 (p = 0.055) which is suggestive of structure. While this
681 shows some differentiation across its distribution, pairwise F_{ST} values were small and on
682 average smaller than those between populations of *P. pungens* suggesting higher
683 population connectivity in *P. rigida*. The SDM prediction for *P. rigida* during the LGM
684 indicated three regions with high habitat suitability: a large region farther south than where
685 it is currently distributed, one just east of the Appalachian Mountains, and another just
686 south of the Laurentide ice sheet. While genetic differences may have accumulated in
687 these separate refugia, the SDM prediction for the HOL were more compact and
688 contiguous for *P. rigida,* as was similarly observed for *P. pungens*, thus providing greater
689 potential for intraspecific gene flow across diverged populations.
690
691 ***Future work and conclusions***
692 Conifers, especially pines, have large and complex genomes (> 20 Gbp), so it is important
693 to first understand their demographic history, the prevalence of interspecific gene flow
694 historically and currently, and the role climate and geography may have played in shaping

695 standing levels of genetic diversity before expending valuable resources into a more
696 elaborate genomic research endeavor. The divergence history of *P. pungens* and *P.*
697 *rigida* involved a complex interplay of recurring interspecific gene flow and dramatic
698 population size reductions associated with changes in climate. We suspect that low
699 population differentiation within each species is linked to climate instability and the
700 mismatch between being long-lived, sessile species struggling to track rapidly shifting
701 climatic optima during complex demographic scenarios and that prezygotic isolating
702 barriers linked to reproductive phenology are involved in the maintenance of species
703 boundaries. The results of our demographic inference warrant more research into these
704 two species and other coniferous species of this region. No studies to date have inferred
705 divergence histories for pine species native to eastern North America, and to our
706 knowledge, only one other inference for conifer speciation has been conducted for this
707 region, *Picea mariana* and *P. rubens*, and it was an interesting, rare account for
708 progenitor-derivative speciation (Perron et al. 2000). Future detailed examinations of
709 hybridization between *P. pungens* and *P. rigida* are needed to elucidate the role
710 hybridization plays (e.g., adaptive or reinforcing) in the maintenance of species
711 boundaries and identify environmental associations with hybrid establishment (e.g.,
712 mountain aspect and elevation). Ideally, future research involving these two species
713 would use a method that sufficiently captures genic regions so population structure in
714 both species may be revealed and investigations into genomic islands of divergence that
715 are often associated with ecological speciation can be performed (Nosil and Feder 2012).
716

While more time, effort, and genomic resources are needed for us to accurately predict gains and losses in biodiversity or describe the development of reproductive isolation in conifer speciation, we must recognize that some montane conifer species will be disproportionally affected by future climate projections (Aitken et al. 2008) and time is of the essence in terms of capturing and understanding current levels of biodiversity. High elevational species such as *P. pungens* may already be experiencing a tipping point, but because *P. pungens* is a charismatic Appalachian tree with populations already threatened by fire suppression practices over the last century, conservation efforts have begun through seed banking (Jetton et al. 2015) and prescribed burning experiments of natural stands (Welch and Waldrop 2001). Our contributions to these conservation efforts include genome-wide population diversity estimates for *P. pungens* and *P. rigida* and a demographic inference scenario that involves a long history of interspecific gene flow and hybridization. In conifer species of the family *Pinaceae*, there are multiple accounts of introgression occurring through hybrid zones (De La Torre et al. 2014; Hamilton et al. 2015; Menon et al. 2018). The implications of introgression are far-reaching as it leads to greater genetic diversity and thus a greater capacity for adaptive evolution. Trees are keystone species, so understanding a population's potential to withstand environmental changes provides some insight into the future stability of an ecosystem.

Author contributions

CB performed field sampling of *Pinus rigida,* data analyses, and modeling. TF processed the genetic data and advised statistical analyses. CF led the field sampling and library prep for *Pinus pungens*, and AJE assisted with data analyses. All authors contributed to the writing of this manuscript.

Data Archiving Statement

Raw reads generated during this study are available at NCBI SRA database under BioProject: PRJNA803632 (Sample IDs: SAMN25684544 – SAMN25684843). Python scripts for demographic modeling and R scripts for genetic analyses and producing SDMs are available at www.github.com/boltece/Speciation_2pines.

Supplemental material:

Online Resource 1: Summary statistics per SNP, sampled tree, and population
Online Resource 2: Metadata and files needed to reproduce SDMs
Online Resource 3: Results from demographic inference and parameter unscaling
Online Resource 4: Supplemental figures

References:

Abbott RJ (2017) Plant speciation across environmental gradients and the occurrence and nature of hybrid zones. J Syst Evol 55:238–258. https://doi.org/10.1111/jse.12267

Aiello-Lammens ME, Boria RA, Radosavljevic A, et al (2015) spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. Ecography 38:541–545. https://doi.org/10.1111/ECOG.01132

Aitken SN, Yeaman S, Holliday JA, et al (2008) Adaptation, migration or extirpation: climate change outcomes for tree populations. Evol Appl 1:95–111. https://doi.org/10.1111/J.1752-4571.2007.00013.X

Baack E, Melo MC, Rieseberg LH, Ortiz-Barrientos D (2015) The origins of reproductive isolation in plants. New Phytol 207:968–984. https://doi.org/10.1111/NPH.13424

Bagley JC, Heming NM, Gutiérrez EE, et al (2020) Genotyping-by-sequencing and ecological niche modeling illuminate phylogeography, admixture, and Pleistocene range dynamics in quaking aspen (*Populus tremuloides*). Ecol Evol 10:4609–4629. https://doi.org/10.1002/ece3.6214

Bolte CE, Eckert AJ (2020) Determining the when, where and how of conifer speciation: a challenge arising from the study 'Evolutionary history of a relict conifer *Pseudotaxus chienii*.' Ann Bot 125:v–vii. https://doi.org/10.1093/AOB/MCZ201

Bonebrake TC, Mastrandrea MD (2010) Tolerance adaptation and precipitation changes complicate latitudinal patterns of climate change impacts. Proc Natl Acad Sci U S A 107:12581–12586. https://doi.org/10.1073/pnas.0911841107

783 Brown AL (2021) Phenotypic characterization of Table Mountain (*Pinus pungens*) and
784 pitch pine (*Pinus rigida*) hybrids along an elevational gradient in the Blue Ridge
785 Mountains, Virginia. Thesis, Virginia Commonwealth University

786 Cannon CH, Petit RJ (2020) The oak syngameon: more than the sum of its parts. New
787 Phytol 226:978–983. https://doi.org/10.1111/nph.16091

788 Capblancq T, Butnor JR, Deyoung S, et al (2020) Whole-exome sequencing reveals a
789 long-term decline in effective population size of red spruce (*Picea rubens*). Evol
790 Appl 13:2190. https://doi.org/10.1111/EVA.12985

791 Carstens BC, Richards CL (2007) Integrating coalescent and ecological niche modeling
792 in comparative phylogeography. Evolution 61:1439–1454.
793 https://doi.org/10.1111/J.1558-5646.2007.00117.X

794 Cavender-Bares J (2019) Diversification, adaptation, and community assembly of the
795 American oaks (*Quercus*), a model clade for integrating ecology and evolution.
796 New Phytol 221:669–692. https://doi.org/10.1111/nph.15450

797 Christe C, Stölting KN, Paris M, et al (2017) Adaptive evolution and segregating load
798 contribute to the genomic landscape of divergence in two tree species connected
799 by episodic gene flow. Mol Ecol 26:59–76. https://doi.org/10.1111/mec.13765
800
801 Critchfield WB (1963) The Austrian x red pine hybrid. Silvae Genet12:187-191

802 Critchfield WB (1967) Crossability and relationships of the closed-cone pines. Silvae
803 Genet 16:89–97

804 Critchfield WB (1986) Hybridization and Classification of the White Pines (*Pinus* Section
805 *Strobus*). Taxon 35:647–656

806 Danecek P, Auton A, Abecasis G, et al (2011) The variant call format and VCFtools.
807 Bioinformatics 27:2156–2158. https://doi.org/10.1093/bioinformatics/btr330

808 De La Torre AR, Birol I, Bousquet J, et al (2014) Insights into conifer giga-genomes.
809 Plant Physiol 166:1724–1732. https://doi.org/10.1104/pp.114.248708

810 De La Torre AR, Li Z, Van De Peer Y, Ingvarsson PK (2017) Contrasting rates of
811 molecular evolution and patterns of selection among gymnosperms and flowering
812 plants. Mol Biol Evol 34:1363–1377. https://doi.org/10.1093/molbev/msx069

813 Della-Bianca L (1990) *Pinus pungens* Lamb., Table Mountain pine. In: Burns, R.M. and
814 B.H. Honkala (eds.). Silvics of North America. Volume 1. Conifers. USDA Forest
815 Service Agriculture Handbook 654, Washington, D.C.

816 Dobzhansky T (1950) Heredity, Environment, and Evolution. Assoc Adv Sci 111:161-

166. https://doi.org/10.1126/science.111.2877.161

Dorman KW, Barber JC (1956) Time of flowering and seed ripening in southern pines. USDA Forest Service, Southeastern Forest Experiment Station, Old Station Paper SE-072, 72.

Dyer RJ, Nason JD (2004) Population Graphs: The graph theoretic shape of genetic structure. Mol Ecol 13:1713–1727. https://doi.org/10.1111/J.1365-294X.2004.02177.X

Dyke AS, Moore A, Robertson L (2003) Deglaciation of North America, Open File 1574, Natural Resources Canada, Ottawa.

Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int J Climatol 37:4302–4315. https://doi.org/10.1002/JOC.5086

Francis RM (2017) pophelper: an R package and web app to analyse and visualize population structure. Mol Ecol Resour 17:27–32. https://doi.org/10.1111/1755-0998.12509

Fu L, Niu B, Zhu Z, et al (2012) CD-HIT: Accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150–3152. https://doi.org/10.1093/bioinformatics/bts565

Gao J, Wang B, Mao JF, et al (2012) Demography and speciation history of the homoploid hybrid pine *Pinus densata* on the Tibetan Plateau. Mol Ecol 21:4811–4827. https://doi.org/10.1111/j.1365-294X.2012.05712.x

Gernandt DS, Aguirre Dugua X, Vázquez-Lobo A, et al (2018) Multi-locus phylogenetics, lineage sorting, and reticulation in *Pinus* subsection *Australes*. Am J Bot 105:711–725. https://doi.org/10.1002/AJB2.1052

Goudet J, Jombart T (2020) hierfstat, estimations and tests of hierarchical *F*-statistics. R package version 0.5-7. https://CRAN.R-project.org/package=hierfstat

Gougherty A V., Chhatre VE, Keller SR, Fitzpatrick MC (2020) Contemporary range position predicts the range-wide pattern of genetic diversity in balsam poplar (*Populus balsamifera* L.). J Biogeogr 47:1246–1257. https://doi.org/10.1111/jbi.13811

Gugger PF, Ikegami M, Sork VL (2013) Influence of late Quaternary climate change on present patterns of genetic variation in valley oak, *Quercus lobata* Née. Mol Ecol 22:3598–3612. https://doi.org/10.1111/MEC.12317

857 Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the
858 joint demographic history of multiple populations from multidimensional SNP
859 frequency data. PLoS Genet 5:1-11. https://doi.org/10.1371/journal.pgen.1000695

860 Hagman M (1975) Incompatibility in forest trees. Proc R Soc London Ser B Biol Sci
861 188:313–326.

862 Hamilton JA, De la Torre AR, Aitken SN (2015) Fine-scale environmental variation
863 contributes to introgression in a three-species spruce hybrid complex. Tree Genet
864 Genomes 11. https://doi.org/10.1007/s11295-014-0817-y

865 Hapke A, Thiele D (2016) GIbPSs: a toolkit for fast and accurate analyses of
866 genotyping-by-sequencing data without a reference genome. Mol Ecol Resour
867 16:979–990. https://doi.org/10.1111/1755-0998.12510

868 Hernández-León S, Gernandt DS, Pérez de la Rosa J a., Jardón-Barbolla L (2013)
869 Phylogenetic relationships and species delimitation in *Pinus* section *Trifoliae*
870 inferrred from plastid DNA. PLoS One 8:1–14.
871 https://doi.org/10.1371/journal.pone.0070501

872 Hernández-Serrano A, Verdú M, González-Martínez SC, Pausas JG (2013) Fire
873 structures pine serotiny at different scales. Am J Bot 100:2349–2356.
874 https://doi.org/10.3732/ajb.1300182

875 Herten K, Hestand MS, Vermeesch JR, Van Houdt JKJ (2015) GBSX: A toolkit for
876 experimental design and demultiplexing genotyping by sequencing experiments.
877 BMC Bioinformatics 16:1-6. https://doi.org/10.1186/s12859-015-0514-3

878 Hewitt GM (2001) Speciation, hybrid zones and phylogeography - Or seeing genes in
879 space and time. Mol Ecol 10:537–549. https://doi.org/10.1046/J.1365-
880 294X.2001.01202.X
881
882 Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary.
883 Philos Trans R Soc Lond B 359:183–195. https://doi.org/10.1098/rstb.2003.1388
884 Hickerson MJ, Carstens BC, Cavender-Bares J, et al (2010) Phylogeography's past,
885 present, and future: 10 years after. Mol Phylogenet Evol 54:291–301.
886 https://doi.org/10.1016/J.YMPEV.2009.09.016
887
888 Hijmans RJ, Cameron SE, Parra JL, et al (2005) Very high resolution interpolated
889 climate surfaces for global land areas. Int J Climatol 25:1965–1978.
890 https://doi.org/10.1002/JOC.1276
891
892 Hijmans RJ (2019). geosphere: Spherical trigonometry. R package version 1.5-10.
893 https://CRAN.R-project.org/package=geosphere

Hohenlohe PA, Day MD, Amish SJ, et al (2013) Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. Mol Ecol 22:3002–3013. https://doi.org/10.1111/mec.12239

Ikeda DH, Max TL, Allan GJ, et al (2017) Genetically informed ecological niche models improve climate change predictions. Glob Chang Biol 23:164–176. https://doi.org/10.1111/gcb.13470

Jetton RM, Crane BS, Whittier WA, Dvorak WS (2015) Genetic resource conservation of Table Mountain pine (*Pinus pungens*) in the central and southern Appalachian Mountains. Tree Plant Notes 58:42–52

Jin WT, Gernandt DS, Wehenkel C, et al (2021) Phylogenomic and ecological analyses reveal the spatiotemporal evolution of global pines. Proc Natl Acad Sci USA 118. https://doi.org/10.1073/PNAS.2022302118/-/DCSUPPLEMENTAL

Ju M-M, Feng L, Yang J, et al (2019) Evaluating population genetic structure and demographic history of *Quercus spinosa* (Fagaceae) based on specific length amplified fragment sequencing. Front Genet 10:965. https://doi.org/10.3389/FGENE.2019.00965

Jump AS, Peñuelas J (2005) Running to stand still: Adaptation and the response of plants to rapid climate change. Ecol Lett 8:1010–1020. https://doi.org/10.1111/j.1461-0248.2005.00796.x

Kass JM, Muscarella R, Galante PJ, et al (2021) ENMeval 2.0: Redesigned for customizable and reproducible modeling of species' niches and distributions. Methods Ecol Evol 12:1602–1608. https://doi.org/10.1111/2041-210X.13628

Keller SR, Olson MS, Salim S, et al (2010) Genomic diversity, population structure, and migration following rapid range expansion in the Balsam poplar, *Populus balsamifera*. Mol Ecol 19:1212–1226. https://doi.org/10.1111/j.1365-294X.2010.04546.x

Keeley JE (2012) Ecology and evolution of pine life histories. Ann For Sci 69:445–453. https://doi.org/10.1007/s13595-012-0201-8

Kim BY, Wei X, Fitz-Gibbon S, et al (2018) RADseq data reveal ancient, but not pervasive, introgression between Californian tree and scrub oak species (*Quercus* sect. *Quercus*: *Fagaceae*). Mol Ecol 27:4556–4571. https://doi.org/10.1111/MEC.14869

Kriebel HB (1972). Embryo development and hybridity barriers in the white pines (Section *Strobus*). Silvae Genet 21:39-44.

Kulmuni J, Butlin RK, Lucek K, et al (2020) Towards the completion of speciation: The evolution of reproductive isolation beyond the first barriers: Progress towards complete speciation. Philos Trans R Soc B Biol Sci 375:20190528. https://doi.org/10.1098/rstb.2019.0528

Łabiszak B, Zaborowska J, Wójkiewicz B, Wachowiak W (2021) Molecular and paleo-climatic data uncover the impact of an ancient bottleneck on the demographic history and contemporary genetic structure of endangered *Pinus uliginosa*. J Syst Evol 59:596–610. https://doi.org/10.1111/jse.12573

Lafontaine G de, Prunier J, Gérardi S, Bousquet J (2015) Tracking the progression of speciation: variable patterns of introgression across the genome provide insights on the species delimitation between progenitor–derivative spruces (*Picea mariana* × *P. rubens*). Mol Ecol 24:5229–5247. https://doi.org/10.1111/MEC.13377

Lascoux M, Palmé AE, Cheddadi R, Latta RG (2004) Impact of Ice Ages on the genetic structure of trees and shrubs. Philos Trans R Soc Lond B Biol Sci 359:197-207. https://doi.org/10.1098/rstb.2003.1390

Ledig FT, Smouse PE, Hom JL (2015) Postglacial migration and adaptation for dispersal in pitch pine (*Pinaceae*). Am J Bot 102:2074–2091. https://doi.org/10.3732/AJB.1500009

Legendre P, Legendre L (2012) Numerical Ecology. Third Edition. Elsevier.

Li L, Abbott RJ, Liu B, et al (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. https://doi.org/10.1111/MEC.12466

Lima JS, Telles MPC, Chaves LJ, et al (2017) Demographic stability and high historical connectivity explain the diversity of a savanna tree species in the Quaternary. Ann Bot 119:645–657. https://doi.org/10.1093/AOB/MCW257

Lind BM, Menon M, Bolte CE, et al (2018) The genomics of local adaptation in trees: are we out of the woods yet? Tree Genet Genomes 14:29. https://doi.org/10.1007/s11295-017-1224-y

Little EL Jr. (1971) Atlas of United States trees, Vol. 1, conifers and important hardwoods: U.S. Department of Agriculture, 1146, 9, p200

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Research 27: 209–220.

McKinney GJ, Waples RK, Seeb LW, Seeb JE (2017) Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-

sequencing data from natural populations. Mol Ecol Resour 17:656–669. https://doi.org/10.1111/1755-0998.12613

McKinney GJ, Waples RK, Pascal CE, et al (2018) Resolving allele dosage in duplicated loci using genotyping-by-sequencing data: A path forward for population genetic analysis. Mol Ecol Resour 18:570–579. https://doi.org/10.1111/1755-0998.12763

McWilliam JR (1959) Interspecific Incompatibility in *Pinus*. Am J Bot 46:425–433

Menon M, Bagley JC, Friedline CJ, et al (2018) The role of hybridization during ecological divergence of southwestern white pine (*Pinus strobiformis*) and limber pine (*P. flexilis*). Mol Ecol 27:1245–1260. https://doi.org/10.1111/MEC.14505

Michaux, FA (1819) The North American Sylva, or A description of the forest trees of the United States, Canada and Nova Scotia considered particularly with respect to their use in the arts, and their introduction into commerce; to which is added a description of the most useful of the European forest trees: illustrated by 156 coloured engravings. C. d'Hautel, Paris https://doi.org/10.5962/bhl.title.48807

Nosil, P (2012) Ecological Speciation. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199587100.001.0001

Nosil P, Feder JL (2012) Widespread yet heterogeneous genomic divergence. Mol Ecol 21:2829–2832. https://doi.org/10.1111/j.1365-294X.2012.05580.x

Oksanen J, Blanchet FG, Friendly M, et al (2020) vegan: Community ecology package. R package version 2.5-7. https://CRAN.R-project.org/package=vegan

Otto-Bliesner BL, Marshall SJ, Overpeck JT, et al (2006) Simulating Arctic climate warmth and icefield retreat in the last interglaciation. Sci 311:1751–1753. https://doi.org/10.1126/science.1120808

Parchman TL, Gompert Z, Mudge J, et al (2012) Genome-wide association genetics of an adaptive trait in lodgepole pine. Mol Ecol 21:2991–3005. https://doi.org/10.1111/j.1365-294X.2012.05513.x

Parchman TL, Jahner JP, Uckele KA, et al (2018) RADseq approaches and applications for forest tree genetics. Tree Genet. Genomes 14. https://doi.org/10.1007/s11295-018-1251-3

Parisien MA, Moritz MA (2009) Environmental controls on the distribution of wildfire at multiple spatial scales. Ecol Monogr 79:127–154. https://doi.org/10.1890/07-1289.1

Park B, Donoghue MJ (2019) Phylogeography of a widespread eastern North American shrub, *Viburnum lantanoides*. Am J Bot 106:389–401. https://doi.org/10.1002/AJB2.1248

Pausas JG, Schwilk D (2012) Fire and plant evolution. New Phytol 193:301–303. https://doi.org/10.1111/j.1469-8137.2011.04010.x

Perron M, Perry DJ, Andalo C, Bousquet J (2000) Evidence from sequence-tagged-site markers of a recent progenitor-derivative species pair in conifers. Proc Natl Acad Sci USA 97:11331–11336. https://doi.org/10.1073/PNAS.200417097

Peterson BK, Weber JN, Kay EH, et al (2012) Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7:e37135 . https://doi.org/10.1371/journal.pone.0037135

Peterson AT, Anamza T (2015) Ecological niches and present and historical geographic distributions of species: A 15-year review of frameworks, results, pitfalls, and promises. Folia Zool 64:207–217. https://doi.org/10.25225/FOZO.V64.I3.A3.2015

Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. Annu Rev Ecol Evol Syst 37:187-214. https://doi.org/10.1146/annurev.ecolsys.37.091305.110215

Phillips SJ, Anderson RP, Dudík M, et al (2017) Opening the black box: an open-source release of Maxent. Ecography 40:887–893. https://doi.org/10.1111/ECOG.03049

Puritz JB, Hollenbeck CM, Gold JR (2014) dDocent: A RADseq, variant-calling pipeline designed for population genomics of non-model organisms. PeerJ 2:e431. https://doi.org/10.7717/peerj.431

R Core Team (2021) R: A Language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Raj A, Stephens M, Pritchard JK (2014) FastSTRUCTURE: Variational inference of population structure in large SNP data sets. Genetics 197:573–589. https://doi.org/10.1534/genetics.114.164350

Richards CL, Carstens BC, Lacey Knowles L (2007) Distribution modelling and statistical phylogeography: An integrative framework for generating and testing alternative biogeographical hypotheses. J. Biogeogr. 34:1833–1845. https://doi.org/10.1111/j.1365-2699.2007.01814.x

Saladin B, Leslie AB, Wüest RO, et al (2017) Fossils matter: improved estimates of divergence times in *Pinus* reveal older diversification. BMC Evol Biol 17:95. https://doi.org/10.1186/S12862-017-0941-Z

Savolainen O, Bokma F, García-Gil R, et al (2004) Genetic variation in cessation of growth and frost hardiness and consequences for adaptation of *Pinus sylvestris* to climatic changes. For Ecol Manage 197:79–89. https://doi.org/10.1016/j.foreco.2004.05.006

Seehausen O, Butlin RK, Keller I, et al (2014) Genomics and the origin of species. Nat Rev Genet 15:176–192. https://doi.org/10.1038/nrg3644

Soltis DE, Morris AB, McLachlan JS, et al (2006) Comparative phylogeography of unglaciated eastern North America. Mol Ecol 15:4261–4293. https://doi.org/10.1111/j.1365-294X.2006.03061.x

Sousa VC, Grelaud A, Hey J (2011) On the nonidentifiability of migration time estimates in isolation with migration models. Mol Ecol 20:3956–3962. https://doi.org/10.1111/j.1365-294X.2011.05247.x

Stetter MG, Thornton K, Ross-Ibarra J (2018) Genetic architecture and selective sweeps after polygenic adaptation to distant trait optima. PLoS Genet 14:e1007794. https://doi.org/10.1101/313247

Svenning JC, Eiserhardt WL, Normand S, et al (2015) The Influence of Paleoclimate on Present-Day Patterns in Biodiversity and Ecosystems. Annu Rev Ecol Evol Syst 46:551–572. https://doi.org/10.1146/annurev-ecolsys-112414-054314

Vallejo-Marín M, Hiscock SJ (2016) Hybridization and hybrid speciation under global change. New Phytol 211:1170–1187. https://doi.org/10.1111/nph.14004

Vasilyeva G, Goroshkevich S (2018) Artificial crosses and hybridization frequency in five-needle pines. Dendrobiology 80:123–130. https://doi.org/10.12657/denbio.080.012

Wang IJ, Bradburd GS (2014) Isolation by environment. Mol Ecol 23:5649–5662. https://doi.org/10.1111/mec.12938

Warren DL, Glor RE, Turelli M (2008) Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. Evolution 62:2868–2883. https://doi.org/10.1111/J.1558-5646.2008.00482.X

Warren DL, Matzke NJ, Cardillo M, et al (2021) ENMTools 1.0: an R package for comparative ecological biogeography. Ecography 44:504–511. https://doi.org/10.1111/ecog.05485

Welch NT, Waldrop TA (2001) Restoring Table Mountain pine (*Pinus pungens* Lamb.) communities with prescribed fire: An overview of current research. Castanea 66:42–49.

Williams JW, Jackson ST (2007) Novel climates, no-analog communities, and ecological surprises. Front Ecol Environ 5:475–482. https://doi.org/10.1890/070037

Wright S (1943) Isolation by distance. Genetics 28:114–138.

Wright JW (1959) Species hybridization in the white pines. Forest Sci 5:210–222.

Yang R-C (1998) Estimating hierarchical *F*-statistics. Evolution 52:950–956. https://doi.org/10.2307/2411227

Yang YX, Zhi LQ, Jia Y, et al (2020) Nucleotide diversity and demographic history of *Pinus bungeana*, an endangered conifer species endemic in China. J Syst Evol 58:282–294. https://doi.org/10.1111/jse.12546

Zobel DB (1969) Factors affecting the distribution of *Pinus pungens*, an Appalachian endemic. Ecol Monogr 39:303–333.

Zou J, Sun Y, Li L, et al (2013) Population genetic evidence for speciation pattern and gene flow between *Picea wilsonii*, *P. morrisonicola* and *P. neoveitchii*. Ann Bot 112:1829–1844. https://doi.org/10.1093/AOB/MCT241

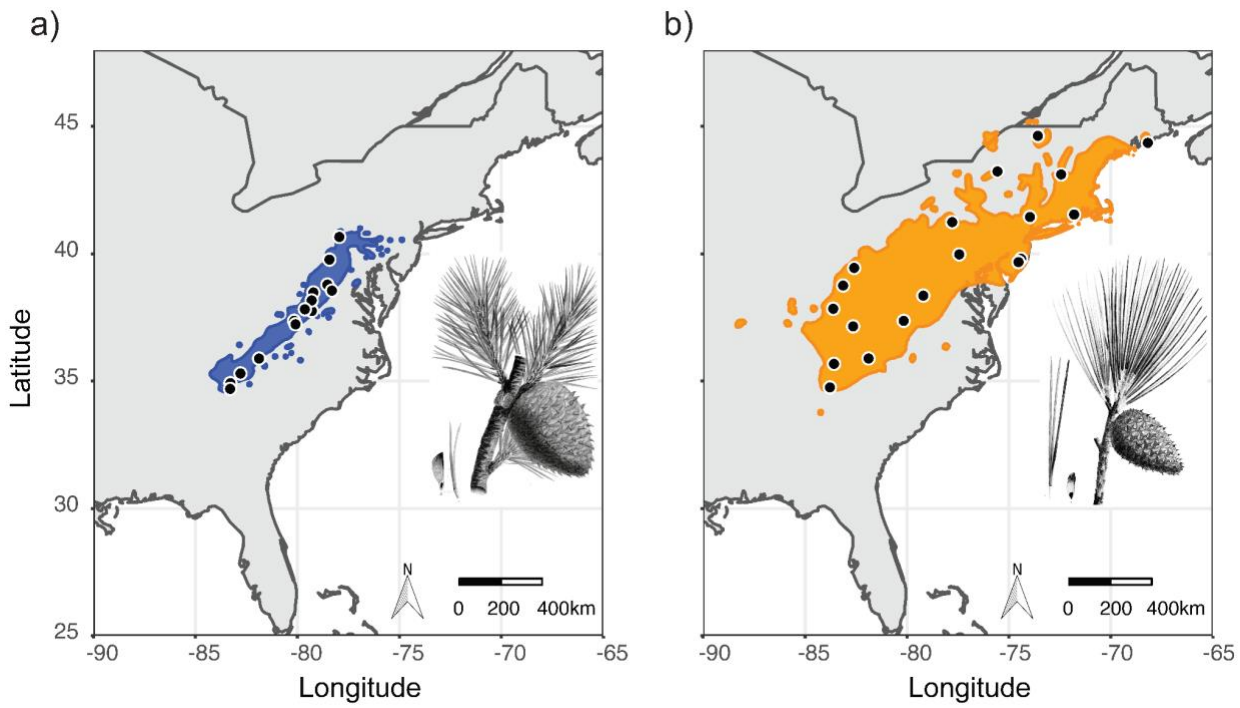

Fig. 1 Known geographical distribution of focal species, a) *Pinus pungens* and b) *P. rigida*, (Little 1971) in relation to populations sampled (black dots) for genetic analysis; Phenotypic characterization of each species was illustrated by Pierre-Joseph Redouté (Michaux 1819)

Table 1 Location of sampled populations, number of trees (*n*) that were sampled, and the observed heterozygosity (H_o) versus the expected heterozygosity ($H_e = 2pq$) for *Pinus pungens* and *P. rigida* populations.

Species	Code	Location	Lat	Long	*n*	H_o	H_e
P. pungens	PU_BB	Briery Branch, VA	38.48	-79.22	8	0.110	0.108
P. pungens	PU_BN	Buchanan State Forest, PA	39.77	-78.43	6	0.141	0.121
P. pungens	PU_BV	Buena Vista, VA	37.76	-79.29	11	0.124	0.120
P. pungens	PU_DT	Dragon's Tooth, VA	37.37	-80.16	7	0.101	0.098
P. pungens	PU_EG	Edinburg Gap, VA	38.79	-78.53	8	0.139	0.124
P. pungens	PU_EK	Elliott Knob, VA	38.17	-79.30	10	0.131	0.123
P. pungens	PU_GA	Walnut Fork, GA	34.92	-83.28	10	0.129	0.123
P. pungens	PU_LG	Looking Glass Rock, NC	35.30	-82.79	8	0.130	0.119
P. pungens	PU_NM	North Mountain, VA	37.82	-79.63	12	0.130	0.121
P. pungens	PU_PM	Poor Mountain, VA	37.23	-80.09	11	0.130	0.125
P. pungens	PU_SC	Pine Mountain, VA	34.70	-83.30	8	0.128	0.122
P. pungens	PU_SH	Shenandoah NP, VA	38.55	-78.31	5	0.160	0.128
P. pungens	PU_SV	Stone Valley Forest, PA	40.66	-77.95	9	0.110	0.110
P. pungens	PU_TR	Table Rock Mountain, NC	35.89	-81.88	12	0.113	0.114
P. rigida	RI_BR	Bass River State Forest, NJ	39.80	-74.41	9	0.101	0.105
P. rigida	RI_CT	Pachaug State Forest, CT	41.54	-71.81	10	0.096	0.107
P. rigida	RI_DT	Dragon's Tooth, VA	37.37	-80.16	10	0.109	0.106
P. rigida	RI_GA	Chattahoochee NF, GA	34.75	-83.78	9	0.096	0.103
P. rigida	RI_GW	George Washington NF, VA	38.36	-79.20	10	0.102	0.103
P. rigida	RI_HH	Hudson Highlands State Park, NY	41.44	-73.97	7	0.102	0.101
P. rigida	RI_JF	Jefferson NF, VA	37.15	-82.64	10	0.095	0.100
P. rigida	RI_KY	Daniel Boone NF, KY	37.84	-83.62	9	0.113	0.110
P. rigida	RI_ME	Acadia NP, ME	44.36	-68.19	10	0.107	0.106
P. rigida	RI_MI	Michaux State Forest, PA	39.98	-77.44	10	0.123	0.114
P. rigida	RI_NJ	Wharton State Forest, NJ	39.68	-74.53	9	0.098	0.101
P. rigida	RI_NY	Macomb State Park, NY	44.63	-73.58	9	0.101	0.104
P. rigida	RI_OH	South Bloomingville, OH	39.45	-82.59	8	0.093	0.096
P. rigida	RI_RS	Rome Sand Plains, NY	43.23	-75.56	9	0.097	0.103
P. rigida	RI_SH	Shawnee State Park, OH	38.75	-83.13	9	0.082	0.094
P. rigida	RI_SP	Sproul State Forest, PA	41.24	-77.78	9	0.106	0.105
P. rigida	RI_TN	Great Smoky Mountains NP, TN	35.68	-83.58	8	0.099	0.104
P. rigida	RI_TR	Table Rock Mountain, NC	35.89	-81.89	10	0.113	0.112
P. rigida	RI_VT	Bellows Falls, VT	43.11	-72.44	10	0.098	0.104

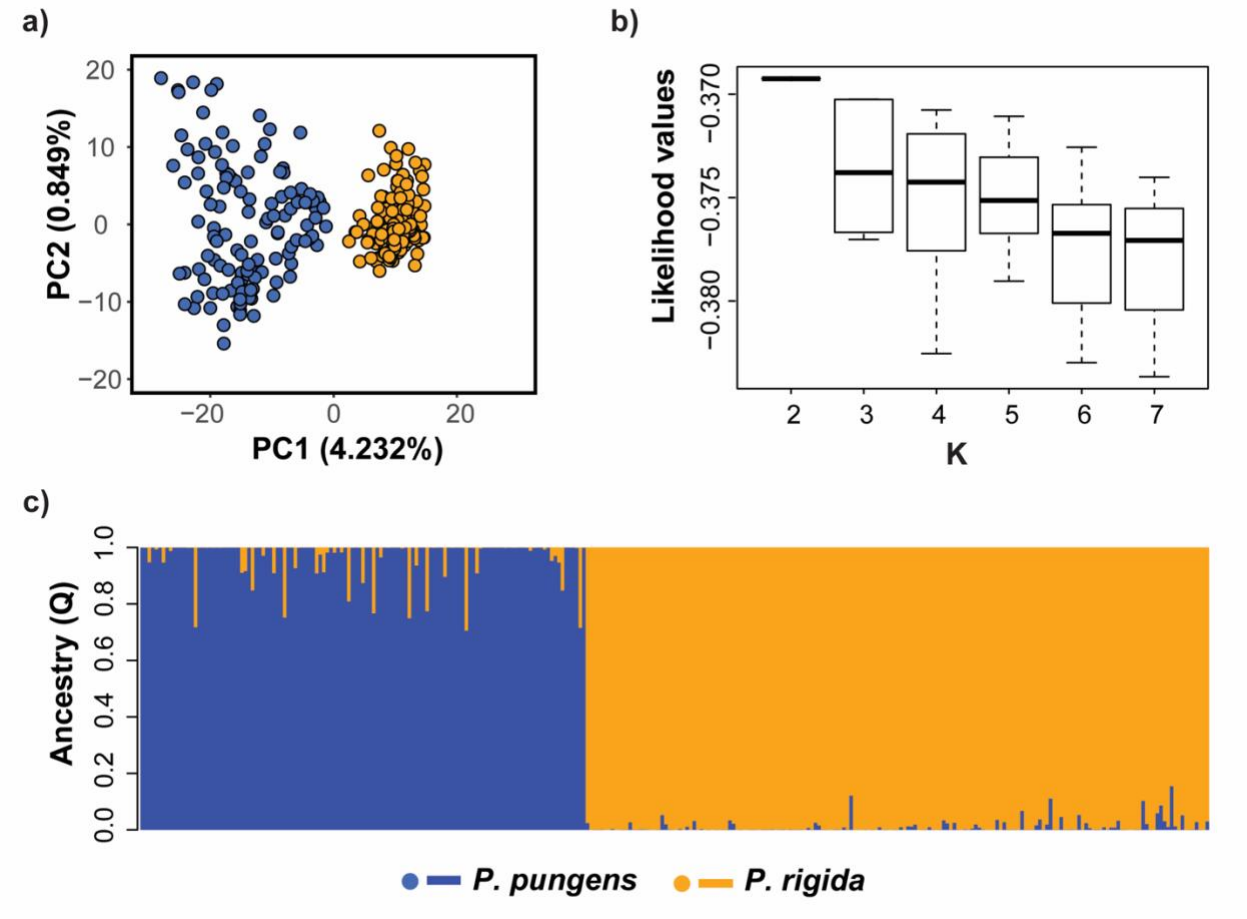

Fig. 2 Measures of genetic differentiation and diversity among sampled trees of *P. pungens* and *P. rigida*: a) Principal components analysis of 2168 genome-wide single nucleotide polymorphism (SNPs) for *Pinus pungens* (blue, left side of PC1) and *P. rigida* (orange, right side of PC1); b) log-likelihood values across ten replicate runs in fastSTRUCTURE for K = 2 through K = 7; c) results of averaged K = 2 ancestry (Q) assignments for each sample arranged latitudinally in each species

Table 2 Summary statistics of genetic differentiation for the sampled populations of *P. rigida* and *P. pungens*. Expected (H_e) and observed heterozygosity (H_o) values are the averages across 2168 SNPs averaged across populations.

Species	F_{ST} (95% CI)	IBD *r* (*p*-value)	IBE *r* (*p*-value)	H_e (range)	H_o (range)
P. pungens	0.0057 (0.0032 - 0.0084)	-0.0789 (0.638)	0.0131 (0.411)	0.118 (0.098-0.129)	0.127 (0.101-0.160)
P. rigida	0.0056 (0.0032 - 0.0082)	0.1758 (0.055)	-0.0669 (0.633)	0.104 (0.094-0.114)	0.102 (0.082 -0.123)

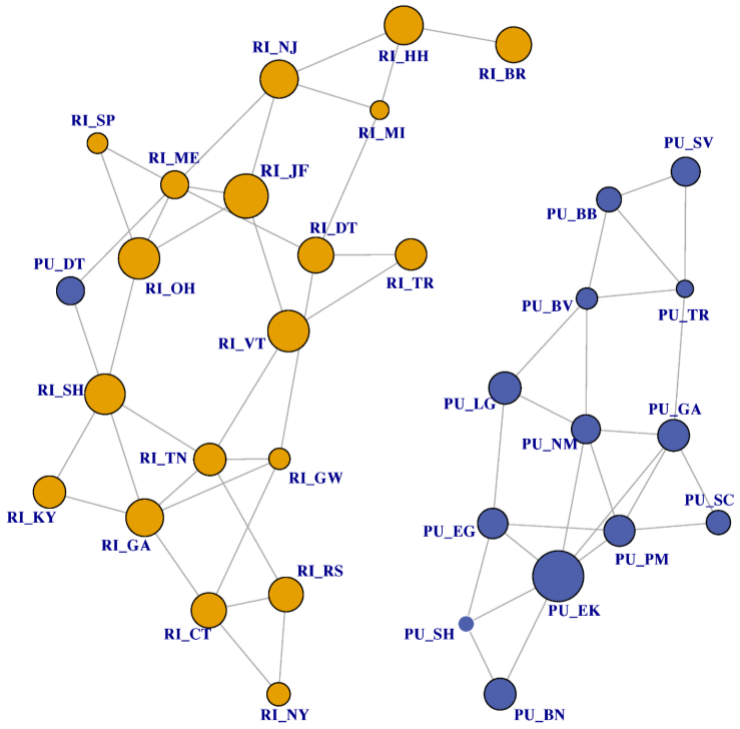

Fig. 3 Population graph reflecting the genetic relationships across all sampled populations of *P. pungens* (blue nodes, PU_xx labels) and *P. rigida* (orange nodes; RI_xx

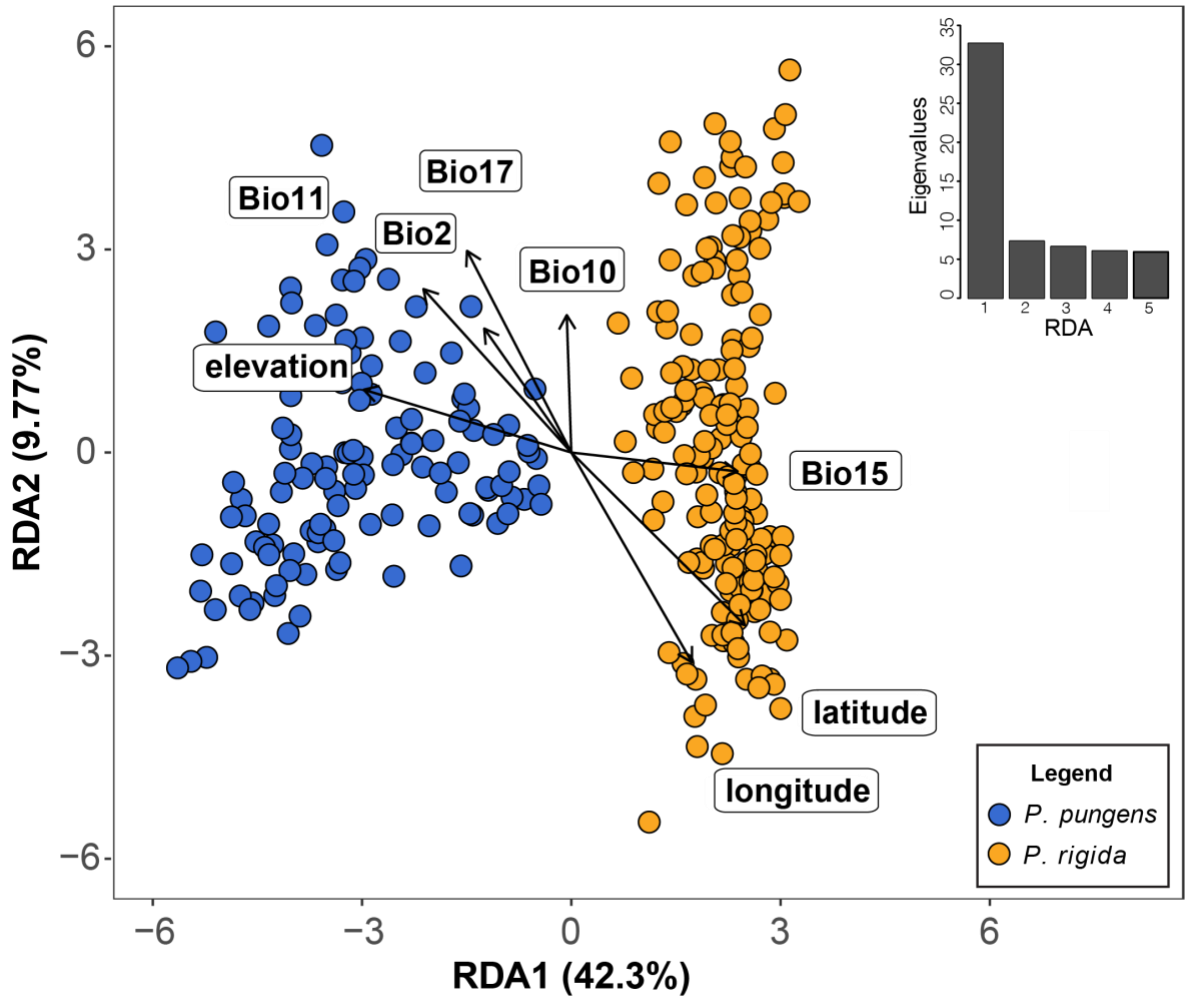

Fig. 4 Redundancy analysis (RDA) of the multilocus genotypes for each tree with climate and geographic predictor variables (full model). Direction and length of arrows on each RDA plot correspond to the loadings of each variable

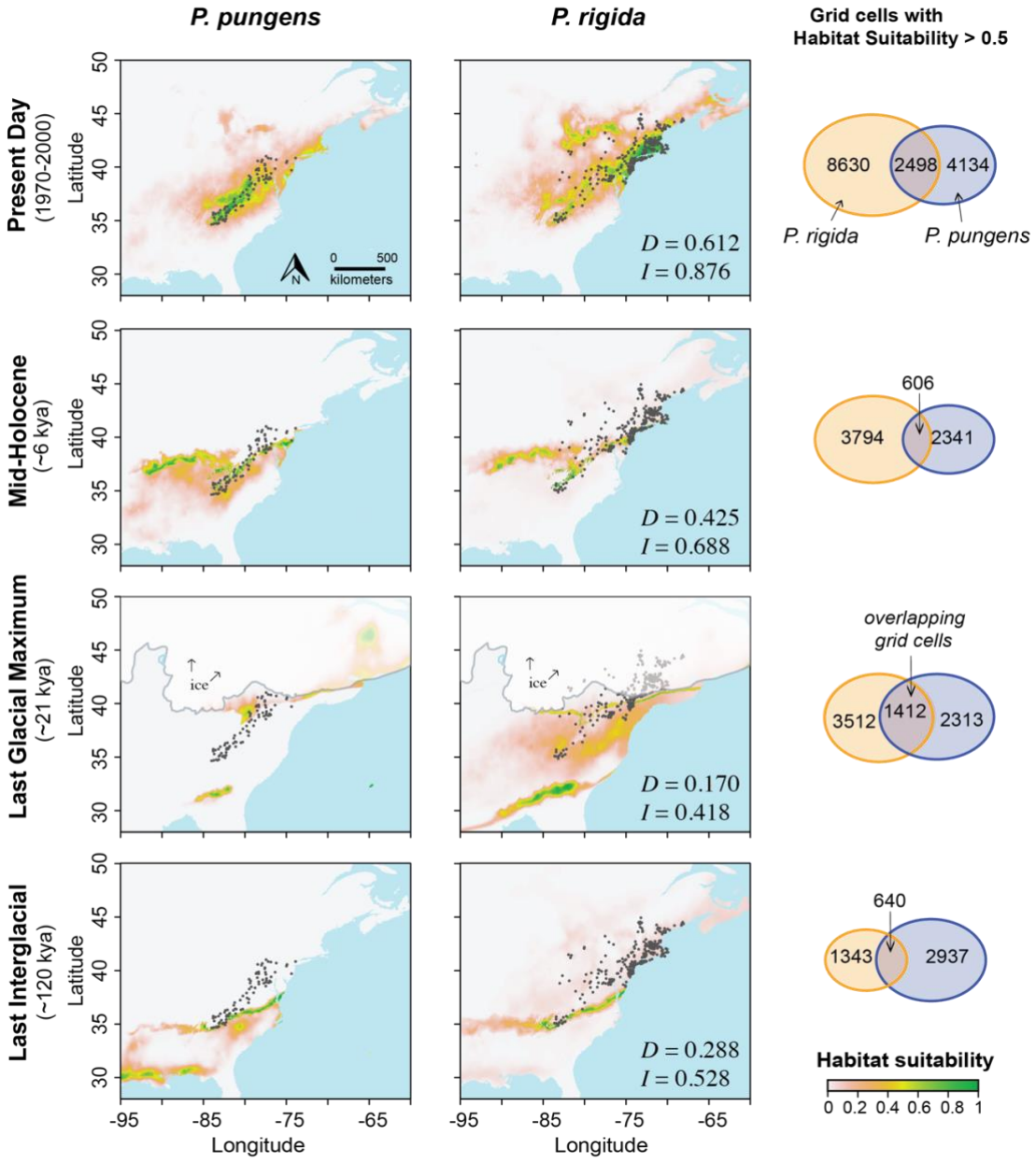

Fig. 5 Species distribution model (SDM) predictions across four time points for *P. pungens* and *P. rigida*. Measures of raster overlap in terms of Schoener's *D* and Warren's *I* index between the models of each species, and at each time point, are presented in the bottom right corner of the prediction plots for *P. rigida.* Venn diagrams illustrate the number of grid cells with moderate to high habitat suitability scores (> 0.5) for each SDM at a given time point, as well as the number of shared, or overlapping, grid cells. Blue Venn diagram ovals show grid cell counts from the *P. pungens* SDM, and orange Venn diagram ovals show grid cell counts from the *P. rigida* SDM for the aligning time point (denoted on the left side). Glacial extent data (labeled ice in LGM plots) for 18 kya was provided by Dyke (2003)

Table 3 Results of model fitting for twelve representative demographic models of divergence. Models are ranked by the number of parameters (*k*). Log-likelihood (log*L*) and Akaike information criterion (AIC) are provided for each model. Model details are given in the footnote.

Model	***k***	**log*L***	**AIC**
SI	3	-2254.18	4,514.37
MIGs	4	-2201.51	4,411.02
MIGa	5	-2210.81	4,431.62
SCs	5	-2213.93	4,437.86
SGFs	5	-2229.65	4,469.30
SCa	6	-2238.03	4,488.06
SGFa	6	-2241.07	4,494.14
PSC	6	-2277.78	4,567.56
PSCSCs	7	-2178.16	4,370.32
PSCMIGs	7	-1866.42	3,746.84
PSCMIGa	8	-2117.91	4,251.82
PSCMIGCs	**9**	**-1853.99**	**3,726.00**

SI, strict isolation; MIGs, symmetrical gene flow; MIGa, asymmetrical gene flow; SCs, secondary contact with symmetrical gene flow; SCa, secondary contact with asymmetrical gene flow; SGFa, speciation with asymmetrical gene flow SGFs, speciation with symmetrical gene flow; PSC, population size change; MIGCs, change in rate of symmetrical gene flow. The best-fit model is in bold.

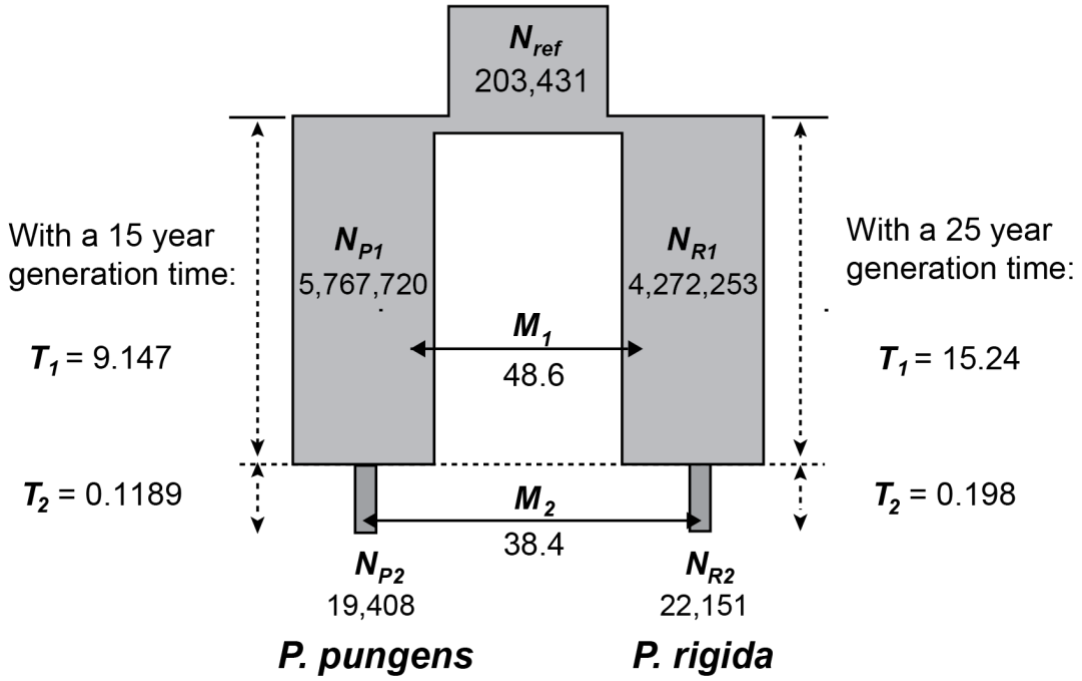

Fig. 6 The best-fit model (PSCMIGCs) and unscaled parameter estimates from $\partial\alpha\partial i$ analysis. Time intervals (T_i) are represented in millions of years and associated with lineage population sizes (N_i) and a specific rate of symmetrical gene flow (M_i)