Received: 29 August 2017 | Revised: 19 December 2017 | Accepted: 21 December 2017

DOI: 10.1111/mec.14505

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

The role of hybridization during ecological divergence of southwestern white pine (*Pinus strobiformis*) and limber pine (*P. flexilis*)

Mitra Menon[1,2] | Justin C. Bagley[2,3] | Christopher J. Friedline[2] | Amy V. Whipple[4] | Anna W. Schoettle[5] | Alejandro Leal-Sàenz[6] | Christian Wehenkel[7] | Francisco Molina-Freaner[8] | Lluvia Flores-Rentería[9] | M. Socorro Gonzalez-Elizondo[10] | Richard A. Sniezko[11] | Samuel A. Cushman[12] | Kristen M. Waring[13] | Andrew J. Eckert[2]

[1]Integrative Life Sciences, Virginia Commonwealth University, Richmond, VA, USA

[2]Department of Biology, Virginia Commonwealth University, Richmond, VA, USA

[3]Departamento de Zoologia, Universidade de Brasília, Brasília, DF, Brazil

[4]Department of Biological Sciences and Merriam Powel Center for Environmental Research, Northern Arizona University, Flagstaff, AZ, USA

[5]Rocky Mountain Research Station, USDA Forest Service, Ft. Collins, CO, USA

[6]Programa Institucional de Doctorado en Ciencias Agropecuarias y Forestales, Universidad Juárez del Estado de Durango, Durango, Mexico

[7]Instituto de Silvicultura e Industria de la Madera, Universidad Juarez del Estado de Durango, Durango, Mexico

[8]Institutos de Geologia y Ecologia, Universidad Nacional Autónoma de Mexico, Estación Regional del Noroeste, Hermosillo, Sonora, Mexico

[9]Department of Biology, San Diego State University, San Diego, CA, USA

[10]Instituto Politecnico Nacional, CIIDIR, Durango, Durango, Mexico

[11]Dorena Genetic Resource Center, USDA Forest Service, Cottage Grove, OR, USA

[12]Rocky Mountain Research Station, USDA Forest Service, Flagstaff, AZ, USA

[13]School of Forestry, Northern Arizona University, Flagstaff, AZ, USA

Correspondence
Mitra Menon, Integrative Life Sciences, Virginia Commonwealth University, Richmond, VA, USA.
Email: menonm2@mymail.vcu.edu
and
Andrew J. Eckert, Department of Biology, Virginia Commonwealth University, Richmond, VA, USA.
Email: aeckert2@vcu.edu

Funding information
USDA Forest Service National Fire Plan; National Science Foundation, Grant/Award Number: EF-1442486, EF-1442597; Northern Arizona University's Technology Research Initiative Program; Virginia Commonwealth University (VCU) Department of Biology and VCU Integrated Life Sciences

Abstract

Interactions between extrinsic factors, such as disruptive selection and intrinsic factors, such as genetic incompatibilities among loci, often contribute to the maintenance of species boundaries. The relative roles of these factors in the establishment of reproductive isolation can be examined using species pairs characterized by gene flow throughout their divergence history. We investigated the process of speciation and the maintenance of species boundaries between *Pinus strobiformis* and *Pinus flexilis*. Utilizing ecological niche modelling, demographic modelling and genomic cline analyses, we illustrated a divergence history with continuous gene flow. Our results supported an abundance of advanced generation hybrids and a lack of loci exhibiting steep transition in allele frequency across the hybrid zone. Additionally, we found evidence for climate-associated variation in the hybrid index and niche divergence between parental species and the hybrid zone. These results are consistent with extrinsic factors, such as climate, being an important isolating mechanism. A build-up of intrinsic incompatibilities and of coadapted gene complexes is also apparent, although these appear to be in the earliest stages of development. This supports previous work in coniferous species demonstrating the importance of extrinsic factors in facilitating speciation. Overall, our findings lend support to the hypothesis that varying strength and direction of selection pressures across the long lifespans of conifers, in combination with their other life history traits, delays the evolution of strong intrinsic incompatibilities.

KEYWORDS
conifers, ecological speciation, extrinsic barriers, hybrid zone, introgression, population genomics

1 | INTRODUCTION

Speciation often occurs along a continuum of divergence such that evolutionary processes leading to species formation initially involve unrestricted gene flow followed by the evolution of reproductive isolation between lineages (Kane et al., 2009; Nosil & Feder, 2012; Roesti, Hendry, Salzburger, & Berner, 2012). Hence, understanding how and when barriers to gene flow arise and are maintained along this continuum is a fundamental goal of evolutionary biology (Losos et al., 2013). Under a model of ecological speciation (Schluter & Conte, 2009), initiation of divergence among populations occurs through disruptive selection leading to the formation of ecotypes. This process results in shifts of allele frequencies correlated with environmental differences between habitats specific to each ecotype. The subsequent transition from ecotypes to reproductively isolated species occurs through the build-up of associations among multiple loci independently experiencing disruptive selection and the action of selection to maintain these coadapted gene complexes (Flaxman, Wacholder, Feder, & Nosil, 2014).

Several studies of speciation have used hybrid zones as windows into the process of divergence between species (reviewed by Petit & Excoffier, 2009). Studies conducted across the entire geographical range of hybridizing species have helped reveal not only the demographic context of speciation, but also the relative importance of intrinsic and extrinsic processes (Ryan et al., 2017; Schield et al., 2017). Specifically, the maintenance of species boundaries has been shown to occur through tension zones (intrinsic incompatibilities *sensu* Barton & Hewitt, 1985; Via, Bouck, & Skillman, 2000; Barton, 2001; Rundle, 2002) and bounded hybrid superiority (extrinsic incompatibilities *sensu* Moore, 1977; Milne, Terzioglu, & Abbott, 2003; Hamilton, Lexer, & Aitken, 2013). The former facilitates divergence through a build-up of genetic incompatibilities among loci causing environmentally independent reduction in hybrid fitness, whereas the latter involves increased hybrid fitness only in an intermediate environment to which the divergent parental allelic combinations confer a putative advantage. These two processes can be coupled, such that genomic regions involved in intrinsic incompatibility coincide with loci exhibiting ecological gradients in allele frequency (Bierne, Welch, Loire, Bonhomme, & David, 2011; Cushman & Landguth, 2016), ensuring the maintenance of species barriers despite the homogenizing effect of gene flow (Kulmuni & Westram, 2017). Thus, the interaction between intrinsic and extrinsic barriers to gene flow generates a genomic mosaic of introgression and differentiation that depends in part upon the demographic context and life history traits of the diverging lineages.

The recent influx of genomic data from nonmodel species has facilitated studies of ecological speciation across varying spatial and temporal scales (Andrew & Rieseberg, 2013; de Lafontaine, Prunier, Gérardi, & Bousquet, 2015; Lackey & Boughman, 2016; Lexer et al., 2010; Marques et al., 2017). The genomic mosaic of introgression noted in these studies has lent support to the genic view of speciation (Wu, 2001). These genomic mosaics can be the result of secondary contact, areas of suppressed recombination, recent divergence without gene flow, allele surfing, sieving of ancestral balanced polymorphisms and selective sweeps specific to each lineage unrelated to the development of reproductive isolation (Cruickshank & Hahn, 2014; Guerrero & Hahn, 2017; Noor & Bennett, 2009). Disentangling these explanations is often complicated because reproductive isolation can progress and be associated with several of these processes, such as with ecological niche partitioning (Agrawal, Feder, & Nosil, 2011).

Species of conifers are known to have ecologically differentiated niches despite the absence of strong morphological differences (e.g., Rehfeldt, 1999). Strong pre- and postzygotic isolating barriers contributing towards morphological disjunctions are often absent in conifers (Buschiazzo, Ritland, Bohlmann, & Ritland, 2012; Critchfield, 1986; Pavy et al., 2012) due to common life history characteristics such as longevity, high dispersal abilities and long generation times (Petit & Hampe, 2006). These contribute towards large effective population sizes and moderate to high levels of genetic diversity, facilitating establishment across an array of ecological conditions. Ecological niche partitioning is thus likely to play a dominant role in facilitating speciation across conifers (e.g., Hamilton et al., 2013).

In this study, we use an integrative approach to investigate processes leading to the divergence of two North American pine species—*Pinus strobiformis* Engelm. (southwestern white pine) and *Pinus flexilis* E. James. (limber pine). Our focal species inhabit a wide latitudinal range in the western part of North America, but display limited differences in morphological and reproductive traits (Benkman, Balda, & Smith, 1984; Bisbee, 2014; Tomback, Samano, Pruett, & Schoettle, 2011). Within a putative area of sympatry, located in the southern Rocky Mountains and Colorado Plateau, morphological evidence points towards the occurrence of hybridization (Steinhoff & Andresen, 1971). To examine the processes influencing species boundaries between these two conifers, we asked three questions: (i) Does the hybrid zone between *P. strobiformis* and *P. flexilis* occupy a niche ecologically divergent from either parent species? (ii) Did the divergence of *P. strobiformis* and *P. flexilis* occur with continual gene flow? (iii) Does a genome-wide mosaic of differentiation characterize divergence between *P. strobiformis* and *P. flexilis* and is this pattern attributed to extrinsic, intrinsic or an interaction of both factors? Our results are consistent with ecological divergence occurring with continual gene flow between the focal species, with several lines of evidence supporting a strong influence of extrinsic factors in reinforcing species boundaries.

2 | MATERIALS AND METHODS

2.1 | Focal taxa and field sampling

Pinus strobiformis and *P. flexilis* are closely related species of white pines that occur in the mountainous areas of western North America. The native range of *P. strobiformis* includes Mexico and the southwestern United States, and its distribution exhibits disjunctions across dry and wet boreal mixed forest ecosystems (Looney & Waring, 2013; Figure 1). *Pinus flexilis* inhabits areas across northern Arizona and northern New Mexico to Alberta, Canada, with a region of putative sympatry with *P. strobiformis* in the southern Rocky Mountains and Colorado Plateau (Figure 1). Across this zone of putative sympatry, cone morphology and dispersal syndromes fall along a continuum of divergence, blending into the characteristics of populations in the allopatric zones of either species (Bisbee, 2014).

We sampled 42 *P. strobiformis* populations encompassing a total of 376 trees (5–13 trees/population) from its entire geographical range. We avoided sampling the southeastern populations of *P. strobiformis*, as this region has been identified as putative hybrid zone with *P. ayacahuite* and trees here have been classified as *P. strobiformis* subspecies *veitchii* (Frankis, 2009). Populations within *P. strobiformis* were classified into “Core” (latitudinal range: 19–30.5°N) and “Periphery” (latitudinal range: 31–33°N), such that Periphery represents the putative hybrid zone between *P. strobiformis* and *P. flexilis*. For *P. flexilis*, a total of 13 populations were sampled, with eight populations sampled from the southern range margin and five sampled closer to the range centre (Figure 1). Across these thirteen populations, we sampled a total of 69 trees (4–10 trees/population). To help minimize relatedness, trees within the same site were sampled with a minimum spacing of 50 m (*P. strobiformis*) and 200 m (*P. flexilis*) from each other.

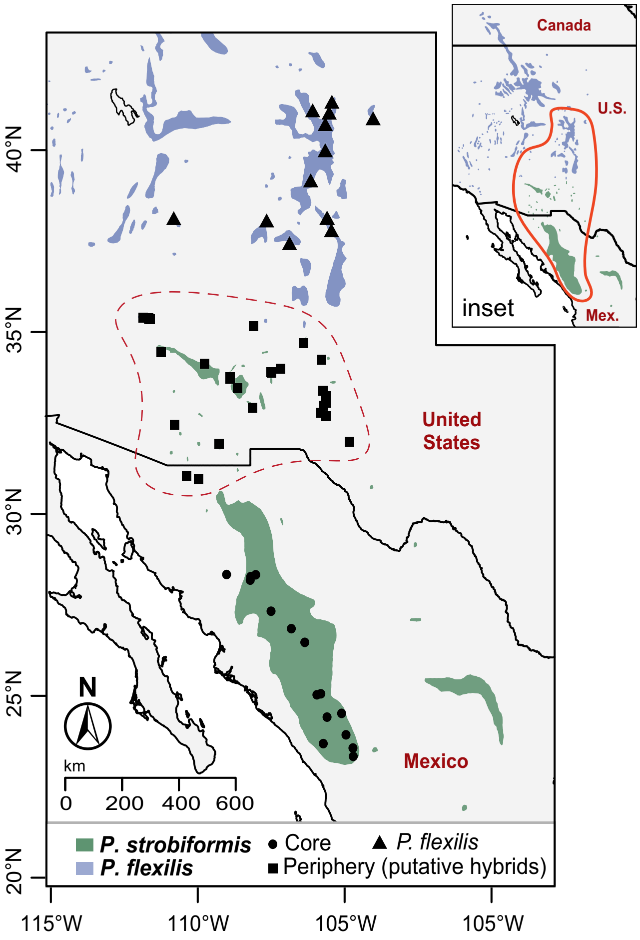

FIGURE 1 Map of sampling localities (black dots) overlaid on polygons showing geographical ranges for *Pinus strobiformis* (green) and *Pinus flexilis* (blue). Peripheral populations (squares) represent the putative hybrid zone. The corresponding locality information is available in Table S2

2.2 | Data generation

2.2.1 | Occurrence data

We assembled a comprehensive data set of occurrences for ecological niche modelling (ENM) by supplementing our field site data with occurrence records downloaded from the Global Biodiversity Information Facility (GBIF), using functions from the dismo package (Hijmans, Phillips, & Elith, 2017) available in the R environment (R Core Team 2017). Using a series of filtering steps to account for observation and sampling biases (Appendix S1A), we obtained a final data set of 254 occurrence records for *P. strobiformis* and 420 for *P. flexilis*. Incorporating intraspecific genetic variation into ENMs can improve model fit and provide more accurate predictions when projecting across time and space (Ikeda et al., 2017; Knowles, Carstens, & Keat, 2007). Thus, we divided presence locations within *P. strobiformis* into the same Core and Periphery groups mentioned above in the “Focal taxa and field sampling” section. These groups likely represent different genetic clusters given the geographically restricted phenotypic evidence of hybridization between *P. flexilis* and *P. strobiformis* (Steinhoff & Andresen, 1971; Bisbee, 2014). We defined three groups that were the focus of our enquiries: (i) populations of *P. flexilis*, (ii) populations of *P. strobiformis* from the northern range periphery (Periphery hereafter) and (iii) populations of *P. strobiformis* from the range core (Core hereafter). Nineteen bioclimatic variables and altitude were used as predictors in the ENMs for all three groups. Present-day geospatial data layers at 30 arc-second resolutions and at 2.5 arc-minute resolutions for the Last Glacial Maximum (LGM) were downloaded from WorldClim v.1.4 (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005). For each of the twenty layers, data were extracted using the raster package (Hijmans et al., 2016) available in r.

2.2.2 | DNA sequence data

We extracted total genomic DNA from 445 individuals sampled across 55 populations of both species using DNeasy Plant Kits (Qiagen). Five ddRADseq libraries (Peterson, Weber, Kay, Fisher, & Hoekstra, 2012), each containing up to 96 multiplexed samples, were prepared using the procedure detailed in Parchman et al. (2012). All libraries were digested using the *EcoR1* and *Mse1* restriction enzymes followed by ligation of adaptors and barcodes. After PCR, we selected DNA fragments in the 300–400 bp size range using agarose gel electrophoresis and we isolated the pooled DNA using a QIAquick Gel Extraction Kit (Qiagen). Single-end sequencing, with one multiplexed library per lane, was used to obtain 105 bp reads, with all sequencing conducted on an Illumina HiSeq 2500 at the Nucleic Acids Research Facility located at Virginia Commonwealth University. The resulting fastq files were processed using the dDocent bioinformatics pipeline (Appendix S1B; Puritz, Hollenbeck, & Gold, 2014). The entire process yielded a total of 51,633 single nucleotide polymorphisms (SNPs), which were used as the starting data set for all subsequent analyses.

2.3 | Data analysis

2.3.1 | Ecological niche modelling and niche divergence

We developed ENMs for each of the following groups: Core, Periphery and *P. flexilis*, using algorithms available in the maximum entropy software program, MaxEnt (Phillips, Anderson, & Schapire, 2006). As MaxEnt was specifically developed for presence-only data, we drew a one-degree rectangular buffer around the known distribution of both species and obtained 100,000 background points at random without duplicates. Data processing, model fitting and model evaluation using 5,000 iterations within MaxEnt were conducted using the dismo, raster, rgdal (Bivand, Keitt, Rowlingson, & Pebesma, 2017) and spThin (Aiello-Lammens, Boria, Radosavljevic, Vilela, & Anderson, 2015) packages available in r. ENMs were constructed from climate variables with an absolute correlation coefficient (r) < .85 (Table S1). Two indices were used to assess model performance for each group: overall regularized training gain (RTG) and area under the curve (AUC). As LGM data were not available at 30 arc-seconds resolution, we built two ENMs for each group (2.5 arc-minutes and 30 arc-seconds), but only used the 2.5 arc-minutes models for hindcasting to infer historical patterns of sympatry between species that could facilitate gene flow. We followed an average projection ensemble approach across three LGM scenarios (CCSM4, MIROC and MPI) to obtain a hindcasted suitability map. Changes in habitat suitability (stability) were assessed by adding MaxEnt-predicted suitability maps across the LGM and present (as in Ortego, Noguerales, Gugger, & Sork, 2015). For these maps, values closer to 2 in a gridded cell are associated with the stability of highly suitable habitat for a given group across time points. In contrast, values closer to 0 are associated with the stability of highly unsuitable habitat for a given group across time points. Suitability scores across the full geographical extent for present conditions at 30 arc-seconds were obtained for all three groups delineated in our study. To investigate patterns of niche evolution, we conducted pairwise comparisons of these suitability scores. We accounted for potential biases towards niche divergence introduced by latitudinally associated environmental variation in the present range of each pair, by performing asymmetric background randomization test, based on Schoener's D, in the r package ENMTools (Warren, Glor, & Turelli, 2008). The two resulting null distributions obtained through this test correspond to the background level of niche divergence for each pair. An observed value of Schoener's D much smaller than expected after accounting for background differences could indicate niche divergence, whereas a value much larger than expected indicates niche conservatism (Warren et al., 2008).

2.3.2 | Population structure and demographic modelling

We assessed the pattern and extent of genetic divergence between *P. strobiformis* and *P. flexilis* using multiple methods. First, we grouped the 42 *P. strobiformis* populations into the same Core and Periphery groups described above (see Data Generation & Figure 1). We conducted principal components analysis (PCA) to visualize grouping of sampled trees into the three groups delineated in our methods (McVean, 2009; Patterson, Price, & Reich, 2006). To complement the PCA, we also conducted an individual-based assignment test using fastSTRUCTURE (Raj, Stephens, & Pritchard, 2014). We set the number of clusters (K) to 2, representing the two parental species investigated here, as we were interested in admixture between two defined species and not the potential number of genetic groups. Lastly, we utilized hierarchical fixation indices (F-statistics) to assess the extent of differentiation between species by nesting trees into populations and populations into species. There are two levels within the hierarchy, with F_{CT} describing differentiation among groups at the highest level of the hierarchy and F_{ST} describing differentiation among groups across all levels of the hierarchy (Yang, 1998). A similar nested model with the highest level of hierarchy being groups within *P. strobiformis* (Core and Periphery) was used to assess intraspecific differentiation. For the former, F-statistics are denoted using the term "species" in the subscript, whereas the latter uses the term "groups" in the subscript. We used a similar hierarchical model with variance partitioning to estimate group specific and pairwise F-statistics for the three groups delineated in this study. We denote pairwise values of F_{ST} using one-letter abbreviations for the groups being compared (e.g., $F_{ST\text{-}CP}$ indicates F_{ST} between Core and Periphery) and group-specific values of F_{ST} with the name of the group in subscript. We constructed 95% confidence intervals of multilocus *F*-statistics using bootstrap resampling ($n = 100$ replicates) in the HIERFSTAT package (Goudet, 2005) available in R. Along with estimation of *F*-statistics, we also assessed overall levels of genetic diversity using multilocus estimates (i.e., means across SNPs) of observed and expected heterozygosities (H_o and H_e, respectively) per population.

Presence of individuals with mixed ancestry, as identified using FASTSTRUCTURE, can be a result of secondary contact, incomplete lineage sorting or the presence of gene flow throughout the divergence history. Disentangling these explanations is important, because it directly influences our understanding of the relative importance of intrinsic and extrinsic factors in facilitating speciation. For instance, when speciation is recent or has occurred with gene flow, we expect to see islands of divergence around regions experiencing strong intrinsic or extrinsic selection (Feder, Egan, & Nosil, 2012; Wu, 2001). However, if hybrids are formed in areas with novel habitats, introgression might be selectively advantageous causing the absence of such islands. To infer the timing and influence of various demographic processes shaping the divergence history of our focal groups, we conducted demographic modelling using Diffusion Approximation for Demographic Inference (∂A∂I v.1.7; Gutenkunst, Hernandez, Williamson, & Bustamante, 2009). We down-sampled the total SNP data set for computational simplicity based on population genetic summary statistics and then randomly sampled one SNP per assembled contig to obtain a final data set of 6,330 SNPs (Appendix S1C).

We compared a model of pure divergence with no gene flow (M_1) against a set of 10 alternative demographic models (M_2–M_7) representing different speciation scenarios including varying timing and directionality of ancient or contemporary gene flow (Figure S1). Complexity was added to the models with gene flow by incorporating heterogeneity in the gene flow parameter across loci (Tine et al., 2014; models M_8–M_{11}, Figure S1), which served as a test for islands of divergence. We ran 10 replicate runs of each model in ∂A∂I, using a 200 × 220 × 240 grid space and the nonlinear Broyden–Fletcher–Goldfarb–Shannon (BFGS) optimization routine. Following Carstens et al. (2013), we conducted model selection in an information-theoretic framework using Akaike information criterion (AIC; Akaike 1974) and ΔAIC ($\text{AIC}_{\text{model } i} - \text{AIC}_{\text{best model}}$) scores (Burnham & Anderson, 2002), calculated using results from the best replicate run (highest composite likelihood) for each model. We performed Fisher Information Matrix (FIM)-based uncertainty analysis on the best-supported model to obtain upper and lower 95% confidence intervals (CIs) for all parameters (Appendix S1D). Unscaled parameter estimates and their 95% CIs, were obtained using a per-lineage substitution rate of 7.28×10^{-10} substitutions/site/year rate estimated for Pinaceae by De La Torre, Li, Van de Peer, and Ingvarsson (2017) and a generation time of 50 years.

2.3.3 | Genomics of interspecific introgression

Analyses of clines across hybrid zones are widely used to identify loci exhibiting exceptional patterns of introgression relative to the average genomic background (Fitzpatrick, 2013; Gompert & Buerkle, 2011; Gompert, Parchman, & Buerkle, 2012; Stankowski, Sobel, & Streisfeld, 2015). We classified our sampled trees into categories corresponding to admixed ($n_A = 111$) and parental species (*P. strobiformis* = 277, *P. flexilis* = 54) based on the *Q*-values from FASTSTRUCTURE. Trees with *Q*-values of 0.9 or higher were classified as pure *P. strobiformis*, those with a *Q* of 0.10 or lower were classified as pure *P. flexilis* and those with intermediate *Q*-values were classified as admixed (e.g., Ortego, Gugger, Riordan, & Sork, 2014). As most loci exhibited little to no differentiation between parental species, we retained only loci with a minor allele frequency (MAF) difference of at least 10% between parental species ($n = 4{,}857$ SNPs). This allowed us to avoid false correlations between cline parameters and fixation indices (Parchman et al., 2013). We used this subset of 4,857 SNPs to perform a Bayesian genomic cline analysis in BGC v1.0 (Gompert & Buerkle, 2011, 2012). Using Markov chain Monte Carlo (MCMC) sampling, BGC estimates the posterior distribution of ancestry for each locus as a function of the genome-wide admixture coefficient. The BGC model includes two genomic cline parameters, α (genomic cline centre) and β (genomic cline rate, i.e., slope), determining the probability of *P. flexilis* ancestry and the rate of transition from *P. flexilis* to *P. strobiformis* given a level of genomic admixture described by the hybrid index, *h*, respectively (Gompert & Buerkle, 2012; Gompert, Parchman et al., 2012). A tree with $h = 0$ was classified as having solely *P. strobiformis* ancestry, whereas a tree with $h = 1$ was classified as having solely *P. flexilis* ancestry. We ran BGC for five replicate runs, each 45,000 steps in length, and, after discarding the first 25,000 steps as burn-in, we thinned the posterior distribution every 20 steps, thus yielding 1,000 samples which were used for inference of model parameters. We used TRACER v1.6 (Rambaut, Suchard, Xie, & Drummond, 2013) to test for convergence among replicated runs, as well as appropriate mixing along MCMC chains. We identified excess ancestry loci (relative to the genome-wide average) as those with posterior α or β credible intervals (CrI; 95% equal-tail intervals) not containing zero. We identified outlier loci as those with posterior mean point estimates of α ($\hat{\alpha}$) or β ($\hat{\beta}$) significantly different from the rest of the genome, as judged by comparison to posterior quantiles of random-effect priors for α and β (Gompert, Parchman et al., 2012). Besides categorizing loci as excess ancestry or outlier, we also tested for correlations among locus-specific $F_{\text{CT-species}}$, α and β, with and without absolute values for α and β. The sign of the cline parameters (specifically β) has direct implications for inferring the processes maintaining species boundaries and hence was incorporated in correlation tests. Specifically, extremely positive values of β reflect strong selection against hybrids or population structure in the hybrid zone (Gompert, Lucas et al., 2012), while extremely negative values of β indicate a wide cline representing easy dispersal across species boundaries (Janoušek et al., 2012).

Although the hybrid index (*h*) obtained from BGC provides information about the age and stability of a hybrid zone, such inferences are limited to only one generation of admixture (Fitzpatrick, 2012). We estimated *h* and interspecific heterozygosity using INTROGRESS

(Gompert & Buerkle, 2010), to extend our interpretations to a historical hybrid zone and categorize individuals into recent (F1s), advanced generation (FNs) and backcrossed hybrids (BCs). This was carried out using a modified classification from Hamilton et al. (2013). Both BGC and INTROGRESS yielded very similar estimates of h (Pearson's $r = .70$, $p < .001$); thus, we used estimates from INTROGRESS due to the availability of interspecific heterozygosity estimates from this software. To test for the influence of extrinsic factors in the maintenance of species boundaries, we performed linear regression analyses with backward variable selection using h against climate and geography as predictor variables. This was carried out using the reduced set of climate variables from the final ENMs (see Table S1).

3 | RESULTS

3.1 | Ecological niche modelling and niche divergence

Ecological niche models for each of the three groups used in this study (Figure 1) had high predictive ability, as indicated by AUC and RTG values (Table 1). For Core and Periphery, several covariates were important, with precipitation seasonality (Bio15) shared between Core and Periphery. For *P. flexilis*, altitude was consistently the most important variable across different measures of variable importance (Table 1). Hindcasting the 2.5 arc-minute model onto LGM data layers supported a recent, post-LGM niche fragmentation and northward expansion in Periphery (Figure S2). A similar post-LGM northward expansion of suitable niche space was observed for *P. flexilis*. Furthermore, there was extensive range overlap between the two species during the LGM, which was greater than what is currently observed (Figure S2). Values of niche similarity based on Schoener's *D* ranged from 0.05 (*P. flexilis* – Core) to 0.17 (Periphery – Core). Background randomization tests revealed statistically significant niche divergence for two of the three comparisons (Figure 2). For the third comparison, however, niche divergence was asymmetrical between Core and Periphery, with the niche of Periphery being conserved relative to the background of Core (Figure 2a). A similar pattern was noted using only the presence points, where each group formed a distinct cluster within the multivariate climate space defined by the top two principal components (PCs) derived from PCA on the climate variables used for construction of the ENMs (Figure S3A).

3.2 | Population structure and divergence history

The PCA using 51,633 SNPs was consistent with trees sampled from Core being differentiated from those of *P. flexilis*, which was most marked along PC1 (Figure 3a). This PC explained 0.90% of the total genetic variance, which was in line with the overall level of differentiation estimated using hierarchical *F*-statistics ($F_{\text{ST-species}} = 0.021$, 95% CI: 0.008–0.031). Trees sampled from Periphery were located between those sampled from Core and *P. flexilis* (Figure 3a), in line with Periphery containing hybrids. There was also a latitudinal gradient in the mean population *Q*-values, as estimated using FASTSTRUCTURE, with Core populations exhibiting little to no ancestry from *P. flexilis* and Periphery being a mixture of *P. flexilis* and Core (Figure 3b). At the individual tree level, we observed a strong negative correlation (Pearson's $r = -.69$, $p < .001$) between *Q*-values of putative hybrids and latitude, which is consistent with a geographical gradient of genomic introgression, such that trees geographically proximal to either parental species contain more ancestry from that parental species. The multilocus estimate of differentiation between species ($F_{\text{CT-species}}$) was 0.01 (95% CI: 0.005–0.018, Figure 4a), while that between groups within *P. strobiformis* ($F_{\text{CT-groups}}$) was 0.003 (95% CI: 0.0007–0.006). Group-specific multilocus F_{ST}, pairwise F_{ST} and heterozygosities differed little among the three groups, with the Core – *P. flexilis* comparison having the highest pairwise $F_{\text{ST-CF}} = 0.019$ (Table 2). Although populations of Periphery exhibited slightly higher heterozygosities and F_{ST} values ($F_{\text{ST-periphery}}$), this pattern was mainly driven by few populations, as indicated by the wider confidence interval around these estimates (Table 2).

The best-supported demographic model was M_4, which is a model of symmetric ancient gene flow between the ancestral *P. strobiformis* and *P. flexilis* lineages, followed by contemporary gene flow between Periphery and *P. flexilis* (Table 3; Figure 4). This model was supported by a large AIC margin of 44.8 information units ($\Delta\text{AIC}_i \geq 44.8$). Converted parameter estimates indicated that the species diverged 18.04 million years ago (Ma) in the Miocene (95% CI: 26.29–9.79 Ma), but that the two groups within *P. strobiformis* diverged 3.63 Ma during the Pliocene (95% CI: 4.44–2.83 Ma) (Figure 5; Table S3). Overall rates of gene flow between species were substantial for both historical and contemporary periods; however, contemporary gene flow between species was geographically restricted to Periphery and *P. flexilis* (Table S3). In addition, *P. flexilis* and Periphery experienced asymmetrical gene flow for which point estimates were larger in the direction of Periphery to *P. flexilis* ($M_{\text{FP}} = 11.53$ migrants/generation with a 95% CI: 0–57.94 vs. $M_{\text{PF}} = 8.80$ with a 95% CI: 0–12.84). Periphery had the largest population size estimate, while *P. flexilis* was inferred to have experienced an approximately 60% reduction in population size through time.

3.3 | Genomics of interspecific introgression

Hybrid index (*h*) values ranged from near zero to 0.80, with values around 0.20 being the most common, thus suggesting overrepresentation of *P. strobiformis* ancestry (Figure 4b). Estimates of interspecific heterozygosity had a narrow range from 0.45 to 0.64, indicating weak reproductive barriers (Hamilton et al., 2013) and a long history of recombination within the hybrid zone (Gompert et al., 2014). Classification of trees into genotypic classes based on *h* and interspecific heterozygosity revealed a dominance of advanced generation hybrids (54%), with some trees being backcrossed into *P. strobiformis* (22%). No recent hybrids (F1s) were apparent. Stepwise linear regression analysis revealed a significant effect of geography and climate on *h* across the putative hybrid zone. Latitude

TABLE 1 Ecological niche model performance and variable importance at 30 arc-second resolution

Groups	AUC	RTG	RTG importance[a]	Permutation importance[b]	Per cent contribution[b]	Regression coefficient importance[c]
Core	0.97	2.51	Bio15[d], Bio4[e]	Bio4	Altitude	Bio4
Periphery	0.99	3.92	Bio9[f]	Bio10[g], Bio9, Bio6[h]	Altitude	Bio15
Pinus flexilis	0.94	1.72	Altitude	Altitude	Altitude	Altitude

AUC, area under the curve; RTG, regularized training gain.

[a] Variables that caused maximum reduction in the total RTG when omitted from the model and the variable with the most contribution to RTG.

[b] Variables with the highest permutation or percentage importance.

[c] Sum of absolute values of regression coefficient (λ) across various predictor transformations or feature classes used in MAXENT.

[d] Precipitation seasonality.

[e] Temperature seasonality.

[f] Mean temperature of the driest quarter.

[g] Mean temperature of the warmest quarter.

[h] Minimum temperature of the coldest month.

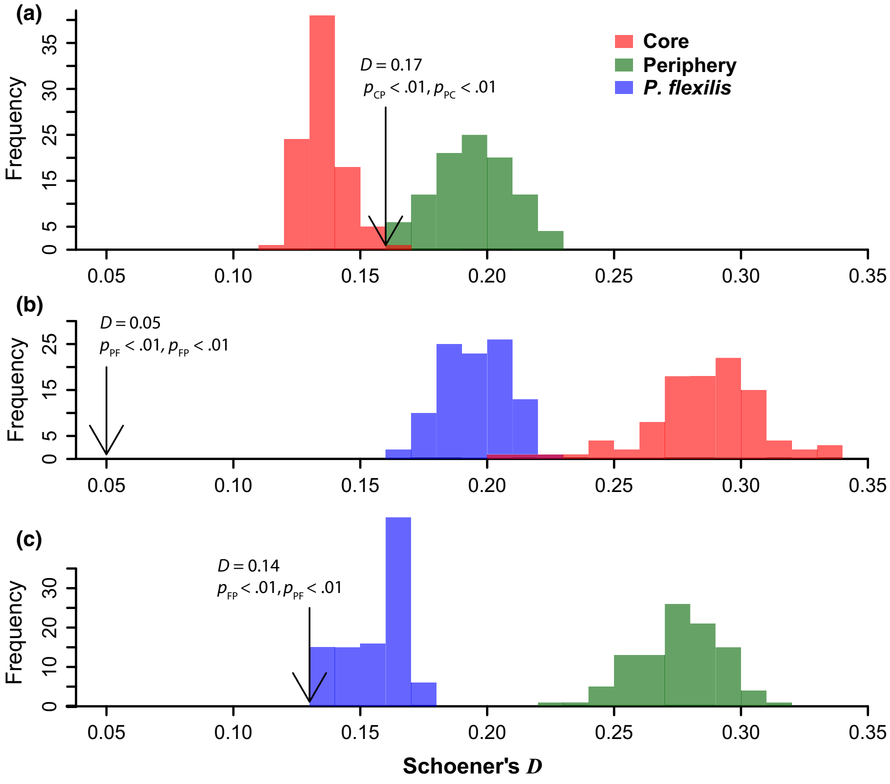

FIGURE 2 Results of niche divergence tests (Schoener's *D*) for a) Core vs. Periphery b) Core vs. *Pinus flexilis* and c) *P. flexilis* vs. Periphery. Histograms indicate the background levels of niche divergence, and arrows indicate the observed value of Schoener's *D* for each pair compared

(Pearson's $r = .41$, $p < .001$), precipitation seasonality (Pearson's $r = -.32$, $p < .01$) and mean temperature of the warmest quarter (Pearson's $r = -.18$, $p < .01$) had a strong influence on *h*, in line with the latter two being important predictor variables for Periphery in our ENM.

Substantial variation was found in estimates of genomic cline parameters (Figure 4c,d), especially for α, with its range (−0.99 to 1.72) being 18.5-fold as wide as that of β (−0.068 to 0.078). Similar to the patterns observed in the distribution of *h*, an asymmetry towards *P. strobiformis* ancestry was noted in the genomic cline estimates. From the posterior distribution of α, we found 3,193 outlier loci, of which 570 (17.9%) had elevated probabilities of *P. flexilis* ancestry (positive $\hat{\alpha}$), and 2,623 (82.1%) had elevated probabilities of *P. strobiformis* ancestry (negative $\hat{\alpha}$). We identified fewer loci with excess ancestry, but in contrast to the pattern for outlier loci, those with excess ancestry favoured *P. flexilis* over *P. strobiformis* ancestry. Among the 287 loci with excess ancestry, 204 (71.1%) had excess *P. flexilis* ancestry (i.e., lower 95% CrI of $\alpha > 0$) and 83 (28.9%) had excess *P. strobiformis* ancestry (i.e., upper 95% CrI of $\alpha < 0$). The multilocus $F_{\text{CT-species}}$ estimate for loci with excess

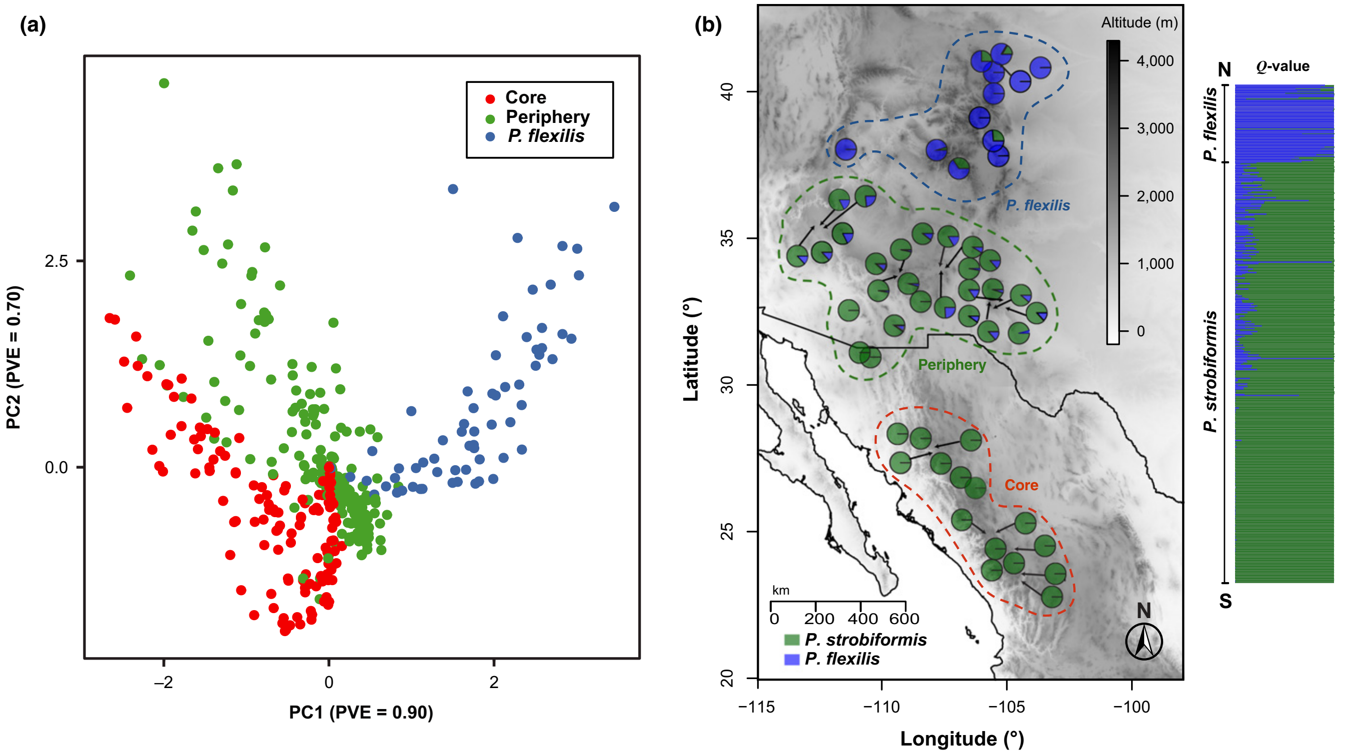

FIGURE 3 (a) Results of population genetic structure analysis using PCA on 51,633 single nucleotide polymorphisms. (b) Results of assignment analyses for each tree in FASTSTRUCTURE for $K = 2$ clusters (right panel) plotted onto a topographic map of the study area (left panel). Each pie chart represents the average ancestry of a population from *Pinus strobiformis* and *Pinus flexilis*

ancestry was 0.12 (95% CrI: 0.09–0.13), while for outlier loci, it was 0.058 (95% CrI: 0.05–0.09). We did not identify any loci that were β outliers or had excess ancestry indicated by β. Hierarchical $F_{\text{CT-species}}$ was negatively correlated with raw values of α (Pearson's $r = -.036$, $p = .01$), positively with raw values of β (Pearson's $r = .048$, $p < .001$) and positively with absolute values of both α (Pearson's $r = .14$, $p < .001$) and β (Pearson's $r = .26$, $p < .001$) (Figure 4e).

4 | DISCUSSION

We identified strong evidence supporting ecological divergence with gene flow between *P. strobiformis* and *P. flexilis*. Our findings are generally consistent with previous reports on the species examined here; however, in contrast to the recent divergence time estimated by Moreno-Letelier, Ortíz-Medrano, and Piñero (2013), our demographic modelling is consistent with deeper divergence, as well as ongoing speciation with gene flow, that is driven and maintained primarily by extrinsic factors. The latter was made possible by explicitly accounting for hybridization as a confounding and contributing factor to local adaptation and speciation.

4.1 | Niche evolution and ecological divergence

Our results indicate that climatic factors have played a major role in driving niche divergence between *P. strobiformis* and *P. flexilis*. Populations within Periphery coincide with the known phenotypic hybrid zone between *P. strobiformis* and *P. flexilis* (Steinhoff & Andresen, 1971; Bisbee, 2014) and formed a distinct group characterized by niche divergence from *P. flexilis* and asymmetrical niche divergence from Core. The asymmetrical pattern of niche divergence between Core and Periphery is likely the result of recent divergence. Under this scenario, we expect niche differentiation to occur primarily along a few environmental variables that strongly influence fitness in the transitional environmental conditions, with little to no differentiation among groups along other environmental axes (Figure S3C). In support of this expectation, precipitation seasonality was an important niche predictor for both Core and Periphery, but they were differentiated along this environmental axis (Figure S3B). While several other bioclimatic variables exhibited as large a difference as precipitation seasonality, they did not significantly contribute towards the niche of both Core and Periphery (Table 1; Figure S3C). These patterns reiterate the presence of hybrid populations in transitional environmental conditions, experiencing early stages of niche divergence.

In line with these results, precipitation seasonality and mean temperature of the warmest quarter had a strong negative association with genomic ancestry and contributed to the niche divergence of Periphery. These two climatic variables influence plant evapotranspiration and affect drought responses (Mishra & Singh, 2010). Drought stress during the active growing season is widely recognized as a limiting factor to plant growth in the western parts of North America (Restaino, Peterson, & Littell, 2016; Williams et al., 2010), and our results are indicative of adaptive divergence along a drought gradient between the three groups (Allen & Breshears,

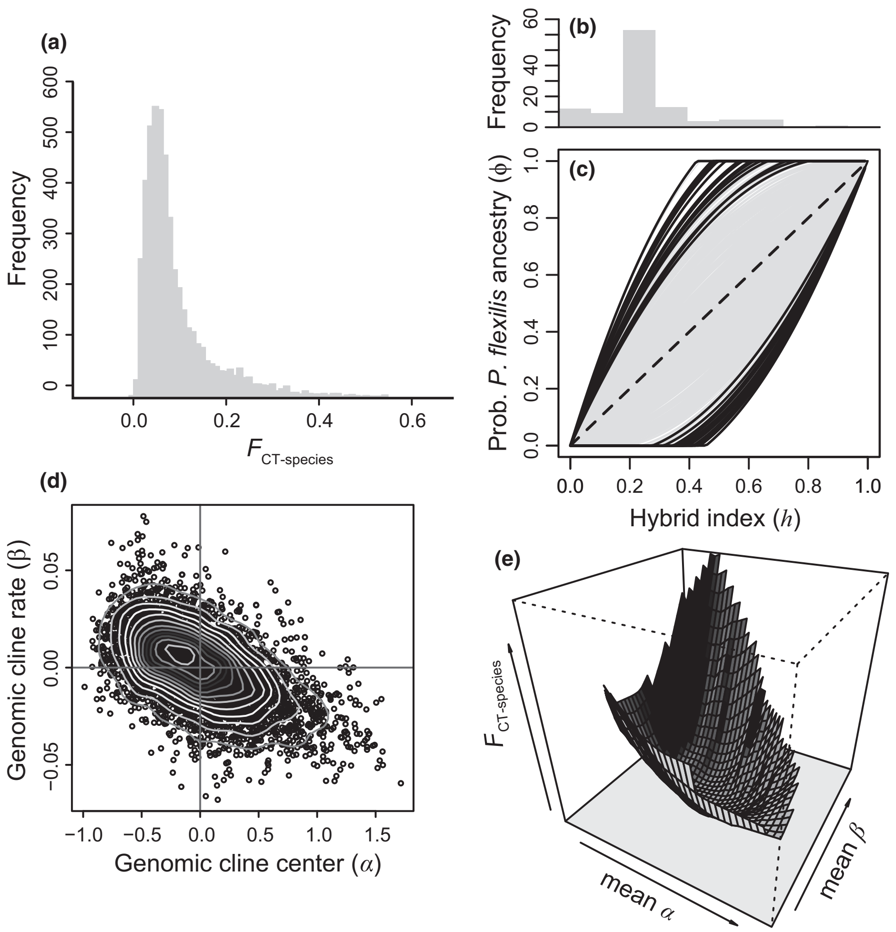

FIGURE 4 (a) Genomic distribution of F_{CT}, (b) frequency distribution of hybrid index, (c) variation in genomic ancestry as a function of hybrid index, (d) correlation between genomic cline parameters and (e) 3D correlation plot of genomic cline parameters and F_{CT}

TABLE 2 Estimates of genetic diversity and divergence within and across the three groups, compared to a genome-wide $F_{\text{ST-species}}$ of 0.02 (95% CI: 0.008–0.03) and $F_{\text{ST-strobiformis}}$ of 0.009 (95% CI: 0.007–0.014)

Group	Multilocus F_{ST} (95% CI)	Pairwise F_{ST} (95% CI)	Mean H_e ± *SD*	Mean H_o ± *SD*
Core	0.003 (0.0025–0.0034)	Periphery: 0.009 (0.001–0.023) *Pinus flexilis*: 0.019 (0.006–0.032)	0.135 ± 0.01	0.111 ± 0.01
Periphery	0.007 (0.0071–0.0073)	*P. flexilis*: 0.015 (0.005–0.024) Core: 0.009 (0.001–0.023)	0.133 ± 0.02	0.105 ± 0.03
P. flexilis	0.003 (0.0025–0.0041)	Core: 0.019 (0.006–0.032) Periphery: 0.015 (0.005–0.024)	0.130 ± 0.01	0.111 ± 0.01

1998; Gitlin et al., 2006). Further, our study broadly agrees with other reports in *P. strobiformis* that indicate precipitation and altitude as important niche predictors (Aguirre-Gutiérrez, Serna-Chavez, Villalobos-Arambula, Pérez de la Rosa, & Raes, 2015; Shirk et al., 2017). Soil and vegetation variables used in previous ENMs, however, were not included in our analyses due to a lack of comparable data for *P. flexilis* and its unclear relationship with divergence history.

Despite fluctuations in suitable range size (Figure S2) and previous studies indicating reduction in genetic diversity at range margins using chloroplast markers (Moreno-Letelier & Piñero, 2009), we find no evidence for this in our study. This could be explained by the asymmetry in gene flow between Periphery and *P. flexilis*, as inferred from the demographic modelling results (Bridle & Vines, 2007; Ortego et al., 2014). Evidence of directional introgression from *P. flexilis* (positive α outliers), moreover, might also have facilitated adaptation to transitional environmental conditions. Such novel allelic combinations have often contributed to the ability of populations to colonize new niches that are intermediate but beyond the climatic conditions experienced by the parental species (De Carvalho et al., 2010; De La Torre, Wang, Jaquish, & Aitken, 2014; Geraldes et al., 2014; Hamilton et al., 2013). Presence of a locally adapted and historical hybrid zone is supported by the absence of β outliers in our genomic cline results (Fouet, Kamdem, Gamez, & White, 2017), as well as by a recent study identifying high Q_{ST} values associated with physiological traits primarily linked to drought tolerance within

TABLE 3 Model composite likelihoods and AIC model selection results for 11 alternative demographic models of *Pinus strobiformis* (Core and Periphery)–*Pinus flexilis* divergence. Results for the best-supported model are underlined, and the two best models are shown in boldface

Model	Model description	ln Composite likelihood	k	AIC	ΔAIC_i
M_1	Strict isolation, no gene flow	−883.143112	6	1,778.29	65.44
M_2	Secondary contact (Periphery–*P. flexilis*)	−886.227416	7	1,786.45	73.60
M_3	Ancient gene flow (speciation with gene flow)	−888.003307	7	1,790.01	77.16
M_4	**Ancient gene flow, plus Periphery–*P. flexilis* gene flow**	**−847.424540**	**9**	**1,712.85**	**0.00**
M_5	Ancient gene flow, plus Core–Periphery gene flow	−885.428135	9	1,788.86	76.01
M_6	Secondary contact (Periphery–*P. flexilis*) and Core–Periphery gene flow	−883.949484	10	1,787.90	75.05
M_7	Ancient gene flow, followed by Periphery–*P. flexilis* gene flow and Core–Periphery gene flow	−892.210862	9	1,806.42	93.57
M_8	**Heterogeneous ancient gene flow**	**−869.824520**	**14**	**1,757.65**	**44.80**
M_9	Heterogeneous ancient gene flow, plus Core–Periphery gene flow	−884.511096	11	1,791.02	78.17
M_{10}	Heterogeneous gene flow during secondary contact (Periphery–*P. flexilis*), and Core-Periphery gene flow	−902.279445	9	1,828.56	115.71
M_{11}	Heterogeneous ancient gene flow, followed by heterogeneous gene flow between Periphery – *P. flexilis* and between Core–Periphery	−922.814525	11	1,873.63	160.78

AIC, Akaike information criterion; k, the number of parameters in the model; ln, natural logarithm.

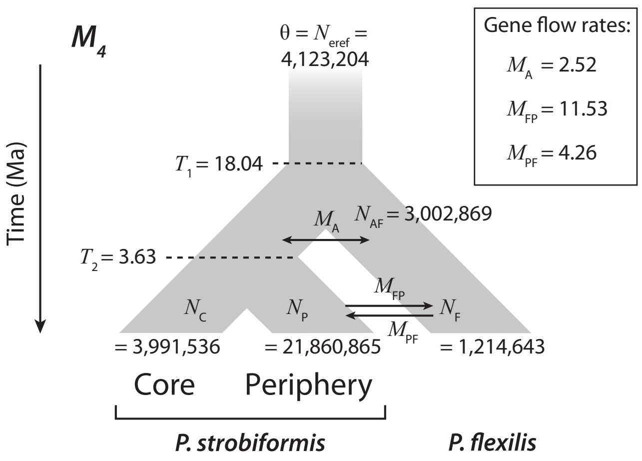

FIGURE 5 The best-supported model from ∂A∂I analysis. This figure shows the parameter estimates for divergence times (T_i) in units of millions of years ago (Ma), reference effective population size (θ; or after conversion, N_{eref}), lineage population sizes (N_i) and rates of gene flow (M_{ij}) for the optimal model determined by Akaike information criterion (AIC) model selection (see Table 3)

Periphery (Goodrich, Waring, & Kolb, 2016). The geographic cline in *h*, asymmetry in excess ancestry loci towards *P. flexilis* and elevated estimates of $F_{ST\text{–periphery}}$, however, indicate the potential for geographically driven neutral introgression to generate biased signals of local adaptation within the peripheral populations (Geraldes et al., 2014). Ongoing investigations using replicate populations in the hybrid zone across gradients of geographic proximity and climate similarity will be able to address this issue in further detail (Lotterhos & Whitlock, 2015; Riquet et al., 2017).

4.2 | Speciation with gene flow without islands of divergence

Demographic modelling indicated that divergence of *P. strobiformis* and *P. flexilis* is not recent (~18 Ma) on an absolute time scale and has occurred with continuous gene flow. The presence of continual gene flow and absence of a period of allopatry, moreover, are also supported by the L-shaped distribution of $F_{CT\text{-species}}$ values (Figure 4a; Nosil & Feder, 2012). Reduction in overlapping niche suitability from LGM to present, between *P. strobiformis* and *P. flexilis*, agrees with the best-supported demographic model indicating continuous but geographically restricted contemporary gene flow. Contemporary reduction in N_e for *P. flexilis* from our demographic modelling is contrary to the predicted post-LGM expansion of suitable habitat. This is likely due to the limited geographical sampling within *P. flexilis* for our genomic analyses or a nonlinear relationship between habitat suitability and realized population sizes. Specifically, due to the geographical bias in the sampling scheme, we were unable to account for further population structure within *P. flexilis*. This may also have biased our inference of gene flow, such that contemporary gene flow between the two species is restricted to geographically proximal genetic groups. However, the primary focus of our study was estimating whether or not divergence occurred with gene flow, which is unlikely to be influenced by sampling biases of this form. Further, based on results from the hindcasted niche models, the extensively sampled southeastern region of *P. flexilis* forms a putative refugium likely representing much of the diversity in southern *P. flexilis* that then expanded northward after the LGM. Thus, regardless of the geographical bias in our sampling scheme, we are likely to have captured a sizeable fraction of the segregating variation within *P. flexilis*.

Despite the potential for islands of divergence under a model of speciation with gene flow (Feder et al., 2012; Nosil, 2008; Tine et al., 2014), as well as niche divergence results consistent with ecological speciation with gene flow between *P. strobiformis* and *P. flexilis*, the best-supported demographic model did not provide evidence for islands of divergence. The absence of elevated islands of divergence in this study, however, does not necessarily indicate an absence of adaptive divergence during speciation with gene flow. Islands of divergence are often expected only under certain genetic architectures and selection scenarios which have been shown to be less prevalent in conifers (Alberto et al., 2013; Lind et al., 2017; Pritchard & Di Rienzo, 2010; Rajora, Eckert, & Zinck, 2016). Alternatively, given the large and complex genomes of conifers (reviewed by De La Torre, Birol et al., 2014), our ddRADseq markers likely underrepresented genic regions, which are often identified as islands of divergence (Marques et al., 2017; Moreno-Letelier & Barraclough, 2015; Nosil & Feder, 2012; Zhou, Zhang, Liu, Wu, & Savolainen, 2014). For example, Moreno-Letelier and Barraclough (2015) demonstrated the potential for islands of divergence at drought-associated genes, which had a high average F_{ST} of 0.33 (0.09–0.40) as compared to the genome-wide estimate from this study ($F_{ST\text{-}species} = 0.02$). Future investigations using exome capture might thus be able to identify islands of divergence, although evidence of adaptation in complex genomes often also appears within intergenic regions (Li et al., 2012), and islands of divergence are not always reflective of speciation genes *sensu stricto* (see Guerrero & Hahn, 2017).

4.3 | Genomic mosaic of introgression

The spatial context of loci within genomes, as well as the temporal scale of divergence between lineages, can influence patterns of introgression and is often depicted by a mosaic landscape of genomic differentiation and ancestry. For instance, Coyne and Orr (1989), Noor and Bennett (2009) and Christe et al. (2017) have all argued that islands of divergence tend to accumulate around regions of reduced recombination such as centromeres and inversions. Extrinsic factors, such as disruptive selection, can also restrict gene flow, but under the observed demographic scenario, these alone are unlikely to generate islands of divergence (Yeaman, Aeschbacher, & Bürger, 2016; Yeaman & Otto, 2011). Extrinsic barriers, however, can often result in the evolution of intrinsic barriers and subsequently become coupled with them, as well as with other loci experiencing similar selection pressures (Agrawal et al., 2011; Flaxman et al., 2014). Thus, given sufficient time, even under a model of speciation with gene flow, such coupling effects will ensure the maintenance of species boundaries relative to the action of either factor alone (Barton & De Cara, 2009). Specifically, in our focal species, previous work using candidate genes for drought stress provides evidence for divergent selection driving speciation, despite low genome-wide levels of differentiation (Moreno-Letelier & Barraclough, 2015). Although a thorough examination of exome-wide variation remains to be carried out, the correlation of *h* with drought related variables when coupled with the work of Moreno-Letelier and Barraclough (2015) implies that adaptive responses to drought stress likely contributed to the origin and maintenance of species boundaries in this system.

A positive correlation between the steepness of genomic clines (β) and F_{CT} points towards coincidence of loci involved in disruptive selection and those involved in reproductive isolation. Such a positive association has been demonstrated across several taxa (*cf*. Gompert et al., 2014; Janoušek et al., 2012; Parchman et al., 2013; Ryan et al., 2017), and we suggest it to be indicative of disruptive selection driving the evolution of intrinsic barriers and its coupling with extrinsic processes. Several empirical and simulation-based studies have demonstrated that both α and β can reflect patterns of selection in the hybrid zone (Gompert, Lucas et al., 2012; Janoušek et al., 2012), but the interpretation of these values is influenced by the underlying demographic scenario (Gompert & Buerkle, 2012; Gompert, Parchman, & Buerkle, 2012; Gompert, Lucas et al., 2012). Under the observed demographic scenario of ongoing gene flow, signatures of selection against hybrids (i.e., underdominance) would be reflected by steep genomic clines (positive β), while selection for hybrids (i.e., overdominance) would be reflected by wide genomic clines (negative β; Gompert & Buerkle, 2011; Janoušek et al., 2012). The observed absence of positive β outliers and of islands of divergence in our demographic analysis indicates that despite some evidence of coupling between intrinsic and extrinsic barriers, widespread intrinsic incompatibilities are absent in this system, at least for the loci examined in this study. This is consistent with studies demonstrating weak reproductive isolation examined through forced crosses among these and other white pine species (Critchfield, 1986). Shared life history strategies among conifers, such as long generation time and high dispersal capacity, are likely to restrict the evolution of post- and prezygotic isolating mechanism (Stacy, Paritosh, Johnson, & Price, 2017). The limited evidence of intrinsic incompatibilities noted in our study supports the above claim, and we suggest that this pattern could be generalized across conifers with similar divergence history. Absence of β outliers and of recent hybrids indicates widespread recombination within the hybrid zone and an intermediate stage of divergence between our focal species (Nosil, Harmon, & Seehausen, 2009). The intermediate stage of divergence between our focal species, despite a long period of divergence in absolute time (i.e., years), is not surprising given the long generation times and large N_e estimates for conifers, which would have reduced the realized period of divergence when measured in coalescent units. Overall, the total absence of β outliers indicates a viable hybrid zone maintained largely through extrinsic factors

(Fouet et al., 2017), which may be the first stage of coupling between intrinsic and extrinsic barriers. Similar patterns of climatic clines in admixture and environmentally dependent maintenance of hybrid zones have been noted in other species of woody perennials in the genera *Quercus* (Dodd & Afzal-Rafii, 2004), *Picea* (De La Torre, Wang et al., 2014; Hamilton et al., 2013), *Rhododendron* (Milne et al., 2003) and *Pinus* (Cullingham, Cooke, & Coltman, 2014).

Contrary to the absence of β outliers, we identified many α outliers which is reflective of a hybrid zone experiencing moderate selection pressure and high levels of gene flow from both of the parental species (Gompert & Buerkle, 2011). Our demographic modelling, however, rejected the latter, thus indicating a moderate influence of natural selection on interspecific gene flow, as has been demonstrated across other conifers (Rehfeldt, 1999). Limited variation in β is associated with a diffuse genomic architecture of isolation (Gompert, Lucas et al., 2012), whereas the high genomic heterogeneity in α, under the estimated demographic scenario, could imply divergent natural selection operating within the hybrid zone (Gompert & Buerkle, 2011). This agrees with the higher values of multilocus F_{ST} within the putative hybrid zone ($F_{ST\text{-periphery}}$) and previous evidence of local adaptation in this region (Goodrich et al., 2016). A similar genomic mosaic of introgression has been noted across several studies (de Lafontaine et al., 2015; Gompert et al., 2014; Lexer et al., 2010; Lindtke, Gompert, Lexer, & Buerkle, 2014; Parchman et al., 2013) and is likely a result of complex interactions between divergence history, selection and genomic features.

Evidence of higher number of outliers with *P. strobiformis* ancestry and a negative association between our cline parameters (α and β) could be explained by three processes: (i) intrinsic incompatibilities resulting from Dobzhansky–Muller effects or complex epistatic effects disproportionally favouring allelic combinations from *P. strobiformis* in the hybrids relative to *P. flexilis* parental background, (ii) widespread directional selection on alleles from *P. strobiformis* in the hybrid zone leading to the formation of coadapted gene complexes and (iii) incomplete lineage sorting resulting from recent divergence between Core and Periphery. In contrast to inferences from the Engelmann–white spruce hybrid zone (De La Torre, Wang et al., 2014), the asymmetry of outlier loci is not due to high rates of gene flow from Core into Periphery, as the best demographic model excluded gene flow between these groups (see Figure 5b). A higher number of outlier loci with introgression favouring *P. strobiformis* are consistent with the strong influence of selection favouring alleles with *P. strobiformis* ancestry in the hybrid zone. Even without a linkage map, the cline results, along with asymmetrical niche divergence between Core and Periphery, point towards widespread directional introgression from *P. strobiformis* into the hybrid zone, which is consistent with local adaptation driving the evolution of coadapted gene complexes from *P. strobiformis* and of emerging intrinsic incompatibilities (Gompert, Lucas et al., 2012). The geographic clines of *h*, despite the absence of current gene flow between the Core and Periphery, also point towards an effect of incomplete lineage sorting. However, higher directional introgression from *P. strobiformis* even after accounting for the skewed pattern of genomic ancestry in the hybrid individuals emphasizes the role of selection over incomplete lineage sorting.

Our results are in accordance with studies in other coniferous species demonstrating that speciation is likely initiated through ecological barriers, and several generations of hybridization might occur before the evolution of intrinsic barriers to gene flow (Hamilton et al., 2013; Stacy et al., 2017; Zhou et al., 2014). Integrating the existing genomic data set with ongoing planting experiments involving climate treatments and measurements of fitness related traits should also help resolve the joint influence of extrinsic and intrinsic isolating mechanisms. Specifically, coincidence between the steepness of genomic, geographic and trait-specific clines would indicate a dominant role of extrinsic factors in facilitating divergence and speciation (De La Torre, Ingvarsson, & Aitken, 2015; Holliday, Ritland, & Aitken, 2010; Ryan et al., 2017; Stankowski et al., 2015). Alternatively, the presence of several loci showing steep clines but lacking climatic or functional associations would indicate a dominance of intrinsic barriers (Ryan et al., 2017). Although the genomic cline analysis used in this study provided key insights into the complexity of species isolation, it lacks sufficient power to account for complex epistatic effects (Gompert & Buerkle, 2011). These have likely played a key role in ecological speciation and in initiating the evolution of reproductive isolation (Flaxman et al., 2014; Lindtke et al., 2012). This study, however, provides concrete evidence of ecological speciation with gene flow, the presence of a historical hybrid zone maintained by extrinsic factors and early stages of coupling between extrinsic and intrinsic barriers contributing towards diversification. Whether these patterns hold generally for speciation within conifers, given their life history characteristics as well as their complex and large genomes, is thus a worthwhile area of future research.

ACKNOWLEDGEMENTS

This research was funded by U.S. National Science Foundation (NSF) grants EF-1442486 (Eckert), PRFB-1306622 (Friedline) and EF-1442597 (Waring), as well as by Northern Arizona University's Technology Research Initiative Program, USDA Forest Service Forest Health Protection Gene conservation programme, USDA Forest Service National Fire Plan Award 01.RMRS.B.6 (Schoettle), Virginia Commonwealth University (VCU) Department of Biology and VCU Integrated Life Sciences. We thank Z. Gompert and T. Parchman for providing R code that aided plotting the BGC results. We also acknowledge the field sampling and laboratory work crew (see Appendix S1E). This work would not have been possible without generous computational resources provided by the VCU Center for High Performance Computing and the Brigham Young University Fulton Supercomputing Lab.

DATA ACCESSIBILITY

Raw reads generated during this study are available at NCBI SRA database (PRJNA419956). The genotype file with sampling location information following SNP calling in 012 format, a python script for demographic modelling in ∂a∂i and the filtered occurrence records for conducting ENMs are available at Dryad (https://doi.org/10.5061/dryad.f6r55).

AUTHOR CONTRIBUTIONS

The study was designed by K.M.W., A.V.W., A.J.E., L.F.R., C.W. and S.C. Field sampling was performed by A.W.S., F.M.F., L.F.R., M.S.G., C.W., A.L.S. and K.M.W. Funding for this study was procured by K.M.W., A.V.W., A.W.S. and A.J.E. M.M., J.C.B. and C.F. performed the data analysis. M.M. generated the genomic data, analysed the data and wrote the manuscript with contributions from J.C.B. All authors edited the manuscript and have approved the version for submission.

ORCID

Mitra Menon http://orcid.org/0000-0003-2263-9764
Justin C. Bagley http://orcid.org/0000-0001-6737-8380

REFERENCES

Agrawal, A. F., Feder, J. L., & Nosil, P. (2011). Ecological divergence and the origins of intrinsic postmating isolation with gene flow. *International Journal of Ecology*, *2011*, e435357. http://doi.org/10.1155/2011/435357

Aguirre-Gutiérrez, J., Serna-Chavez, H. M., Villalobos-Arambula, A., Pérez de la Rosa, J. A., & Raes, N. (2015). Similar but not equivalent: Ecological niche comparison across closely–related Mexican white pines. *Diversity and Distributions*, *21*, 245–257. https://doi.org/10.1111/ddi.12268

Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., & Anderson, R. P. (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

Alberto, F. J., Aitken, S. N., Alía, R., González-Martínez, S. C., Hänninen, H., Kremer, A., ... Savolainen, O. (2013). Potential for evolutionary responses to climate change – Evidence from tree populations. *Global Change Biology*, *19*, 1645–1661. https://doi.org/10.1111/gcb.12181

Allen, C. D., & Breshears, D. D. (1998). Drought-induced shift of a forest–woodland ecotone: Rapid landscape response to climate variation. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 14839–14842. https://doi.org/10.1073/pnas.95.25.14839

Andrew, R. L., & Rieseberg, L. H. (2013). Divergence is focused on few genomic regions early in speciation: Incipient speciation of sunflower ecotypes. *Evolution*, *67*, 2468–2482. https://doi.org/10.1111/evo.12106

Barton, N. H. (2001). The role of hybridization in evolution. *Molecular Ecology*, *10*, 551–568. https://doi.org/10.1046/j.1365-294x.2001.01216.x

Barton, N. H., & De Cara, M. A. R. (2009). The evolution of strong reproductive isolation. *Evolution*, *63*, 1171–1190. https://doi.org/10.1111/j.1558-5646.2009.00622.x

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, *16*, 113–148. https://doi.org/10.1146/annurev.es.16.110185.000553

Benkman, C. W., Balda, R. P., & Smith, C. C. (1984). Adaptations for seed dispersal and the compromises due to seed predation in limber pine. *Ecology*, *65*, 632–642. https://doi.org/10.2307/1941426

Bierne, N., Welch, J., Loire, E., Bonhomme, F., & David, P. (2011). The coupling hypothesis: Why genome scans may fail to map local adaptation genes: The coupling hypothesis. *Molecular Ecology*, *20*, 2044–2072. https://doi.org/10.1111/j.1365-294X.2011.05080.x

Bisbee, J. (2014). Cone morphology of the *Pinus ayacahuite-flexilis* complex of the southwestern United States and Mexico. *Bulletin of the Cupressus Conservation Project*, *3*, 3–33.

Bivand, R., Keitt, T., Rowlingson, B., & Pebesma, E. (2017). *rgdal: Bindings for the Geospatial Data Abstraction Library. R package version 1.2-6*. Retrieved from https://CRAN.R-project.org/package=rgdal

Bridle, J. R., & Vines, T. H. (2007). Limits to evolution at range margins: When and why does adaptation fail? *Trends in Ecology & Evolution*, *22*, 140–147. https://doi.org/10.1016/j.tree.2006.11.002

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. New York, NY, USA: Springer.

Buschiazzo, E., Ritland, C., Bohlmann, J., & Ritland, K. (2012). Slow but not low: Genomic comparisons reveal slower evolutionary rate and higher dN/dS in conifers compared to angiosperms. *BMC Evolutionary Biology*, *12*, 8. https://doi.org/10.1186/1471-2148-12-8

Carstens, B. C., Brennan, R. S., Chua, V., Duffie, C. V., Harvey, M. G., Koch, R. A., ... Sullivan, J. (2013). Model selection as a tool for phylogeographic inference: An example from the willow *Salix melanopsis*. *Molecular Ecology*, *22*, 4014–4028. https://doi.org/10.1111/mec.12347

Christe, C., Stölting, K. N., Paris, M., Fraïsse, C., Bierne, N., & Lexer, C. (2017). Adaptive evolution and segregating load contribute to the genomic landscape of divergence in two tree species connected by episodic gene flow. *Molecular Ecology*, *26*, 59–76. https://doi.org/10.1111/mec.13765

Coyne, J., & Orr, H. A. (1989). Patterns of speciation in Drosophilla. *Evolution*, *43*, 362–381. https://doi.org/10.1111/j.1558-5646.1989.tb04233.x

Critchfield, W. B. (1986). Hybridization and classification of the white pines (*Pinus* section *Strobus*). *Taxon*, *35*, 647–656. https://doi.org/10.2307/1221606

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*, 3133–3157. https://doi.org/10.1111/mec.12796

Cullingham, C. I., Cooke, J. E. K., & Coltman, D. W. (2014). Cross-species outlier detection reveals different evolutionary pressures between sister species. *The New Phytologist*, *204*, 215–229. https://doi.org/10.1111/nph.12896

Cushman, S. A., & Landguth, E. L. (2016). Spatially heterogeneous environmental selection strengthens evolution of reproductively isolated populations in a Dobzhansky-Muller system of hybrid incompatibility. *Frontiers in Genetics*, *7*, 209. https://doi.org/10.3389/fgene.2016.00209

De Carvalho, D., Ingvarsson, P. K., Joseph, J., Suter, L., Sedivy, C., Macaya-Sanz, D., ... Lexer, C. (2010). Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula* L.), a widespread forest tree. *Molecular Ecology*, *19*, 1638–1650. https://doi.org/10.1111/j.1365-294X.2010.04595.x

De La Torre, A. R., Birol, I., Bousquet, J., Ingvarsson, P. K., Jansson, S., Jones, S., ... Bohlmann, J. (2014). Insights into conifer giga-genomes. *Plant Physiology*, *166*, 1724–1732. https://doi.org/10.1104/pp.114.248708

De La Torre, A. R., Ingvarsson, P. K., & Aitken, S. N. (2015). Genetic architecture and genomic patterns of gene flow between hybridizing species of *Picea*. *Heredity*, *115*, 153–164. https://doi.org/10.1038/hdy.2015.19

De La Torre, A. R., Li, Z., Van de Peer, Y., & Ingvarsson, P. K. (2017). Contrasting rates of molecular evolution and patterns of selection among gymnosperms and flowering plants. *Molecular Biology and Evolution*, *34*, 1363–1377. https://doi.org/10.1093/molbev/msx069

De La Torre, A. R., Wang, T., Jaquish, B., & Aitken, S. N. (2014). Adaptation and exogenous selection in a *Picea glauca* × *Picea engelmannii* hybrid zone: Implications for forest management under climate change. *The New Phytologist*, *201*, 687–699. https://doi.org/10.1111/nph.12540

de Lafontaine, G., 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*). *Molecular Ecology*, *24*, 5229–5247. https://doi.org/10.1111/mec.13377

Dodd, R. S., & Afzal-Rafii, Z. (2004). Selection and dispersal in a multispecies oak hybrid zone. *Evolution*, *58*, 261–269. https://doi.org/10.1111/j.0014-3820.2004.tb01643.x

Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics*, *28*, 342–350. https://doi.org/10.1016/j.tig.2012.03.009

Fitzpatrick, B. M. (2012). Estimating ancestry and heterozygosity of hybrids using molecular markers. *BMC Evolutionary Biology*, *12*, 131. https://doi.org/10.1186/1471-2148-12-131

Fitzpatrick, B. M. (2013). Alternative forms for genomic clines. *Ecology and Evolution*, *3*, 1951–1966. https://doi.org/10.1002/ece3.609

Flaxman, S. M., Wacholder, A. C., Feder, J. L., & Nosil, P. (2014). Theoretical models of the influence of genomic architecture on the dynamics of speciation. *Molecular Ecology*, *23*, 4074–4088. https://doi.org/10.1111/mec.12750

Fouet, C., Kamdem, C., Gamez, S., White, B. J. (2017). Genomic insights into adaptive divergence and speciation among malaria vectors of the Anopheles nili group. *Evolutionary applications*, *10*, 897–906. https://doi.org/10.1111/eva.12492

Frankis, M. (2009). The high altitude white pines of Mexico and the adjacent SW USA (*Pinus* L. subgenus *Strobus* Lemmon, Pinaceae). International Dendrology Society. Yearbook 63–72.

Geraldes, A., Farzaneh, N., Grassa, C. J., McKown, A. D., Guy, R. D., Mansfield, S. D., . . . Cronk, Q. C. B. (2014). The landscape genomics of *Populus trichocarpa*: The role of hybridization, limited gene flow, and natural selection in shaping patterns of population structure: Population differentiation in *Populus trichocarpa*. *Evolution*, *68*, 3260–3280. https://doi.org/10.1111/evo.12497

Gitlin, A. R., Sthultz, C. M., Bowker, M. A., Stumpf, S., Paxton, K. L., Kennedy, K., & Whitham, T. G. (2006). Mortality gradients within and among dominant plant populations as barometers of ecosystem change during extreme drought. *Conservation Biology*, *20*, 1477–1486. https://doi.org/10.1111/j.1523-1739.2006.00424.x

Gompert, Z., & Buerkle, C. A. (2010). introgress: A software package for mapping components of isolation in hybrids. *Molecular Ecology Resources*, *10*, 378–384. https://doi.org/10.1111/j.1755-0998.2009.02733.x

Gompert, Z., & Buerkle, C. A. (2011). Bayesian estimation of genomic clines. *Molecular Ecology*, *20*, 2111–2127. https://doi.org/10.1111/j.1365-294X.2011.05074.x

Gompert, Z., & Buerkle, C. A. (2012). bgc: Software for Bayesian estimation of genomic clines. *Molecular Ecology Resources*, *12*, 1168–1176. https://doi.org/10.1111/1755-0998.12009.x

Gompert, Z., Lucas, L. K., Buerkle, C. A., Forister, M. L., Fordyce, J. A., & Nice, C. C. (2014). Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. *Molecular Ecology*, *23*, 4555–4573. https://doi.org/10.1111/mec.12811

Gompert, Z., Lucas, L. K., Nice, C. C., Fordyce, J. A., Forister, M. L., & Buerkle, C. A. (2012). Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution*, *66*, 2167–2181. https://doi.org/10.1111/j.1558-5646.2012.01587.x

Gompert, Z., Parchman, T. L., & Buerkle, C. A. (2012). Genomics of isolation in hybrids. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*, 439–450. https://doi.org/10.1098/rstb.2011.0196

Goodrich, B. A., Waring, K. M., & Kolb, T. E. (2016). Genetic variation in *Pinus strobiformis* growth and drought tolerance from southwestern US populations. *Tree Physiology*, *36*, 1219–1235. https://doi.org/10.1093/treephys/tpw052

Goudet, J. (2005). hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*, 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Guerrero, R. F., & Hahn, M. W. (2017). Speciation as a sieve for ancestral polymorphism. *Molecular Ecology*, *26*, 5362–5368. https://doi.org/10.1111/mec.14290

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLOS Genetics*, *5*, e1000695. https://doi.org/10.1371/journal.pgen.1000695

Hamilton, J. A., Lexer, C., & Aitken, S. N. (2013). Genomic and phenotypic architecture of a spruce hybrid zone (*Picea sitchensis* × *P. glauca*). *Molecular Ecology*, *22*, 827–841. https://doi.org/10.1111/mec.12007

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*, 1965–1978. https://doi.org/10.1002/(ISSN)1097-0088

Hijmans, R. J., vanEtten, J., Cheng, J., Mattiuzzi, M., Sumner, M., Greenberg, J. A., & Shortridge, A. (2016). *raster: Geographic Data Analysis and Modeling. R package version*. 2.5-8. Retrieved from https://CRAN.R-project.org/package=raster

Hijmans, R. J., Phillips, S., & Elith, J. L. (2017). *dismo: Species Distribution Modeling. R package version 1.1-4*. Retrieved from https://CRAN.R-project.org/package=dismo

Holliday, J. A., Ritland, K., & Aitken, S. N. (2010). Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (*Picea sitchensis*). *The New Phytologist*, *188*, 501–514. https://doi.org/10.1111/j.1469-8137.2010.03380.x

Ikeda, D. H., Max, T. L., Allan, G. J., Lau, M. K., Shuster, S. M., & Whitham, T. G. (2017). Genetically informed ecological niche models improve climate change predictions. *Global Change Biology*, *23*, 164–176. https://doi.org/10.1111/gcb.13470

Janoušek, V., Wang, L., Luzynski, K., Dufková, P., Vyskočilová, M. M., Nachman, M. W., . . . Tucker, P. K. (2012). Genome-wide architecture of reproductive isolation in a naturally occurring hybrid zone between *Mus musculus musculus* and *M. m. domesticus*. *Molecular Ecology*, *21*, 3032–3047. https://doi.org/10.1111/j.1365-294X.2012.05583.x

Kane, N. C., King, M. G., Barker, M. S., Raduski, A., Karrenberg, S., Yatabe, Y., . . . Rieseberg, L. H. (2009). Comparative genomic and population genetic analyses indicate highly porous genomes and high levels of gene flow between divergent *Helianthus* species. *Evolution*, *63*, 2061–2075. https://doi.org/10.1111/j.1558-5646.2009.00703.x

Knowles, L. L., Carstens, B. C., & Keat, M. L. (2007). Coupling genetic and ecological-niche models to examine how past population distributions contribute to divergence. *Current Biology*, *17*, 940–946. https://doi.org/10.1016/j.cub.2007.04.033

Kulmuni, J., & Westram, A. M. (2017). Intrinsic incompatibilities evolving as a by-product of divergent ecological selection: Considering them in empirical studies on divergence with gene flow. *Molecular Ecology*, *26*, 3093–3103. https://doi.org/10.1111/mec.14147

Lackey, A. C. R., & Boughman, J. W. (2016). Evolution of reproductive isolation in stickleback fish: Evolution of isolation in stickelback fish. *Evolution*, *71*, 357–372. https://doi.org/10.1111/evo.13114

Lexer, C., Joseph, J. A., Loo, M. V., Barbará, T., Heinze, B., Bartha, D., . . . Buerkle, C. A. (2010). Genomic admixture analysis in European *Populus* spp. reveals unexpected patterns of reproductive isolation and mating. *Genetics*, *186*, 699–712. https://doi.org/10.1534/genetics.110.118828

Li, X., Zhu, C., Yeh, C. T., Wu, W., Takacs, E. M., Petsch, K. A., . . . Yu, J. (2012). Genic and nongenic contributions to natural variation of quantitative traits in maize. *Genome Research*, *22*, 2436–2444. https://doi.org/10.1101/gr.140277.112

Lind, B. M., Freidline, C. J., Wegrzyn, J. L., Maloney, P. E., Vogler, D. R., Neale, D. B., & Eckert, A. J. (2017). Water availability drives signatures of local adaptation in whitebark pine (*Pinus albicaulis* Engelm.) across fine spatial scales of the Lake Tahoe Basin, USA. *Molecular Ecology*, *26*, 3168–3185. https://doi.org/10.1111/mec.14106

Lindtke, D., Buerkle, C. A., Barbará, T., Heinze, B., Castiglione, S., Bartha, D., & Lexer, C. (2012). Recombinant hybrids retain heterozygosity at many loci: New insights into the genomics of reproductive isolation in *Populus*. *Molecular Ecology*, *21*, 5042–5058. https://doi.org/10.1111/j.1365-294X.2012.05744.x

Lindtke, D., Gompert, Z., Lexer, C., & Buerkle, C. A. (2014). Unexpected ancestry of *Populus* seedlings from a hybrid zone implies a large role for postzygotic selection in the maintenance of species. *Molecular Ecology*, *23*, 4316–4330. https://doi.org/10.1111/mec.12759

Looney, C. E., & Waring, K. M. (2013). *Pinus strobiformis* (southwestern white pine) stand dynamics, regeneration, and disturbance ecology: A review. *Forest Ecology and Management*, *287*, 90–102. https://doi.org/10.1016/j.foreco.2012.09.020

Losos, J. B., Arnold, S. J., Bejerano, G., Brodie 3rd, E. D., Hibbett, D., Hoekstra, H. E., . . . Turner, T. L. (2013). Evolutionary biology for the 21st century. *PLoS Biology*, *11*, e1001466. https://doi.org/10.1371/journal.pbio.1001466

Lotterhos, K. E., & Whitlock, M. C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology*, *24*, 1031–1046. https://doi.org/10.1111/mec.13100

Marques, D. A., Lucek, K., Haesler, M. P., Haesler, M. P., Feller, A. F., Meier, J. I., & Seehausen, O. (2017). Genomic landscape// of early ecological speciation initiated by selection on nuptial colour. *Molecular Ecology*, *26*, 7–24. https://doi.org/10.1111/mec.13774

McVean, G. (2009). A genealogical interpretation of principal components analysis. *PLOS Genetics*, *5*, e100068. https://doi.org/10.1371/journal.pgen.1000686

Milne, R. I., Terzioglu, S., & Abbott, R. J. (2003). A hybrid zone dominated by fertile F1s: Maintenance of species barriers in *Rhododendron*. *Molecular Ecology*, *12*, 2719–2729. https://doi.org/10.1046/j.1365-294X.2003.01942.x

Mishra, A. K., & Singh, V. P. (2010). A review of drought concepts. *Journal of Hydrology*, *391*, 202–216. https://doi.org/10.1016/j.jhydrol.2010.07.012

Moore, W. S. (1977). An evaluation of narrow hybrid zones in vertebrates. *The Quarterly Review of Biology*, *52*, 263–277. https://doi.org/10.1086/409995

Moreno-Letelier, A., & Barraclough, T. G. (2015). Mosaic genetic differentiation along environmental and geographic gradients indicate divergent selection in a white pine species complex. *Evolutionary Ecology*, *29*, 733–748. https://doi.org/10.1007/s10682-015-9785-4

Moreno-Letelier, A., Ortíz-Medrano, A., & Piñero, D. (2013). Niche divergence versus neutral processes: Combined environmental and genetic analyses identify contrasting patterns of differentiation in recently diverged pine species. *PLoS ONE*, *8*, e78228. https://doi.org/10.1371/journal.pone.0078228

Moreno-Letelier, A., & Piñero, D. (2009). Phylogeographic structure of *Pinus strobiformis* Engelm. across the Chihuahuan Desert filter-barrier. *Journal of Biogeography*, *36*, 121–131. https://doi.org/10.1111/j.1365-2699.2008.02001.x

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

Nosil, P. (2008). Speciation with gene flow could be common. *Molecular Ecology*, *17*, 2103–2106. https://doi.org/10.1111/j.1365-294X.2008.03715.x

Nosil, P., & Feder, J. L. (2012). Genomic divergence during speciation: Causes and consequences. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*, 332–342. https://doi.org/10.1098/rstb.2011.0263

Nosil, P., Harmon, L. J., & Seehausen, O. (2009). Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*, *24*, 145–156. https://doi.org/10.1016/j.tree.2008.10.011

Ortego, J., Gugger, P. F., Riordan, E. C., & Sork, V. L. (2014). Influence of climatic niche suitability and geographical overlap on hybridization patterns among southern Californian oaks. *Journal of Biogeography*, *41*, 1895–1908. https://doi.org/10.1111/jbi.12334

Ortego, J., Noguerales, V., Gugger, P. F., & Sork, V. L. (2015). Evolutionary and demographic history of the Californian scrub white oak species complex: An integrative approach. *Molecular Ecology*, *24*, 6188–6208. https://doi.org/10.1111/mec.13457

Parchman, T. L., Gompert, Z., Braun, M. J., Brumfield, R. T., McDonald, D. B., Uy, J. A., . . . Buerkle, C. A. (2013). The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. *Molecular Ecology*, *22*, 3304–3317. https://doi.org/10.1111/mec.12201

Parchman, T. L., Gompert, Z., Mudge, J., Schilkey, F. D., Benkman, C. W., & Buerkle, C. A. (2012). Genome-wide association genetics of an adaptive trait in lodgepole pine: Association mapping of serotiny. *Molecular Ecology*, *21*, 2991–3005. https://doi.org/10.1111/j.1365-294X.2012.05513.x

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLOS Genetics*, *2*, e190. https://doi.org/10.1371/journal.pgen.0020190

Pavy, N., Pelgas, B., Laroche, J., Rigault, P., Isabel, N., & Bousquet, J. (2012). A spruce gene map infers ancient plant genome reshuffling and subsequent slow evolution in the gymnosperm lineage leading to extant conifers. *BMC Biology*, *10*, 84. https://doi.org/10.1186/1741-7007-10-84

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (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

Petit, R. J., & Excoffier, L. (2009). Gene flow and species delimitation. *Trends in Ecology & Evolution*, *24*, 386–393. https://doi.org/10.1016/j.tree.2009.02.011

Petit, R. J., & Hampe, A. (2006). Some evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution, and Systematics*, *37*, 187–214. https://doi.org/10.1146/annurev.ecolsys.37.091305.110215

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. MAXENT v3.4.1. *Ecological Modelling*, *190*, 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Pritchard, J. K., & Di Rienzo, A. (2010). Adaptation – Not by sweeps alone. *Nature Reviews Genetics*, *11*, 665–667. https://doi.org/10.1038/nrg2880

Puritz, J. B., Hollenbeck, C. M., & Gold, J. R. (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 (2017). *R v.3.3.2: A language and environment for statistical computing*. Vienna, Austria : The R Foundation for Statistical Computing.

Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE v. 1.0: Variational inference of population structure in large SNP data sets. *Genetics*, *197*, 573–589. https://doi.org/10.1534/genetics.114.164350

Rajora, O. P., Eckert, A. J., & Zinck, J. W. R. (2016). Single-locus versus multilocus patterns of local adaptation to climate in eastern white pine (*Pinus strobus*, Pinaceae). *PLoS ONE*, *11*, e0158691. https://doi.org/10.1371/journal.pone.0158691

Rambaut, A., Suchard, M. A., Xie, D., & Drummond, A. J. (2013). *Tracer, version 1.6*. Edinburgh, UK: University of Edinburgh.

Rehfeldt, G. E. (1999). Systematics and genetic structure of Ponderosae taxa (Pinaceae) inhabiting the mountain islands of the Southwest. *American Journal of Botany*, *86*, 741–752. https://doi.org/10.2307/2656584

Restaino, C. M., Peterson, D. L., & Littell, J. (2016). Increased water deficit decreases Douglas fir growth throughout western US forests. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 9557–9562. https://doi.org/10.1073/pnas.1602384113

Riquet, F., Haag-Liautard, C., Woodall, L., Bouza, C., Patrick, L., Hamer, B., . . . Bierne, N. (2017). Parallel use of a shared genomic island of speciation in clinal and mosaic hybrid zones between cryptic seahorse lineages. *bioRxiv*, 161786. https://doi.org/10.1101/161786

Roesti, M., Hendry, A. P., Salzburger, W., & Berner, D. (2012). Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Molecular Ecology*, *21*, 2852–2862. https://doi.org/10.1111/j.1365-294X.2012.05509.x

Rundle, H. D. (2002). A test of ecologically dependent postmating isolation between sympatric sticklebacks. *Evolution*, *56*, 322–329. https://doi.org/10.1111/j.0014-3820.2002.tb01342.x

Ryan, S., Fontaine, M. C., Scriber, J. M., Pfrender, M. E., O' Neil, S. T., & Hellmann, J. J. (2017). Patterns of divergence across the geographic and genomic landscape of a butterfly hybrid zone associated with a climatic gradient. *Molecular Ecology*, *26*, 4725–4742. https://doi.org/10.1111/mec.14236

Schield, D. R., Adams, R. H., Card, D. C., Perry, B. W., Pasquesi, G. M., Jezkova, T., . . . Castoe, T. A. (2017). Insight into the roles of selection in speciation from genomic patterns of divergence and introgression in secondary contact in venomous rattlesnakes. *Ecology and Evolution*, *7*, 3951–3966. https://doi.org/10.1002/ece3.2996

Schluter, D., & Conte, G. L. (2009). Genetics and ecological speciation. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 9955–9962. https://doi.org/10.1073/pnas.0901264106

Shirk, A. J., Cushman, S. A., Waring, K. M., Wehenkel, C. A, Leal-Saenz, A., Toney, C., & Lopez-Sanchez, C. A. (2018). Southwestern white pine (*Pinus strobiformis*) species distribution models predict large range shift and contraction due to climate change. *Forest Ecology & Management*, *411*, 176–186. https://doi.org/10.1016/j.foreco.2018.01.025

Stacy, E. A., Paritosh, B., Johnson, M. A., & Price, D. K. (2017). Incipient ecological speciation between successional varieties of a dominant tree involves intrinsic postzygotic isolating barriers. *Ecology and Evolution*, *7*, 2501–2512. https://doi.org/10.1002/ece3.2867

Stankowski, S., Sobel, J. M., & Streisfeld, M. A. (2015). The geography of divergence with gene flow facilitates multitrait adaptation and the evolution of pollinator isolation in *Mimulus aurantiacus*. *Evolution*, *69*, 3054–3068. https://doi.org/10.1111/evo.12807

Steinhoff, R. J., & Andresen, J. W. (1971). Geographic variation in *Pinus flexilis* and *Pinus strobiformis* and its bearing on their taxonomic status. *Silvae Genetica*, *20*, 159–167.

Tine, M., Kuhl, H., Gagnaire, P. A., Louro, B., Desmarais, E., Martins, R. S., . . . Reinhardt, R. (2014). European sea bass genome and its variation provide insights into daptation to euryhalinity and speciation. *Nature Communications*, *5*, 5770. https://doi.org/10.1038/ncomms6770

Tomback, D. F., Samano, S., Pruett, E. L., & Schoettle, A. W. (2011). Seed dispersal in limber and southwestern white pine: Comparing core and peripheral populations. In The future of high-elevation, five-needle white pines in Western North America: Proceedings of the High Five Symposium. Proceedings RMRS-P-63. (pp. 69–71). Fort Collins, CO, USA: US Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Via, S., Bouck, A. C., & Skillman, S. (2000). Reproductive isolation between divergent races of pea aphids on two hosts. II. Selection against migrants and hybrids in the parental environments. *Evolution*, *54*, 1626–1637. https://doi.org/10.1111/j.0014-3820.2000.tb00707.x

Warren, D. L., Glor, R. E., & 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

Williams, A. P., Allen, C. D., Millar, C. I., Swetnam, T. W., Michaelsen, J., Still, C. J., & Leavitt, S. W. (2010). Forest responses to increasing aridity and warmth in the southwestern United States. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 21289–21294. https://doi.org/10.1073/pnas.0914211107

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

Yang, R. C. (1998). Estimating hierarchical *F*-statistics. *Evolution*, *52*, 950–956. https://doi.org/10.1111/j.1558-5646.1998.tb01824.x

Yeaman, S., Aeschbacher, S., & Bürger, R. (2016). The evolution of genomic islands by increased establishment probability of linked alleles. *Molecular Ecology*, *25*, 2542–2558. https://doi.org/10.1111/mec.13611

Yeaman, S., & Otto, S. P. (2011). Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution*, *65*, 2123–2129. https://doi.org/10.1111/j.1558-5646.2011.01277.x

Zhou, Y., Zhang, L., Liu, J., Wu, G., & Savolainen, O. (2014). Climatic adaptation and ecological divergence between two closely related pine species in Southeast China. *Molecular Ecology*, *23*, 3504–3522. https://doi.org/10.1111/mec.12830

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

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