bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 **Running title: Ecological speciation in pines**

2

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

5 MITRA MENON[1,2*], JUSTIN C. BAGLEY[2,3], CHRISTOPHER FRIEDLINE[2], AMY V. WHIPPLE[4], ANNA W.
6 SCHOETTLE[5], ALEJANDRO LEAL-SAENZ[13], CHRISTIAN WEHENKEL[6], FRANCISCO MOLINA-
7 FREANER[7], LLUVIA FLORES-RENTERIA[8], M. SOCORRO GONZALEZ-ELIZONDO[9], RICHARD A.
8 SNIEZKO[10], SAMUEL A. CUSHMAN[11], KRISTEN M. WARING[12], AND ANDREW J. ECKERT[2*]

9 [1]*Integrative Life Sciences, Virginia Commonwealth University, Richmond, VA*
10 *23284, USA.* [2]*Department of Biology, Virginia Commonwealth University, Richmond, VA 23284,*
11 *USA.* [3]*Departamento de Zoologia, Universidade de Brasília, 70910-900Brasília, DF,*
12 *Brazil.*[4]*Department of Biological Sciences and Merriam Powel Center for Environmental*
13 *Research, Northern Arizona University, Flagstaff, AZ 86011, USA.* [5] *Rocky Mountain Research*
14 *Station, USDA Forest Service, Flagstaff, AZ 86001, USA.* [6]*Instituto de Silvicultura e Industria de*
15 *la Madera, Universidad Juarez del Estado de Durango, 34120 Durango, MX.* [7]*Institutos de*
16 *Geologia y Ecologia, Universidad Nacional Autonoma de Mexico, Estacion Regional del*
17 *Noroeste, Hermosillo, Sonora CP 83250 MX.* [8]*Department of Biology, San Diego State*
18 *University, San Diego, CA 92182, USA.* [9]*Instituto Politecnico Nacional, CIIDIR, Sigma 119*
19 *Fracc. 20 de Noviembre II, Durango, Dgo., 34234 MX.* [10] *USDA Forest Service, Dorena Genetic*
20 *Resource Center, Cottage grove, OR, 97424, USA.* [11] *USDA Forest Service, Flagstaff, AZ86001,*
21 *USA.* [12]*School of Forestry, Northern Arizona University, Flagstaff, AZ 86011, USA.* [13] *Programa*
22 *Institucional de Doctorado en Ciencias Agropecuarias y Forestales, Universidad Juárez del*
23 *Estado de Durango, 34120 Durango, MX.*

24

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

25 *Authors for correspondence:
26 Mitra Menon,
27 1000 West Cary Street, Integrative Life Sciences, Virginia Commonwealth University,
28 Richmond, Virginia 23284 USA, E-mail: menonm2@mymail.vcu.edu;

29 Andrew J. Eckert,
30 1000 West Cary Street, Life Science Bldg. - Department of Biology Rm 126, Virginia
31 Commonwealth University, Richmond, Virginia 23284 USA, E-mail: aeckert2@vcu.edu.

32 *Keywords:* conifers, ecological speciation, hybrid zone, introgression, population genomics,

33 extrinsic barriers

34 *Contents:* ~8 066 words (abstract, main text, acknowledgments, figure and table captions); 3

35 tables; 5 figures.

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract (234 words)

Interactions between extrinsic factors, such as disruptive selection, and intrinsic factors, such as genetic incompatibilities among loci, can 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 *P. flexilis*. Utilizing ecological niche modeling, demographic modeling, and genomic cline analyses, we illustrated a history of divergence with continuous gene flow between these species. We found an abundance of advanced generation hybrids and a lack of loci exhibiting large allele frequency differences 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. Our results are consistent with extrinsic factors, such as climate, being an important isolating mechanism for these species. A buildup of intrinsic incompatibilities and of co-adapted gene complexes is also apparent in our results, 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 creating and enforcing species boundaries. Overall, we lend support to the hypothesis that varying strengths and directions of selection pressures across the long lifespans of conifers, in combination with their life history strategies, delay the evolution of strong intrinsic incompatibilities.

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 *et al.* 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 between several loci independently experiencing disruptive selection, and the action of selection to maintain these co-adapted divergent gene complexes (Flaxman *et al*. 2014).

Several studies of speciation have used hybrid zones as windows into the process of divergence among 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 (Schield *et al*. 2017; Ryan *et al*. 2017). Specifically, the maintenance of species boundaries has been shown to occur through models of tension zones (intrinsic incompatibilities *sensu* Barton & Hewitt 1985; Via *et al.* 2000; Barton 2001; Rundle 2002) and bounded hybrid superiority (extrinsic incompatibilities *sensu* Moore 1977; Milne *et al.* 2003; Hamilton *et al.* 2013).The former facilitates divergence through a buildup of genetic incompatibilities among loci causing environmentally independent reduction in fitness of the hybrids, whereas the latter involves

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

increased hybrid fitness only in a novel environment to which the divergent parental allelic combinations confer a putative advantage. These two models are often coupled, such that genomic regions involved in intrinsic incompatibility coincide with loci exhibiting ecological gradients in allele frequency (Bierne *et al*. 2011; Cushman & Landguth 2016), ensuring the maintenance of species barriers under 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 non-model species has facilitated studies of ecological speciation across varying spatial and temporal scales (Lexer *et al*. 2010; Andrew & Rieseberg 2013; de Lafontaine *et al*. 2015; Lackey & Boughman 2016; Marques *et al*. 2017). Studies using genome scans often lend support to the genic view of speciation (Wu 2001), which predicts that a handful of genomic regions experiencing strong selection pressures will exhibit high differentiation against a background of lower differentiation driven by unrestricted gene flow. Varying levels of differentiation across loci thus generates a mosaic of genomic differentiation. Methodological approaches designed to detect the processes underlying this mosaic, however, are confounded by demographic histories of secondary contact, genomic areas of suppressed recombination, recent divergences without gene flow, allele surfing, and selective sweeps specific to each lineage unrelated to the development of reproductive isolation (Noor & Bennett 2009; Cruickshank & Hahn 2014). For example, ecological speciation could result in genomic regions of elevated differentiation (i.e. genomic islands of differentiation) that are associated with niche partitioning, but not necessarily with reproductive isolation. Further, these islands of differentiation are mostly expected when adaptation occurs from moderate- to large-

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

effect *de novo* mutations (Lackey & Boughman 2016). Genomic approaches used to identify islands of differentiation may thus be biased against identifying polygenic regions associated with species divergence, or towards identifying loci that contribute to ecological niche divergence of hybrids relative to both parental species and hence restrict gene flow between them. To avoid these biases, ecological niche divergence should be evaluated in the hybrid zone with respect to both parental species and correlated to patterns of hybrid ancestry (e.g. Hamilton *et al*. 2013).

Species of conifers are known to have ecologically differentiated niches despite the absence of strong morphological differences (e.g. Rehfeldt 1999). Strong pre- and post-zygotic isolating barriers contributing towards morphological disjunctions are often absent in conifers (Critchfield 1986; Buschiazzo *et al*. 2012; Pavy *et al*. 2012) due to their life history characteristics, such as longevity, high dispersal abilities, and long generation times (Petit & Hampe 2006; Neale & Kremer 2011). 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 arising from extrinsic barriers is thus likely to play a dominant role in facilitating speciation within 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 *P. flexilis* E. James. (limber pine). Our focal species inhabit a wide latitudinal range in the western part of North America, with a putative area of sympatry located in the southern Rocky Mountains and Colorado Plateau in which morphological evidence points towards the occurrence of hybridization (Steinhoff & Andresen 1971; Tomback & Achuff 2010). These species also display limited differences in morphological and reproductive traits

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Benkman *et al.* 1984; Tomback *et al.* 2011; Bisbee 2014) and show evidence of local adaptation to the heterogeneous climatic conditions across their geographical range and also within the area of putative hybridization (Steinhoff & Andresen 1971; Moreno-Letelier *et al.* 2013; Borgman *et al.* 2015; Moreno-Letelier & Barraclough 2015; Goodrich *et al.* 2016). To examine the processes influencing species boundaries between these two conifer species, we asked three questions: (1) Is there niche divergence among *P. strobiformis, P. flexilis* and the putative hybrid zone? (2) Did the divergence of *P. strobiformis* and *P. flexilis* occur with continual gene flow? (3) 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 among the focal species, with several lines of evidence supporting the strong influence of extrinsic factors in reinforcing species boundaries.

Materials and Methods

Focal taxa and field sampling

Pinus strobiformis and *P. flexilis* are closely related species of white pines that occur across broad temperature and precipitation gradients 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; Fig. 1). *Pinus flexilis* occurs in mountainous regions from northern Arizona and northern New Mexico to Alberta, with a region of putative sympatry with *P. strobiformis* in the southern Rocky Mountains and Colorado Plateau (Fig. 1). Across this zone of putative sympatry, cone morphology and dispersal syndromes fall along a continuum of

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

148 divergence blending into the characteristics of populations in the allopatric zones of either
149 species (Bisbee 2014).

150 We sampled 42 *P. strobiformis* populations encompassing a total of 376 trees (5-13
151 trees/population) from its entire geographical range. For *P. flexilis*, a total of 13 populations were
152 sampled, with eight populations sampled from the southern periphery of the geographical range
153 and five sampled closer to the range center (Fig. 1). Across these thirteen populations, we
154 sampled a total of 69 trees (4–10 trees/population). To help minimize relatedness, trees within
155 the same site were sampled with a minimum spacing of 50 m (*P. strobiformis*) and 200 m (*P.*
156 *flexilis*) from each other.

157 **Data generation**

158 *Occurrence data*

159 We assembled a comprehensive dataset of occurrences for ecological niche modeling
160 (ENM) by supplementing our field site data with occurrence records downloaded from the
161 Global Biodiversity Information Facility (GBIF) using functions from the DISMO package
162 (Hijmans *et al*. 2017) available in the R environment (R Core Team 2017). Using a series of
163 filtering steps to account for observation and sampling biases (Supporting Information, Appendix
164 S1.A), we obtained a final dataset of 254 occurrence records for *P. strobiformis* and 420 for *P.*
165 *flexilis*. Incorporating intraspecific genetic variation into ENMs can improve model fit and
166 provide more robust predictions when projecting across time and space (Knowles *et al*. 2007;
167 Ikeda *et al*. 2017). Thus, we divided presence locations within *P. strobiformis* into core
168 (latitudinal range: 19–30.5 $^{\circ}$N) and northern periphery (latitudinal range: 31–33 $^{\circ}$N). These groups
169 likely represent different genetic clusters given the geographically restricted hybridization
170 between *P. flexilis* and *P. strobiformis* (Steinhoff & Andresen 1971; Tomback & Achuff 2010;

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Bisbee 2014). We defined three groups that were the focus of our enquiries – (1) populations of *P. flexilis*, (2) populations of *P. strobiformis* from the northern range periphery (Periphery hereafter), and (3) 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 for these variables were downloaded from WorldClim v.1.4 (Hijmans *et al*. 2005) at 30 arc-second resolutions and at 2.5 arc-minute resolutions for the Last Glacial Maximum (LGM) and data were extracted from each layer using the RASTER package (Hijmans *et al.* 2016) available in R.

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 *et al*. 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, barcodes, and primers. Following PCR, we selected DNA fragments in the 300–400 bp size range using agarose gel electrophoresis followed by isolation of pooled DNA from these gels using QIAquick Gel Extraction Kits (Qiagen). Single-end sequencing, with one multiplexed library per lane, was used to obtain 100bp reads, with all sequencing conducted with Illumina HiSeq 2500 at the Nucleic Acids Research Facility located at Virginia Commonwealth University. The resulting FASTQ files were processed using dDocent bioinformatics pipeline (details in Supporting Information, Appendix S1.B; Puritz *et al.* 2014). The entire process yielded a total of 51 633 single nucleotide polymorphisms (SNPs), which were used as the starting dataset for all subsequent analyses.

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Data analysis

Ecological niche modeling and niche divergence

We developed ENMs for each of the three groups: Core, Periphery, and *P. flexilis*, using the algorithms available in Maximum Entropy (MAXENT; Phillips *et al*. 2006). Since 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 in a cell. Data processing, model fitting, and model evaluation using 5,000 iterations within MAXENT were conducted using the DISMO, RASTER, RGDAL (Bivand *et al*. 2017), and SPTHIN (Aiello-Lammens *et al*. 2015) packages available in R. ENMs were constructed from climate variables with an absolute correlation coefficient (r) less than 0.85. Two indices were used to assess model performance for each group: overall regularized training gain (RTG) and area under the curve (AUC). Since 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 & 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 *et al.* 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 obtained across the full geographical extent for the present conditions at 30 arc-seconds were compared across all three pairs of groups (Core–Periphery, Core–*P. flexilis*, Periphery–*P. flexilis*) to investigate

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

patterns of niche evolution. To account for potential biases towards niche divergence introduced by latitudinally associated environmental variation in the present range of each pair, we performed an asymmetric background randomization test, based on Schoener's *D*, in the R package ENMTOOLS (Warren *et al.* 2008). The two resulting null distributions of niche divergence obtained through this test correspond to the background of each group compared against the other. An observed value of Schoener's *D* much smaller than expected after accounting for background differences is indicative of niche divergence, whereas a value much larger than expected indicates niche conservatism (Warren *et al*. 2008).

Population structure and demographic modeling

We assessed the pattern and the 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 discussed above (see Data Generation & Fig. 1). We conducted principal components analysis (PCA) to visualize grouping of sampled trees into the three groups delineated in our methods (Patterson *et al.* 2006; McVean 2009). To complement the PCA, we also conducted an individual-based assignment test using FASTSTRUCTURE, a variational Bayesian version of STRUCTURE designed for use with large SNP datasets (Raj *et al*. 2014). For any given number of predefined clusters (*K*), FASTSTRUCTURE assigns a *Q*-value representing the proportion of a sample's ancestry derived from each cluster. We set *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 groups within our genetic data. 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

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

239 the highest level of the hierarchy and F_{ST} describing differentiation among groups across all
240 levels of the hierarchy (see Yang 1998). A similar nested model with the highest level of
241 hierarchy being groups within *P. strobiformis* was used to assess intraspecific differentiation. For
242 the former, *F*-statistics are denoted using the term 'species' in the subscripts, whereas the latter
243 uses the term 'groups' in the subscripts. We used a similar hierarchical model with variance
244 partitioning to estimate group specific and pairwise *F*-statistics for the three groups delineated in
245 this study. We denote pairwise values of F_{ST} using one-letter abbreviations for the groups being
246 compared (e.g. $F_{ST\text{-CP}}$ indicates F_{ST} between Core and Periphery), and group specific values of
247 F_{ST} with the name of the group in subscripts. We constructed 95% confidence intervals of
248 multilocus *F*-statistics using bootstrap resampling (n = 100 replicates) in the HIERFSTAT package
249 (Goudet 2005) available in R. Along with estimation of *F*-statistics, we also assessed overall
250 levels of genetic diversity using multilocus estimates (i.e. means across SNPs) of observed and
251 expected heterozygosities (H_o and H_e) per population.
252 Presence of individuals with mixed ancestry, as identified using FASTSTRUCTURE, can
253 be a result of secondary contact, recent divergence causing incomplete lineage sorting, or the
254 presence of gene flow throughout the divergence history. Disentangling these explanations is
255 important, because it directly influences our understanding of the relative importance of intrinsic
256 and extrinsic factors in facilitating speciation. For instance, when speciation is recent or has
257 occurred with gene flow, we expect a heterogeneous landscape of genomic differentiation, such
258 that while most of the genome is freely introgressed between species, only a few genomic
259 regions associated with intrinsic or extrinsic factors inhibiting gene flow will exhibit elevated
260 differentiation (Wu 2001; Feder *et al*. 2012). However, if hybrids are formed in areas with novel
261 habitats, introgression might be selectively advantageous, such that heterozygotes at climate-

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

associated loci will confer higher fitness, and will lack elevated islands of differentiation. To infer the timing and influence of various demographic processes shaping the divergence history of our focal groups, we conducted demographic modeling using Diffusion Approximation for Demographic Inference (∂A∂I; Gutenkunst *et al.* 2009).We down-sampled the total SNP dataset for computational simplicity based on population genetic summary statistics and then randomly sampled one SNP per assembled contig to obtain a final dataset of 4,800 SNPs that were used in subsequent ∂A∂I analyses (Supporting Information, Appendix S1.C).

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 (Supporting Information, Fig. 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}, Fig. S1). 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 ($AIC_{model\ i} - AIC_{best\ model}$) scores (Burnham & Anderson 2002), calculated using results from the best replicate run (highest composite likelihood) for each model. Unscaled parameter estimates were obtained using a per-site substitution rate of 7.28×10^{-10} substitutions/site/year rate estimated for Pinaceae by De La Torre *et al.* (2017) and a generation time of 50 years.

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

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2013; Gompert*et al*. 2012a; Gompert & Buerkle 2011; Stankowski*et al*. 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 *Q* of 0.1 or lower were classified as pure *P. flexilis*, and those with intermediate *Q*-values were classified as admixed (e.g. Ortego *et al*. 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 2012; Gompert & Buerkle 2011). 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 center) 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 *et al.* 2012a). 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 *et al.* 2013) to test for convergence among replicated runs, as well as appropriate mixing along MCMC chains. We identified loci with excess ancestry (relative to the genome-wide average) as those with posterior α or β

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

credible intervals (CI; 95% equal-tail intervals) not containing zero. Moreover, 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 *et al*. 2012a). Several empirical and simulation based studies have demonstrated that both α and β can reflect patterns of selection in the hybrid zone (Gompert et al. 2012b; Janoušek *et al*. 2012), but the interpretation of these values is influenced by the underlying demographic scenario (Gompert & Buerkle 2012; Gompert *et al*. 2012a,2012b). Besides categorizing loci, we also tested for correlations among locus-specific $F_{CT\text{-}species}$, α, and β, with and without absolute values for α and β. The sign of the cline parameters (specifically β) have direct implications for inferring the processes maintaining species boundaries and hence were incorporated in correlation tests. Specifically, extremely positive values of β reflect strong selection against hybrids or population structure in the hybrid zone (Gompert *et al*. 2012b), 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), in order to extend our interpretations to a historical hybrid zone and categorize individuals into recent (F1s), advanced generation (FNs), and backcrossed hybrids (BC). This was done using a modified classification from Hamilton *et al.* (2013). Both BGC and INTROGRESS yielded very similar estimates of *h* (Pearson's $r = 0.7$, $p = 0.00042$), thus we used estimates from INTROGRESS due to the availability of inter-specific heterozygosity estimates

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

from this software. To test for the influence of exogenous 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.

Results

Ecological niche modeling and niche divergence

ENMs for each of the three groups used in this study (Fig. 1) had high predictive ability, as indicated by AUC and RTG values (Table 1). For Core and Periphery, several covariates stood out as important with precipitation seasonality (Bio15) being shared between Core and Periphery; however, altitude was consistently the most important variable for *P. flexilis* 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 (Supporting Information Fig. 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 (Supporting Information Fig. S2). The 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 (Fig. 2). Specifically, there was asymmetrical niche divergence between Core and Periphery, with the niche of periphery being conserved relative to the background of core. 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 (Supporting information, Fig. S3.A).

Population structure and divergence history

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

353 The PCA using 51 633 SNPs was consistent with trees sampled from Core being
354 differentiated from those of *P. flexilis*, which was most marked along PC1 (Fig. 3A). This PC
355 explained 0.9% of the total genetic variance, which was in line with the overall level of
356 differentiation estimated using hierarchical *F*-statistics ($F_{ST\text{-species}}$ = 0.021, 95% CI: 0.008–0.031).
357 Trees sampled from Periphery were located between those sampled from Core and *P. flexilis*
358 (Fig. 3A), in line with peripheral populations containing hybrids between the two parental
359 species. There was also a latitudinal gradient in the mean population *Q*-values, as estimated
360 using FASTSTRUCTURE, with Core populations exhibiting little to no ancestry from *P. flexilis*
361 and Periphery being a mixture of *P. flexilis* and Core (Fig. 3B). At the individual tree level, we
362 observed a strong negative correlation (Pearson's $r = -0.69$, p = 4.087e-07) between *Q*-values of
363 putative hybrids and latitude, which is consistent with a geographical gradient of genomic
364 introgression where trees geographically proximal to either parental species contain more
365 ancestry from that parental species. Multilocus estimation of differentiation between species
366 ($F_{\text{CT-species}}$) was 0.01 (95% CI: 0.005–0.018, Fig. 5A), while that between groups within *P.*
367 *strobiformis* ($F_{\text{CT-groups}}$) was 0.003 (95% CI: 0.0007–0.006). Multilocus F_{ST} within each group,
368 pairwise F_{ST} between each pair of group, and heterozygosities differed little among the three
369 groups, with the Core–*P. flexilis* comparison having the highest pairwise $F_{ST\text{-CF}}$ = 0.019 (Table
370 2). Although populations of Periphery exhibited slightly higher heterozygosities and F_{ST} values
371 ($F_{ST\text{-periphery}}$), this pattern was mainly driven by few populations, as indicated by the wider
372 confidence interval around these estimates (Table 2).

373 The best-supported demographic model was M_4, which is a model of symmetric ancient
374 gene flow between ancestral *P. strobiformis* and *P. flexilis* lineages, followed by contemporary
375 gene flow between Periphery and *P. flexilis* (Table 3; Fig 4). This model was supported by a

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

distinct minimum AIC score that was better than that of all other ∂A∂I models by a margin of 44.8 information units (ΔAIC_i = 44.8 or greater). The next best model, M_8, was similar to that of M_4, but without contemporary gene flow between *P. strobiformis* and *P. flexilis*, and a heterogeneous ancient gene flow between ancestral populations of the two parental species. Converted parameter estimates indicated that the species diverged 11.36 million years ago (Ma) in the Miocene, but that groups within *P. strobiformis* diverged 2.29 Ma during early Pleistocene (Fig. 4). Overall rates of gene flow between species were substantial for both historical and contemporary periods, however the contemporary gene flow between species was geographically restricted Periphery (Supporting Information Table S2). In addition, *P. flexilis* and Periphery experienced asymmetrical gene flow, which was larger in the direction of *P. flexilis* to Periphery (M_{PF} = 8.81 migrants/generation versus M_{FP} = 4.35). Periphery had the largest population size estimate, while *P. flexilis* was inferred to have experienced a reduction in population size through time.

Genomics of interspecific introgression

Values of *h* ranged from near zero to 0.80, with values around 0.20 being the most common thus suggesting overrepresentation of *P. strobiformis* ancestry (Fig. 5B). 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%), but 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 (Pearson's r = 0.41, p = 4.087e-07), precipitation

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

399 seasonality (Pearson's $r = -0.32$, $p = 0.006$), and mean temperature of the warmest quarter
400 (Pearson's $r = -0.18$, $p = 0.005$) had a strong influence on h, in line with the latter two being
401 important predictor variables for Periphery in our ENM.

402 The influence of selection is reflected in both α and β; however, exceptional values of α
403 are more indicative of selection favoring directional introgression, while extreme β values reflect
404 strong selection against hybrids or presence of population structure within the hybrid zone
405 (Gompert *et al.* 2011). Substantial variation was found in estimates of genomic cline parameters
406 (Fig. 5D,C), especially for α, with its range (−0.99 to 1.72) being 18.5-fold as wide as that of the
407 range for β (−0.068 to 0.078). Similar to the patterns observed in the distribution of h, an
408 asymmetry towards *P. strobiformis* ancestry was noted in the genomic cline estimates. From the
409 posterior distribution of α, we found 3,193 outlier loci, of which 570 (17.9%) had elevated
410 probabilities of *P. flexilis* ancestry (positive $\hat{\alpha}$), and 2,623 (82.1%) had elevated probabilities of
411 *P. strobiformis* ancestry (negative $\hat{\alpha}$). We identified fewer loci with excess ancestry, but in
412 contrast to the pattern for outlier loci those with excess ancestry favored *P. flexilis* over *P.*
413 *strobiformis* ancestry. Among the 287 loci with excess ancestry, 204 (71.1%) had excess *P.*
414 *flexilis* ancestry (i.e. lower 95% CI of $\alpha > 0$) and 83 (28.9%) had excess *P. strobiformis* ancestry
415 (i.e. upper 95% CI of $\alpha < 0$). The multilocus $F_{CT\text{-species}}$ estimate for loci with excess ancestry was
416 0.12 (95% CI: 0.09–0.13) while for outlier loci it was 0.058 (95% CI: 0.05–0.09). We did not
417 identify any loci that were β outliers or had excess ancestry indicated by β. Hierarchical $F_{\text{CT-species}}$
418 between species was negatively correlated with raw values of α (Pearson's r =−0.036, $p = 0.01$),
419 positively with raw values of β (Pearson's r = 0.048, $p = 0.0007$) and positively with absolute
420 values of both α (Pearson's r = 0.14, p = 2.2e-16) and β ((Pearson's r = 0.26, p = 2.2e-16) (Fig.
421 5E).

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discussion

We identified strong evidence supporting ecological divergence between *P. strobiformis* and *P. flexilis* despite extensive gene flow. Our findings are generally consistent with previous reports on the species examined here; however, in contrast to the recent divergence time estimates arrived at using chloroplast and mitochondrial loci (Moreno-Letelier *et al.* 2013), our demographic modeling results reject a hypothesis of extremely recent divergence between the two species. Instead, we support a model of ongoing speciation with gene flow that is driven and maintained primarily by extrinsic factors.

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 hybrid zone between *P. strobiformis* and *P. flexilis* (Steinhoff & Andresen 1971; Tomback & Achuff 2010; 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 a result of recent divergence. Under this scenario, we expect that niche differentiation would occur primarily along a few environmental variables that strongly influence fitness in the transitional environmental conditions, with little to no differentiation among groups on the other environmental axes. For example, precipitation seasonality was an important niche predictor for both Core and Periphery, although they were differentiated along this environmental axis (Supporting information, Fig. S3.B). This pattern reiterates the presence of hybrid populations in transitional environmental conditions and experiencing early stages of niche divergence from both parents.

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

444 In line with these results, precipitation seasonality and mean temperature of warmest
445 quarter had a strong negative association with genomic ancestry and contributed to the niche
446 divergence of Periphery. These two climatic variables influence plant evapotranspiration and
447 thus affect drought responses (Mishra & Singh 2010). Drought stress during the active growing
448 season is widely recognized as a limiting factor to plant growth in the western parts of North
449 America (Williams *et al*. 2010; Restaino *et al*. 2016) and our results are indicative of adaptive
450 divergence along a drought tolerance gradient between the three groups (Gitlin *et al* 2006; Allen
451 & Breshears 1998). Further, our study broadly agrees with other reports in *P. strobiformis*
452 indicating precipitation and altitude to be some key niche predictors (Aguirre-Gutiérrez *et al*.
453 2015; Shirk *et al*. 2017). Climatic clines of admixture and environmentally-dependent
454 maintenance of hybrid zones have been noted in other species of woody perennials in the genera
455 *Quercus* (Dodd & Afzal-Rafii 2004), *Picea* (Hamilton *et al.* 2013; De La Torre *et al*. 2014b),
456 *Rhododendron* (Milne *et al*. 2003), and *Pinus* (Cullingham *et al.* 2014*)*. An important variable
457 unaccounted for in our study, however, is white pine blister rust infestation (*Cronartium ribicola*
458 Fisch J.C), which influences stand dynamics of both species (Looney & Waring 2013) and drives
459 trade-offs between drought-tolerance and resistance in *P. flexilis* (Vogan & Schoettle 2015) .
460 However, given the low and recent incidence of infestation (Looney e*t al*. 2015), it is unlikely
461 that *C. ribicola* contributed towards niche divergence of the focal taxa. It is more likely that
462 under projected scenarios of environmental change, interactions between these selective forces
463 may occur in the future to influence ongoing speciation dynamics.

464 Despite fluctuations in suitable range size (Supporting information, Fig. S2) and previous
465 studies indicating reduction in genetic diversity at range margins using chloroplast markers
466 (Moreno-Letelier & Piñero 2009), we find no evidence for this in our study. This might be

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

467 explained by the asymmetry in gene flow between Periphery and *P. flexilis*, as inferred from our
468 demographic modeling results (Bridle & Vines 2007; Ortego *et al*. 2014). Moreover, evidence of
469 directional introgression from *P. flexilis* (positive α outliers) might also have facilitated
470 adaptation to transitional environmental conditions. Such novel allelic combinations have often
471 contributed to the ability of populations to colonize new niches that are intermediate to the
472 climatic conditions experienced by the parental species (De Carvalho *et al*. 2010; Hamilton *et al*.
473 2013; De La Torre *et al*. 2014b; Geraldes *et al*. 2014). Presence of a locally adapted and
474 historical hybrid zone is supported by the absence of β outliers in our genomic cline results
475 (Kamdem *et al*. 2016) and a recent study uncovering high Q_{ST} values associated with
476 physiological traits primarily linked to drought tolerance within the group Periphery (Goodrich *et*
477 *al*. 2016). The geographic cline in *h*, asymmetry in excess ancestry loci towards *P. flexilis*, and
478 elevated estimates of $F_{ST\text{–periphery,}}$ however, indicate the potential for geographically driven
479 neutral introgression to generate biased signals of local adaptation within the peripheral
480 populations (Geraldes *et al*. 2014). Ongoing investigations using replicate populations in the
481 hybrid zone across gradients of geographic proximity and climate similarity will be able to
482 address this issue in further detail (Lotterhos & Whitlock 2015; Riquet *et al*. 2017).

483 **Speciation with gene flow without islands of divergence**

484 Demographic modeling indicated that divergence of *P. strobiformis* and *P. flexilis* is not
485 recent (~11 Ma) on an absolute time scale and has occurred with continuous gene flow. The
486 presence of continual gene flow and absence of a period of allopatry, moreover, is also supported
487 by the L-shaped distribution of $F_{\text{CT-species}}$ values (Fig. 5A. Nosil & Feder 2012). Reduction in
488 overlapping niche suitability from LGM to present, between *P. strobiformis* and *P. flexilis*,
489 agrees with the best-supported demographic model indicating continuous and geographically

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

restricted contemporary gene flow (also see Moreno-Letelier & Piñero 2009). Contemporary reduction in N_e for *P. flexilis* from our demographic modeling 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, the two modeling approaches estimating population sizes across very different temporal scales, and a nonlinear relationship between habitat suitability and realized population sizes. Despite the potential for islands of divergence under a model of speciation with gene flow (Nosil 2008; Feder *et al.* 2012; 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 herein 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. The lack of islands of divergence is expected in conifers, given the prevalence of polygenic architectures defining continuous trait variation across species boundaries and the expected prevalence of soft sweeps (Pritchard & Rienzo 2010; Alberto *et al*. 2013; Rajora *et al*. 2016; Lind *et al.* 2017). Alternatively, given the large and complex genomes of conifers (reviewed by De La Torre *et al.* 2014a) our ddRADseq markers underrepresented genic regions, which are often identified as islands of divergence (Nosil & Feder 2012; Zhou *et al*. 2014; Moreno-Letelier & Barraclough 2015; Marques *et al.* 2017). For example, Moreno-Letelier & Barraclough (2015) demonstrated the potential for islands of divergence at drought-associated genes, which had a high average F_{ST} value of 0.33 (0.09–0.4) compared to the genome-wide estimate from this study ($F_{\text{ST-species}} = 0.02$). Future investigations using candidate gene approaches or exome capture

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

might thus be able to identify islands of divergence in conifers, although evidence of adaptation in complex genomes often also appears within intergenic regions (Li *et al.* 2012).

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 are often depicted by a mosaic landscape of genomic differentiation and ancestry. For instance, Coyne & Orr (1989), Noor & 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 & Otto 2011; Yeaman *et al*. 2016). However, extrinsic barriers can often result in the evolution of intrinsic barriers and subsequently become coupled with intrinsic barriers and with several 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 done, the correlation of *h* with drought related variables when coupled with the work of Moreno-Letelier & Barraclough (2015) implies that adaptive responses to drought stress likely contributed to the origin and maintenance of species boundaries in this system.

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.* Janoušek *et al.* 2012; Parchman *et al.* 2013; Gompert *et al.* 2014; 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. Under the 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 known patterns of forced crosses for these and other white pine species (Critchfield 1986). The limited evidence of intrinsic incompatibilities noted in our study could be generalized across conifers with similar divergence history, owing to their life history strategies such as long generation time and high dispersal capacity, which will restrict the evolution of post- and pre-zygotic isolating mechanism (Stacy *et al*. 2017). Absence of negative β outliers and of recent hybrids indicates widespread recombination within the hybrid zone and an intermediate stage of divergence between our focal species (Nosil *et al*. 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

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

557 extrinsic factors (Kamdem *et al.* 2016), which may be the first stage of coupling between
558 intrinsic and extrinsic barriers.

559 Contrary to the absence of β outliers, we identified many α outliers which is reflective of
560 a hybrid zone experiencing moderate selection pressure and high levels of gene flow from the
561 parental species (Gompert & Buerkle 2011). Limited variation in β is associated with a diffuse
562 genomic architecture of isolation (Gompert *et al.* 2012b), whereas the high genomic
563 heterogeneity in α, under the estimated demographic scenario, could imply divergent natural
564 selection operating within the hybrid zone (Gompert & Buerkle 2011). This agrees with the
565 higher values of multilocus F_{ST} within the putative hybrid zone ($F_{ST\text{-periphery}}$) and previous
566 evidence of local adaptation in this region (Goodrich *et al.* 2016). A similar genomic mosaic of
567 introgression has been noted across several studies (Lexer *et al.* 2010; Parchman *et al.* 2013;
568 Gompert *et al.* 2014; Lindtke *et al.* 2014; de Lafontaine *et al.* 2015) and is likely a result of
569 complex interactions between divergence history, selection, and genomic features.

570 Evidence of higher number of outliers from *P. strobiformis* and a negative association
571 between our cline parameters (α and β) could be explained by three processes: (i) intrinsic
572 incompatibilities resulting from Dobzhansky–Muller effects or more complex epistatic effects
573 disproportionally favoring allelic combinations from *P. strobiformis* in the hybrids relative to *P.*
574 *flexilis* parental background, (ii) widespread directional selection on alleles from *P. strobiformis*
575 in the hybrid zone leading to the formation of co-adapted gene complexes, and (iii) incomplete
576 lineage sorting resulting from recent divergence between Core and Periphery. In contrast to
577 inferences from the Engelmann–white spruce hybrid zone (De La Torre *et al.* 2014b), the
578 asymmetry of outlier loci is not due to high rates of gene flow from Core into Periphery, as the
579 best demographic model excluded gene flow between these groups (see Figure 5B). A higher

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

number of outlier loci with introgression favoring *P. strobiformis* is consistent with the strong influence of selection favoring alleles with *P. strobiformis* ancestry in the hybrid zone. Even without a linkage map, the cline results, along with precipitation seasonality being a strong shared niche predictor for Core and Periphery, points towards widespread directional introgression from *P. strobiformis* into the hybrid zone, which is consistent with local adaptation driving the evolution of co-adapted gene complexes from *P. strobiformis* and of emerging intrinsic incompatibilities (Gompert *et al.* 2012b). The geographic clines of *h*, despite the absence of current gene flow between the Core and Periphery, also points 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. Further, directional introgression of alleles from Core into Periphery might also have led to the asymmetrical niche divergence between these groups.

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; Zhou *et al*. 2014; Stacy *et al*. 2017). Integrating the existing genomic dataset 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, co-incidence between the steepness of genomic, geographic, and trait specific clines would indicate a dominant role of extrinsic factors in facilitating divergence and speciation (Holliday *et al.* 2010; De La Torre *et al*. 2015; Stankowski *et al*. 2015; Ryan *et al*. 2017). Alternatively, the presence of several loci showing steep clines, but lacking climatic or functional associations

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

603 would indicate a dominance of intrinsic barriers (Ryan *et al*. 2017). Although the genomic cline
604 analysis used in this study provided key insights into a complexity of species isolation, it lacks
605 sufficient power to account for complex epistatic effects (Gompert & Buerkle 2011). These have
606 likely played a key role in ecological speciation and in initiating the evolution of reproductive
607 isolation (Lindtke *et al*. 2012; Flaxman *et al.* 2014). Ultimately, furthering genomic resources
608 will help test whether absolute measures of divergence are correlated with recombination rate,
609 and will account for the non-random genomic distribution of climate-associated genes and their
610 tendency to co-localize in areas of reduced recombination (Wolf & Ellegren 2016). This study,
611 however, provides concrete evidence of ecological speciation with gene flow, the presence of a
612 historical hybrid zone maintained by extrinsic factors, and early stages of coupling between
613 disruptive selection and intrinsic barriers contributing towards diversification. Whether these
614 patterns hold generally for speciation within conifers, given their life history characteristics as
615 well as their complex and large genomes, is thus a worthwhile area of future research.
616

617 ## Acknowledgements

618 This research was funded by U.S. National Science Foundation (NSF) grants EF-1442486
619 (Eckert), and EF-1442597 (Waring), as well as by Northern Arizona University's Technology
620 Research Initiative Program, USDA Forest Service Forest Health Protection Gene conservation
621 program, USDA Forest Service National Fire Plan Award 01.RMRS.B.6 (Schoettle), Virginia
622 Commonwealth University (VCU) Department of Biology, and VCU Integrated Life Sciences.
623 We thank Z. Gompert and T. Parchman for providing R code that aided plotting the BGC results.
624 We also acknowledge our field sampling and lab work crew (see Supporting information,
625 Appendix S1.D). This work would not have been possible without generous computational

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

626 resources provided by the VCU Center for High Performance Computing and the Brigham
627 Young University Fulton Supercomputing Lab.

References

628
629
630 Agrawal AF, Feder JL, Nosil P (2011) Ecological Divergence and the Origins of Intrinsic Postmating
631 Isolation with Gene Flow. *International Journal of Ecology*, **2011**, e435357.
632 Aguirre-Gutiérrez J, Serna-Chavez HM, Villalobos-Arambula A et al (2015) Similar but not equivalent:
633 ecological niche comparison across closely–related Mexican white pines. *Diversity and*
634 *Distributions*, **21**, 245-257.
635 Aiello-Lammens ME, Boria RA, Radosavljevic A, Vilela B, Anderson RP (2015) spThin: an R package
636 for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*,
637 **38**, 541–545.
638 Allen CD & Breshears DD (1998). Drought-induced shift of a forest–woodland ecotone: Rapid landscape
639 response to climate variation. *Proc. Nat. Acad Sci.* **95**, 14839–14842.
640 Alberto FJ, Aitken SN, Alía R *et al.* (2013) Potential for evolutionary responses to climate change –
641 evidence from tree populations. *Global Change Biology*, **19**, 1645–1661.
642 Andrew RL & Rieseberg LH (2013). Divergence is focused on few genomic regions early in speciation:
643 incipient speciation of sunflower ecotypes. *Evolution*, **67**, 2468–2482.
644 Barton NH (2001) The role of hybridization in evolution. *Molecular Ecology*, **10**, 551–568.
645 Barton NH, De Cara MAR (2009) The Evolution of Strong Reproductive Isolation. *Evolution*, **63**, 1171–
646 1190.
647 Barton NH, Hewitt GM (1985) Analysis of Hybrid Zones. *Annual Review of Ecology and Systematics*, **16**,
648 113–148.
649 Benkman CW, Balda RP, Smith CC (1984) Adaptations for Seed Dispersal and the Compromises Due to
650 Seed Predation in Limber Pine. *Ecology*, **65**, 632–642.
651 Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why genome scans
652 may fail to map local adaptation genes: The coupling hypothesis. *Molecular Ecology*, **20**, 2044–
653 2072.
654 Bisbee J (2014) Cone morphology of the *Pinus ayacahuite-flexilis* complex of the southwestern United
655 States and Mexico. *Bulletin of the Cupressus conservation project*, **3**, 3–33.
656 Bivand R, Keitt T, Rowlingson B *et al.* (2017) *rgdal: Bindings for the Geospatial Data Abstraction*
657 *Library*.

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Borgman EM, Schoettle AW, Angert AL (2015) Assessing the potential for maladaptation during active management of limber pine populations: a common garden study detects genetic differentiation in response to soil moisture in the Southern Rocky Mountains. *Canadian Journal of Forest Research*, **45**, 496–505.

Bridle JR, Vines TH (2007) Limits to evolution at range margins: when and why does adaptation fail? *Trends in Ecology & Evolution*, **22**, 140–147.

Burnham KP, Anderson DR (2002) *Model selection and multimodel inference: a practical information-theoretic approach*. Springer, New York.

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.

Carstens BC, Brennan RS, Chua V *et al.* (2013) Model selection as a tool for phylogeographic inference: an example from the willow *Salix melanopsis*. *Molecular Ecology*, **22**, 4014–4028.

Christe C, Stölting KN, Paris M *et al.* (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.

Coyne J, Orr HA (1989) Patterns of speciation in Drosophilla. *Evolution*, **43**, 362–381.

Critchfield WB (1986). Hybridization and classification of the white pines (*Pinus* section *Strobus*). *Taxon*, **35**, 647-656.

Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, **23**, 3133–3157.

Cullingham CI, Cooke JEK, Coltman DW (2014) Cross-species outlier detection reveals different evolutionary pressures between sister species. *The New Phytologist*, **204**, 215–229.

De Carvalho D, Ingvarsson PK, Joseph J *et al.* (2010) Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula* L.), a widespread forest tree. *Molecular Ecology*, **19**, 1638–1650.

De La Torre AR, Birol I, Bousquet J *et al.* (2014a) Insights into Conifer Giga-Genomes. *Plant Physiology*, **166**, 1724-1732.

De La Torre AR, Wang T, Jaquish B, Aitken SN (2014b) 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.

De La Torre AR, Ingvarsson PK, & Aitken SN (2015). Genetic architecture and genomic patterns of gene flow between hybridizing species of *Picea*. *Heredity*, **115**, 153–164.

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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*(6), 1363-1377. Chicago

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.

Dodd RS, Afzal-Rafii Z (2004) Selection and dispersal in a multispecies oak hybrid zone. *Evolution*, **58**, 261–269.

Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends in Genetics*, **28**, 342–350.

Fitzpatrick BM (2012) Estimating ancestry and heterozygosity of hybrids using molecular markers. *BMC Evolutionary Biology*, **12**, 131.

Fitzpatrick BM (2013) Alternative forms for genomic clines. *Ecology and Evolution*, **3**, 1951–1966.

Flaxman SM, Wacholder AC, Feder JL, Nosil P (2014) Theoretical models of the influence of genomic architecture on the dynamics of speciation. *Molecular Ecology*, **23**, 4074–4088.

Geraldes A, Farzaneh N, Grassa CJ *et al.* (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.

Gitlin AR, Sthultz CM, Bowker MA *et al* (2006) Mortality Gradients within and among Dominant Plant Populations as Barometers of Ecosystem Change During Extreme Drought. *Conservation Biology*, **20**, 1477–1486.

Gompert Z, Buerkle CA (2010) introgress: a software package for mapping components of isolation in hybrids. *Molecular Ecology Resources*, **10**, 378–384.

Gompert Z, Buerkle CA (2011) Bayesian estimation of genomic clines. *Molecular Ecology*, **20**, 2111–2127.

Gompert Z, Buerkle CA (2012) bgc: Software for Bayesian estimation of genomic clines. *Molecular Ecology Resources*, **12**, 1168–1176.

Gompert Z, Lucas LK, Buerkle CA *et al.* (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.

Gompert Z, Parchman TL, Buerkle CA (2012a) Genomics of isolation in hybrids. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 439–450.

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

724 Gompert Z, Lauren LK, Chris NC *et al.* (2012b) Genomic regions with a history of divergent selection
725 affect fitness of hybrids between two Butterfly Species. *Evolution*, **66**, 2167-2181.
726 Goodrich BA, Waring KM, Kolb TE (2016) Genetic variation in *Pinus strobiformis* growth and drought
727 tolerance from southwestern US populations. *Tree Physiology*, **36**, 1219–1235.
728 Goudet J (2005) hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular*
729 *Ecology Notes*, **5**, 184–186.
730 Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the Joint Demographic
731 History of Multiple Populations from Multidimensional SNP Frequency Data. *PLOS Genetics*, **5**,
732 e1000695.
733 Hamilton JA, Lexer C, Aitken SN (2013) Genomic and phenotypic architecture of a spruce hybrid zone (
734 *Picea sitchensis* × *P. glauca*). *Molecular Ecology*, **22**, 827–841.
735 Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate
736 surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.
737 Hijmans RJ, Etten J van, Cheng J *et al.* (2016) *raster: Geographic Data Analysis and Modeling*.
738 Hijmans RJ, Phillips S, Elith JL and J (2017) *dismo: Species Distribution Modeling*.
739 Holliday JA, Ritland K, & Aitken SN (2010). Widespread, ecologically relevant genetic markers
740 developed from association mapping of climate-related traits in Sitka spruce (*Picea sitchensis*).
741 *The New Phytologist*,**188**, 501–514
742 Ikeda DH, Max TL, Allan GJ *et al.* (2017) Genetically informed ecological niche models improve climate
743 change predictions. *Global Change Biology*, **23**, 164–176.
744 Janoušek V, Wang L, Luzynski K *et al.* (2012) Genome-wide architecture of reproductive isolation in a
745 naturally occurring hybrid zone between *Mus musculus musculus* and *M. m. domesticus*.
746 *Molecular Ecology*, **21**, 3032–3047.
747 Kamdem C, Fouet C, Gamez S, White BJ (2016) Genomic signatures of introgression at late stages of
748 speciation in a malaria mosquito. *bioRxiv*, 068239.
749 Kane NC, King MG, Barker MS *et al.* (2009) Compartive genomic and population genetic analyses
750 indicate highly porous genomes and high levels of gene flow beween divergent Helianthus
751 species. *Evolution; international journal of organic evolution*, **63**, 2061–2075.
752 Knowles LL, Carstens BC, Keat ML (2007) Coupling Genetic and Ecological-Niche Models to Examine
753 How Past Population Distributions Contribute to Divergence. *Current Biology*, **17**, 940–946.
754 Kulmuni J, Westram AM (2017) Intrinsic incompatibilities evolving as a by-product of divergent
755 ecological selection: Considering them in empirical studies on divergence with gene flow.
756 *Molecular Ecology*, **26**, 3093–3103.

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Lackey ACR, Boughman JW (2016) Evolution of reproductive isolation in stickleback fish: Evolution of isolation in stickelback fish. *Evolution*, **71**, 357–372.

Lexer C, Joseph JA, Loo MV *et al*. (2010). Genomic Admixture Analysis in European *Populus* spp. Reveals Unexpected Patterns of Reproductive Isolation and Mating. *Genetics*, **186**, 699–712.

Li X, Zhu C, Yeh CT *et al*. (2012). Genic and nongenic contributions to natural variation of quantitative traits in maize. *Genome Research*, **22**, 2436–2444.

Lindtke D, Buerkle CA, Barbará T, *et al*. (2012). Recombinant hybrids retain heterozygosity at many loci: new insights into the genomics of reproductive isolation in *Populu*s. *Molecular Ecology*, **21**, 5042–5058.

Lindtke D, Gompert Z, Lexer C, Buerkle CA (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.

Lind BM, Freidline CJ, Wegrzyn JL *et al*. (2017) Water availability drives signatures of local adaptation in whitebark pine (*Pinus albicaulis* Englm.) across fine spatial scales of the Lake Tahoe Basin, USA. *Molecular Ecology*, **26**, 3168-3185.

Looney CE, Waring KM (2013) *Pinus strobiformis* (southwestern white pine) stand dynamics, regeneration, and disturbance ecology: A review. *Forest Ecology and Management*, **287**, 90–102.

Looney CE, Waring KM, Fairweather ML (2015) Monitoring the health of *Pinus strobiformis*: early impacts of white pine blister rust invasion. In: Forest Health Monitoring: National Status, Trends, and Analysis. USDA Forest service, pp:167-175.

Losos JB, Arnold SJ, Bejerano G *et al.* (2013) Evolutionary Biology for the 21st Century. *PLoS Biology*, **11**.

Lotterhos KE & Whitlock MC (2015) The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology*, **24,** 1031-1046.

Marques DA, Lucek K, Haesler MP *et al*. (2017). Genomic landscape of early ecological speciation initiated by selection on nuptial colour. *Molecular Ecology*, **26**, 7–24.

Mastretta-Yanes A, Moreno-Letelier A, Piñero D, Jorgensen TH, Emerson BC (2015) Biodiversity in the Mexican highlands and the interaction of geology, geography and climate within the Trans-Mexican Volcanic Belt. *Journal of Biogeography*, **42**, 1586–1600.

McVean G (2009). A Genealogical Interpretation of Principal Components Analysis. *PLOS Genetics*, **5**, e100068

Milne RI, Terzioglu S, Abbott RJ (2003) A hybrid zone dominated by fertile F1s: maintenance of species barriers in *Rhododendron*. *Molecular Ecology*, **12**, 2719–2729.

Mishra AK, Singh VP (2010). A review of drought concepts. *Journal of Hydrology*, **391,** 202-216.

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Moore WS (1977) An Evaluation of Narrow Hybrid Zones in Vertebrates. *The Quarterly Review of Biology*, **52**, 263–277.

Moreno-Letelier A & Piñero D (2009) Phylogeographic structure of *Pinus strobiformis* Engelm. across the Chihuahuan Desert filter-barrier. *Journal of Biogeography*, **36**, 121-131.

Moreno-Letelier A, Barraclough TG (2015) Mosaic genetic differentiation along environmental and geographic gradients indicate divergent selection in a white pine species complex. *Evolutionary Ecology*, **29**, 733–748.

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.

Neale DB, Kremer A (2011) Forest tree genomics: growing resources and applications. *Nature Reviews Genetics*, **12**, 111–122.

Noor MF, Bennett SM (2009) Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*, **103**, 439–444.

Nosil P (2008) Speciation with gene flow could be common. *Molecular Ecology*, **17**, 2103–2106.

Nosil P, Feder JL (2012) Genomic divergence during speciation: causes and consequences. *Phil. Trans. R. Soc. B*, **367**, 332–342.

Nosil P, Harmon LJ, Seehausen O (2009) Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*, **24**, 145–156.

Ortego J, Gugger PF, Riordan EC, Sork VL (2014) Influence of climatic niche suitability and geographical overlap on hybridization patterns among southern Californian oaks (B Emerson, Ed,). *Journal of Biogeography*, **41**, 1895–1908.

Ortego J, Noguerales V, Gugger PF, Sork VL (2015) Evolutionary and demographic history of the Californian scrub white oak species complex: an integrative approach. *Molecular Ecology*, **24**, 6188–6208.

Parchman TL, Gompert Z, Braun MJ *et al.* (2013) The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. *Molecular Ecology*, **22**, 3304–3317.

Parchman TL, Gompert Z, Mudge J *et al.* (2012) Genome-wide association genetics of an adaptive trait in lodgepole pine: Association mapping of serotiny. *Molecular Ecology*, **21**, 2991–3005.

Patterson N, Price AL, Reich D (2006) Population Structure and Eigenanalysis. *PLOS Genetics*, **2**, e190

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.

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

824 Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double Digest RADseq: An
825 Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model
826 Species. *PLOS ONE*, **7**, e37135.
827 Petit RJ, Excoffier L (2009) Gene flow and species delimitation. *Trends in Ecology & Evolution*, **24**,
828 386–393.
829 Petit RJ, Hampe A (2006) Some Evolutionary Consequences of Being a Tree. *Annual Review of Ecology,*
830 *Evolution, and Systematics*, **37**, 187–214.
831 Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic
832 distributions. *Ecological Modelling*, **190**, 231–259.
833 Pritchard JK, Rienzo AD (2010) Adaptation – not by sweeps alone. *Nature Reviews Genetics*, **11**, 665–
834 667.
835 Puritz JB, Hollenbeck CM, Gold JR (2014) *dDocent*□: a RADseq, variant-calling pipeline designed for
836 population genomics of non-model organisms. *PeerJ*, **2**, e431.
837 R Core Team (2017) R: A Language and Environment for Statistical Computing. Vienna, Austria□: the R
838 Foundation for Statistical Computing. *R Foundation for Statistical Computing, Vienna, Austria*.
839 Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: Variational Inference of Population Structure
840 in Large SNP Data Sets. *Genetics*, **197**.
841 Rajora OP, Eckert AJ, Zinck JWR (2016) Single-Locus versus Multilocus Patterns of Local Adaptation to
842 Climate in Eastern White Pine (*Pinus strobus*, Pinaceae). *PLOS ONE*, **11**, e0158691.
843 Rambaut, Suchard, Marc, Drummond, Alexei (2013) Tracer, Version 1.6.
844 Rehfeldt GE (1999) Systematics and genetic structure of Ponderosae taxa (Pinaceae) inhabiting the
845 mountain islands of the Southwest. *Am J Bot*, **86**, 741-752.
846 Restaino CM, Peterson DL, Littell J (2016). Increased water deficit decreases Douglas fir growth
847 throughout western US forests. *Proceedings of the National academy of Sciences*, **113,** 9557-
848 9562.
849 Riquet F, Haag-Liautard C, Woodall L *et al*. (2017) Parallel use of a shared genomic island of speciation
850 in clinal and mosaic hybrid zones between cryptic seahorse lineages. *bioRxiv*, 161786
851 Roesti M, Hendry AP, Salzburger W, Berner D (2012) Genome divergence during evolutionary
852 diversification as revealed in replicate lake-stream stickleback population pairs. *Molecular*
853 *Ecology*, **21**, 2852–2862.
854 Rundle HD (2002) A test of ecologically dependent postmating isolation between sympatric sticklebacks.
855 *Evolution*, **56**, 322–329.
856 Ryan S, Fontaine MC, Scriber JM *et al.* (2017) Patterns of divergence across the geographic and genomic
857 landscape of a butterfly hybrid zone associated with a climatic gradient. *bioRxiv*, 149179.

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

858 Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. *Nature Reviews*
859 *Genetics*, **14**, 807–820.
860 Schield DR, Adams RH, Card DC *et al.* (2017) Insight into the roles of selection in speciation from
861 genomic patterns of divergence and introgression in secondary contact in venomous rattlesnakes.
862 *Ecology and Evolution*, **7**, 3951–3966.
863 Schluter D, Conte GL (2009) Genetics and ecological speciation. *Proceedings of the National Academy of*
864 *Sciences*, **106**, 9955–9962.
865 Shirk AJ, Cushman SA, Wehenkel CA *et al* (2017) Southwestern white pine (*Pinus strobiformis*) species
866 distribution models predict large range shift and contraction due to climate change. *Forest*
867 *Ecology & Management (in review)*.
868 Sork VL, Aitken SN, Dyer RJ *et al.* (2013) Putting the landscape into the genomics of trees: approaches
869 for understanding local adaptation and population responses to changing climate. *Tree Genetics*
870 *& Genomes*, **9**, 901–911.
871 Stacy EA, Paritosh B, Johnson MA, Price DK (2017) Incipient ecological speciation between
872 successional varieties of a dominant tree involves intrinsic postzygotic isolating barriers. *Ecology*
873 *and Evolution*, **7**, 2501–2512.
874 Stankowski S, Sobel JM, Streisfeld MA (2015) The geography of divergence with gene flow facilitates
875 multitrait adaptation and the evolution of pollinator isolation in *Mimulus aurantiacus*. *Evolution*,
876 **69**, 3054–3068.
877 Steinhoff RJ, Andresen JW (1971) Geographic variation in *Pinus flexilis* and *Pinus strobiformis* and its
878 bearing on their taxonomic status. Silvae Genet, **20**, 159–167.
879 Tine M, Kuhl H, Gagnaire P-A *et al.* (2014) European sea bass genome and its variation provide insights
880 into adaptation to euryhalinity and speciation. *Nature Communications*, **5**.
881 Tomback DF, Achuff P (2010) Blister rust and western forest biodiversity: ecology, values and outlook
882 for white pines. *Forest Pathology*, **40**, 186–225.
883 Tomback DF, Samano S, Pruett EL, Schoettle AW, others (2011) Seed dispersal in limber and
884 southwestern white pine: comparing core and peripheral populations. In: *The Future of High-*
885 *Elevation, Five-Needle White Pines in Western North America: Proceedings of the High Five*
886 *Symposium. Proceedings RMRS-P-63. Fort Collins, CO: US Department of Agriculture, Forest*
887 *Service, Rocky Mountain Research Station*, pp. 69–71.
888 Via S, Bouck AC, Skillman S (2000) Reproductive isolation between divergent races of pea aphids on
889 two hosts. II. Selection against migrants and hybrids in the parental environments. *Evolution*, **54**,
890 1626–1637.

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Vogan PJ, Schoettle AW (2015). Selection for resistance to white pine blister rust affects the abiotic stress tolerances of limber pine. *Forest Ecology and Management*, **344**, 110-119.

Warren DL, Glor RE, Turelli M (2008) Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution*, **62**, 2868–2883.

Williams AP, Allen CD, Millar CI, Swetnam TW *et al.* (2010). Forest responses to increasing aridity and warmth in the southwestern United States. *Proceedings of the National academy of Sciences*. **107**, 21289-21294.

Wolf JBW & Ellegren H (2016). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*,**18**, 87–100.

Wu CI (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–865.

Yang RC (1998). Estimating hierarchical *F*-statistics. *Evolution*, **52**, 950-956.

Yeaman S, Otto SP (2011). Establishment and Maintenance of Adaptive Genetic Divergence Under Migration, Selection, and Drift. *Evolution*, **65**, 2123–2129.

Yeaman S, Aeschbacher S, Bürger R (2016) The evolution of genomic islands by increased establishment probability of linked alleles. *Molecular Ecology*, **25**, 2542–2558.

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.

Data accessibility

Raw reads generated during this study are available at NCBI SRA database (SRXXXXXXX). Genotype file following SNP calling in 012 format and admixture proportions per individual tree are available at Dryad (doi: XXXX).

Author contributions

The study was designed by KMW, AVW, AJE, LFR, CW, and SC. Field sampling was performed by AWS, FMF, LFR, MSG, CW, ALS, and KMW. Funding for this study was procured by KMW, AWS, and AJE. MM, JCB and CF performed the data analysis. MM generated the genomic data and wrote the manuscript. All authors edited the article and have approved the version for submission.

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

Groups	AUC	RTG	RTG importance #	Permutation importance¢	Percent contribution¢	Regression coefficient importance *
Core	0.97	2.51	Bio15[1], Bio4[2]	Bio4	Altitude	Bio4
Periphery	0.99	3.92	Bio9[3]	Bio10[4], Bio9, Bio6[5]	Altitude	Bio15
P. flexilis	0.94	1.72	Altitude	Altitude	Altitude	Altitude

AUC: Area under the curve; RTG: Regularized training gain

#: Variables that caused maximum reduction in the total RTG when omitted from the model and the variable with the most contribution to RTG

*: Sum of absolute values of regression coefficient (λ) across various predictor transformations or feature classes used in MAXENT

¢: Variables with the highest permutation or percentage importance

[1]Precipitation seasonality, [2]Temperature seasonality, [3]Mean temperature of the driest quarter, [4]Mean temperature of the warmest quarter, [5]Minimum temperature of the coldest month

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

928 **Table 2.** Estimates of genetic diversity and divergence within and across the three groups, compared to a genome-wide $F_{ST\text{-}species}$ of
929 0.02 (95% CI: 0.008–0.03) and $F_{ST\text{-}strobiformis}$ of 0.009 (95% CI: 0.007–0.014).

Group	**Multilocus F_{ST} (95% CI)**	**Pairwise F_{ST} (95% CI)**	**Mean H_e ± s.d.**	**Mean H_o ± s.d.**
Core	0.003 (0.0025–0.0034)	periphery: 0.009(0.001–0.023) *P. 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

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 3. Model composite likelihoods and AIC model selection results for 11 alternative demographic models of *P. strobiformis* (core and periphery)–*P. flexilis* divergence. The best supported model, that with the minimum AIC score (hence, $\Delta AIC_i = 0$), is 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	1778.29	65.44
M_2	Secondary contact (Periphery–*P. flexilis*)	−886.227416	7	1786.45	73.60
M_3	Ancient gene flow (speciation with gene flow)	−888.003307	7	1790.01	77.16
***M*₄**	**Ancient gene flow, plus Periphery–*P. flexilis* gene flow**	**−847.424540**	**9**	**1712.85**	**0.00**
M_5	Ancient gene flow, plus Core–periphery gene flow	−885.428135	9	1788.86	76.01
M_6	Secondary contact (Periphery–*P. flexilis*) and Core–Periphery gene flow	−883.949484	10	1787.90	75.05
M_7	Ancient gene flow, followed by Periphery–*P. flexilis* gene flow, and Core–Periphery gene flow	−892.210862	9	1806.42	93.57
***M*₈**	**Heterogeneous ancient gene flow**	**−869.824520**	**14**	**1757.65**	**44.80**
M_9	Heterogeneous ancient gene flow, plus Core–Periphery gene flow	−884.511096	11	1791.02	78.17
M_{10}	Heterogeneous gene flow during secondary contact (Periphery–*P. flexilis*), and Core-Periphery gene flow	−902.279445	9	1828.56	115.71
M_{11}	Heterogeneous ancient gene flow, followed by heterogeneous gene flow between Periphery–*P. flexilis*, and between Core–Periphery	−922.814525	11	1873.63	160.78

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

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure Captions

Fig. 1. Map of sampling localities (black dots) overlaid on polygons showing geographical ranges of *Pinus strobiformis* (green) and *P. flexilis* (blue). The corresponding locality information is available in Table S1 of the Supporting Information.

Fig. 2. Results of niche divergence tests (Schoener's *D*) for all pairwise comparisons among groups. Histograms indicate the background levels of niche divergence and arrows indicate the observed value of Schoener's *D* for each pair compared.

Fig. 3A) Results of population genetic structure analysis using PCA on 51 633 SNPs. **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 *P. strobiformis* and *P. flexilis.*

Fig. 4. The best-supported model from $\partial A\partial I$ analysis. This figure shows the optimized parameter estimates for divergence times (T_i) in units of millions of years ago (Ma), reference effective population size (theta; or after conversion, N_{eref}), lineage population sizes (N_i), and rates of gene flow (M_{ij}) for the optimal model determined by AIC model selection (see results in Table 2).

Fig. 5A) 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}

bioRxiv preprint doi: https://doi.org/10.1101/185728; this version posted September 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

959

960

961 ## Supporting Information

962 **Table S1.** Sampling locations used in this study, their classification into Core and Periphery, and

963 mean admixture proportions.

964 **Table S2**. Raw and converted parameter estimates from the $\partial A \partial I$ model that was best-supported

965 by AIC model selection.

966 **Fig. S1.** Schematics and parameter details for each of the 11 demographic models for the

967 divergence of core and periphery groups within *P. strobiformis* and *P. flexilis* run in our $\partial A \partial I$

968 analysis. Parameters include divergence times (T_i), population sizes (N_i), homogeneous rates of

969 gene flow (M_{ij}, gene flow from lineage j to i) and genomically heterogeneous rates of gene flow

970 (M_{ijh}).

971 **Fig. S2.** Ecological niche model projections for Core, Periphery, and *P. flexilis*, under present

972 and past climate

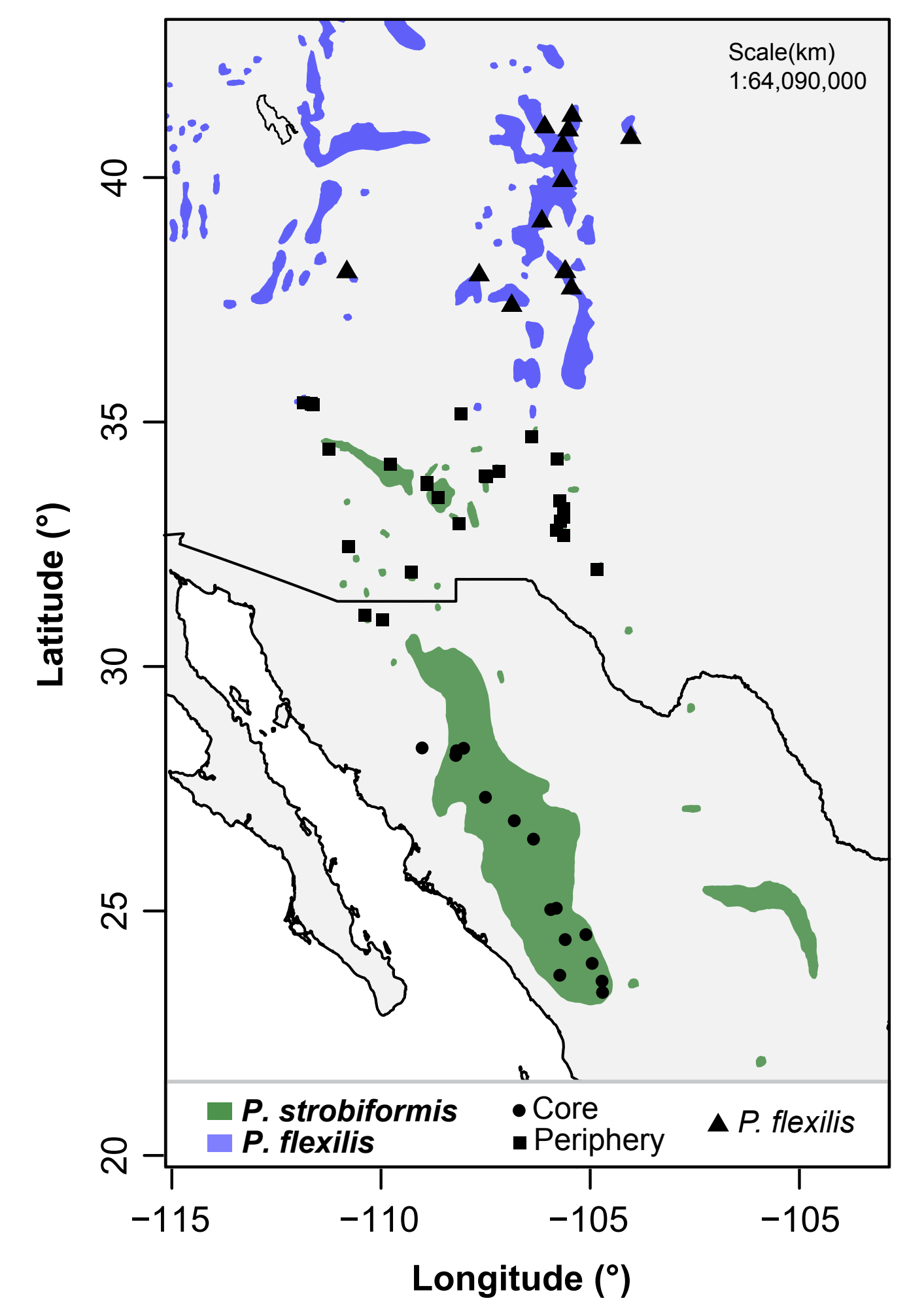
973 **Fig. S3. A**) Climate PCA with variables used in the ENMs **B)** Distribution of precipitation

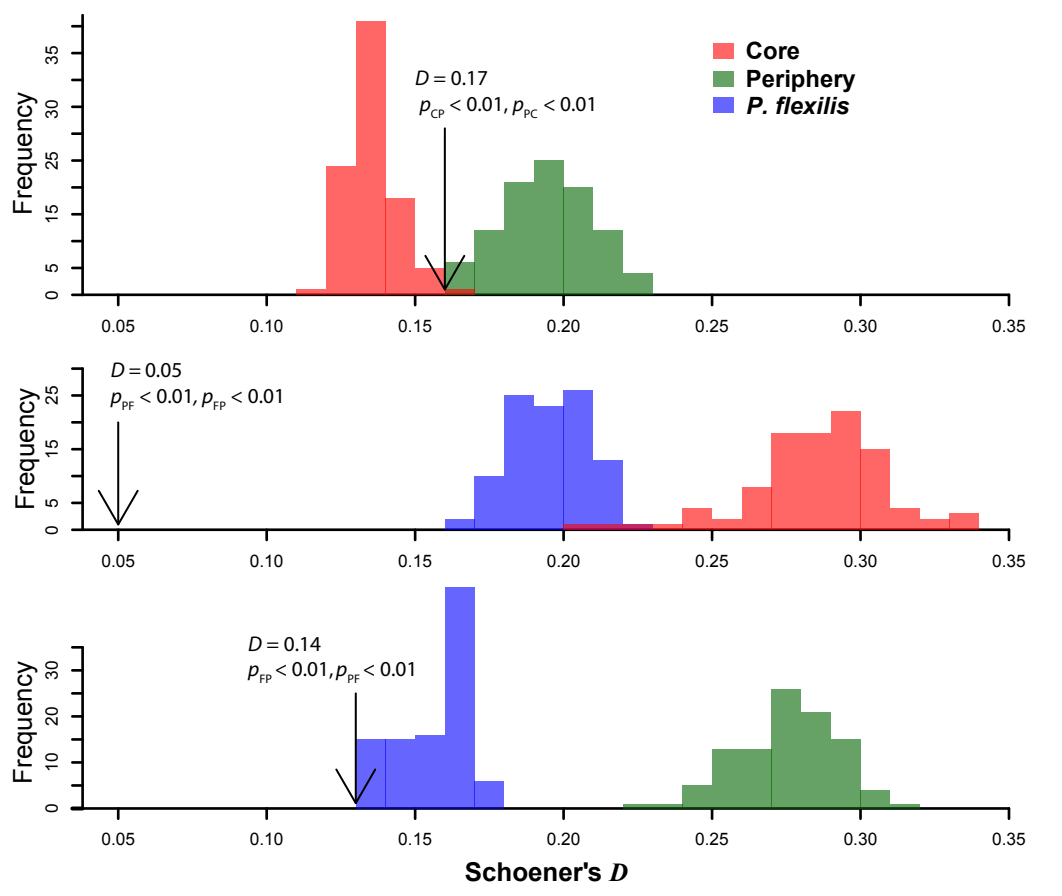
974 seasonality (Bio15) at presence locations of Core & Periphery.

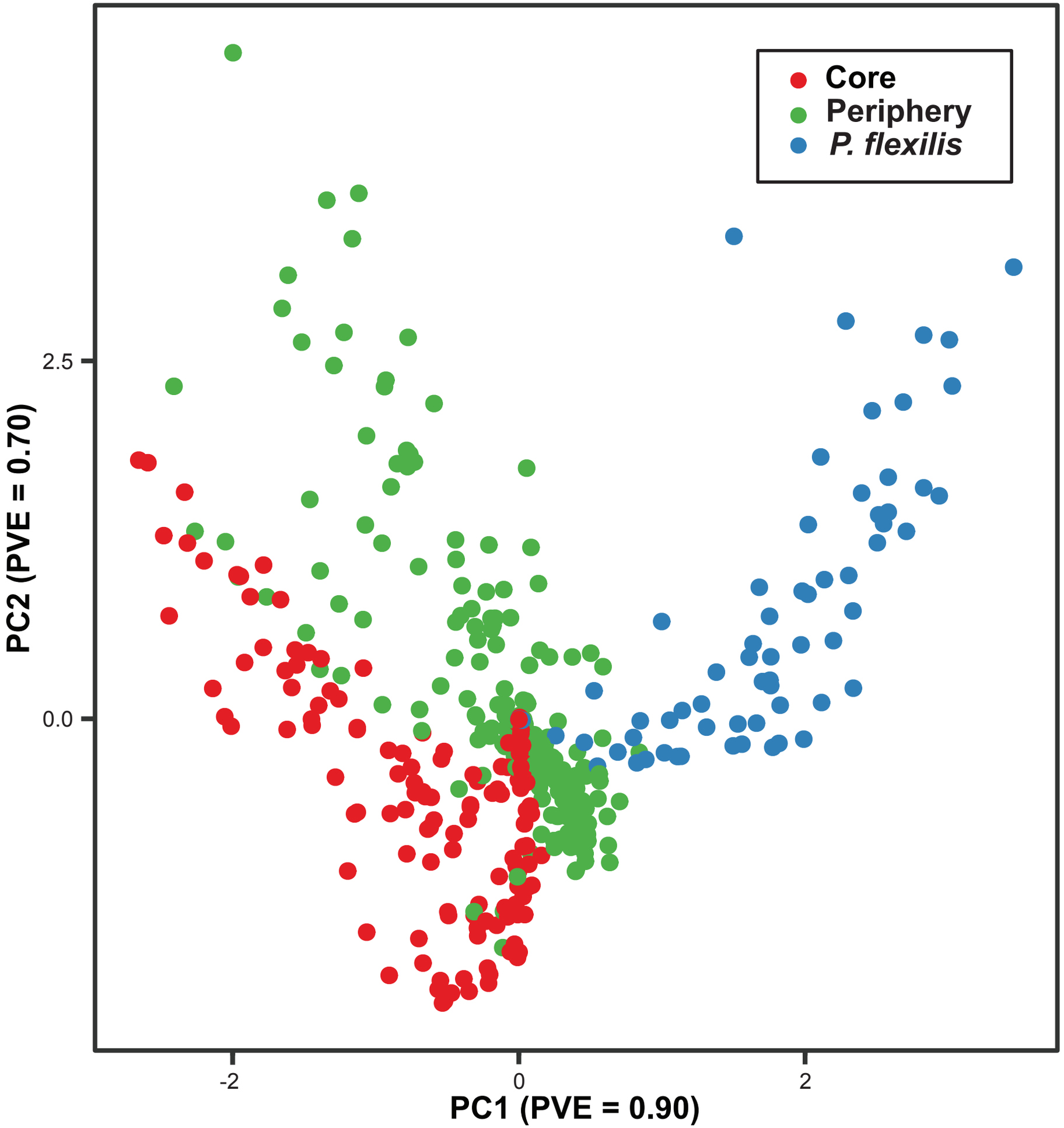
975 **Appendix S1.A)** Filtering of occurrence records and ENM details, **B)** Bioinformatic pipeline for

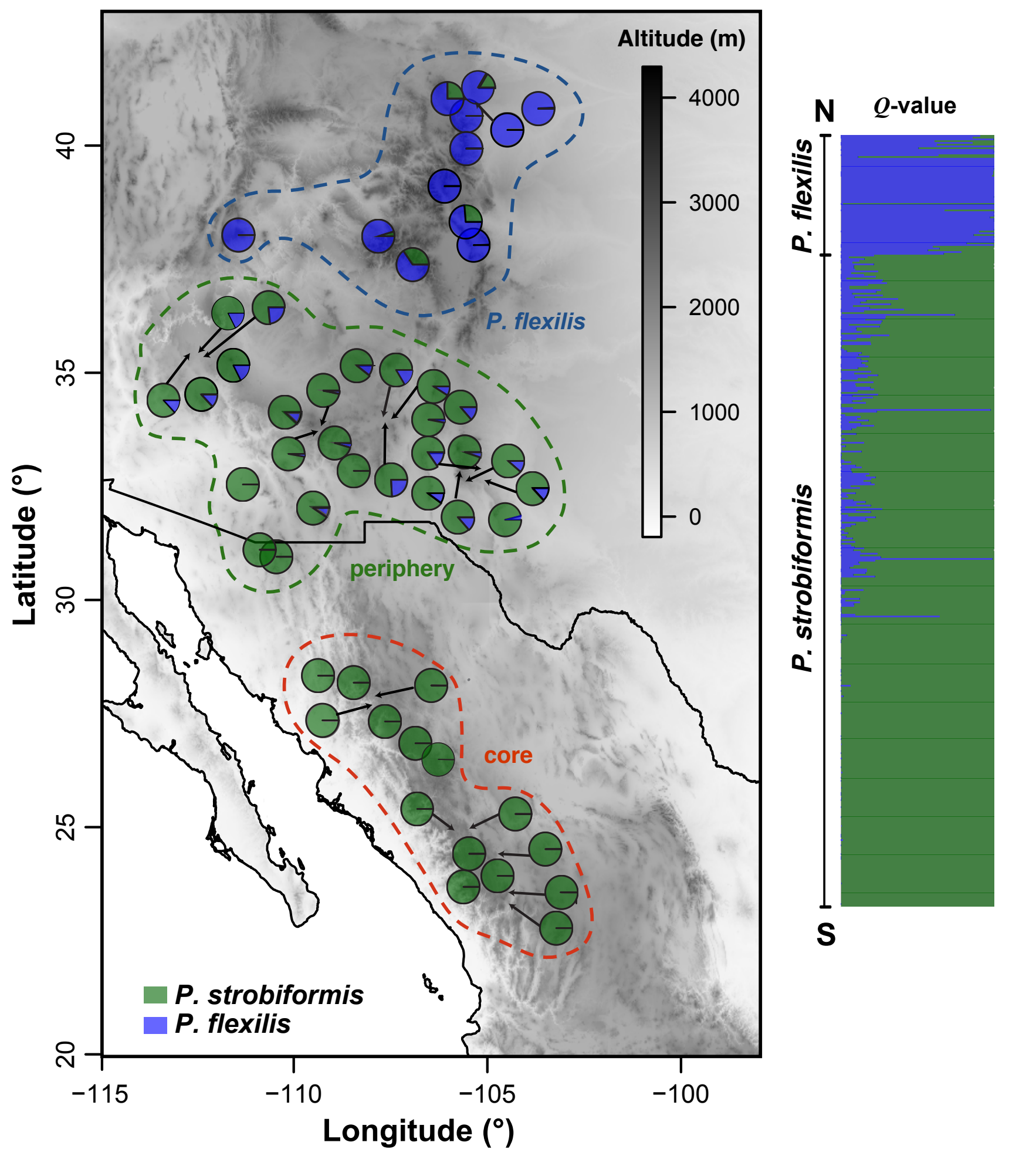
976 processing raw ddRADseq reads, and **C)** Down-sampling of SNP loci for demographic modeling

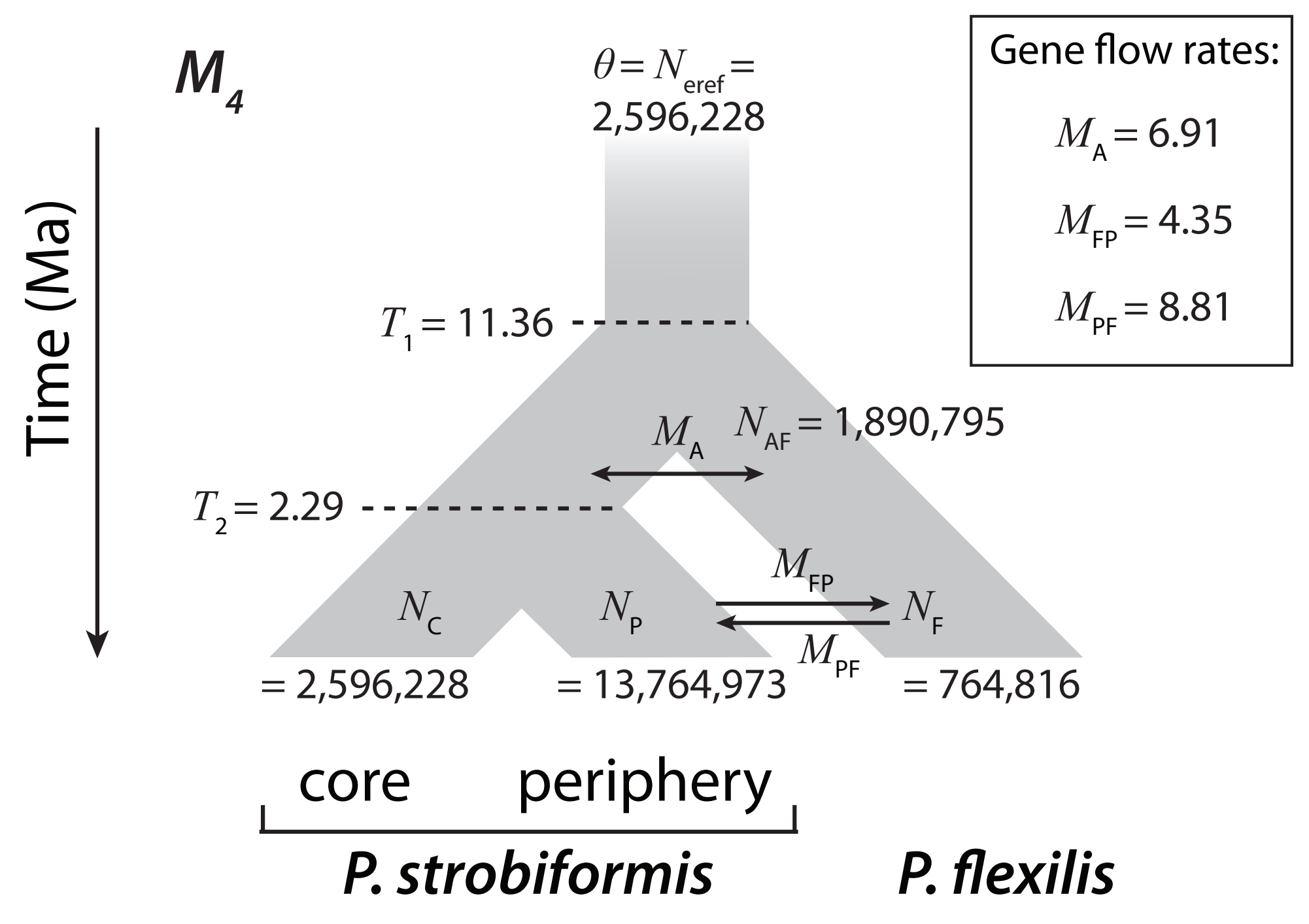
977 using $\partial A \partial I$. **D)** List of volunteers and technicians involved in sampling and library preparation.

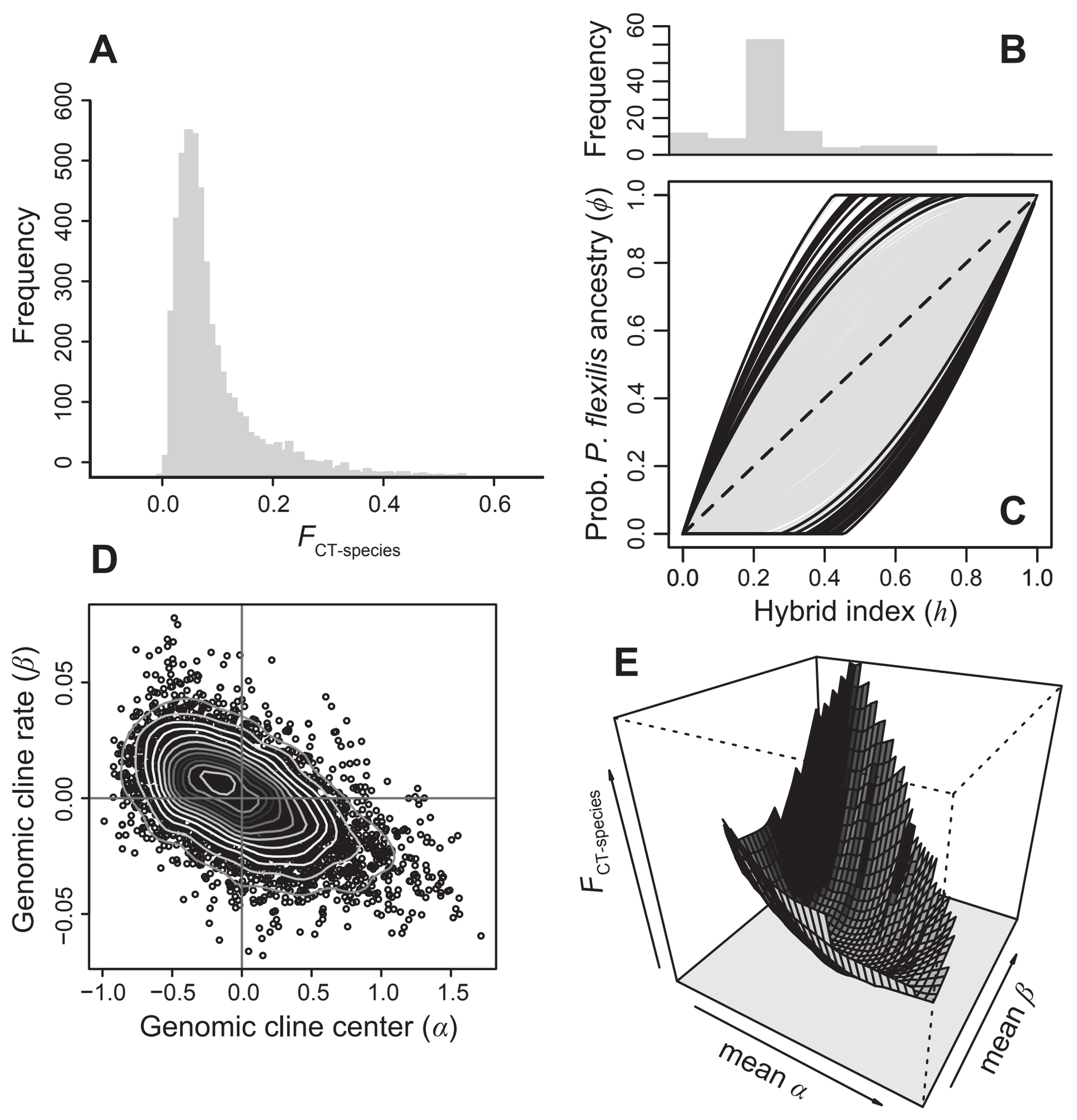

Scale(km)
1:64,090,000
Latitude (°)
40
35
30
25
20
−115
−110
−105
−105
Longitude (°)
P. strobiformis
P. flexilis
Core
Periphery
P. flexilis


Core
Periphery
P. flexilis
D = 0.17
p_CP < 0.01, p_PC < 0.01
D = 0.05
p_PF < 0.01, p_FP < 0.01
D = 0.14
p_FP < 0.01, p_PF < 0.01
Frequency
Schoener's D


Core
Periphery
P. flexilis
2.5
0.0
PC2 (PVE = 0.70)
-2
0
2
PC1 (PVE = 0.90)


Altitude (m)
4000
3000
2000
1000
0
P. flexilis
periphery
core
P. strobiformis
P. flexilis
Latitude (°)
Longitude (°)
40
35
30
25
20
−115
−110
−105
−100
N
S
Q-value
P. flexilis
P. strobiformis


M4
θ = Neref = 2,596,228
Gene flow rates:
MA = 6.91
MFP = 4.35
MPF = 8.81
Time (Ma)
T1 = 11.36
MA
NAF = 1,890,795
T2 = 2.29
MFP
NC
NP
NF
MPF
= 2,596,228
= 13,764,973
= 764,816
core
periphery
P. strobiformis
P. flexilis


A
Frequency
0
100
200
300
400
500
600
0.0
0.2
0.4
0.6
$F_{\text{CT-species}}$
B
Frequency
0
20
40
60
C
Prob. P. flexilis ancestry (ϕ)
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
Hybrid index (h)
D
Genomic cline rate (β)
−0.05
0.00
0.05
−1.0
−0.5
0.0
0.5
1.0
1.5
Genomic cline center (α)
E
$F_{\text{CT-species}}$
mean α
mean β