Genomic insights into hybrid zone formation: the role of climate, landscape, and demography in the emergence of a novel hybrid lineage between *Populus trichocarpa* and *P. balsamifera*

Authors: Constance E. Bolte, Tommy Phannareth, Michelle Zavala-Paez, Brianna N. Sutara, Muhammed F. Can, Matthew C. Fitzpatrick, Jason A. Holliday, Stephen R. Keller, Jill A. Hamilton

Corresponding author:
Jill A. Hamilton, jvh6349@psu.edu

Abstract

Population demographic changes, alongside landscape, geographic, and climate heterogeneity, can influence the timing, stability, and extent of introgression where species hybridize. Thus, quantifying interactions across diverged lineages, and the relative contributions of interspecific genetic exchange and selection to divergence at the genome-wide level is needed to better understand the drivers of hybrid zone formation and maintenance. We used seven latitudinally arrayed transects to quantify the contributions of climate, geography, and landscape features to broad patterns of genetic structure across the hybrid zone of *Populus trichocarpa* and *P. balsamifera* and evaluated the demographic context of hybridization over time. We found genetic structure differed among the seven transects. While ancestry was structured by climate, landscape features influenced gene flow dynamics. Demographic models indicated a secondary contact event may have influenced contemporary hybrid zone formation with the origin of a putative hybrid lineage that inhabits regions with higher aridity than either of the ancestral groups. Phylogenetic relationships based on chloroplast genomes support the origin of this hybrid lineage inferred from demographic models based on nuclear data. Our results point towards the importance of climate and landscape patterns in structuring the contact zones between *P. trichocarpa* and *P. balsamifera* and emphasize the value whole genome sequencing can have to advancing our understanding of how neutral processes influence divergence across space and time.

Keywords: Natural hybridization, speciation, forest trees, gene flow, climate, *Populus trichocarpa, Populus balsamifera*

Introduction

Understanding the processes influencing the formation and maintenance of species is a central goal of evolutionary biology and is crucial to the management and conservation of biodiversity in a rapidly changing environment (Frankham, 2010; Hoffmann et al., 2015; Payseur & Rieseberg, 2016). As such, hybridization and introgression have long been of interest (Dobzhansky, 1936; Heiser, 1949; Jeffrey, 1916; Stebbins, 1959), and recent advances in population-scale genome sequencing have revolutionized our understanding of the frequency and pervasiveness of hybridization in nature (Hamilton & Miller, 2016; Janes & Hamilton, 2017; VanWallendael et al., 2022). Multiple taxa are known to form natural hybrid zones; however, they are particularly prevalent among forest tree species (Abbott, 2017; Suarez-Gonzalez et al., 2018; Swenson & Howard, 2004). Climatic oscillations during glacial and interglacial periods have influenced the distributions of many temperate and boreal trees, facilitating opportunities for interspecific gene flow during times of secondary contact (Hamilton et al., 2015; Hewitt, 2004; Jump & Peñuelas, 2005; Soltis et al., 2006; Swenson & Howard, 2004). However, despite widespread observations of hybridization across forest tree species, gaps remain in our understanding of how demographic, neutral and nonneutral evolutionary processes influence the formation and maintenance of natural hybrid zones in forested landscapes across space and time. Teasing apart these processes requires an assessment of the historical, landscape, and climatic factors that underlay hybrid zone formation. Ultimately, understanding how population demographic changes, alongside landscape heterogeneity and climate adaptation contribute to the timing, stability, and extent of hybridization will be critical to managing species in a rapidly changing climate.

Population demographics, including expansion and contraction of a species' range can have substantial influence on standing genetic variation, impacting the evolutionary trajectory of populations across space and time. Hybrid zones often form when previously isolated lineages come into contact and the impacts of these processes, including genetic drift and gene flow, likely play an important role shaping the genetic structure of a contact zone (Abbott, 2017; N. H. Barton & Hewitt, 1989). Shifts between isolation and contact can coincide with changes in effective population size (N_e) influencing the direction and extent of genetic exchange, while fine-scale variance at the population or genome-level can further influence elimination or fixation of alleles. In forest trees, genetic exchange often appears to be largely asymmetrical favoring the genomic background of one species due solely or in part to differential dispersal capacity, wind-patterns, and unidirectional reproductive incompatibilities (El Mujtar et al., 2017; Hamilton et al., 2013a; Lepais et al., 2009; Lexer et al., 2005, 2006). Given that an influx of genetic variants via interspecific gene flow not only buffers N_e, but can generate novel genetic recombinants upon which natural selection can act, characterizing genetic structure and the impact genetic exchange may have across space and time is needed to predict evolution.

Demographic inference can be particularly useful for characterizing processes underlying patterns of genetic exchange that can influence species evolutionary relationships, distributional shifts, or adaptive potential (Bacilieri et al., 1996; Petit et al., 2004). Several forest tree hybrid zones have characterized patterns of introgression, including oaks (Cannon & Petit, 2020; Eaton et al., 2015; McVay et al., 2017), poplars (Chhatre et al., 2018; Christe et al., 2017; Lexer et al., 2005; Suarez-Gonzalez et al., 2016; Suarez-Gonzalez et al., 2018), and spruce (Hamilton et al., 2013a, 2013b). However, few studies have leveraged whole genome sequencing to tease apart the temporal and spatial dynamics underlying the evolutionary history of contact zones using both biparentally and uniparentally inherited genomes. Viewing distinct evolutionary trajectories associated with biparentally and uniparentally inherited genomes through the lens of demographic change and landscape heterogeneity enables assessment of the role these evolutionary processes may play to the formation and maintenance of natural hybrid zones.

Here, we use *Populus*, an ecologically and economically important model system in forest trees that naturally hybridize with congeners where geographical ranges overlap. *Populus trichocarpa,* the first tree to have its genome fully sequenced (Tuskan et al., 2006), has played an extensive role in our understanding of tree genome biology, comparative genomics, and adaptive introgression (e.g., Jansson & Douglas, 2007; Shang et al., 2020; Suarez-Gonzalez et al., 2016). Natural hybrid zones exist between *P. trichocarpa* and *P. balsamifera*, largely associated with geographically and climatically steep transitions from maritime to continental climates west and east of the Rocky Mountains. However, divergence has likely involved a combination of intrinsic and extrinsic factors, including a dynamic history of population size change throughout glacial and interglacial periods, landscape-related barriers to gene flow, and climatically-related responses to selection (Geraldes et al., 2014; Keller et al., 2010; Levsen et al., 2012; Slavov et al., 2012). We compare whole-genome sequence data for biparentally (nuclear) and uniparentally (chloroplast) inherited genomes to quantify divergence and the evolutionary relationship between species across space and time. This will provide an understanding of the role neutral and non-neutral processes play in the formation of these long-lived hybrid zones (Petit et al. 2005).

Using a sampling design of latitudinally arrayed transects, we sequenced whole genomes for 576 trees to capture parental-types and hybrids across repeated zones of contact between *P. trichocarpa* and *P. balsamifera* distributed across their entire range of overlap. With these data, we observed genetic structure within and across each contact zone and asked (1) How has climate, geography, and landscape-level barriers influenced genetic structure, and (2) What is the history and extent of interspecific gene flow? These data provide new insights into how different evolutionary processes contribute to the formation and persistence of hybrid zones over varying spatial and temporal scales.

127

128 ## Materials and Methods

129

130 ***Sampling of plant material***

131 In January 2020, vegetative branch cuttings were collected from 576 poplar trees
132 spanning repeated natural contact zones between *Populus trichocarpa* and *Populus*
133 *balsamifera* (Figure 1a). These represent individual collections from seven latitudinally
134 distributed transects (hereafter referred to as Alaska_1, Alaska_2, Cassiar, Chilcotin,
135 Jasper, Crowsnest, and Wyoming) spanning much of the species' overlapping
136 distributions, from 40°N to 65°N and -100°W to -150°W longitude. On average the
137 distance between samples was approximately 43 kilometers (range: <1 - 962 km).
138 Vegetative branch cuttings were transported back to a greenhouse at Virginia Tech
139 (Blacksburg, VA, USA) for propagation. Details associated with propagation and
140 greenhouse conditions are provided in Appendix 1 of Supplementary.S1. Young leaf
141 tissue was sampled from each of the 576 propagated branch cuttings, placed
142 immediately on dry ice, and then transferred to a -80°C freezer for long-term
143 preservation.

144

145 ***Genomic Library Preparation and Bioinformatics***

146 For each of the 576 samples, DNA was extracted from approximately 100mg of leaf
147 tissue using the Qiagen plant DNeasy kit, with one modification (details available in
148 Supplementary.S1; Appendix 2). Genomic DNA libraries were constructed at the Duke
149 University Center for Genomic and Computational Biology, using an Illumina DNA Prep
150 kit (Illumina Inc., San Diego, USA). Genomic libraries were sequenced on an S4 flow
151 cell in 2x150bp format on an IlluminaNovaSeq 6000 instrument with 64 samples per
152 lane. De-indexing, QC, trimming adapter sequences, and sequence preprocessing were
153 completed by the sequencing facility.

154

155 Subsequent bioinformatic analysis was performed on Virginia Tech's Advanced
156 Research Computing System (ARC) using Burrows-Wheeler Aligner (BWA) to map
157 reads to the *P. trichocarpa* reference genome (v4.0), and the resulting SAM files were

158 converted to BAM format with SAMtools (Li & Durbin, 2010). The Genome Analysis
159 Toolkit (v3.7) Haplotype Caller algorithm was then used to generate individual gVCF
160 files, which were merged in a single VCF file with the GATKGenotype GVCFs function.
161 This raw VCF had ~82 million variant calls and was quality-filtered for variants that had
162 poor map quality (MQ < 40.00), elevated strand bias (FS > 40.000, SOR > 3.0),
163 differential map quality between reference and alternative alleles (MQRankSum < -
164 12.500), positional bias between reference and alternate alleles (ReadPosRankSum < -
165 8.000), or low coverage depth (QD < 2.0).
166
167 After quality filtering, ~34 million single nucleotide polymorphisms (SNPs) and ~4.8
168 million indels remained in the variant call set. Filtering removed all of the ~28 million
169 multiallelic sites (defined as having more than two allelic states), of which ~5 million were
170 SNP sites (i.e., did not include indels). For 575 of the 576 samples, missing data was on
171 average 0.7 % (SD: 0.1 - 1.3 %). One sample, GPR-14_S50_L001, had 99.9% missing
172 data and was therefore not retained for further analysis.
173
174 ***Analysis of population structure***
175 To account for potential sampling of non-focal poplar species and remove those
176 samples, we merged the genetic data for our 575 samples with genetic data available
177 for 48 samples of *P. angustifolia*, which overlaps in portions of the southern distribution
178 of our sampling region (Chhatre et al., 2018). Methods in Supplementary.S1 (Appendix
179 3) detail the analyses used. While we found no need to remove samples due to
180 admixture with *P. angustifolia*, we removed 29 of the 575 samples with unknown
181 ancestry (Supplementary.S1; Figure S1 and S2), leaving 546 samples for subsequent
182 genetic analyses.
183
184 To assess the genetic structure of the contact zones between *P. balsamifera* and *P.*
185 *trichocarpa*, we first thinned variant sites in high linkage disequilibrium using *Plink* (v1.9)
186 based on 10,000 bp windows sliding in increments of 1000 bp along each chromosome,
187 flagging site pairs with an $r^2 > 0.1$. One site for each pair exceeding the r^2 threshold (i.e.,
188 'prune.out' file), along with singletons (--mac 2) and those with missing data (--max-

189 missing 1.0), were removed using *vcftools*, version 0.1.16. The resulting 299,335 low LD
190 variant sites were used to visualize the scale and spatial extent of hybridization.
191 Genomic structure was assessed using a principal component analysis (PCA) with the
192 'prcomp' function of the *stats* package in R, version 4.2.1. We then performed an
193 ADMIXTURE analysis (Alexander et al., 2009) using cluster assignments ranging from
194 *K*=2 to *K*=6. The cluster assignment (*K*) with the lowest cross validation error (k-folds =
195 10) was determined to best fit the data (Figure 1c).
196
197 ***Assessing evolutionary relationships from nuclear data***
198 To evaluate the evolutionary relationships between genetic clusters, we used the three
199 genetic clusters identified from ADMIXTURE (*K* = 3; see Results) to determine the
200 topology with the greatest support. We identified parental-type samples for each genetic
201 cluster, hereafter referred to as *P. balsamifera*, *coastal P. trichocarpa*, and *interior P.*
202 *trichocarpa*, based on K=3 using an ancestry coefficient threshold of Q > 0.90. Of the
203 546 samples included in the analysis, 185 parental-types were identified (112 *P.*
204 *balsamifera*, 51 *coastal P. trichocarpa*, and 22 *interior P. trichocarpa*). Using parental-
205 types, we evaluated three different topologies based on the proposed divergence history
206 between *interior P. trichocarpa* and the other two parental-type genetic lineages (Table
207 1). The first topology evaluated expectations based on a recent divergence from *coastal*
208 *P. trichocarp*a, while the second tested for a recent divergence from *P. balsamifera*, and
209 the third evaluated whether *interior P. trichocarpa* was a hybrid lineage formed from an
210 admixture event between *coastal P. trichocarpa* and *P. balsamifera*.
211
212 The evolutionary relationship of the three genetic clusters was evaluated using the
213 genotypes from the 185 parental-types (112 *P. balsamifera*, 51 *coastal P. trichocarpa*,
214 and 22 *interior P. trichocarpa*) and Diffusion Approximation for Demographic Inference
215 ($\partial\alpha\partial i$ v.2.1.1;(Gutenkunst et al., 2009). We used the unfolded joint site frequency
216 spectrum (jSFS) of derived allele counts following polarization of genotypes following the
217 approach of (Luqman et al., 2023). Details pertaining to data sources and program
218 implementation are provided in Supplementary_S1, Appendix 4. We assigned allelic
219 states as either ancestral or derived based on the consensus ancestral state: 6495 sites

(2.260%) had the reference allele assigned to the derived state and 275,751 sites (95.959%) had the reference allele assigned to the ancestral state. The remaining 1.781% of sites had undefined ancestral alleles. To increase computational efficiency, we used the derived allele frequencies for a random sample of 10 diploid individuals per parental-type cluster to build an unfolded 3-dimensional joint site frequency spectrum (jSFS) using the function 'GenerateFs' in the dadi-cli program (X. Huang et al., 2023). We performed ten replicate runs of each model in $\partial\alpha\partial i$ with a 40 x 50 x 60 grid space and the nonlinear Broyden-Fletcher-Goldfarb-Shannon (BFGS) optimization routine. Akaike information Criterion (AIC; Akaike 1974) was employed for model selection, with the best replicate run (highest log composite likelihood) for each model used to calculate ΔAIC ($AIC_{model\ i}$ – $AIC_{best\ model}$) scores (Burnham & Anderson, 2002).

To estimate the timing of divergence and quantify the extent and directionality of gene flow between genetic clusters, we tested four hypotheses, or models, of connectivity and divergence based on the best supported topology (Table 1). The first model assumed a scenario of recurrent gene flow, while the other three models assumed scenarios of isolation that varied based on the timing and frequency of secondary contact (Supplementary.S1, Figure S4). We performed ten replicate runs of each model in $\partial\alpha\partial i$ with a 40 x 50 x 60 grid space and the nonlinear Broyden-Fletcher-Goldfarb-Shannon (BFGS) optimization routine. Akaike information Criterion (AIC; Akaike 1974) was employed for model selection, and Fisher Information Matrix (FIM)-based uncertainty analysis was then conducted for the best supported model to obtain upper and lower 95% confidence intervals (CIs) for the inferred parameter estimates. To unscale parameter estimates and their 95% CIs we used a mutation rate of 1.33×10^{-10} substitutions/site/year rate for *Populus trichocarpa* (Hofmeister et al., 2020), a generation time of 15 years, and a calculated effective genome length (Supplementary.S3). To calculate effective genome length, we multiplied the proportion of variant sites used in the modeling (count of filtered sites in each pairwise inference with no missing data and singletons removed / count of variant sites prior to filtering) by the number of base pairs from the *P. trichocarpa* reference genome considered in variant calling (389,204,664 bp).

251 ***Assessing evolutionary relationships from chloroplast data***

252 To gain insight into the evolutionary relationships among the three genetic clusters
253 identified from nuclear data, we used the 185 parental-types to examine differentiation
254 across uni-parentally inherited chloroplast genomes. Chloroplast (cp) genomes were
255 assembled using NOVOPlasty (Dierckxsens et al., 2017). The published Nisqaully-1 *P.*
256 *trichocarpa* chloroplast genome (NC_009143.1) was used as the reference sequence for
257 genome assembly. The ribulose-bisphosphate carboxylase (*rbc*L) gene was manually
258 extracted and used as a seed, serving as an initial reference point for the chloroplast
259 genome assembly. Assembled chloroplast genomes were annotated using Geneious
260 v7.1.4 (Dierckxsens et al. 2016). Following annotation, inverted repeats and coding
261 regions were manually checked for each annotated individual. In total, 184 out of the 185
262 parental-types were successfully assembled, with one individual (genotype 831) excluded
263 due to incomplete assembly.

264

265 A maximum likelihood (ML) phylogenetic tree was built to assess the relationships among
266 the chloroplast genomes for each parental type. The chloroplast genomes from two *P.*
267 *angustifolia* samples (NC_037413 and MW376761) were downloaded from NCBI and
268 selected as an outgroup. Chloroplast genomes for all 184 samples and outgroups were
269 aligned using MAFFT V.7.453 (Katoh & Standley, 2013) to perform a ML phylogeny in IQ-
270 tree 1.6.11 (Trifinopoulos et al., 2016). The best fit evolutionary model for the sequence
271 alignment was calculated by ModelFinder (Kalyaanamoorthy et al., 2017), with the
272 selected model being the transversion model with empirical base frequencies and
273 FreeRate model (TVM+F+R2). A ML tree was generated following 1,000 bootstrap
274 replications and a consensus phylogenetic tree was created based on all parental-type
275 samples using the FigTree software v1.4.13.

276

277 ***Associating genetic structure with climate and geography***

278 To quantify the multivariate relationships between genetic structure, climate, and
279 geographic variation across the hybrid zone, a redundancy analysis (RDA) was
280 conducted within the *vegan* package, version 2.6-2 (Oksanen et al., 2020) in R version
281 4.2.1. Climate data at 30 second resolution averaged across the years 1961-1990 for

282 twenty-five variables associated with geographic origin were extracted using *ClimateNA*
283 (version 5.2; (Wang et al., 2016) for each sampled tree (Supplementary.S2) and then
284 tested for collinearity using variance inflation factor (VIF) analysis implemented in the
285 *usdm* R package (version 1.1-18; (Naimi et al., 2014). Six climatic variables with a VIF
286 less than 10 were retained for subsequent analysis, including continentality (TD), mean
287 annual temperature (MAT), mean annual precipitation (MAP), climate moisture deficit
288 (CMD), relative humidity (RH), and precipitation as snow (PAS). Geographic variables
289 included for subsequent analysis were latitude, longitude, and elevation.
290
291 Genotype data (299,335 nuclear SNPs and 2,828 chloroplast SNPs) for all 546 samples
292 was converted to minor allele counts per sample using vcftools (version 0.1.16). A
293 permutation-based analysis of variance (ANOVA) procedure with 999 permutations
294 (Legendre & Legendre, 2012) was used to test the overall statistical significance of the
295 RDA model ($\alpha = 0.05$) as well as the significance of each RDA axis separately. Variance
296 partitioning was then performed using the 'varpart' function of the *vegan* package to
297 quantify the influence of predictor variables and their confounding effects to observed
298 genetic variation. This analysis was repeated without hybrid samples ($0.10 < Q < 0.90$) to
299 compare the multivariate relationships among the 185 parental-type samples.
300
301 ***Visualizing barriers to gene flow***
302 To evaluate genetic connectivity across the contact zones between *P. trichocarpa* and *P.*
303 *balsamifera,* we used EEMS (Estimated Effective Migration Surfaces; (Petkova et al.,
304 2016) and observed correspondence between landscape features, ancestry, and patterns
305 of genetic connectivity. A genetic dissimilarity matrix was generated using the 'bed2diffs
306 script' included within the EEMS program informed by the 299,335 genome-wide nuclear
307 SNPs across 546 individuals. Two additional geographic files were created using
308 geographic coordinates; including the location data of each individual (.coord extension)
309 and a large general habitat polygon (.outer extension) which was manually obtained using
310 Google Maps API v3 Tool. The geographical coordinates for each sample alongside the
311 parameterized grid space (500 demes with each grid cell ~21,000 km^2) assigned
312 individuals to one of 49 demes, defined as pseudo populations of close geographical

proximity to a vertex of the defined grid space (Supplementary.S1, Figure S3). We then estimated the effective migration rate using a stepping-stone model. Model parameters included a burn-in of 25,000,000 with an MCMC length of 50,000,000 and thinning to sample 1 in every 100,000 iterations. The results of three replicate runs were averaged to visualize effective diversity (q) and effective migration rate (m) surfaces (in $\log_{10}$ scale) across the contact zone using the R package *rEEMSplots* v 0.0.1 (Petkova et al., 2016).

Results

Genomic structure of hybrid zone identifies three distinct genetic clusters

Using all 546 individuals, the nuclear genomic PCA ascribed genetic structure along PC1 (1.764% variance explained) to species-specific differences between *P. trichocarpa* (far left of PC1) and *P. balsamifera* (far right of PC1) and PC2 (0.925% variance explained) to genetic structure within *P. trichocarpa* (Figure 1b). Admixture analysis best supported the presence of three genetic clusters (K = 3), with admixed individuals present across all transects (Figure 1c). Two of the three genetic groups clustered geographically within the range of *P. trichocarpa* (Little 1971) – this included a *coastal P. trichocarpa* cluster which had a northern, coastal distribution, and an *interior P. trichocarpa* cluster, which had a southern, interior distribution (Figure 1d). The third genetic cluster was *P. balsamifera* in the K = 3 admixture analysis. Pairwise estimates of genome-wide genetic differentiation (F_{ST}) were similar across the three genetic clusters, ranging from 0.053 for *coastal - interior P. trichocarpa,* 0.054 for *interior P. trichocarpa - P. balsamifera*, and 0.056 for *coastal P. trichocarpa - P. balsamifera* comparisons.

Of the 546 samples included in the analysis, 185 parental-types associated with K = 3 genetic clusters were identified based on Q > 0.90 (112 individuals identified as *P. balsamifera*, 51 individuals identified as *coastal P. trichocarpa*, and 22 individuals identified as *interior P. trichocarpa). Coastal P. trichocarpa* and *P. balsamifera* genetic ancestry was observed in Alaska and Cassiar transects, although the Alaska transect did

344 not appear to have any *coastal P. trichocarpa* parental-types. All three genetic clusters
345 were observed in Chilcotin, Jasper, Crowsnest, and Wyoming transects (Figure 1c).
346 Shared ancestry associated with *P. balsamifera* along the Crowsnest transect was
347 restricted to the eastern edge of the *P. trichocarpa* distribution. The Wyoming transect
348 was dominated by *interior P. trichocarpa* parental-type ancestry (Q > 0.90 at K = 3;
349 Supplementary.S2), whereas individuals sampled near the western edge of this transect
350 exhibited mixed ancestry between the coastal and interior genetic clusters of *P.*
351 *trichocarpa*. Consistent with the Crowsnest transect, *P. balsamifera* ancestry was limited
352 to the far eastern distribution of *P. trichocarpa,* with *P. balsamifera* parental-types
353 restricted to fragmented populations geographically distant from the main distribution of
354 either species.
355
356 ***Nuclear and chloroplast genomes identify the persistence of three genetic lineages***
357 Given the similar estimates of genome-wide nuclear differentiation (F_{ST}), we performed a
358 topology test to better understand the relationships among the three genetic clusters. The
359 best fit topology as determined by AIC (Table 1a) supported *interior P. trichocarpa* as an
360 ancient lineage of admixed origin. The second-best fit topology placed *interior P.*
361 *trichocarpa* as a split from *coastal P. trichocarpa*, but delta AIC was 4,943 units higher.
362
363 To complement assessment of nuclear topologies, evolutionary relationships were also
364 assessed using ML analysis of chloroplast genomes. The alignment of the 184 individuals
365 and outgroup genomes produced a single matrix with 164,183 nucleotide sites, including
366 1,020 polymorphic sites. Three distinct clades were detected using the ML tree, including
367 *coastal P. trichocarpa*, *interior P. trichocarpa*, and *P. balsamifera* respectively. The
368 *coastal P. trichocarpa* clade appeared ancestral to both *interior P. trichocarpa* and *P.*
369 *balsamifera*, with all three exhibiting robust bootstrap support. Lineage assignments were
370 congruent between nuclear and chloroplast genomes, except for individual 757, which
371 clustered with *interior P. trichocarpa* in the nuclear admixture analysis but in the
372 chloroplast ML tree belonged to the *P. balsamifera* clade.
373
374 ***Climate strongly associates with genetic structure***

375 Climate and geography explained 3.79% (r^2) of the genetic variance across the 546
376 sampled trees and 299,335 SNP dataset. Most of the explainable genetic variance (PVE)
377 was accounted for by the first RDA axis (38.26%, Figure 2a) with continentality (TD) and
378 mean annual temperature (MAT) having the highest predictor loadings (eigenvectors).
379 Individuals identified as *P. balsamifera* were associated with higher TD and lower MAT
380 on average. The second RDA axis captured 19.23% (PVE) of the explainable genetic
381 variance, with the highest predictor loadings on the second RDA axis being relative
382 humidity (RH), elevation, and longitude. These variables were associated with
383 differentiation within *P. trichocarpa* across the seven transects. Associations between
384 climate, geography and genetic variation were not observed along RDA2 for *P.*
385 *balsamifera* (Figure 2). Combining loadings for RDA1 and RDA2, we found latitude, TD,
386 and climate moisture deficit (CMD) had the strongest associations with genetic variation
387 within and among sampled transects.
388
389 Because geographic and climatic predictor variables are often correlated, we performed
390 variance partitioning to tease apart the respective contributions of climate, geography,
391 and their interactions to genomic variance. We observed a large, confounded effect
392 (58.69%) between the climatic and geographic variables chosen for analysis. Removing
393 confounded effects through conditioning revealed climate was a stronger predictor of
394 genetic variance. Climate, conditioned on geography, accounted for 27.28% of
395 explainable genetic variance, whereas geography, conditioned on climate, accounted for
396 just 14.03% of explainable genetic variance (Supplementary.S1; Table S2).
397
398 Using the same suite of analyses to investigate associations between environment and
399 ancestry for the 185 parental-types revealed higher contributions of climate and
400 geography to explainable genetic structure. From side-by-side comparisons of nuclear
401 and chloroplast genetic structure in the PCA, we observed similar clustering (Figure 2).
402 Considering the differentiation along PC1 axis, *coastal P. trichocarpa* and *interior P.*
403 *trichocarpa* appear less diverged based on nuclear data when compared to chloroplast.
404 Similar clustering and environmental associations were also observed in the RDA (Figure
405 2) and aligned with the associations presented for the 546 samples (Figure 2a). However,

406 with hybrids removed from the analysis, the explanatory variance increased (adj. r^2 =
407 7.8%; PVE = 51.65%). The explanatory variance based on chloroplast structure was the
408 highest observed (adj. r^2 = 30.1%; PVE = 66.26%). Variance partitioning was also
409 conducted with the hybrids excluded, and with the confounding effects of geography
410 removed, climate explained approximately 40-42% of the genetic variance in both the
411 chloroplast and nuclear data. Conversely, after conditioning on climate, geography
412 explained only 9.46% of the nuclear genetic variance and 2.87% of the chloroplast genetic
413 variance (Supplementary.S1, Table S3).
414
415 ***Gene flow dynamics shaped by physical barriers***
416 Using departures from isolation-by-distance (IBD) to indicate barriers to gene flow, EEMS
417 analysis identified the continental divide of the Rocky Mountains as a large barrier
418 between *P. trichocarpa* and *P. balsamifera*. Additional barriers to gene flow were
419 identified between transects associated with elevational gradients (Figure 3). For
420 example, *interior P. trichocarpa* within the most southern transect was observed in a
421 region of high local connectivity but was isolated from neighboring individuals identified
422 as *coastal P. trichocarpa* and *P. balsamifera* in other analyses. The EEMS results
423 indicated both Alaska transects were more isolated from the other northern transect,
424 Cassiar, than expected based on IBD. In contrast, individuals spanning *coastal P.*
425 *trichocarpa* parental-type and hybrids across the Chilcotin, Jasper, and Crowsnest
426 transects were more similar than expected under IBD, suggesting substantial intraspecific
427 migration across latitudes within the southern range of this hybrid zone. This pattern was
428 not observed for *P. balsamifera,* as only individuals located east of the Rocky Mountains
429 exhibited evidence of greater effective migration than expected.
430
431 ***Demographic inference reveals periods of interspecific vicariance, gene flow, and***
432 ***the emergence of a hybrid lineage***
433 Of the four demographic models of connectivity and divergence evaluated (Table 1b), the
434 model with the greatest support inferred an initial divergence between *coastal P.*
435 *trichocarpa* - *P. balsamifera* approximately 2.12 million years ago (95% CI: 2.11 – 2.13
436 million years ago). Following initial divergence, limited gene flow (*m*A ~ 1.90e-04)

437 continued until approximately 860,000 years ago when gene flow ceased, during which
438 time the two lineages were isolated (Figure 4). Following approximately 96,000 years of
439 isolation, the two lineages experienced secondary contact and 764,000 years ago formed
440 the *interior P. trichocarpa* lineage via admixture. During the initial formation of *interior P.*
441 *trichocarpa* the proportions of each progenitor lineage were inferred to be approximately
442 34.8% *coastal P. trichocarpa* (parameter *f*; Table 2) and 65.2% *P. balsamifera* (1 - *f*).
443 Following secondary contact, asymmetrical gene flow was inferred until present, with the
444 highest rates of gene flow observed between *coastal P. trichocarpa* and *interior P.*
445 *trichocarpa* (*m13* ~ 5.08e-05 and *m31* ~ 4.58e-05*;* Table 2), with a slightly greater rate of
446 genetic exchange from *interior* into *coastal* (*m13*). Gene flow appeared to be strongly
447 asymmetrical, with more movement from *P. balsamifera* into *coastal* and *interior P.*
448 *trichocarpa* (*m12* ~ 3.64e-05, *m32* ~ 3.13e-05, respectively) and less *coastal* or *interior*
449 *P. trichocarpa* movement into *P. balsamifera* (*m21* ~ 1.30e-05 and *m23* ~ 9.08e-06; Table
450 3). The highest effective population size inferred was for *P. balsamifera* (N_2 ~ 100,000)
451 followed by *interior P. trichocarpa* (N_3 ~ 53,800) and *coastal P. trichocarpa* (N_1 ~ 36,500).
452 The residuals between the model and data fit were small and evenly distributed,
453 suggesting good model fit (Supplementary.S1, Figure S5). In addition, confidence
454 intervals for each parameter were tightly bound to the estimations inferred (Table 2).
455
456
457
458 **Discussion**
459
460 This study provides valuable insights into how complex spatial and temporal dynamics
461 influence hybridization, highlighting the role neutral processes can play in shaping the
462 genetic structure of hybrid zones. A dynamic history between glacial and interglacial
463 cycles suggests that secondary contact over time has had a substantial influence on
464 hybrid zone structure. We identified three distinct genetic clusters geographically
465 distributed across our seven repeated zones of species overlap. Differentiation within and
466 across nuclear and chloroplast genomes supported the presence of three genetic

467 lineages, with tests of topology suggesting one genetic lineage reflects an ancient
468 admixture event between *coastal P. trichocarpa* and *P. balsamifera* leading to the origin
469 of a new hybrid lineage, *interior P. trichocarpa*. While physical geographic barriers have
470 played a significant role influencing historical and contemporary asymmetrical genetic
471 exchange, our results highlight the important role climate gradients alongside geography
472 have played influencing hybrid zone structure. Together, this suggests that genetic
473 patterns observed could be sensitive to climatic shifts, which has implications for
474 understanding the future of these contact zones in a changing climate.
475
476 ***Demographic shifts associate with shifts in historical climate***
477 According to our best fit 3D demographic model, initial divergence between *P.*
478 *balsamifera and coastal P. trichocarpa (*~ 2.12 million years ago) aligns with a shift from
479 the Pliocene to the Pleistocene Climate Era (Tan et al., 2018). This initial divergence was
480 inferred to occur amid recurrent gene flow for approximately 1.25 million years, followed
481 by a period of isolation lasting ~ 96,000 years. Interestingly, this period of isolation not
482 only aligns with the Mid-Pleistocene Transition, when glacial cycles became longer (from
483 ~40,000 to ~100,000 years long) and shifts to interglacial climates become abrupt and
484 extreme, but it also coincides in timing with a glacial maximum (~800,000 years ago;
485 (Chalk et al., 2017; Ellis & Palmer, 2016; Willeit et al., 2019). Secondary contact between
486 *P. balsamifera* and *coastal P. trichocarpa*, along with the origin of *interior P. trichocarpa,*
487 was estimated to have occurred ~764,000 years ago and coincides with a transition to
488 interglacial climate (Ellis & Palmer, 2016). This suggests opportunities for genetic
489 exchange were regained following expansion from isolated refugia.
490
491 While climate has continued to oscillate since the inferred secondary contact event, which
492 might be expected to reduce opportunities for gene flow given the different climatic
493 affinities of each lineage, our demographic model suggested continued genetic exchange.
494 One explanation may be that species distributions have moved in tandem with climate
495 such that populations were proximal enough to maintain gene flow, resulting in a stable
496 hybrid zone. Indeed, the long-distance dispersal of seed and pollen in *Populus* and other
497 forest trees blurs definitions for how geographically isolated species need to be to disrupt

498 gene flow. An alternative explanation may be that periods of strict isolation did occur
499 during extreme climatic shifts but were not sustained long enough for genomic patterns
500 of divergence associated with allopatry to develop, or patterns were broken down
501 following subsequent contact. This latter possibility may characterize hybridizing species
502 with porous genomes whose species boundaries vary temporally with repeated cycles of
503 reinforcement and homogenization (e.g., Mix-Isolation-Mix hypothesis; He et al., 2019) or
504 due to changes in the strength and direction of selection over historically varying climatic
505 optima (e.g., Shang et al., 2020). Among the divergence histories inferred for forest tree
506 species, there is growing evidence for divergence with recurrent gene flow despite abrupt
507 shifts in climate (e.g., Bolte et al., 2022; He et al., 2019; Lexer et al., 2006; Menon et al.,
508 2018).
509
510 ***Gene flow asymmetries in the context of future climate change***
511 Gene flow was nearly three times higher from *P. balsamifera* into *coastal P. trichocarpa*
512 than from *coastal P. trichocarpa* into *P. balsamifera.* This observation is corroborated by
513 previous estimates of genetic exchange in this hybrid zone by Suarez-Gonzalez et al.
514 (2018), who noted asymmetrical introgression from *P. balsamifera* into *P. trichocarpa*.
515 This asymmetry may suggest *P. balsamifera* has a demographic advantage within the
516 ecotonal transition (e.g., dispersal down an elevational gradient or with river flow from
517 inland to coast). On the other hand, lower rates of introgression between *P. balsamifera*
518 and *interior P. trichocarpa* may be indicative of stronger geographical isolation, a lack of
519 hybrids geographically positioned between these genetic clusters to facilitate
520 backcrossing, or even more likely, asynchronous reproductive phenology. According to
521 descriptions within (Haeussler et al., 1990) flowering times for interior populations of *P.*
522 *trichocarpa* can be nearly one month later than those reported for populations of *P.*
523 *trichocarpa* (coastal) and *P. balsamifera* in British Columbia, Canada.
524
525 Gene flow between *coastal P. trichocarpa* and *interior P. trichocarpa* were greatest
526 relative to all inferred between-lineage contemporary estimates (Figure 4b). Indeed,
527 *interior P. trichocarpa* is genetically more similar to *coastal P. trichocarpa* based on
528 patterns of nuclear differentiation. However, despite this, the demographic model that

529 received the greatest support suggested that in the formation of *interior P. trichocarpa*
530 (~764,000 years ago) only 35% of the hybrid genome reflected ancestry from *coastal P.*
531 *trichocarpa*. Taken together, these results may indicate that higher rates of contemporary
532 gene flow between *coastal* and *interior P. trichocarpa* homogenized population
533 differences following the initial divergence of the *interior P. trichocarpa* lineage.
534 Regardless, the influx of new genetic recombinants through introgression is evident
535 across all three parental-types and may impact adaptation to a changing climate (Aitken
536 et al., 2008; N. Barton & Bengtsson, 1986; Martinsen et al., 2001).
537
538 ***Climate and landscape heterogeneity contribute to genetic structure***
539 Landscape-induced restrictions to gene flow have contributed to patterns of differentiation
540 across the hybrid zone. Estimated effective migration surfaces (EEMS) suggest gene flow
541 has likely been suppressed by elevational barriers through the Rocky Mountains of
542 northwestern North America. West of the Rocky Mountains, we observed variation in
543 effective migration rates within the *P. trichocarpa* distribution described by (EL Little Jr.,
544 1971). For example, effective migration was greater within the northern transects of
545 Alaska and Cassiar but not between these two transects. Similarly, there appeared to be
546 substantial gene flow between hybrids within the 'mid-latitudinal' transects (Chilcotin,
547 Jasper, and Crowsnest), but migration was restricted along the coastal regions of these
548 three transects where parental-types of *coastal P. trichocarpa* were identified. This
549 suggests gene flow between *P. trichocarpa* populations is lower than expected based on
550 expectations of isolation-by-distance. Thus, landscape barriers to gene flow likely
551 contribute to strong population structure observed within the range of *P. trichocarpa*
552 (Geraldes et al., 2014; Slavov et al., 2012).
553
554 Redundancy analysis (RDA) and variance partitioning revealed genetic structure and
555 climate are significantly associated across the hybrid zone. The Rocky Mountains present
556 environmental transitions from drier and colder conditions on the eastern side (favoring
557 *P. balsamifera*) to more maritime climates on the western side of the Rockies (favoring
558 *coastal P. trichocarpa*). The distributions and climate associations we observed between
559 *P. balsamifera, coastal P. trichocarpa*, and *interior P. trichocarpa* (Figure 2a) are

comparable to other hybrid zones within the region, including *Picea glauca, P. sitchensis*, and *P. engelmannii* hybrid zones (De La Torre et al., 2015; Hamilton et al., 2013b, 2015). In both systems, maritime to continental climate transitions were associated with genomic ancestry. Because the frequency of hybrid genotypes is largely associated with ecotonal transitions, this suggests these hybrid zones likely reflect a model of bounded hybrid superiority. However, quantifying fitness in reciprocal transplant experiments is needed to test the bounded hybrid superiority model.

The admixed lineage, *interior P. trichocarpa*, appears to be bounded to a drier and warmer climate than either of its progenitors. The genetic structure of *interior P. trichocarpa* is associated with high climate moisture deficit. If selection maintains genetic structure among the lineages observed with our study, we may find trait divergence, especially those underlying water-use, across genetic lineages. Interestingly, *Populus* hybrid lineages of the Qinghai-Tibet Plateau (QTP) also inhabit landscapes with higher aridity relative to their progenitors. However, survival and growth measurements within common garden experiments suggested phenotypic plasticity may have contributed to adaptive divergence and persistence in environments beyond the parental species' range (Shi et al., 2024). A similar experimental design could be employed for *coastal P. trichocarpa, interior P. trichocarpa,* and *P. balsamifera* to test if the admixed lineage (*interior P. trichocarpa*) exhibit transgressive traits relative to parents while also evaluating the role of plasticity to success in novel environments. In the absence of trait data, we can leverage genomic data to differentiate between, and functionally assess, genomic regions associated with restricted or excess introgression that may underlay adaptive divergence.

Nuclear and chloroplast genomes provide complementary insights into hybrid zone dynamics

The estimates observed herein suggest 2.11 - 2.13 million years ago *P. trichocarpa* and *P. balsamifera* diverged and are congruent with observations for this species-pair from (Sanderson et al., 2023). The differentiation we observed in the chloroplast not only aligns with patterns of nuclear differentiation among the three assigned parental-types but also supports the observations and interpretations of previous chloroplast assessments of

divergence in *Populus*. (D. I. Huang et al., 2014) ascribed the deep chloroplast divergence of *P. trichocarpa* from *P. balsamifera* to chloroplast capture of an ancient, now extinct, lineage. While the genome-wide patterns of nuclear differentiation (e.g., PCA, F_{ST}) suggests *coastal P. trichocarpa* and *interior P. trichocarpa* are more closely related, our phylogenetic inference from the chloroplast placed *coastal P. trichocarpa* outside the clade of *interior P. trichocarpa* and *P. balsamifera*. Only one individual out of 185 did not have a matching chloroplast - nuclear parental-type assignment. Considering our demographic modeling and the fact that our focal taxa exhibit dioecy with maternal inheritance of the chloroplast genome, the origin of *interior P. trichocarpa* likely resulted from admixture between *P. balsamifera* seed parents and *coastal P. trichocarpa* pollen parents. Given the inferred length of time that has passed since the admixed origin (~764,000 years), the *interior P. trichocarpa* chloroplast genome has likely evolved following divergence. Co-evolution and co-adaptation between the chloroplast and nuclear genomes may have contributed to chloroplast divergence of *interior P. trichocarpa* from the original seed parent (Greiner & Bock, 2013) and would be an interesting future avenue for research.

Conclusions

Here, we provide insight into the evolutionary processes shaping divergence and hybridization across space and time in a *Populus* system of western North America. Demographic models support the origin of a hybrid lineage following secondary contact of ancestral lineages and a history of asymmetrical gene flow between focal taxa. Our results highlight the influence of climate and geography to genetic structure and divergence, as well as the influence of landscape-level attributes to gene flow dynamics. Overall, our study sheds light on the role demographic processes can play within a long-lived hybrid zone across space and time and highlight the need for fine-scale assessment of regions of the genome important to introgression and adaptation for species management in a changing climate.

Acknowledgements:
We thank Joe Braasch, Lionel Di Santo, Sonia DeYoung, Sara Klopf, Jack Woods, Ben Woods, Raju Soolanayakanahally for assisting with field sample collection. We also thank Kyle Peer, Clay Sawyers, and Deborah Bird from the Virginia Tech Reynold's Homestead Forestry Research Station for their assistance with plant propagation. We thank Ryan Gutenkunst and Andrew Eckert for advice and communications related to demographic inference. We acknowledge the valuable questions from and conversations with our team members during project calls and meetings: Cigdem Kansu, Susanne Lachmuth, Baxter Worthing, and Alayna Mead.

Competing interests: All authors claim no competing interests.

Author contributions:
MCF, JH, SRK, and JAH conceived the study and collected the samples. MC extracted DNA. TP processed genomic data. CEB analyzed genomic data, conducted demographic inference, and performed genetic-environment associations. MZ and BS assembled chloroplast genomes and performed chloroplast analyses. SRK, JH and JAH advised analyses. CEB and JAH led the writing of the manuscript. SRK, JH, and MCF provided valuable feedback.

Data availability: Sequence reads for the 546 samples analyzed in this manuscript have been archived on NCBI (PRJNA996882). Chloroplast genomes are archived under BankIt2793276 (Accession numbers PP274128 - PP274690). Both nuclear and chloroplast sequences will remain under embargo until publication. Scripts for variance partitioning, RDA, and demographic inference can be found at https://github.com/boltece/Poplar_divergence.

Funding: National Science Foundation grant IOS-1856450, USDA National Institute of Food and Agriculture project and Hatch Appropriations (PEN04809 and Accession 7003639), NIFA project VA-136641, Schatz Center for Tree Molecular Genetics

Supporting Information:

Supplementary.S1 - Supplementary methods (appendices) and figures

Supplementary.S2 - Sample summaries for climate and geographic data, admixture coefficients, and PCA and RDA coordinates

Supplementary.S3 - Summarized results from demographic inference including AIC calculations, unscaling of $\partial\alpha\partial i$ parameter estimates, and confidence interval calculations

References:

Abbott, R. J. (2017). Plant speciation across environmental gradients and the occurrence and nature of hybrid zones. *Journal of Systematics and Evolution*, *55*(4), 238–258. https://doi.org/10.1111/JSE.12267

Aitken, S. N., Yeaman, S., Holliday, J. A., Wang, T., & Curtis-McLane, S. (2008). Adaptation, migration or extirpation: Climate change outcomes for tree populations. *Evolutionary Applications*, *1*(1), 95–111. https://doi.org/10.1111/J.1752-4571.2007.00013.X

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9). https://doi.org/10.1101/gr.094052.109

Bacilieri, R., Ducousso, A., & Kremer, A. (1996). Comparison of morphological characters and molecular markers for the analysis of hybridization in sessile and pedunculate oak. *Annales Des Sciences Forestières*, *53*(1), 79–91. https://doi.org/10.1051/forest:19960106

Barton, N., & Bengtsson, B. O. (1986). The barrier to genetic exchange between hybridising populations. *Heredity*, *57*(3), 357–376. https://doi.org/10.1038/hdy.1986.135

Barton, N. H., & Hewitt, G. M. (1989). Adaptation, speciation and hybrid zones. *Nature*, *341*(6242), 497–503. https://doi.org/10.1038/341497a0

Bolte, C. E., Faske, T. M., Friedline, C. J., & Eckert, A. J. (2022). Divergence amid recurring gene flow: Complex demographic histories for two North American pine species (Pinus pungens and P. rigida) fit growing expectations among forest trees. *Tree Genetics & Genomes*, *18*(5), 35. https://doi.org/10.1007/s11295-022-01565-8

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

Cannon, C. H., & Petit, R. J. (2020). The oak syngameon: More than the sum of its parts. *New Phytologist*, *226*(4), 978–983. https://doi.org/10.1111/nph.16091

Chalk, T. B., Hain, M. P., Foster, G. L., Rohling, E. J., Sexton, P. F., Badger, M. P. S., Cherry, S. G., Hasenfratz, A. P., Haug, G. H., Jaccard, S. L., Martínez-García, A., Pälike, H., Pancost, R. D., & Wilson, P. A. (2017). Causes of ice age intensification across the Mid-Pleistocene Transition. *Proceedings of the National Academy of Sciences*, *114*(50), 13114–13119. https://doi.org/10.1073/pnas.1702143114

Chhatre, V. E., Evans, L. M., DiFazio, S. P., & Keller, S. R. (2018). Adaptive introgression and maintenance of a trispecies hybrid complex in range-edge

populations of Populus. *Molecular Ecology*, *27*(23), 4820–4838. https://doi.org/10.1111/mec.14820

Christe, C., Stölting, K. N., Paris, M., Fraïsse, C., Bierne, N., & Lexer, C. (2017). Adaptive evolution and segregating load contribute to the genomic landscape of divergence in two tree species connected by episodic gene flow. *Molecular Ecology*, *26*(1), 59–76. https://doi.org/10.1111/mec.13765

De La Torre, A., Ingvarsson, P. K., & Aitken, S. N. (2015). Genetic architecture and genomic patterns of gene flow between hybridizing species of picea. *Heredity*, *115*(2), 153–164. https://doi.org/10.1038/hdy.2015.19

Dierckxsens, N., Mardulyn, P., & Smits, G. (2017). NOVOPlasty: *De novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Research*, gkw955. https://doi.org/10.1093/nar/gkw955

Dobzhansky, T. H. (1936). STUDIES ON HYBRID STERILITY. 11. LOCALIZATION OF STERILITY FACTORS I N DROSOPHILA PSEUDOOBSCURA HYBRIDS. *Genetics*, *21*(113).

Eaton, D. A. R., Hipp, A. L., González-Rodríguez, A., & Cavender-Bares, J. (2015). Historical introgression among the American live oaks and the comparative nature of tests for introgression: INTROGRESSION IN THE AMERICAN LIVE OAKS. *Evolution*, *69*(10), 2587–2601. https://doi.org/10.1111/evo.12758

EL Little Jr. (1971). *Atlas of United States trees. Vol. 1. Conifers and important hardwoods*. USDA Misc. Publ. 1146 (200 maps).

El Mujtar, V., Sola, G., Aparicio, A., & Gallo, L. (2017). Pattern of natural introgression in a Nothofagus hybrid zone from South American temperate forests. *Tree Genetics & Genomes*, *13*(2), 49. https://doi.org/10.1007/s11295-017-1132-1

Ellis, R., & Palmer, M. (2016). Modulation of ice ages via precession and dust-albedo feedbacks. *Geoscience Frontiers*, *7*(6), 891–909. https://doi.org/10.1016/j.gsf.2016.04.004

Frankham, R. (2010). Challenges and opportunities of genetic approaches to biological conservation. *Biological Conservation*, *143*(9), 1919–1927. https://doi.org/10.1016/j.biocon.2010.05.011

Geraldes, A., Farzaneh, N., Grassa, C. J., Mckown, A. D., Guy, R. D., Mansfield, S. D., Douglas, C. J., & Cronk, Q. C. B. (2014). Landscape genomics of populus trichocarpa: The role of hybridization, limited gene flow, and natural selection in shaping patterns of population structure. *Evolution*, *68*(11), 3260–3280. https://doi.org/10.1111/evo.12497

Greiner, S., & Bock, R. (2013). Tuning a ménage à trois: Co-evolution and co-adaptation of nuclear and organellar genomes in plants. *BioEssays*, *35*(4), 354–365. https://doi.org/10.1002/bies.201200137

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, *5*(10). https://doi.org/10.1371/journal.pgen.1000695

Haeussler, S., Coates, D., & Mather, J. (1990). *Autecology of Common Plants in British Columbia: A Literature Review*. Forestry Canada, Pacific Forestry Center; British Columbia Ministry of Forests, Research Branch.

Hamilton, J. A., De la Torre, A. R., & Aitken, S. N. (2015). Fine-scale environmental variation contributes to introgression in a three-species spruce hybrid complex. *Tree Genetics and Genomes*, *11*(1). https://doi.org/10.1007/s11295-014-0817-y

Hamilton, J. A., Lexer, C., & Aitken, S. N. (2013a). Differential introgression reveals candidate genes for selection across a spruce (Picea sitchensis × P. glauca) hybrid zone. *New Phytologist*, *197*(3), 927–938. https://doi.org/10.1111/nph.12055

Hamilton, J. A., Lexer, C., & Aitken, S. N. (2013b). Genomic and phenotypic architecture of a spruce hybrid zone (Picea sitchensis × P. glauca). *Molecular Ecology*, *22*(3), 827–841. https://doi.org/10.1111/mec.12007

Hamilton, J. A., & Miller, J. M. (2016). Adaptive introgression as a resource for management and genetic conservation in a changing climate. *Conservation Biology*, *30*(1), 33–41. https://doi.org/10.1111/cobi.12574

He, Z., Li, X., Yang, M., Wang, X., Zhong, C., Duke, N. C., Wu, C. I., & Shi, S. (2019). Speciation with gene flow via cycles of isolation and migration: Insights from multiple mangrove taxa. *National Science Review*, *6*(2), 275–288. https://doi.org/10.1093/nsr/nwy078

Heiser, C. (1949). Natural hybridization with particular reference to introgression. *The Botanical Review*, *15*, 645–687.

764 Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary.
765 *Philosophical Transactions of the Royal Society B: Biological Sciences*,
766 *359*(1442), 183–195. https://doi.org/10.1098/rstb.2003.1388
767 Hoffmann, A., Griffin, P., Dillon, S., Catullo, R., Rane, R., Byrne, M., Jordan, R.,
768 Oakeshott, J., Weeks, A., Joseph, L., Lockhart, P., Borevitz, J., & Sgrò, C.
769 (2015). A framework for incorporating evolutionary genomics into biodiversity
770 conservation and management. *Climate Change Responses*, *2*(1).
771 https://doi.org/10.1186/s40665-014-0009-x
772 Hofmeister, B. T., Denkena, J., Colomé-Tatché, M., Shahryary, Y., Hazarika, R.,
773 Grimwood, J., Mamidi, S., Jenkins, J., Grabowski, P. P., Sreedasyam, A., Shu,
774 S., Barry, K., Lail, K., Adam, C., Lipzen, A., Sorek, R., Kudrna, D., Talag, J.,
775 Wing, R., … Schmitz, R. J. (2020). A genome assembly and the somatic genetic
776 and epigenetic mutation rate in a wild long-lived perennial Populus trichocarpa.
777 *Genome Biology*, *21*(1). https://doi.org/10.1186/s13059-020-02162-5
778 Huang, D. I., Hefer, C. A., Kolosova, N., Douglas, C. J., & Cronk, Q. C. B. (2014).
779 Whole plastome sequencing reveals deep plastid divergence and cytonuclear
780 discordance between closely related balsam poplars, Populus balsamifera and P.
781 trichocarpa (Salicaceae). *New Phytologist*, *204*(3), 693–703.
782 https://doi.org/10.1111/nph.12956
783 Huang, X., Struck, T. J., Davey, S. W., & Gutenkunst, R. N. (2023). *dadi-cli: Automated*
784 *and distributed population genetic model inference from allele frequency spectra*
785 [Preprint]. Bioinformatics. https://doi.org/10.1101/2023.06.15.545182

Janes, J. K., & Hamilton, J. A. (2017). Mixing it up: The role of hybridization in forest management and conservation under climate change. *Forests*, *8*(7). https://doi.org/10.3390/f8070237

Jansson, S., & Douglas, C. J. (2007). Populus: A Model System for Plant Biology. *Annual Review of Plant Biology*, *58*(1), 435–458. https://doi.org/10.1146/annurev.arplant.58.032806.103956

Jeffrey, E. C. (1916). Hybridism and the rate of evolution in angiosperms. *The American Naturalist*.

Jump, A. S., & Peñuelas, J. (2005). Running to stand still: Adaptation and the response of plants to rapid climate change. *Ecology Letters*, *8*(9), 1010–1020. https://doi.org/10.1111/j.1461-0248.2005.00796.x

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Keller, S. R., Olson, M. S., Salim, S., William, S. A., & Peter, T. (2010). Genomic diversity, population structure, and migration following rapid range expansion in the Balsam Poplar, Populus balsamifera. *Molecular Ecology*, *19*(6), 1212–1226. https://doi.org/10.1111/j.1365-294X.2010.04546.x

Legendre, P., & Legendre, L. (2012). *Numerical ecology*. Elsevier.

Lepais, O., Petit, R. J., Guichoux, E., Lavabre, J. E., Alberto, F., Kremer, A., & Gerber, S. (2009). Species relative abundance and direction of introgression in oaks. *Molecular Ecology*, *18*(10), 2228–2242. https://doi.org/10.1111/j.1365-294X.2009.04137.x

Levsen, N. D., Tiffin, P., & Olson, M. S. (2012). Pleistocene speciation in the genus populus (salicaceae). *Systematic Biology*, *61*(3), 401–412. https://doi.org/10.1093/sysbio/syr120

Lexer, C., Fay, M. F., Joseph, J. A., Nica, M. S., & Heinze, B. (2005). Barrier to gene flow between two ecologically divergent Populus species, P. alba (white poplar) and P. tremula (European aspen): The role of ecology and life history in gene introgression. *Molecular Ecology*, *14*(4), 1045–1057. https://doi.org/10.1111/j.1365-294X.2005.02469.x

Lexer, C., Kremer, A., & Petit, R. J. (2006). Shared alleles in sympatric oaks: Recurrent gene flow is a more parsimonious explanation than ancestral polymorphism. *Molecular Ecology*, *15*(7), 2007–2012. https://doi.org/10.1111/j.1365-294X.2006.02896.x

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, *26*(5), 589–595. https://doi.org/10.1093/bioinformatics/btp698

Luqman, H., Wegmann, D., Fior, S., & Widmer, A. (2023). Climate-induced range shifts drive adaptive response via spatio-temporal sieving of alleles. *Nature Communications*, *14*(1), 1080. https://doi.org/10.1038/s41467-023-36631-9

Martinsen, G. D., Whitham, T. G., Turek, R. J., & Keim, P. (2001). HYBRID POPULATIONS SELECTIVELY FILTER GENE INTROGRESSION BETWEEN SPECIES. In *Evolution* (Vol. 55, Issue 7, pp. 1325–1335).

McVay, J. D., Hipp, A. L., & Manos, P. S. (2017). A genetic legacy of introgression confounds phylogeny and biogeography in oaks. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1854). https://doi.org/10.1098/rspb.2017.0300

Menon, M., Bagley, J. C., Friedline, C. J., Whipple, A. V., Schoettle, A. W., Leal-Sàenz, A., Wehenkel, C., Molina-Freaner, F., Flores-Rentería, L., Gonzalez-Elizondo, M. S., Sniezko, R. A., Cushman, S. A., Waring, K. M., & Eckert, A. J. (2018). The role of hybridization during ecological divergence of southwestern white pine (Pinus strobiformis) and limber pine (P. flexilis). *Molecular Ecology*, *27*(5), 1245–1260. https://doi.org/10.1111/MEC.14505

Naimi, B., Hamm, N. A. S., Groen, T. A., Skidmore, A. K., & Toxopeus, A. G. (2014). Where is positional uncertainty a problem for species distribution modelling? *Ecography*, *37*(2), 191–203. https://doi.org/10.1111/j.1600-0587.2013.00205.x

Oksanen, J., Blanchet, F., Friendly, M., Kindt, R., & Legendre, P. (2020). Package "vegan"—Community Ecology Package Version 2.5-6. *Cran. Ism. Ac. Jp.*

Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology*, *25*(11), 2337–2360. https://doi.org/10.1111/mec.13557

Petit, R. J., Bodénès, C., Ducousso, A., Roussel, G., & Kremer, A. (2004). Hybridization as a mechanism of invasion in oaks. *New Phytologist*, *161*(1), 151–164. https://doi.org/10.1046/j.1469-8137.2003.00944.x

Petkova, D., Novembre, J., & Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, *48*(1), 94–100. https://doi.org/10.1038/ng.3464

Sanderson, B. J., Ghambir, D., Feng, G., Hu, N., Cronk, Q. C., Percy, M., Molina Freaner, F., Johnson, M. G., Smart, L. B., Yin, T., Ma, T., DiFazio, S. P., & Liu, J. (2023). Phylogenomics reveals patterns of ancient hybridization and differential diversification that contribute to phylogenetic conflict in willows, poplars, and close relatives. *Systematic Biology*, *syad042*. https://doi.org/10.1093/sysbio/syad042

Shang, H., Hess, J., Pickup, M., Field, D. L., Ingvarsson, P. K., Liu, J., & Lexer, C. (2020). Evolution of strong reproductive isolation in plants: Broad-scale patterns and lessons from a perennial model group: Evolution of strong barriers in plants. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1806). https://doi.org/10.1098/rstb.2019.0544

Shi, Y.-J., Huang, J.-L., Mi, J.-X., Li, J., Meng, F.-Y., Zhong, Y., He, F., Tian, F.-F., Zhang, F., Chen, L.-H., Yang, H.-B., Hu, H.-L., & Wan, X.-Q. (2024). A model of hybrid speciation process drawn from three new poplar species originating from distant hybridization between sections. *Molecular Phylogenetics and Evolution*, *190*, 107966. https://doi.org/10.1016/j.ympev.2023.107966

Slavov, G. T., Difazio, S. P., Martin, J., Schackwitz, W., Muchero, W., Rodgers-Melnick, E., Lipphardt, M. F., Pennacchio, C. P., Hellsten, U., Pennacchio, L. A., Gunter, L. E., Ranjan, P., Vining, K., Pomraning, K. R., Wilhelm, L. J., Pellegrini, M., Mockler, T. C., Freitag, M., Geraldes, A., … Tuskan, G. A. (2012). Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest tree Populus trichocarpa. *New Phytologist*, *196*(3), 713–725. https://doi.org/10.1111/j.1469-8137.2012.04258.x

Soltis, D. E., Morris, A. B., McLachlan, J. S., Manos, P. S., & Soltis, P. S. (2006). Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, *15*(14), 4261–4293. https://doi.org/10.1111/j.1365-294X.2006.03061.x

Stebbins, G. L. (1959). The Role of Hybridization in Evolution. *Proceeding of the American Philosophical Society*, *103*, 231–251.

Suarez-Gonzalez, A., Hefer, C. A., Christe, C., Corea, O., Lexer, C., Cronk, Q. C. B., & Douglas, C. J. (2016). Genomic and functional approaches reveal a case of adaptive introgression from Populus balsamifera (balsam poplar) in P. trichocarpa (black cottonwood). *Molecular Ecology*, *25*(11), 2427–2442. https://doi.org/10.1111/mec.13539

Suarez-Gonzalez, A., Hefer, C. A., Lexer, C., Cronk, Q. C. B., & Douglas, C. J. (2018). Scale and direction of adaptive introgression between black cottonwood (*Populus trichocarpa*) and balsam poplar (*P. balsamifera*). *Molecular Ecology*, *27*(7), 1667–1680. https://doi.org/10.1111/mec.14561

Suarez-Gonzalez, A., Lexer, C., & Cronk, Q. C. B. (2018). Adaptive introgression: A plant perspective. *Biology Letters*, *14*(3). https://doi.org/10.1098/rsbl.2017.0688

Swenson, N. G., & Howard, D. J. (2004). Do suture zones exist? *Evolution*, *58*(11), 2391–2397. https://doi.org/10.1111/j.0014-3820.2004.tb00869.x

Tan, N., Ladant, J.-B., Ramstein, G., Dumas, C., Bachem, P., & Jansen, E. (2018). Dynamic Greenland ice sheet driven by pCO2 variations across the Pliocene Pleistocene transition. *Nature Communications*, *9*(1), 4755. https://doi.org/10.1038/s41467-018-07206-w

Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A., & Minh, B. Q. (2016). W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research*, *44*(W1), W232–W235. https://doi.org/10.1093/nar/gkw256

Tuskan, G. A., DiFazio, S., Jansson, S., & Bohlmann, J. (2006). The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). *Science*, *313*, 1596–1605.

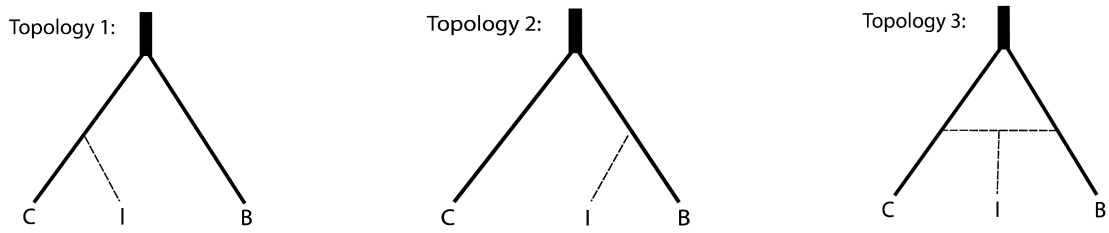
VanWallendael, A., Lowry, D. B., & Hamilton, J. A. (2022). One hundred years into the study of ecotypes, new advances are being made through large-scale field experiments in perennial plant systems. *Current Opinion in Plant Biology*, *66*, 102152. https://doi.org/10.1016/j.pbi.2021.102152

Wang, T., Hamann, A., Spittlehouse, D., & Carroll, C. (2016). Locally downscaled and spatially customizable climate data for historical and future periods for North America. *PLoS ONE*, *11*(6). https://doi.org/10.1371/journal.pone.0156720

Willeit, M., Ganopolski, A., Calov, V., & Brovkin, V. (2019). Mid-Pleistocene transition in glacial cycles explained by declining CO2 and regolith removal. *SCIENCE ADVANCES*, *5*(4), 1–8. https://doi.org/DOI: 10.1126/sciadv.aav7337

Table 1. Demographic inference AIC comparisons across 3 population models. Model abbreviations are defined in the footnote. The lower the AIC and delta AIC the better the model fit to the data. The first three rows (Table 1a) are results from the topology test to identify the evolutionary relationship of *interior P. trichocarpa* (I) to *P. balsamifera* (B) and *coastal P. trichocarpa* (C). The following four rows (Table 1b) are results of four models with gene flow parameters tested in various scenarios under the best fit topology (Topology 3; Figure S6) of admixed origin. Bolded models in each of the two inference routines signify best fit with delta AIC = 0.00.

a) **Model**	***k***	**AIC**	**delta AIC**
Topology 1: SI	6	68,945.17	4,942.96
Topology 2: SI	6	75,772.18	11,769.97
Topology 3: SI	6	64,002.21	**0.00**



b) **Model**	***k***	**AIC**	**delta AIC**
Topology 3: T2_SC	12	34,935.00	136.26
Topology 3: T2_IM	12	43,391.55	8,592.80
Topology 3: T3_IM_SC	14	34,798.75	**0.00**
Topology 3: T4_SC_SC	14	35,426.09	627.34

*Models; SI (strict isolation), IM (isolation-with-migration), SC (secondary contact), T# (number of time intervals considered)

Table 2. Parameter estimates and 95% confidence intervals (CI) associated with the best fit 3 population demographic model (Topology 3 - T3_IM_SC_IM). Calculations used to unscaling parameter estimates (using theta, mutation rate per generation, and genome length) to meaningful units are provided in Supplementary S3. Each value for T (time) is an interval. Interpretation of each migration (gene flow) parameter is provided in the footnote of this table. The parameter *f* is an estimate of the proportion of *coastal P. trichocarpa* in the formation of *interior P. trichocarpa* at the start of T3. The proportion of *P. balsamifera* is thus 1 - *f.*

Parameter	$\partial\alpha\partial i$ **scaled**	$\partial\alpha\partial i$ **unscaled**	**95% CI: (lower)**	**95% CI: (upper)**
N_ref (ancestral, theta)	4938	14,031	13,650	14,413
N_1 (*coastal trichocarpa*)	2.60	36,501	35,851	37,150
N_2 (*balsamifera*)	7.12	99,958	96,125	103,792
N_3 (*interior trichocarpa*)	3.83	53,809	51,106	56,512
T1 (ancient gene flow)	2.99	1,259,951	1,254,996	1,264,907
T2 (no gene flow)	0.23	96,645	93,272	100,018
T3 (admixture)	1.82	764,327	763,623	765,031
$M_A \rightarrow mA$	5.34	1.90E-04	1.58E-04	2.22E-04
$M_{12} \rightarrow m12$	1.02	3.64E-05	3.56E-05	3.71E-05
$M_{21} \rightarrow m21$	0.36	1.30E-05	1.15E-05	1.45E-05
$M_{23} \rightarrow m23$	0.25	9.08E-06	7.48E-06	1.07E-05
$M_{32} \rightarrow m32$	0.88	3.13E-05	2.89E-05	3.37E-05
$M_{13} \rightarrow m13$	1.43	5.08E-05	5.01E-05	5.14E-05
$M_{31} \rightarrow m31$	1.28	4.58E-05	4.28E-05	4.88E-05
f	0.348	34.8%	0.346	0.350

**M is a scaled estimate of the number of migrants per generation. *m* is an estimate of gene flow rate. After M (or *m*) the notation of A is ancestral migration, 12 is migrations from *balsamifera* into *coastal trichocarpa*, 21 is migration from *coastal trichocarpa* into *balsamifera*, 23 is migration from *interior trichocarpa* into *balsamifera*, 32 is migration from *balsamifera* into *interior trichocarpa*, 13 is migrations from *interior trichocarpa* into *coastal trichocarpa*, and 31 is migration from *coastal trichocarpa* into *interior trichocarpa.*

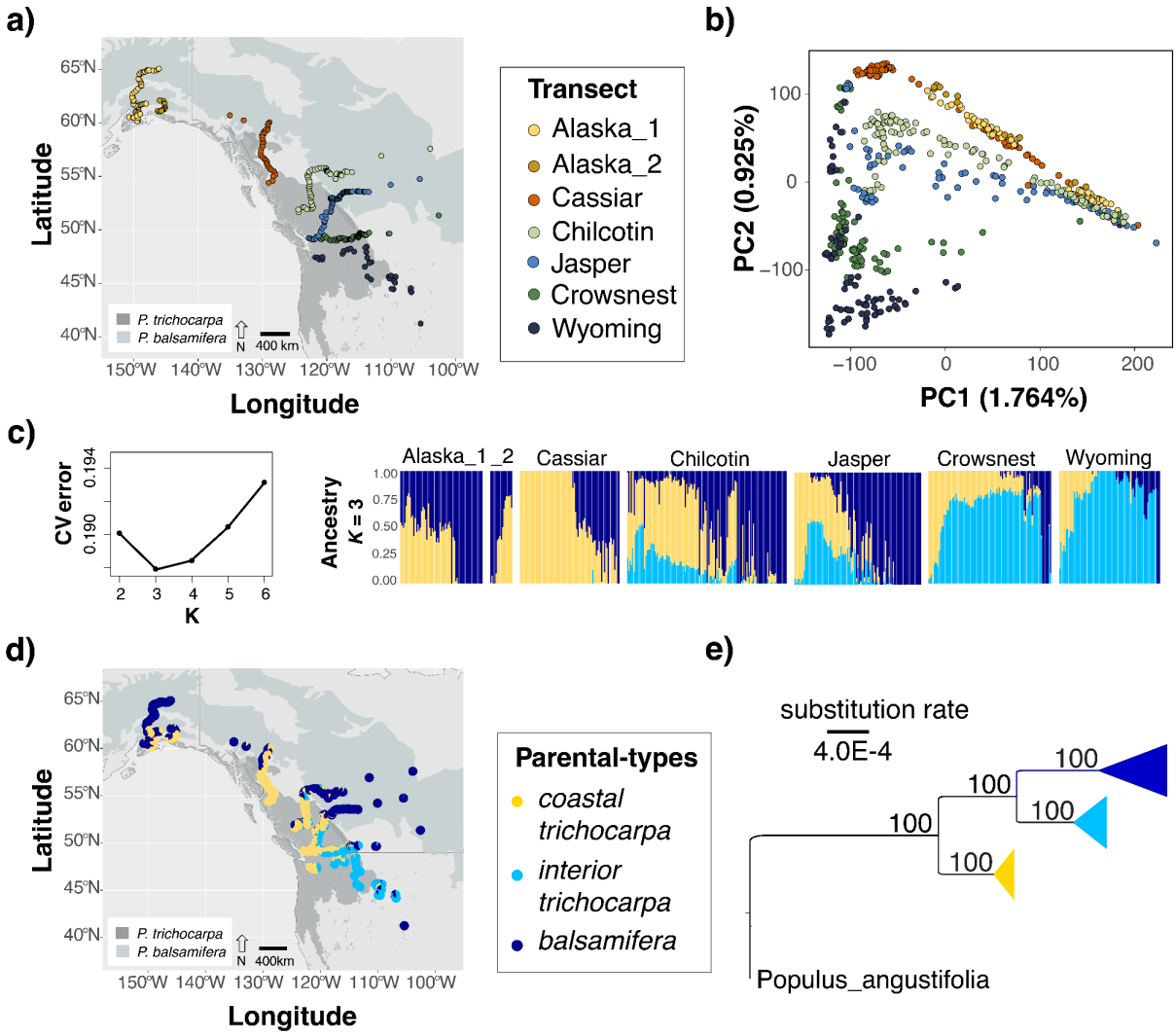


Figure 1. Spatial patterns of genetic structure and ancestry across 546 trees sampled from a) seven transects of the *Populus balsamifera* and *P. trichocarpa* hybrid zone, labeled by transect; b) transect-specific genomic structure based on Principal Component Analysis for PC1 and PC2 c) results of admixture analysis for each sample at *K* = 3, the best supported cluster assignment on the basis on lowest cross validation (CV) error; d) geographic distribution of *K* = 3 ancestry proportions illustrated as pie charts; e) ML phylogeny of chloroplast evolutionary relationships among parental-type samples from *K* = 3 admixture results (Q > 0.90 for each cluster). Bootstrap confidence is provided on each branch.

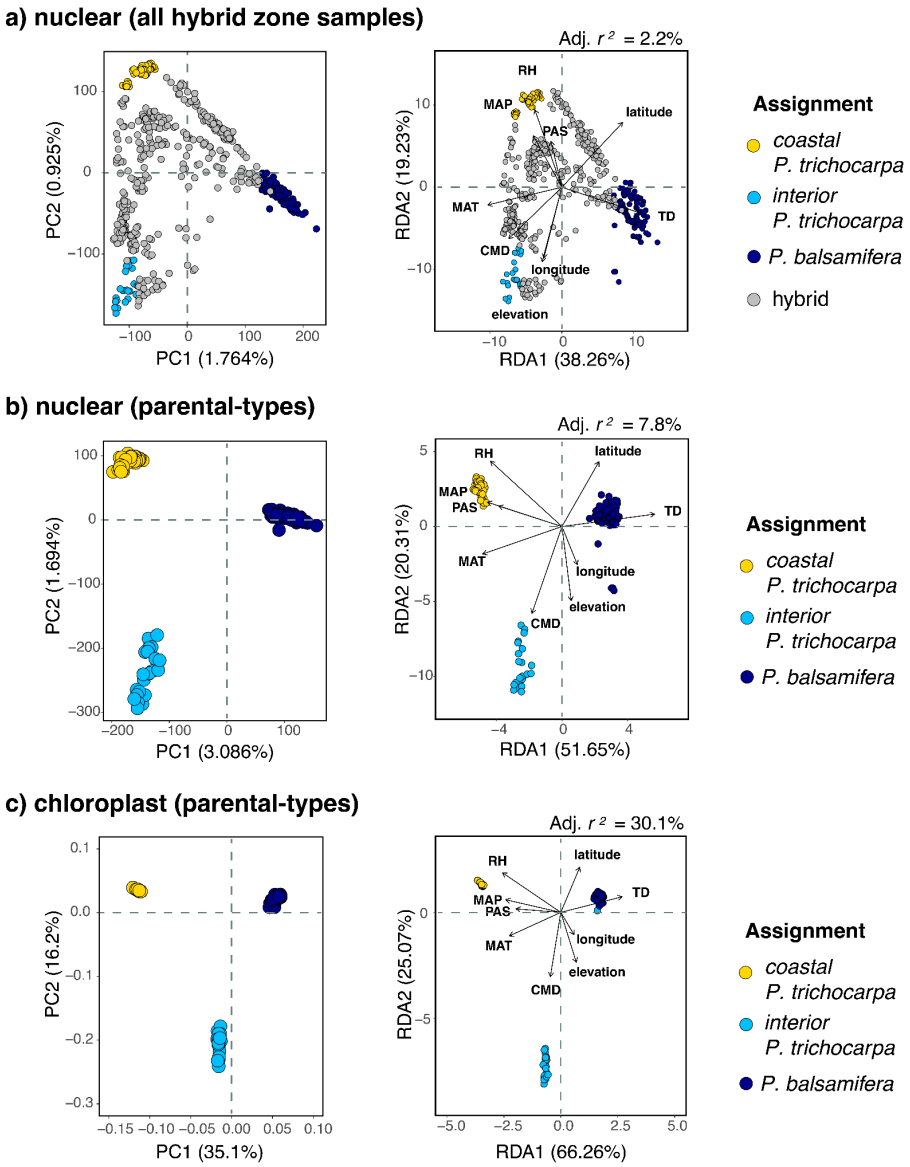


Figure 2. Genetic structure (PCA) and Redundancy analysis (RDA) based on associations between 6 climate, 3 geographic variables for a) genome-wide nuclear data of all 546 samples analyzed, versus b) genome-wide nuclear and c) chloroplast differentiation across parental-types. In each RDA plot, the direction and length of each arrow corresponds with the strength of prediction (i.e. loading) along each RDA axis. Climate and geographic variables include MAT (mean annual temperature), TD (continentality), PAS (precipitation as snow), RH (relative humidity), MAP (mean annual precipitation), CMD (climate moisture deficit), latitude, longitude, and elevation. Patterns between chloroplast and nuclear DNA are congruent with the exception of one sample (ID= 757_S202_L004) which has *interior P. trichocarpa* nuclear genetic background with *P. balsamifera* chloroplast identity.

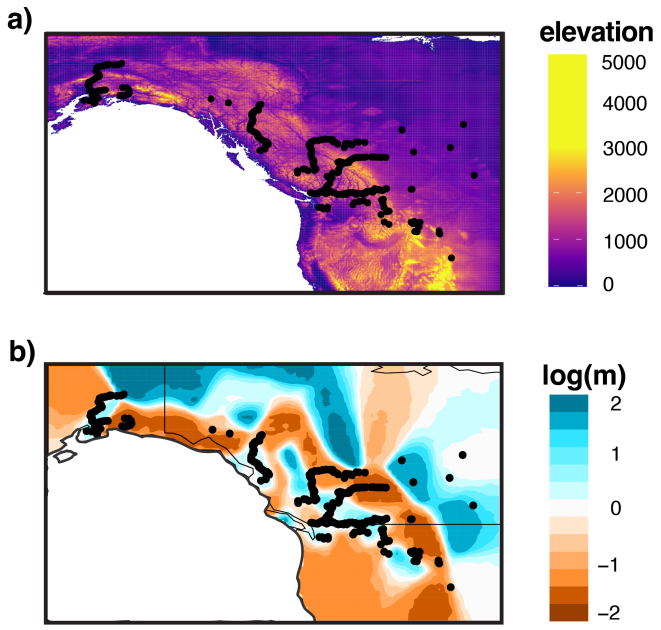


Figure 3. Estimated effective migration surfaces (EEMS) analysis of the relationship between genetic distance and geographical distance for 546 sampled trees across seven transects of the *P. trichocarpa* and *P. balsamifera* hybrid zone to visualize potential barriers to gene flow. Panel a) is a raster map of elevational differences (in meters) across the studied region, and panel b) illustrates putative barriers to gene flow within and among sampled transects. The black dots represent the geographic location of each of the 546 samples. Regions with negative migration rate predictions (brown colors on the log 10 scale) indicate lower genetic connectivity (gene flow) than expected. Regions with positive migration rate predictions (blue colors on the log 10 scale) indicate higher genetic connectivity than expected.

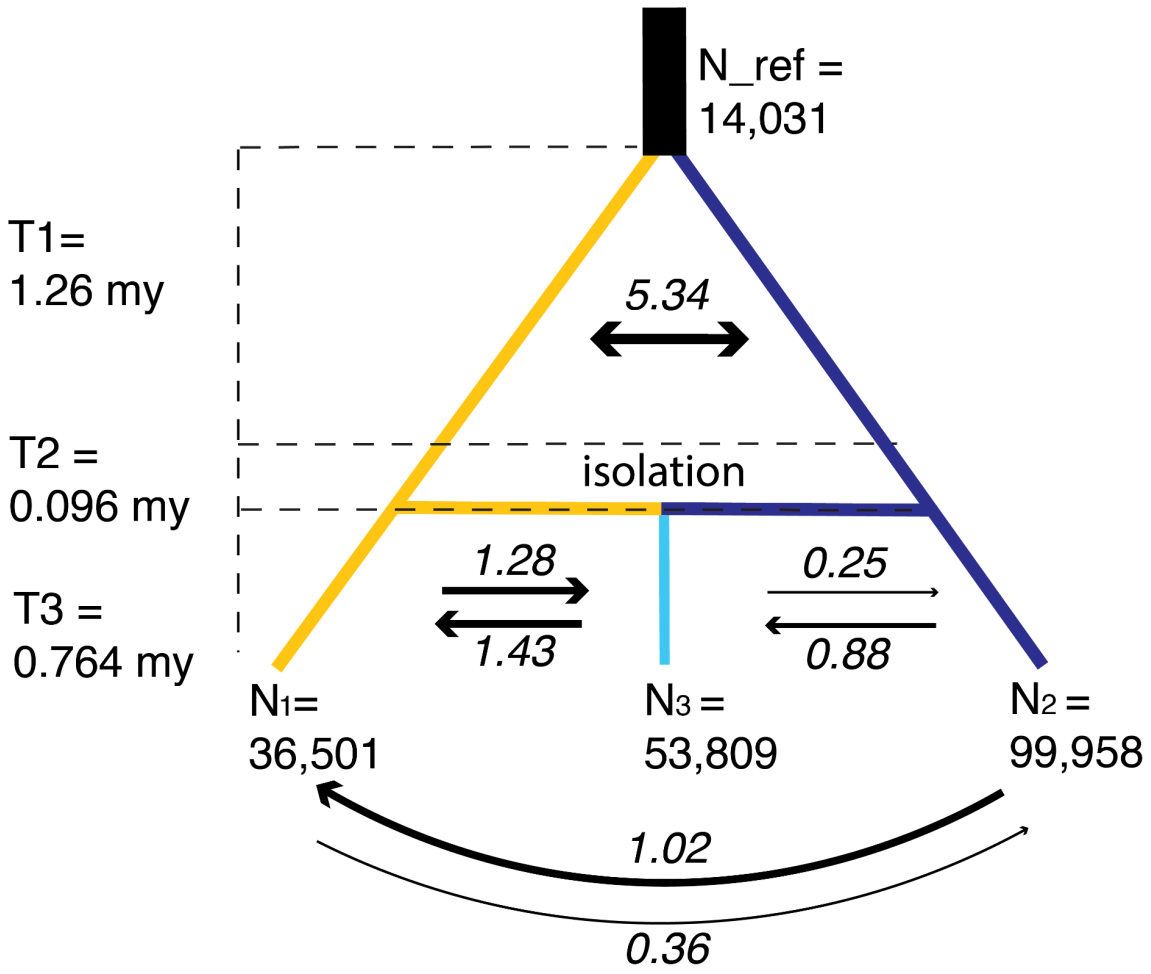


Figure 4. Distribution of samples used in three population demographic inference (a) and the best fit model of divergence among *coastal P. trichocarpa* (in yellow), *P. balsamifera* (in dark blue), and *interior P. trichocarpa* (in light blue). Parameter estimates and confidence intervals for the 3-population model are provided in Table 2. Note that T1, T2, and T3 are time intervals in millions of years (my).