# Climate and geography shape hybrid zone dynamics across space and time: insights from *Populus trichocarpa* x *P. balsamifera*

**Authors**: Constance E. Bolte, Tommy Phannareth, Matthew C. Fitzpatrick, Jason A. Holliday, Stephen R. Keller, Jill A. Hamilton

**Corresponding author:**
Constance E. Bolte, connie bolte@gmail.com

**ORCiDs:**
Constance E. Bolte, 0000-0003-0438-7102
Department of Ecosystem Science and Management, Pennsylvania State University, University Park, PA, 16801

Tommy Phannareth, 0009-0007-1042-6098
Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg VA 24061

Matthew C. Fitzpatrick, 0000-0003-1911-8407
Appalachian Laboratory, University of Maryland Center for Environmental Science, Frostburg, MD 21532

Jason A. Holliday, 0000-0002-2662-8790
Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg VA 24061

Stephen R. Keller, 0000-0001-8887-9213
Department of Plant Biology, University of Vermont, Burlington, VT 05465

Jill A. Hamilton, 0000-0002-3041-0508
Department of Ecosystem Science and Management, Pennsylvania State University, University Park, PA, 16801

## Summary

- Quantifying the mechanisms underlying hybrid zone formation requires an evaluation of both neutral and nonneutral evolutionary processes. Population demographic changes, alongside landscape heterogeneity and climate adaptation, 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.
- We quantified the contributions of climate and geography to the genetic structure across replicate contact zones between *Populus trichocarpa* and *P. balsamifera,* evaluating the demographic context of hybridization and the role of genetic connectivity over time.
- We found genomic structure differed among the seven latitudinally arrayed transects. While ancestry was strongly structured by climate, geography appears to have restricted gene flow in certain regions of the hybrid zone. The best fitting demographic model indicated high rates of gene flow alongside selection have likely influenced the establishment and maintenance of the hybrid zone.
- Our results point towards the importance of climate in structuring the contact zones between *P. trichocarpa* and *P. balsamifera* and emphasize the value of sampling replicated zones of contact to understand how neutral and non-neutral processes influence hybrid zone formation across space and time.

**Keywords:** Natural hybridization, selection, climate, demographic history, forest trees, gene flow, *Populus*

## Introduction

Understanding the processes influencing the formation and maintenance of species is a central goal of evolutionary biology and crucial to species management in a rapidly changing environment (Frankham 2010; Hoffmann *et al.,* 2015; Payseur & Rieseberg, 2016). As such, hybridization and hybrid formation have long been of interest (Jeffrey, 1916; Dobzhansky, 1936; Heiser, 1949; 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). Natural hybrid zones exist across taxonomic groups; however, they are particularly prevalent among forest tree species (Swenson & Howard, 2005; Abbott, 2017; Suarez-Gonzalez *et al.*, 2018). Climatic oscillations during glacial and interglacial periods have influenced the distribution of many temperate and boreal trees, facilitating opportunities for interspecific gene flow (Hewitt, 2004; Jump & Penuelas, 2005; Swenson & Howard, 2005; Soltis *et al.,* 2006; Hamilton *et al.,* 2015). Despite widespread occurrence of hybridization, gaps remain in our understanding of how neutral and nonneutral evolutionary processes interact to influence genomic divergence, shaping the extent and distribution of hybrid zones in forested landscapes. Differentiating these processes requires an assessment of historical, geographical, and climatic factors contributing to hybridization. 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.

Both neutral and nonneutral processes influence standing genetic variation, impacting the evolutionary trajectory of populations across space and time. The formation of hybrid zones often occurs when previously isolated lineages come into contact (Barton & Hewitt, 1989; Abbott, 2017). The shift between isolation and contact can coincide with changes in effective population size ($N_e$). While range-wide variance in $N_e$ can influence the direction and extent of genetic exchange, fine-scale variance in $N_e$ at the population or genome-level can further influence elimination or fixation of alleles. This points towards

103 the important role neutral processes (e.g., drift, gene flow) can play in shaping the genetic
104 structure of the hybrid zone. However, population genetic and macroevolutionary patterns
105 suggest that isolation, in the strict sense, is uncommon in forest trees, which often
106 continue to exchange genes following divergence (Petit & Hampe, 2006). For example,
107 interspecific and intraspecific gene flow can continue to occur either within or across
108 refugia via pollen dispersal (Hampe & Petit, 2005; Hampe & Jump, 2011; Woolbright *et*
109 *al*., 2014). Genetic exchange often appears to be largely asymmetrical favoring the
110 genomic background of one species due solely or in part to differential dispersal capacity,
111 wind-patterns, and unidirectional reproductive incompatibilities (Lexer *et al*., 2005; Lexer
112 *et al*., 2006; Lepais *et al*., 2009; Hamilton *et al*., 2013b; El Mujtar *et al*., 2017).
113 Characterizing directionality of gene flow is needed to predict evolutionary change given
114 that an influx of genetic variants via interspecific gene flow not only buffers $N_e$ but
115 generates novel genetic recombinants upon which natural selection can act. Despite
116 variance in demographic history across the landscape, forest tree populations often retain
117 substantial genetic variation resulting from genetic exchange while exhibiting signatures
118 of local adaptation (reviewed in Kremer *et al*., 2012). This highlights the important role
119 natural selection plays in the evolution and diversification of forest tree populations.
120
121 Demographic inference can be particularly useful for characterizing processes underlying
122 patterns of genetic exchange that may have influenced species evolutionary change,
123 distributional shifts, or adaptive potential (Bacilieri *et al.,* 1996; Petit *et al.,* 2004). For
124 example, it is now possible to test the differential contributions of neutral and non-neutral
125 processes to divergence by parameterizing models to evaluate the rate of gene flow
126 subjected to selection (i.e., a test for heterogeneous gene flow; first employed in Tine *et*
127 *al.,* 2014). From a theoretical standpoint, during the early stages of speciation when
128 species boundaries may still be semi-permeable, it is expected that heterogeneous rates
129 of gene flow will occur across the genome (Harrison & Larson, 2014). Several forest tree
130 hybrid zones have characterized patterns of introgression, including oaks (Eaton *et al*.,
131 2015; McVay *et al*., 2017; Cannon & Petit, 2020), poplars (Lexer *et al.*, 2005; Suarez-
132 Gonzales *et al*., 2016; Christe *et al*., 2017; Chhatre *et al*., 2018; Suarez-Gonzales *et al*.
133 2018), and spruce (Hamilton *et al.*, 2013a,b). However, few studies have leveraged whole

genome sequencing data to characterize the role of neutral and non-neutral processes to the formation and maintenance of natural hybrid zones through the lens of demographic change, landscape heterogeneity, and environmental selection in forest trees (Christe *et al.*, 2017; Leroy *et al.*, 2020).

Here, we use *Populus*, a model system in forest trees that comprises ecologically and economically important species that hybridize where their 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 biology, comparative genomics, and adaptive introgression (e.g., Jansson & Douglas, 2007; Suarez-Gonzalez *et al*., 2016; Shang *et al.,* 2020). Natural hybrid zones exist between *P. trichocarpa* and *P. balsamifera*, which has prompted previous investigations to assess the mechanisms underlying hybrid zone maintenance (Geraldes *et al*., 2014; Suarez-Gonzalez *et al*., 2018). Species contact zones are largely associated with geographically and climatically steep transitions from maritime to continental climates west and east of the Rocky Mountains, suggesting divergence has likely involved a combination of intrinsic and extrinsic factors, including landscape-related barriers to gene flow and climatically-related responses to selection. In areas where they currently or historically overlap, introgression from *P. trichocarpa* and *P. angustifolia* into the genetic background of *P. balsamifera* is thought to have promoted adaptation in *P. balsamifera,* particularly along the southern and western range edges of its contemporary distribution (Chhatre *et al.*, 2018). Similarly, Geraldes *et al*. (2014) proposed that introgressed variants of *P. balsamifera* enabled *P. trichocarpa* to disperse into geographically and climatically novel regions. Indeed, Suarez-Gonzalez *et al*. (2018) concluded that asymmetrical introgression of *P. balsamifera* into *P. trichocarpa* background was adaptive and likely facilitated the expansion of *P. trichocarpa* into colder climates. These findings suggest hybridization is a mechanism through which *Populus* adapts to environmental change, including and expansion and persistence into marginal or novel environments (Cannon & Petit, 2020).

Using a latitudinally replicated transect sampling design, we focused on the natural hybrid zone between *P. trichocarpa* and *P. balsamifera* through the Rocky Mountains. We

sampled and sequenced whole genomes for 576 trees across seven replicate contact zones. With these data, we (1) described the genomic structure of replicate contact zones, (2) tested the role of climate, geography, and landscape-level barriers to genomic structure, and (3) disentangled the contributions of neutral and non-neutral interspecific gene flow to divergence. We discuss our findings in the light of historical and contemporary climate change and emphasize the role whole genome resequencing can play in understanding how neutral and non-neutral processes influence hybrid zone formation across space and time.

## Methods

### *Sampling of plant material*

Vegetative branch cuttings were collected from 576 poplar trees spanning the natural contact zone between *Populus trichocarpa* and *Populus balsamifera* in January 2020 (Figure 1). These represent individual collections spanning seven latitudinal transects (hereafter referred to as Alaska_1, Alaska_2, Cassiar, Chilcotin, Jasper, Crowsnest, and Wyoming) across much of the species' overlapping latitudinal ranges, from 40° N to 65° N and -100° W to -150° W longitude. All trees from which cuttings were taken were provided a unique identifier and latitude, longitude, and elevation of origin were recorded. Vegetative branch cuttings were propagated at Virginia Tech and exposed to a 30 second Zerotol 2-0 (27.1% hydrogen peroxide, 2% peroxyacetic acid) fungal dip treatment at 1:100 ratio of Zerotol 2-0 to water. Cuttings were then dipped in Garden Safe Take Root hormone, rooted into a standard rooting mix (peat moss, vermiculite, perlite, osmocote, bone meal, gypsum, epsom salt, and dolomite), and subsequently placed on a mist bench (Phytotronics Nova mist controller, DGT blue/white mist .53 GPM). Mist was applied for 38 days (March 9 - April 17, 2020); for the first 31 days, every 10 minutes for 8 seconds, then the last 7 days, every 60 minutes for 10 seconds. Greenhouse conditions included a day temperature of 24 °C, a night temperature of 16 °C without supplemental lighting. Young leaves were collected from each genotype for

195 subsequent genomic DNA extraction. Tissue samples were placed immediately on dry
196 ice and then transferred to a -80°C freezer for long-term preservation.
197
198 ***Genomic Library Preparation and Bioinformatics***
199 For each of the 576 samples, approximately 100mg of leaf tissue was weighed and
200 frozen using liquid nitrogen, ground into fine powder with mortar and pestle, and
201 transferred into a microcentrifuge tube for genomic DNA extraction. Details for DNA
202 extraction, library preparation, filtering and quality control during variant calling are
203 provided in Supplementary.S1; Appendix 1. After quality filtering, ~34 million single
204 nucleotide polymorphisms (SNPs) and ~4.8 million indels remained in the variant call
205 set. Filtering removed all of the ~28 million multiallelic sites, of which ~5 million were
206 SNP sites. For 574 samples of the 576 collected, there was 0.7 ± 0.6 % missing data.
207 The two samples with higher missing data were removed. Genetic differentiation among
208 individuals was visualized using principal component analysis (PCA), from which we
209 identified 28 additional sample outliers (Supplementary.S1, Figure S1) that probably
210 reflect ancestry from different *Populus species*. We suspect this reflects the inclusion of
211 *P. deltoides* and *P. angustifolia*, whose distributions overlap marginally with our southern
212 sampling transects and which may have been mis-identified in the field during our winter
213 sampling. For the purposes of this study, we excluded the 30 outlier samples, leaving
214 546 individuals for subsequent genetic analyses.
215
216 ***Analyses of genomic structure***
217 To assess the genetic structure of the contact zones between *P. balsamifera* and *P.*
218 *trichocarpa*, we first thinned variant sites in high linkage disequilibrium with *Plink* (v1.9),
219 using 10,000 bp windows sliding in increments of 1000 bp along each chromosome, and
220 flagging site pairs with an $r^2 > 0.1$. These sites, along with singletons (--mac 2) and those
221 with missing data (--max-missing 1.0), were removed using *vcftools*, version 0.1.16. The
222 resulting 299,335 low LD variant sites were used to visualize the scale and spatial extent
223 of hybridization. We first explored genomic structure using a principal component
224 analysis (PCA), using the 'prcomp' function of the *stats* package in R, version 4.2.1. We
225 then performed admixture analysis using the program ADMIXTURE (Alexander *et al.,*

226 2009) at cluster assignments ranging from K=2 to K=8. The cluster assignment (K) with
227 the lowest cross validation error (k-folds = 10) was determined to best fit the data
228 (Supplementary.S1, Figure S2).
229
230 ***Associating genetic structure with climate and geography***
231 Climate data averaged across the years 1991-2020 for twenty-five variables associated
232 with geographic origin were extracted using *ClimateNA (*version 5.2; Wang *et al*. 2016)
233 for each sampled tree (Supplementary.S2) and then tested for collinearity using variance
234 inflation factor (VIF) analysis implemented in the *usdm* R package (version 1.1-18; Naimi
235 *et al.*, 2014). Six climatic variables with a VIF less than 10 were retained for subsequent
236 analysis, including TD (continentality), MAT (mean annual temperature), MAP (mean
237 annual precipitation), CMD (climate moisture deficit), RH (relative humidity), and PAS
238 (precipitation as snow).
239
240 To quantify the multivariate relationships between genetic variation, climate, and
241 geographic variation across the hybrid zone, a redundancy analysis (RDA) was
242 conducted within the *vegan* package, version 2.6-2 (Oksanen *et al*., 2020) in R version
243 4.2.1. Genotype data in variant call format for all 546 samples based on 299,335 variants
244 sites was converted to a 012 file (i.e., minor allele counts for each sample) using vcftools
245 (version 0.1.16). A permutation-based analysis of variance (ANOVA) procedure with 999
246 permutations (Legendre & Legendre, 2012) was used to test the overall statistical
247 significance of the RDA model ($\alpha$ = 0.05) as well as the significance of each RDA axis
248 separately. Variance partitioning was then performed using the 'varpart' function of the
249 *vegan* package to quantify the influence of predictor variables and their confounding
250 effects to observed genetic variation.
251
252 ***Visualizing barriers to gene flow***
253 We evaluated genetic connectivity and barriers to gene flow across the contact zones
254 between *P. trichocarpa* and *P. balsamifera* using EEMS (Estimated Effective Migration
255 Surfaces; Petkova *et al*., 2016). A genetic dissimilarity matrix was generated using the
256 bed2diffs script included within the EEMS program informed by the 299,335 genome-

wide variant sites across 546 individuals. Two additional geographic files were created using geographic coordinates; including the location data of each individual (.coord extension) and a large general habitat polygon (.outer extension) which was manually obtained using Google Maps API v3 Tool (http://www.birdtheme.org/useful/v3tool.html). The geographical coordinates for each sample alongside the parameterized grid space (500 demes with each grid cell ~21,000 $km^2$) assigned individuals to one of 49 demes, defined as pseudo populations of close geographical proximity to a vertex of the defined grid space. 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 thinned iterations by 99,999 between writing steps. The results of three replicate runs correspond to effective diversity (q) and effective migration rate (m) surfaces and were visualized using the R package rEEMSplots v 0.0.1 (Petkova *et al.*, 2016). An average across the three replicate runs was used to visualize effective migration rates (in log 10 scale) across the contact zone.

***Demographic inference***

To quantify demographic change across the hybrid zone over time, we used the unfolded joint site frequency spectrum (jSFS) of derived allele counts based on polarizing genotypes based on a consensus ancestral state fasta file. The consensus fasta file was made following the approach of Luqman *et al.* (2023). Details pertaining to data sources and program implementation are provided in Supplementary_S1, Appendix 2. Demographic inference was performed using parental genotypes of each of the three genetic clusters identified in admixture analysis: coastal *P. trichocarpa*, interior *P. trichocarpa*, and *P. balsamifera.* We first filtered the 546 samples to 318 parental genotypes (116 *P. balsamifera* with Q > 0.90; 202 *P. trichocarpa* with Q < 0.10) using a Q > 0.9 ancestral coefficient threshold from $K = 2$ admixture analysis. We then filtered out singletons (*vcftools*; --mac 2) which reduced the number of variant sites from 299,335 to 287,362 for the 318 parental-type trees. We then assigned allelic 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

288 undefined ancestral alleles. Following polarization, the data were further subset for
289 pairwise demographic inference based on three genetic clusters (*K* = 3). Using a Q > 0.9
290 ancestral coefficient threshold from *K* = 3 admixture analysis, there were 112 parental-
291 type samples for *P. balsamifera*, 51 for coastal *P. trichocarpa*, and 21 for interior *P.*
292 *trichocarpa*. Three pairwise assessments were used in demographic inference: 1) coastal
293 *P. trichocarpa* - *P. balsamifera* divergence, 2) interior *P. trichocarpa* - *P. balsamifera*
294 divergence, and 3) coastal *P. trichocarpa* - interior *P. trichocarpa* divergence. The derived
295 allele frequencies of shared polymorphic sites were used to build three respective
296 unfolded joint site frequency spectrums (jSFS) in the program easySFS. The jSFS for
297 inference of coastal *P. trichocarpa* and *P. balsamifera* divergence was produced from
298 shared allele frequencies across 238,168 variant sites. The jSFS for interior *P. trichocarpa*
299 and *P. balsamifera* was informed from 226,097 variant sites, and for the jSFS between
300 coastal and interior *P. trichocarpa* 169,758 variant sites were informative. Each set of
301 genetic data for pairwise demographic inferences were subject to the same filtering
302 (*vcftools*; –mac 2 –max-missing 1.0).
303
304 Demographic modeling was conducted using Diffusion Approximation for Demographic
305 Inference ($\partial\alpha\partial i$ v.2.1.1; Gutenkunst *et al.,* 2009) to provide historical context to the
306 formation of this hybrid zone and quantify the extent and directionality of interspecific
307 gene flow between genetic clusters. We tested nine demographic models representing
308 different hybridization scenarios (recurring versus secondary contact; with and without
309 heterogeneity in rates of gene flow) and effective population size change (with or without
310 change across time periods) against a model of strict isolation (Supplementary.S1, Figure
311 S3). We performed ten replicate runs of each model in $\partial\alpha\partial i$ with a 250 x 260 x 270 grid
312 space and the nonlinear Broyden-Fletcher-Goldfarb-Shannon (BFGS) optimization
313 routine. Akaike information Criterion (AIC; Akaike, 1974) was employed for model
314 selection, with the best replicate run (highest log composite likelihood) for each model
315 used to calculate ΔAIC ($AIC_{model\ i} - AIC_{best\ model}$) scores (Burnham & Anderson, 2002).
316 Fisher Information Matrix (FIM)-based uncertainty analysis was then conducted for the
317 best supported model to obtain upper and lower 95% confidence intervals (CIs) for the
318 inferred parameter estimates. To unscale parameter estimates and their 95% CIs we

319 used a mutation rate of 1.33 x $10^{-10}$ substitutions/site/year rate for *Populus trichocarpa*
320 (Hofmeister *et al*., 2020), a generation time of 15 years, and a calculated effective genome
321 length (Supplementary.S3). To calculate effective genome length, we multiplied the
322 proportion of variant sites used in the inference (count of filtered sites in each pairwise
323 inference with no missing data and singletons removed / count of variant sites prior to
324 filtering) to the number of base pairs from the *P. trichocarpa* reference genome
325 considered in variant calling (389,204,664 bp).
326
327
328 **Results**
329
330 ***Genetic structure across the hybrid zone***
331 Admixture analysis assuming two clusters indicated admixed individuals (0.10 < Q < 0.90)
332 were present across each transect (Figure 1). However, the best-fit cluster assignment of
333 *K*=3 indicated additional structure within *P. trichocarpa* that clustered geographically (i.e.,
334 north - coastal versus south - interior; Figure 1b). Pairwise estimates of genetic
335 differentiation ($F_{ST}$) across the three genetic clusters ranged from 0.053 - 0.056. Based
336 on *K*=2 and *K*=3, *P. trichocarpa* parental-types (Q < 0.10) were not present in either
337 Alaska transect, but hybrids of mixed ancestry were observed in the region (0.29 < Q <
338 0.90 at *K=2*; Supplementary.S2). The Cassiar transect had a high proportion of coastal
339 *P. trichocarpa* parental types and hybrids with a greater proportion of *P. balsamifera*
340 ancestry. Chilcotin, Jasper, and Crowsnest transects exhibited genetic structure that
341 included all three genetic clusters (Figure 1b), indicating a contact zone between northern
342 and southern genetic clusters associated with *P. trichocarpa* also involved hybridization
343 with *P. balsamifera.* Shared ancestry associated with *P. balsamifera* along the Crowsnest
344 transect was restricted to the eastern edge of the *P. trichocarpa* distribution, as defined
345 by Little (1971). The Wyoming transect was dominated by interior *P. trichocarpa* parental-
346 types (Q > 0.90 at *K*=3; Supplementary.S2), whereas individuals sampled near the
347 western edge of this transect exhibited mixed ancestry between the coastal and interior
348 genetic clusters of *P. trichocarpa*. Consistent with the Crowsnest transect, *P. balsamifera*
349 ancestry was limited to the far eastern distributional limit of *P. trichocarpa, with* pure *P.*

350 *balsamifera* parental types restricted to fragmented populations geographically distant
351 from the main distribution of either species. Information related to each sampled tree
352 (geographical information, PC loadings, and admixture coefficients) are summarized in
353 Supplementary.S2.
354
355 Principal component analysis (PCA) largely ascribed genetic differences along PC1
356 (1.764% variance explained) to differences between species and PC2 (0.925% variance
357 explained) to genetic structure along north-south gradients in *P. trichocarpa* (Figure 1c).
358 Individuals labeled by ancestry coefficients from *K*=2 largely account for genetic
359 differences between *P. trichocarpa* (far left of PC1) and *P. balsamifera* (far right of PC1)
360 (Figure 1d). In addition, substantial genetic structure was captured for *P. trichocarpa*
361 along the PC2 axis, primarily separating the transects by latitude.
362
363 ***Associations of climate and geography to genetic structure***
364 Climate and geography explained 3.79% ($r^2$) of the genetic variance across the 546
365 sampled trees and 299,335 SNP dataset. The majority of the explainable genetic variance
366 (PVE) was accounted for by the first RDA axis (38.26%, Figure 2) with continentality (TD)
367 and mean annual temperature (MAT) having the highest predictor loadings
368 (eigenvectors). Individuals identified as *P. balsamifera* were associated with higher TD
369 and lower MAT on average. We observed the pattern of ancestry as strongly
370 environmentally structured by climate (Figure 2) The second RDA axis explained 19.23%
371 (PVE) of the genetic variance in the dataset. The variables with the highest predictor
372 loadings on the second RDA axis were relative humidity (RH), elevation, and longitude.
373 These variables were associated with differentiation within *P. trichocarpa* across the
374 seven transects. Associations between climate, geography and genetic variation were not
375 observed along RDA2 for *P. balsamifera* (Figure 2). Combining loadings for RDA1 and
376 RDA2, we found latitude, TD, and climate moisture deficit (CMD) had the strongest
377 associations with genetic variation within and among sampled transects.
378
379 Because geographic and climatic predictor variables are often correlated, we performed
380 variance partitioning to tease apart the respective contributions of climate, geography,

381 and their interactions to genomic variance. We observed a large confounded effect
382 (58.69%) between the climatic and geographic variables chosen for analysis. Removing
383 confounded effects through conditioning revealed climate was a stronger predictor of
384 genetic variance. Climate, conditioned on geography, accounted for 27.28% of genetic
385 variance, whereas geography, conditioned on climate, accounted for just 14.03% of
386 genetic variance (Table 1).
387
388 **Table 1.** Partitioning of Variance of the full and partial contributions of climatic and
389 geographic variation on genetic variation. Both predictors explained 38.26% of the total
390 genetic variation. Adjusted $r^2$ represents the percent of genetic variance that could be
391 explained by the predictors after removing correlative effects. The proportion of variance
392 explained (PVE) represents the overall contribution of the predictors, both full and
393 partial, to genetic structure. Models are ordered by the greatest PVE.
394
395

| **Models** | **Adj. $r^2$ (%)** | **PVE (%)** |
|---|---|---|
| Genetics ~ Climate + Geo | 2.17* | 38.26 |
| Genetics ~ Climate \| Geo | 0.59* | 27.28 |
| Genetics ~ Geo \| Climate | 0.31* | 14.03 |
| Residuals | 97.83 | - |

396
397
398 ***Barriers to gene flow***
399 Using departures from isolation-by-distance (IBD) to indicate barriers to gene flow, EEMS
400 analysis identified the continental divide of the Rocky Mountains as a large barrier
401 between *P. trichocarpa* and *P. balsamifera*. Additional barriers to gene flow were
402 identified between transects associated with elevational gradients (Figure 3). The EEMS
403 results indicated both Alaska transects were more isolated from the other northern
404 transect, Cassiar, than expected based on IBD. In contrast, individuals spanning *P.*
405 *trichocarpa* parental-type and hybrids across the Chilcotin, Jasper, and Crowsnest
406 transects were more similar than expected under IBD, suggesting substantial intraspecific

407 migration across latitudes within the southern range of this hybrid zone. Interestingly, this
408 pattern was not observed for *P. balsamifera,* as only individuals located east of the Rocky
409 Mountains exhibited higher effective migration than expected.
410
411 ***Demographic inference***
412 Comparing the ten replicate runs for each of the nine demographic models examined for
413 the divergence between coastal *P. trichocarpa* and *P. balsamifera*, the best fit model
414 allowed for heterogeneous rates of gene flow to recur following divergence and population
415 size changes during the second time period (PSCIM_het; Table 2). According to this
416 model (Figure 4a), species diverged approximately **929,717** years ago (T1 + T2; Figure
417 4a) and for the first **921,291** years (95% CI: 887,185 - 955,398) relatively large effective
418 population sizes ($N_e$) were maintained across species. However, around **8,426** years ago
419 (95% CI: 8,361 - 8,491), a significant reduction in $N_e$ was inferred. For *P. balsamifera*, $N_e$
420 decreased from approximately **5.07 million** to **9,684** individuals, while for coastal *P.*
421 *trichocarpa*, $N_e$ decreased from **1.10 million** to **9,629**. Gene flow was predominantly
422 asymmetric, with alleles moving primarily from coastal *P. trichocarpa* into the *P.*
423 *balsamifera* genomic background. From this model, the observed rates of gene flow from
424 coastal *P. trichocarpa* (Ne2) into *P. balsamifera* (Ne1) for loci under selection ($m12_s$=
425 *1.89e-05*) were marginally lower than neutral gene flow rates ($m12_n$ = 2.19e-05). In
426 contrast, gene flow from *P. balsamifera* into coastal *P. trichocarpa* was nearly 2 orders of
427 magnitude lower for regions under selection than for neutral regions of the genome ($m21_s$
428 = 4.18e-08 versus $m21_n$ = 3.83e-06), implying the existence of a strong asymmetry in
429 gene flow rates between species driven primarily by selection. Consistent with this,
430 selection-driven divergence between coastal *P. trichocarpa* and *P. balsamifera* was the
431 highest among the 3 pairwise demographic comparisons, with only 18.23% (95% CI:
432 13.39 - 23.08%) of the genomes evolving under neutral processes (inferred by parameter
433 *P*). The narrow confidence intervals for all parameter estimates indicated good model fit
434 (Table 3).
435

436 **Table 2.** Demographic inference AIC comparisons across models. Model abbreviations
437 are defined in the footnote. The lower the AIC and delta.AIC the better the model fit to
438 the data.
439

| | | balsamifera – coastal trichocarpa | | balsamifera – interior trichocarpa | | coastal trichocarpa - interior trichocarpa | |
|---|---|---|---|---|---|---|---|
| **Model** | ***k*** | **AIC** | **delta.AIC** | **AIC** | **delta.AIC** | **AIC** | **delta.AIC** |
| SI | 3 | 231,047.37 | 15,024.48 | 183,411.33 | 11.63 | 128,494.69 | 13.20 |
| IM | 5 | 233,704.43 | 17,681.54 | 186,041.87 | 2,642.17 | 159,260.35 | 30,778.85 |
| SC | 6 | 229,100.67 | 13,077.78 | 185,638.23 | 2,238.53 | 129,217.07 | 735.57 |
| PSCSC | 8 | 229,764.16 | 13,741.27 | 185,874.52 | 2,474.82 | 128,481.50 | **0.00** |
| PSCIM | 8 | 230,487.57 | 14,464.68 | 185,373.75 | 1,974.05 | 137,636.07 | 9,154.58 |
| SC | 9 | 233,262.65 | 17,239.76 | 185,744.80 | 2,345.10 | 128,627.54 | 146.04 |
| IM | 8 | 237,924.83 | 21,901.94 | 234,282.84 | 50,883.14 | 160,842.28 | 32,360.79 |
| PSCSC_het | 11 | 230,295.22 | 14,272.33 | 185,104.34 | 1,704.64 | 128,860.78 | 379.29 |
| PSCIM_het | 11 | 216,022.89 | **0.00** | 183,399.70 | **0.00** | 128,736.92 | 255.42 |

440
441 *Models; SI (strict isolation), IM (isolation-with-migration), SC (secondary contact), PSC
442 (population size change). Models_het (heterogeneous gene flow parameters inferred).
443
444

445 Similar to the divergence history of coastal *P. trichocarpa* and *P. balsamifera,* the best fit
446 model for interior *P. trichocarpa* and *P. balsamifera* also allowed for heterogeneous rates
447 of gene flow to recur following divergence and an effective population size change at T2
448 (PSCIM_het; Table 2). According to this model (Figure 4b), interior *P. trichocarpa* and *P.*
449 *balsamifera* diverged from each other approximately **26,369** years ago and reductions to
450 $N_e$ occurred **3,462** years ago (95% CI: 3,439 - 3,485). For *P. balsamifera*, $N_e$ was reduced
451 from **117,186** (95% CI: 114,987 - 119,384) to **6,620** (95% CI: 6,571 - 6,669), while for
452 interior *P. trichocarpa*, $N_e$ was reduced from **42,376** (95% CI: 41,993 - 42,759) to **3,171**
453 (95% CI: 3,146 - 3,195). However, unlike the inference for coastal *P. trichocarpa* and *P.*
454 *balsamifera* divergence, gene flow was nearly symmetrical, with low rates and wide
455 confidence intervals. The most narrow confidence interval was related to neutral gene
456 flow of interior *P. trichocarpa* variants into *P. balsamifera* ($m12_n$ = 8.03e-07; 95% CI:

457 9.29e-08 - 1.51e-06) and was the only inference for introgression that had positive values
458 at the lower 95% CI bounds (Table 3). Based on the patterns in the jSFS for *P.*
459 *balsamifera* and interior *P. trichocarpa*, these two genetic clusters exhibited the second
460 highest level of divergence among pairwise inferences, with only 22.30% (95% CI: 0 -
461 49.54) of the genomes evolving under neutral processes.
462
463 Demographic inference indicated a recent divergence (approximately 2,397 years ago,
464 Figure 4c) between the two genetic clusters observed within *P. trichocarpa* (coastal and
465 interior). The divergence history was characterized by a period of isolation lasting
466 approximately 2,366 years (95% CI: 2,307 - 2,425), followed by very recent gene flow
467 beginning within the last several decades (95% CI: -81.61-144.1). Interestingly, the onset
468 of gene flow coincided with drastic increases to $N_e$. For the interior genetic cluster, $N_e$
469 increased from 1,958 (95% CI: 1,940 - 1,978) to 12,802 (95% CI: 12,802.2 - 12,802.3).
470 For the coastal genetic cluster, $N_e$ increased substantially from 2,651 (CI: 2,6514 - 2,687)
471 to 50,608 (CI: 24,254 - 76,963). The direction and extent of gene flow indicate
472 introgression from the interior to the coastal genetic cluster ($m12_n$ = 2.98e-05; 95% CI:
473 2.91e-05 - 3.04e-05). However, the small and not significant estimate of gene flow from
474 the coastal to the interior genetic cluster suggests that introgression is limited in the
475 opposite direction ($m21_n$ = 7.57e-07; 95% CI: -5.67e-06 - 7.18e-06).
476
477 Based on our data, contemporary gene flow is primarily moving from coastal *P.*
478 *trichocarpa* into *P. balsamifera* and from interior *P. trichocarpa* into coastal *P. trichocarpa*
479 (Figure 4e). The residuals between the model and data fit were primarily associated with
480 the under-representation of low frequency alleles (Supplementary.S1, Figure S3).
481 Calculations for AIC, parameter unscaling, and confidence intervals from FIM uncertainty
482 tests are available in Supplementary.S2. Python scripts for each model evaluated are
483 available on GitHub (https://github.com/boltece/poplar_hybrids).
484
485
486
487
488

**Table 3.** Parameter estimates and 95% confidence intervals (CIs) associated with each of the best fit pairwise models of divergence. Parameters related to introgression are denoted with “m” and numbers indicate directionality such that m12 is interpreted as Ne2 moving into Ne1. The extra notation of (n) is for neutral introgression and (s) is for introgression experiencing selection. Some CIs match the unscaled parameter estimates due to scientific notation. Supplementary S3 allows interaction with the table cells to view more precise reporting of lower and upper CIs.

| model | balsamifera (Ne1) - coastal trichocarpa (Ne2) | | | balsamifera (Ne1) - interior trichocarpa (Ne2) | | | coastal trichocarpa (Ne1) - interior trichocarpa (Ne2) | | |
|---|---|---|---|---|---|---|---|---|---|
| **parameter** | **unscaled** | **lower_CI** | **upper_CI** | **unscaled** | **lower_CI** | **upper_CI** | **unscaled** | **lower_CI** | **upper_CI** |
| **Ne1_T1** | 5.07E+06 | 4.90E+06 | 5.24E+06 | 1.17E+05 | 1.15E+05 | 1.19E+05 | 2.65E+03 | 2.61E+03 | 2.69E+03 |
| **Ne2_T1** | 1.10E+06 | 9.97E+05 | 1.20E+06 | 4.24E+04 | 4.20E+04 | 4.28E+04 | 1.96E+03 | 1.94E+03 | 1.98E+03 |
| **T1** | 9.21E+05 | 8.87E+05 | 9.55E+05 | 2.29E+04 | 2.29E+04 | 2.29E+04 | 2.37E+03 | 2.31E+03 | 2.42E+03 |
| **Ne1_T2** | 9.68E+03 | 9.68E+03 | 9.68E+03 | 6.62E+03 | 6.57E+03 | 6.67E+03 | 5.06E+04 | 2.43E+04 | 7.70E+04 |
| **Ne2_T2** | 9.63E+03 | 9.63E+03 | 9.63E+03 | 3.17E+03 | 3.15E+03 | 3.20E+03 | 1.28E+04 | 1.28E+04 | 1.28E+04 |
| **T2** | 8.43E+03 | 8.36E+03 | 8.49E+03 | 3.46E+03 | 3.44E+03 | 3.48E+03 | 3.13E+01 | -8.16E+01 | 1.44E+02 |
| **m12 (n)** | 2.19E-05 | 2.18E-05 | 2.20E-05 | 8.03E-07 | 9.29E-08 | 1.51E-06 | 2.98E-05 | 2.91E-05 | 3.04E-05 |
| **m21 (n)** | 3.83E-06 | 3.15E-06 | 4.52E-06 | 8.30E-07 | -9.06E-07 | 2.57E-06 | 7.57E-07 | -5.67E-06 | 7.18E-06 |
| **m12 (s)** | 1.89E-05 | 1.88E-05 | 1.90E-05 | 6.29E-07 | -1.43E-07 | 1.40E-06 | -- | -- | -- |
| **m21 (s)** | 4.18E-08 | 3.13E-08 | 5.22E-08 | 7.53E-07 | -2.66E-06 | 4.16E-06 | -- | -- | -- |
| **P** | 1.82E-01 | 1.34E-01 | 2.31E-01 | 2.23E-01 | -4.93E-02 | 4.95E-01 | -- | -- | -- |
| **theta** | 7.56E+05 | 7.53E+05 | 7.58E+05 | 9.72E+05 | 9.63E+05 | 9.82E+05 | 8.26E+05 | 8.15E+05 | 8.37E+05 |

## Discussion

This study provides valuable insights into the complex dynamics underlying hybridization, highlighting the interplay between neutral and non-neutral processes shaping the genetic structure of hybrid zones. Our results point to the different contributions of climate,

geography, and landscape-level features to timing, extent, and maintenance of genetic differences across natural hybrid zones. Geographic barriers have played a significant role in limiting historical and contemporary gene flow, but where regions of connectivity exist, asymmetries in gene flow characterize *Populus* genetic exchange. These findings shed light on how the direction and extent of gene flow varies across contact zones and provide insights into how much of the genome is selectively porous to introgression. Furthermore, our results indicate a strong association between the genomic structure of this dynamic hybrid zone and climatic gradients. This suggests that genetic patterns observed could be sensitive to climatic shifts, which has implications for understanding the future of these contact zones in a changing climate.

***Climate and landscape heterogeneity contribute to genetic structure***

Neutral and non-neutral processes, including landscape-induced restrictions to gene flow and selection have contributed to patterns of differentiation across the hybrid zone between *P. balsamifera* and *P. trichocarpa*. Estimated effective migration surfaces (EEMS) suggest gene flow has likely been suppressed by elevational barriers through the Rocky Mountains of northwestern North America. While mountains are generally considered disruptive to gene flow, *P. balsamifera* and *P. trichocarpa* are riparian trees that grow along river drainage basins, and thus river valleys provide potential conduits of connectivity. Our replicate transects largely followed major drainage systems across east-west mountain barriers where the species distributions are continuous and overlapping. Genetic differences identified through principal component analysis were species and transect-dependent, with hybrids positioned according to their degree of ancestry between parental types within each transect. Elevational barriers explained genetic divergence among transects, especially those of Alaska, Cassiar, Chilcotin and Wyoming (Figure 3). Patterns of connectivity inferred from EEMS suggest coastal and interior *P. trichocarpa* genetic clusters have experienced lower rates of gene flow than would be predicted from their continuous distribution. While the Coast and Cascade ranges create barriers to gene flow between these two clusters, extreme shifts in historical climate may have accelerated divergence and effective population size reductions. The recent divergence of coastal and interior *P. trichocarpa* may reflect the interplay of historical

535 climate and drift given the relatively small effect sizes inferred through demographic
536 modeling during a period of isolation. At present, these diverged genetic clusters are
537 distributionally adjacent and can undergo genetic exchange. Combined, the impacts of
538 historical climate to effective population size and distributional extent, in conjunction with
539 geographic barriers to gene flow, illustrate the interaction between non-neutral and
540 neutral processes in driving divergence.
541
542 Similar to genetic drift, selection can rapidly influence allele frequencies. Redundancy
543 analysis (RDA) and variance partitioning revealed a significant association between
544 genetic structure and climate across the hybrid zone. The Rocky Mountains present
545 environmental transitions from drier and colder conditions on the eastern side (favoring
546 *P. balsamifera*) to more maritime climates on the western side of the Rockies (favoring *P.*
547 *trichocarpa*). In addition, the distribution of both species align with north-south gradients
548 in average relative humidity and growing degree days relevant to phenology and growth
549 (Keller *et al.*, 2011; Olson *et al.*, 2013; Montgomery *et al*., 2020). These distributions are
550 comparable with hybrid zones between *Picea glauca* and *P. sitchensis* (Hamilton *et al*.,
551 2013b). The strong association of continentality to the genetic structure across *Populus*
552 aligns with conclusions in Hamilton *et al*. (2013b), which indicated maritime to continental
553 climate transitions were associated with genomic ancestry. This suggests these hybrid
554 zones likely reflect a model of bounded hybrid superiority where strong associations with
555 ecotonal transitions underscore the important role of selection in maintaining species
556 boundaries.
557
558 The best-fit demographic models indicate that approximately 80% of the genome has
559 been influenced by nonneutral evolutionary processes during species divergence. While
560 this estimate may appear high, the genome of *P. trichocarpa* has extensive linkage
561 disequilibrium (Shang *et al*., 2020) suggesting direct selection on a single locus can have
562 indirect, but genome-wide effects (Butlin & Smadja, 2018). Furthermore, many adaptive
563 traits have complex genetic architectures involving multiple genes of small individual
564 effect (e.g., phenology, drought, and cold tolerance; Alberto *et al*., 2013; Yeaman *et al*.,
565 2016; Lind *et al.*, 2018; Zhang *et al.*, 2019; De La Torre *et al*., 2022). With environmental

566 selection impacting many loci distributed across the genome, we expect genome-trait
567 interactions to be the dominant mode of lineage divergence in *Populus* (e.g., Kremer &
568 Hipp, 2020; Jin *et al*., 2021), and with variation in recombination rate that exists across
569 the genome and the evolution of intrinsic barriers (Shang *et al.*, 2020), porosity across
570 the genome is likely highly variable. The genomic landscape itself is characterized by a
571 heterogeneous distribution of linkage with varying effective population sizes (inferred in
572 Christe *et al.*, 2017 for *Populus*) and as such the effectiveness of selection or drift to
573 change allele frequencies is likely genomic-region dependent.
574
575 ***Heterogeneous gene flow and the role of selection in contact zones***
576 Demographic inference is commonly used to understand the impact neutral processes
577 have on divergence history. While gene flow is generally viewed as neutral, it also
578 introduces genetic variation that may be adaptive or even maladaptive. Our demographic
579 analysis evaluated gene flow to examine if and to what extent interspecific gene flow was
580 influenced by selection. Of the three pairwise demographic inferences evaluated, two
581 demonstrated the importance of heterogeneous gene flow, particularly involving *P.*
582 *balsamifera*. We observed high rates of gene flow from the coastal *P. trichocarpa* genetic
583 cluster into the genetic background of *P. balsamifera*, while the converse rate was
584 relatively low. Notably, nearly half of the introgressed variation was influenced by either
585 negative or positive selection (Figure 4a), but we suspect much of what was detected as
586 negative selection may indicate barriers to reproduction as recently found for other
587 *Populus* species relationships (Liu *et al.*, 2022). This suggests that intrinsic barriers to
588 reproduction have likely evolved between *P. trichocarpa* and *P. balsamifera* reducing or
589 concentrating genome porosity to select regions of the genome. Inferred asymmetry in
590 gene flow may suggest coastal *P. trichocarpa* has either a demographic or a selective
591 advantage within the ecotonal transition facilitating introgression. This pattern is similar
592 to that observed between *Picea glauca* x *P. sitchensis*, where wind patterns and
593 differential timing of pollen receptivity were identified as factors contributing to gene flow
594 directionality (Hamilton *et al*., 2013b). The influx of new genetic recombinants through
595 introgression with *P. balsamifera* following hybrid zone filtering could enhance adaptation
596 to a changing climate (Barton & Bengtsson, 1986; Martinsen *et al.,* 2001; Aitken *et al*.,

597 2008). Given the congruent selection and gene flow dynamics for two highly diverged
598 pairs of forest tree taxa within the same region, we expect these patterns to be
599 transferable to other hybridizing forest tree taxa of this region.
600
601 The demographic model for interior *P. trichocarpa* and *P. balsamifera* divergence also
602 revealed heterogeneous gene flow, albeit low and with wide confidence intervals. Low
603 migration rates are likely indicative of geographical isolation or a general lack of hybrids
604 geographically positioned between these genetic clusters to facilitate backcrossing.
605 Interestingly, our third demographic model, involving coastal *P. trichocarpa* and interior
606 *P. trichocarpa* divergence, inferred a very recent secondary contact (95% CI: 0 - 141
607 years ago) after ~2,400 years of divergence in isolation. While heterogeneity in gene flow
608 did not improve model fit, there are important take-aways associated with the interactions
609 between these two genetic clusters. Gene flow was largely unidirectional, favoring the
610 genetic background of coastal *P. trichocarpa*, and coincides with drastic effective
611 population size increases (i.e., 19x increase to $N_e$ for coastal versus 6.5x increase for the
612 interior). This suggests gene flow can act as a mechanism to increase species and
613 population-level genetic diversity (Suarez-Gonzalez *et al.*, 2018; Burgarella *et al.,* 2019).
614
615 ***Isolation-with-migration shapes the divergence history of P. balsamifera and P.***
616 ***trichocarpa***
617 According to the best fit demographic models of divergence between *P. balsamifera and*
618 *P. trichocarpa*, speciation occurred amid ongoing gene flow. We see two possible
619 explanations for this result. Species distributions may have moved in tandem with climate
620 such that populations were proximal enough to maintain gene flow, resulting in a stable
621 hybrid zone. Alternatively, periods of strict isolation may have occurred during extreme
622 climatic shifts but were not sustained long enough for genomic patterns of divergence
623 that associate with allopatry to develop, or patterns were broken down following
624 subsequent contact. This latter possibility may characterize hybridizing species with
625 porous genomes whose species boundaries vary temporally with repeated cycles of
626 reinforcement and homogenization (e.g., MIM hypothesis; He *et al.,* 2019) or due to

627 changes in the strength and direction of selection over historically varying climatic optima
628 (e.g., Shang *et al.*, 2020).
629
630 Divergence between *P. balsamifera* and coastal *P. trichocarpa* was initiated ~930,000
631 years ago, aligning with the Mid-Pleistocene Transition, when glacial cycles became
632 longer and shifts to interglacial climates became more abrupt and extreme (Willeit *et al.*,
633 2019). Our divergence time estimates are much older than the 75,000 years estimated
634 by Levsen *et al.* (2012) for *P. trichocarpa* and P*. balsamifera* but are congruent with
635 observations for this species-pair from Sanderson *et al.* (2023). Furthermore, the
636 divergence timing between the interior genetic cluster of *P. trichocarpa* from *P.*
637 *balsamifera* aligned with the last glacial maxima (~26,000 years ago), and yet still low
638 levels of gene flow persisted. Observed patterns of divergence with recurrent gene flow
639 between *P. trichocarpa* - *P. balsamifera* fit a growing consensus for common divergence
640 histories among forest tree species (e.g., Lexer *et al*., 2006; Menon *et al.,* 2018; He *et al.,*
641 2019; Bolte *et al.,* 2022).
642
643 Our demographic inference was designed primarily to evaluate rate, directionality, and
644 timing of gene flow following divergence. However, an unexpected observation surfaced
645 when allowing instantaneous change to effective population sizes. All three genetic
646 groups exhibited dramatic reductions to effective population size ($N_e$) during the early to
647 late Holocene climate (~4,400 - 8,400 years ago), a period of warming known as the
648 Holocene Climatic Optimum that was known to dry river basins, which may have affected
649 the distribution and genetic diversity of our focal riparian species (Vance *et al.,* 1983;
650 Zoltai & Vitt, 1990; Schneider, 2013).
651
652 ***Conclusions***
653 Here, we provide insight into the complex interplay between neutral and non-neutral
654 processes in shaping replicate contact zones between *P. balsamifera* and *P. trichocarpa*.
655 We found evidence for the influence of climate, geography, and landscape-level attributes
656 to genomic structure and connectivity within and between these species across latitudinal
657 gradients. The results highlight the role of geographic barriers in limiting historic and

contemporary gene flow, as well as the importance of regions of connectivity to genetic exchange. Heterogeneity in gene flow rates suggests selection has and continues to play an important role in shaping the contact zone, particularly across ecotonal transitions. Overall, our study sheds light on the role different evolutionary processes can play within a long-lived hybrid zone across space and time and highlights 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: Michelle Zavala Paez, 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. TP processed genomic data. CEB analyzed genomic data and conducted demographic inference while SRK, JH and JAH advised. 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 are currently being archived on NCBI and under embargo until publication (SRAXXXXXXX; we will update the SRA number once all reads are uploaded). Scripts for variance partitioning, RDA, and demographic inference can be found at https://github.com/boltece/poplar_hybrids

**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

## References:

**Abbott RJ. 2017.** Plant speciation across environmental gradients and the occurrence and nature of hybrid zones. *Journal of Systematics and Evolution* 55: 238–258.

**Alberto FJ, Aitken SN, Alía R, González-Martínez SC, Hänninen H, Kremer A, Lefèvre F, Lenormand T, Yeaman S, Whetten R, *et al.* 2013.** Potential for evolutionary responses to climate change - evidence from tree populations. *Global Change Biology* 19: 1645–1661.

**Alexander DH, Novembre J, Lange K. 2009**. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19: 1655–1664.

**Baack E, Melo MC, Rieseberg LH, Ortiz-Barrientos D. 2015.** The origins of reproductive isolation in plants. *New Phytologist* 207: 968–984.

**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: 79–91.

**Barton N, Bengtsson BO. 1986.** The barrier to genetic exchange between hybridizing populations. *Heredity* 57: 357–376.

**Barton NH, Hewitt GM. 1989.** Adaptation, speciation and hybrid zones. *Nature* 341: 497–503.

**Bolte CE, Faske TM, Friedline CJ, Eckert AJ. 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: 35.

**Burgarella C, Barnaud A, Kane NA, Jankowski F, Scarcelli N, Billot C, Vigouroux Y, Berthouly-Salazar C. 2019.** Adaptive introgression: An untapped evolutionary mechanism for crop adaptation. *Frontiers in Plant Science* 10.

**Butlin RK, Smadja CM. 2018.** Coupling, Reinforcement, and Speciation. *The American Naturalist* 191: 155–172.

**Cannon CH, Petit RJ. 2020.** The oak syngameon: more than the sum of its parts. *New Phytologist* 226: 978–983.

**Chhatre VE, Evans LM, DiFazio SP, Keller SR. 2018.** Adaptive introgression and maintenance of a trispecies hybrid complex in range-edge populations of Populus. *Molecular Ecology* 27: 4820–4838.

**Christe C, Stölting KN, Paris M, Fraïsse C, Bierne N, Lexer C. 2017.** Adaptive evolution and segregating load contribute to the genomic landscape of divergence in two tree species connected by episodic gene flow. *Molecular Ecology* 26: 59–76.

**De La Torre AR, Sekhwal MK, Puiu D, Salzberg SL, Scott AD, Allen B, Neale DB, Chin ARO, Buckley TN. 2022.** Genome-wide association identifies candidate genes for drought tolerance in coast redwood and giant sequoia. *The Plant Journal* 109: 7–22.

**Dobzhansky TH. 1936.** Studies on hybrid sterility. 11. Localization of sterility factors in Drosophila pseudoobscura hybrids. *Genetics* 21: 113

**Eaton DAR, Hipp AL, 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: 2587–2601.

**El Muztar 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: 49.

**Frankham R. 2010.** Challenges and opportunities of genetic approaches to biological conservation. *Biological Conservation* 143: 1919–1927.

**Geraldes A, Farzaneh N, Grassa CJ, Mckown AD, Guy RD, Mansfield SD, Douglas CJ, Cronk QCB. 2014.** Landscape genomics of populus trichocarpa: The role of hybridization, limited gene flow, and natural selection in shaping patterns of population structure. *Evolution* 68: 3260–3280.

**Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009.** Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics* 5.

**Hamilton JA, Lexer C, Aitken SN. 2013a.** Differential introgression reveals candidate genes for selection across a spruce (Picea sitchensis × P. glauca) hybrid zone. *New Phytologist* 197: 927–938.

**Hamilton JA, Lexer C, Aitken SN. 2013b.** Genomic and phenotypic architecture of a spruce hybrid zone (Picea sitchensis × P. glauca). In: Molecular Ecology. 827–841.

**Hamilton JA, De la Torre AR, Aitken SN. 2015.** Fine-scale environmental variation contributes to introgression in a three-species spruce hybrid complex. *Tree Genetics and Genomes* 11.

**Hamilton JA, Miller JM. 2016.** Adaptive introgression as a resource for management and genetic conservation in a changing climate. *Conservation Biology* 30: 33–41.

**Hampe A, Jump AS. 2011.** Climate Relicts: Past, Present, Future. *Annual Review of Ecology, Evolution, and Systematics* 42: 313–333.

**Hampe A, Petit RJ. 2005.** Conserving biodiversity under climate change: The rear edge matters. *Ecology Letters* 8: 461–467.

**Harrison RG, Larson EL. 2014.** Hybridization, introgression, and the nature of species boundaries. In: Journal of Heredity. Oxford University Press, 795–809.

**He Z, Li X, Yang M, Wang X, Zhong C, Duke NC, Wu CI, Shi S. 2019.** Speciation with gene flow via cycles of isolation and migration: Insights from multiple mangrove taxa. *National Science Review* 6: 275–288.

**Heiser CB. 1949.** Natural hybridization with particular reference to introgression. *The Botanical Review* 15: 645-687.

**Hewitt GM. 2004.** Genetic consequences of climatic oscillations in the Quaternary. In: Philosophical Transactions of the Royal Society B: Biological Sciences. 359:183–195.

**Hoffmann A, Griffin P, Dillon S, Catullo R, Rane R, Byrne M, Jordan R, Oakeshott J, Weeks A, Joseph L, *et al.* 2015.** A framework for incorporating evolutionary genomics into biodiversity conservation and management. *Climate Change Responses* 2.

**Hofmeister BT, Denkena J, Colomé-Tatché M, Shahryary Y, Hazarika R, Grimwood J, Mamidi S, Jenkins J, Grabowski PP, Sreedasyam A, *et al.* 2020**. A genome assembly and the somatic genetic and epigenetic mutation rate in a wild long-lived perennial Populus trichocarpa. *Genome Biology* **21**.

**Janes JK, Hamilton JA. 2017.** Mixing it up: The role of hybridization in forest management and conservation under climate change. *Forests* 8.

**Jansson S, Douglas CJ. 2007.** Populus: A model system for plant biology. *Annual Review of Plant Biology* 58: 435–458.

**Jeffrey EC. 1916.** Hybridism and the rate of evolution in angiosperms. *The American Naturalist.*

**Jin WT, Gernandt DS, Wehenkel C, Xia XM, Wei XX, Wang XQ. 2021.** Phylogenomic and ecological analyses reveal the spatiotemporal evolution of global pines. *Proceedings of the National Academy of Sciences of the United States of America* 118.

**Jump AS, Peñuelas J. 2005.** Running to stand still: Adaptation and the response of plants to rapid climate change. *Ecology Letters* 8: 1010–1020.

**Kremer A, Ronce O, Robledo-Arnuncio JJ, Guillaume F, Bohrer G, Nathan R, Bridle JR, Gomulkiewicz R, Klein EK, Ritland K, *et al.* 2012.** Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecology Letters* 15: 378–392.

**Kremer A, Hipp AL. 2020.** Oaks: an evolutionary success story. *New Phytologist* 226: 987–1011.

**Legendre, P. and Legendre, L., 2012.** *Numerical ecology*. Elsevier.

**Lepais O, Petit RJ, Guichoux E, Lavabre JE, Alberto F, Kremer A, Gerber S. 2009.** Species relative abundance and direction of introgression in oaks. *Molecular Ecology* 18: 2228–2242.

**Leroy T, Rougemont Q, Dupouey JL, Bodénès C, Lalanne C, Belser C, Labadie K, Le Provost G, Aury JM, Kremer A, *et al.* 2020.** Massive postglacial gene flow between European white oaks uncovered genes underlying species barriers. *New Phytologist* 226: 1183–1197.

**Levsen ND, Tiffin P, Olson MS. 2012.** Pleistocene speciation in the genus populus (salicaceae). *Systematic Biology* 61: 401–412.

**Lexer C, Fay MF, Joseph JA, Nica MS, 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: 1045–1057.

**Lexer C, Kremer A, Petit RJ. 2006.** Shared alleles in sympatric oaks: Recurrent gene flow is a more parsimonious explanation than ancestral polymorphism. *Molecular Ecology* 15: 2007–2012.

**Lind BM, Menon M, Bolte CE, Faske TM, Eckert AJ. 2018.** The genomics of local adaptation in trees: are we out of the woods yet? *Tree Genetics and Genomes* 14.

870 **Liu S, Zhang L, Sang Y, Lai Q, Zhang X, Jia C, Long Z, Wu J, Ma T, Mao K, *et al.***
871 **2022.** Demographic History and Natural Selection Shape Patterns of Deleterious Mutation
872 Load and Barriers to Introgression across *Populus* Genome (Y Kim, Ed.). *Molecular*
873 *Biology and Evolution* 39: msac008.
874
875 **Luqman H, Wegmann D, Fior S, Widmer A. 2023.** Climate-induced range shifts drive
876 adaptive response via spatio-temporal sieving of alleles. *Nature Communications* 14:
877 1080.
878
879 **Martinsen GD, Whitham TG, Turek RJ, Keim P. 2001.** Hybrid populations selectively
880 filter gene introgression between species. *Evolution* 55: 1325–1335.
881
882 **McVay JD, Hipp AL, Manos PS. 2017.** A genetic legacy of introgression confounds
883 phylogeny and biogeography in oaks. *Proceedings of the Royal Society B: Biological*
884 *Sciences* 284.
885
886 **Menon M, Bagley JC, Friedline CJ, Whipple AV, Schoettle AW, Leal-Sàenz A,**
887 **Wehenkel C, Molina-Freaner F, Flores-Rentería L, Gonzalez-Elizondo MS, *et al.***
888 **2018.** The role of hybridization during ecological divergence of southwestern white pine
889 (Pinus strobiformis) and limber pine (P. flexilis). *Molecular Ecology* 27: 1245–1260.
890
891 **Montgomery RA, Rice KE, Stefanski A, Rich RL, Reich PB. 2020.** Phenological
892 responses of temperate and boreal trees to warming depend on ambient spring
893 temperatures, leaf habit, and geographic range. *Proceedings of the National Academy of*
894 *Sciences* 117: 10397–10405.
895
896 **Naimi B, Hamm NAS, Groen TA, Skidmore AK, Toxopeus AG. 2014.** Where is
897 positional uncertainty a problem for species distribution modelling? *Ecography* 37: 191–
898 203.
899

**Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., Minchin, P.R., O'hara, R.B., Simpson, G.L., Solymos, P. and Henry, M., 2020.** Package "vegan" Title Community Ecology Package Version 2.5-6. *Cran. Ism. Ac. Jp*.

**Payseur BA, Rieseberg LH. 2016.** A genomic perspective on hybridization and speciation. *Molecular Ecology* 25: 2337–2360.

**Petit RJ, Bodénès C, Ducousso A, Roussel G, Kremer A. 2004**. Hybridization as a mechanism of invasion in oaks. *New Phytologist* 161: 151–164.

**Petit RJ, Hampe A. 2006.** Some Evolutionary Consequences of Being a Tree. *Annual Review of Ecology, Evolution, and Systematics* 37: 187-214.

**Petkova D, Novembre J, Stephens M. 2016.** Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics* 48: 94–100.

**Sanderson BJ, Ghambir D, Feng G, et al. 2023.** Phylogenomics reveals patterns ancient hybridization and differential diversification that contribute to phylogenetic conflict in willows, poplars, and close relatives. *Systematic Biology* syad042.

**Schneider R. 2013.** Alberta's Natural Subregions Under a Changing Climate: Past, Present, and Future. *University of Alberta*: 1–86.

**Shang H, Hess J, Pickup M, Field DL, Ingvarsson PK, 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.

**Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS. 2006.** Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology* 15: 4261–4293.

**Stebbins GL. 1959.** The Role of Hybridization in Evolution. *Proceeding of the American Philosophical Society* 103: 231-251

**Suarez-Gonzalez A, Hefer CA, Christe C, Corea O, Lexer C, Cronk QCB, Douglas CJ. 2016.** Genomic and functional approaches reveal a case of adaptive introgression from Populus balsamifera (balsam poplar) in P. trichocarpa (black cottonwood). *Molecular ecology* 25: 2427–2442.

**Suarez-Gonzalez A, Lexer C, Cronk QCB. 2018.** Adaptive introgression: A plant perspective. *Biology Letters* 14.

**Swenson NG, Howard DJ. 2005.** Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. *The American Naturalist* 166: 581–591.

**Tine M, Kuhl H, Gagnaire PA, Louro B, Desmarais E, Martins RST, Hecht J, Knaust F, Belkhir K, Klages S, *et al.* 2014.** European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature communications* 5: 5770.

**Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov * A, *et al.* 2006.** *The genome of black cottonwood,*.

**Vance RE, Emerson D, Habgood T. 1983.** A mid-Holocene record of vegetative change in central Alberta. *Canadian Journal of Earth Sciences* 20: 364–376.

**VanWallendael A, Lowry DB, Hamilton JA. 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.

**Wang T, Hamann A, Spittlehouse D, Carroll C. 2016a.** Locally downscaled and spatially customizable climate data for historical and future periods for North America. *PLoS ONE* 11.

**Willeit M, Ganopolski A, Calov R, Brovkin V. 2019.** Mid-Pleistocene transition in glacial cycles explained by declining $CO_2$ and regolith removal. *Science Advances* 5: 1-8

**Woolbright SA, Whitham TG, Gehring CA, Allan GJ, Bailey JK. 2014.** Climate relicts and their associated communities as natural ecology and evolution laboratories. *Trends in Ecology & Evolution* 29: 406–416.

**Yeaman S, Hodgins KA, Lotterhos KE, Suren H, Nadeau S, Degner JC, Nurkowski KA, Smets P, Wang T, Gray LK, *et al.* 2016.** Convergent local adaptation to climate in distantly related conifers. *Science* 353: 1431–1433.

**Zhang M, Suren H, Holliday JA. 2019.** Phenotypic and Genomic Local Adaptation across Latitude and Altitude in Populus trichocarpa (B Gaut, Ed.). *Genome Biology and Evolution* 11: 2256–2272.

**Zoltai SC, Vitt DH. 1990.** Holocene Climatic Change and the Distribution of Peatlands in Western Interior Canada. *Quaternary Research* 33: 231-240

## 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.

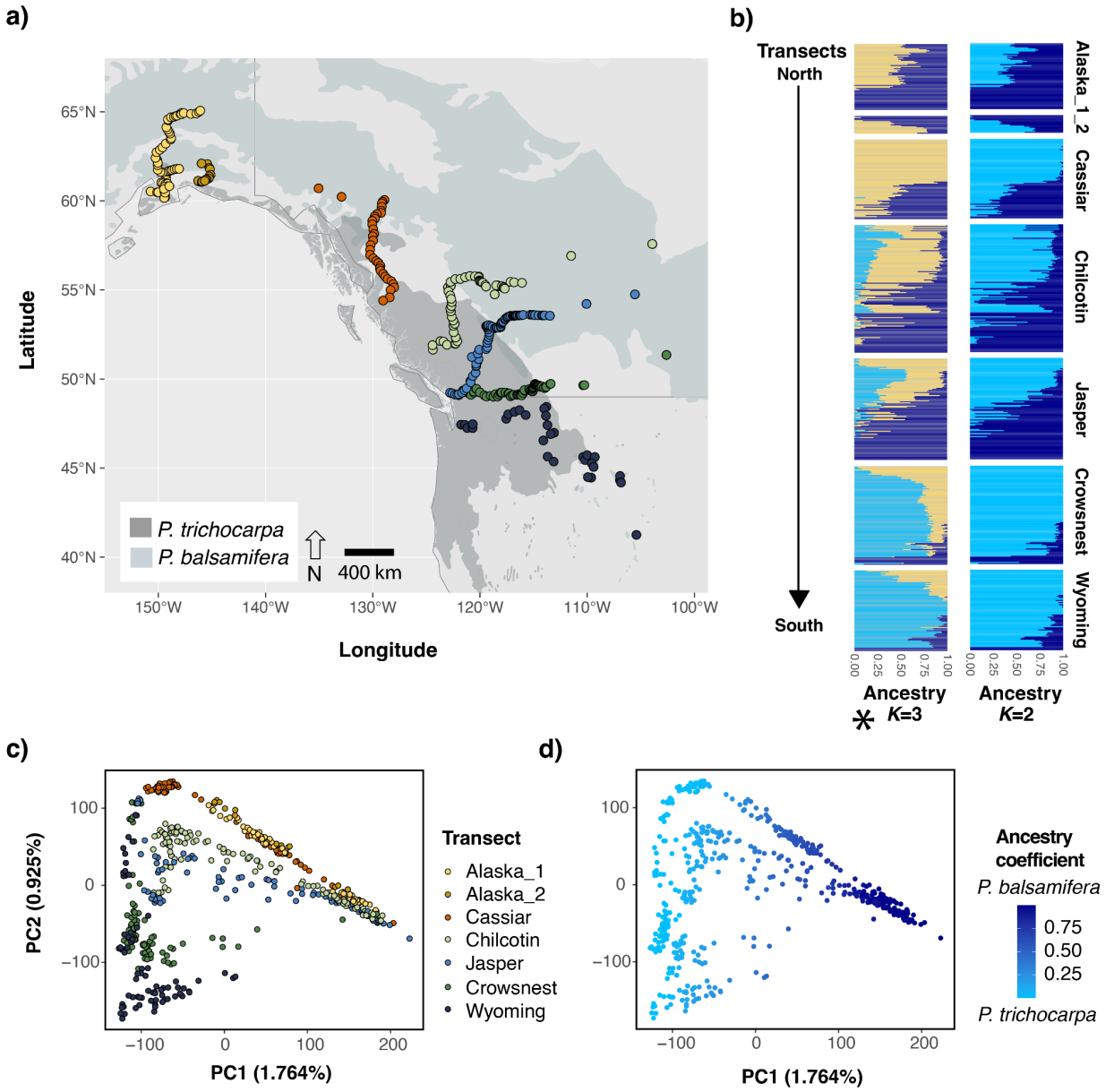


**Figure 1.** Spatial patterns of genetic diversity and ancestry across a) 546 trees sampled from seven transects of the *Populus balsamifera* and *P. trichocarpa* hybrid zone. Transect colors and assigned names are consistent with those listed in the legend of panel c). Results from admixture analysis for each sample at *K*=2 and *K*=3 are illustrated in panel b). Patterns from Principal Component Analysis for PC1 and PC2 are provided by c) transect assignment and d) ancestry coefficient at *K*=2.

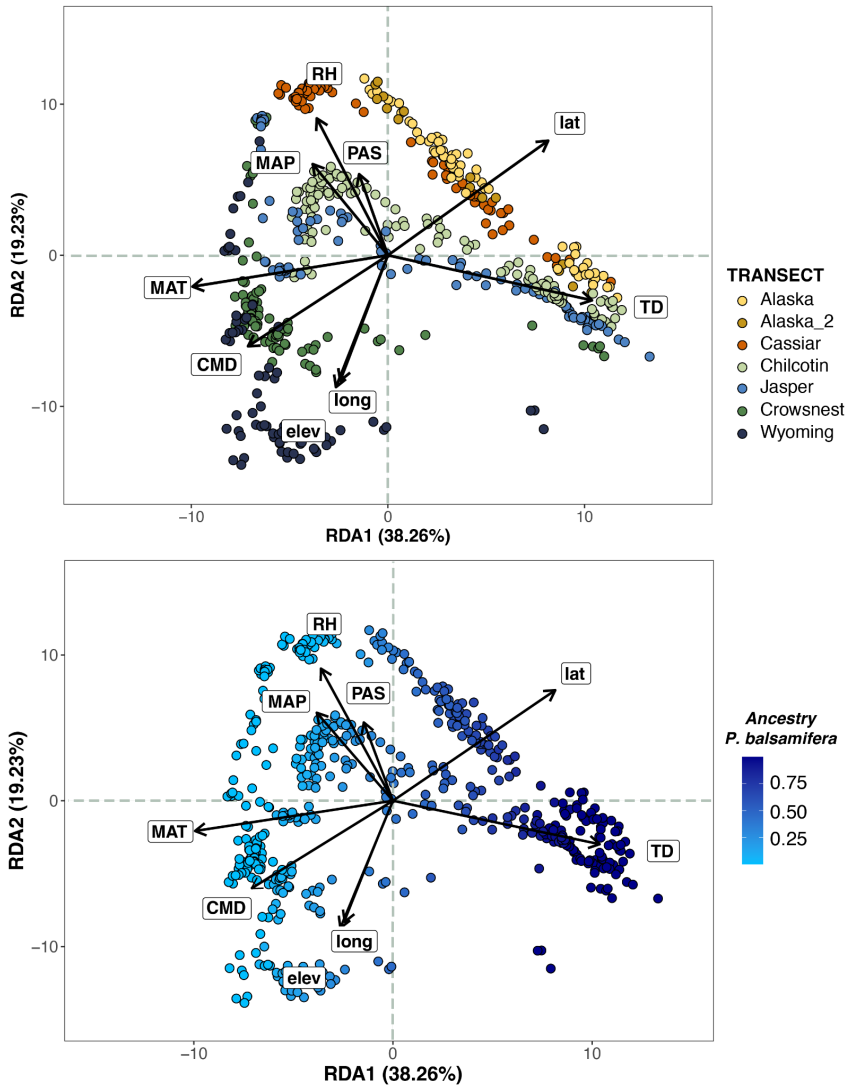


**Figure 2.** Redundancy analysis (RDA) based on associations between climate, geography, and genome-wide estimates of genetic variation for 546 trees sampled across the contact zone between *Populus trichocarpa* x *P. balsamifera*. In the top panel, the color of the individual tree (circles) in RDA space corresponds with transect identity. In the bottom panel, individual trees are colored by ancestry coefficient. 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), lat (latitude), long (longitude), and elev (elevation).

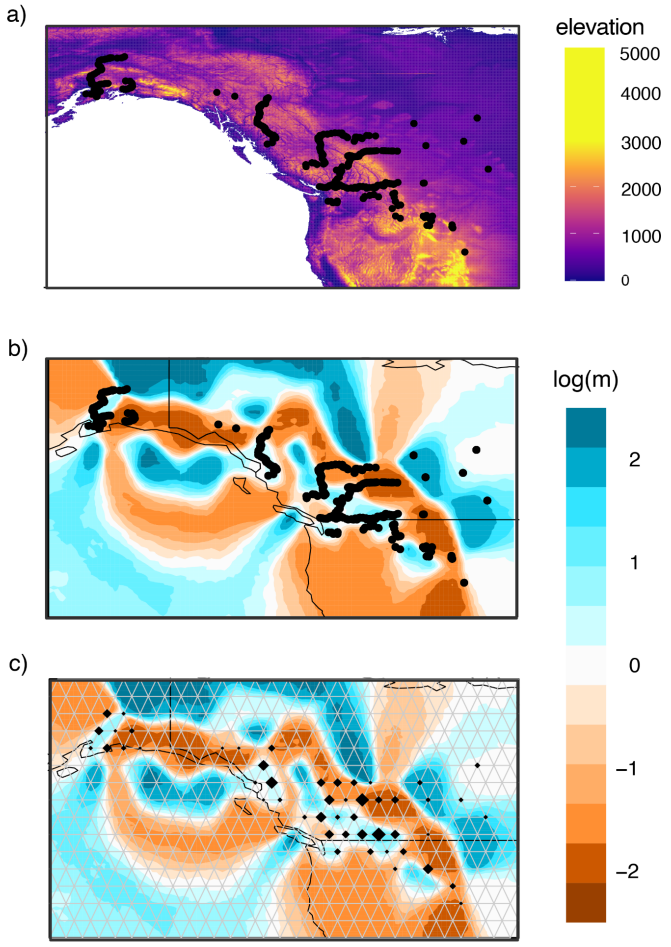


**Figure 3.** Estimated effective migration surfaces (EEMS) analysis of the relationship between genetic distance and geographical distance for 546 sampled trees across 7 transects of the *P. trichocarpa* and *P. balsamifera* hybrid zone to visualize potential barriers to gene flow. Panel a) is a raster 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. In panel c) the same EEMS result is plotted but the distribution of 49 demes across the sampled hybrid zone is provided. The size of the black diamond corresponds with the number of samples of that region with low genetic distance.

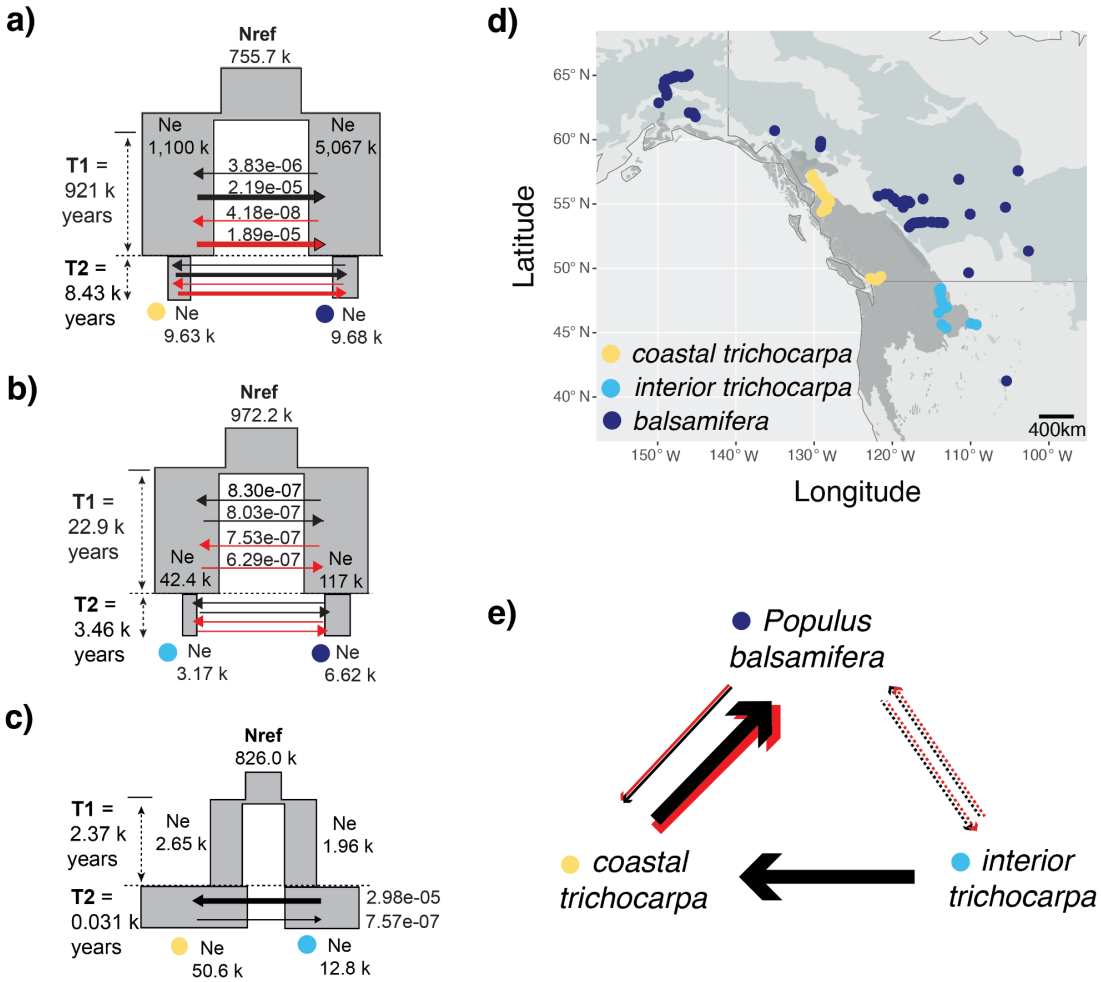


**Figure 4.** Best fit models for each pairwise demographic. Panel a) illustrates the demographic history inferred for coastal *P. trichocarpa* (in yellow) and *P. balsamifera* (in dark blue). Panel b) illustrates the demographic history inferred for interior *P. trichocarpa* (in light blue) and *P. balsamifera* (in dark blue). Panel c) illustrates the demographic history inferred for coastal and interior *P. trichocarpa*. The demographic illustrations are not to scale. Each model provides effective population sizes ($N_e$), time intervals involved in divergence history (T, T1, and T2), and gene flow rate and directionality (*m* and arrows) are provided. Panel d) shows the distribution of each parental-type sample and panel e) is a schematic representation of gene flow across the three genetic groupings taken from pairwise demographic inferences. Black arrows represent neutral gene flow, while red arrows represent gene flow under selection. Size and direction of arrows indicate the relative direction and rate of gene flow.