



VCU

Virginia Commonwealth University
VCU Scholars Compass

Theses and Dissertations


Graduate School

2020

Interspecific gene flow potentiates adaptive evolution in a hybrid zone formed between *Pinus strobiformis* and *Pinus flexilis*

Mitra Menon
Virginia Commonwealth University

Follow this and additional works at: <https://scholarscompass.vcu.edu/etd>

 Part of the [Bioinformatics Commons](#), [Botany Commons](#), [Evolution Commons](#), [Genomics Commons](#), [Integrative Biology Commons](#), and the [Other Forestry and Forest Sciences Commons](#)

© Mitra Menon

Downloaded from

<https://scholarscompass.vcu.edu/etd/6221>

This Dissertation is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

© Mitra Menon 2020

All Rights Reserved

**Interspecific gene flow potentiates adaptive evolution in a hybrid zone formed between
Pinus strobiformis and *Pinus flexilis***

By Mitra Menon (PhD, MS)

A Dissertation

Submitted in Partial Fulfillment
of the Requirments for the Degree of
Doctor of Philosophy in
Integrative Life Sciences

Virginia Commonwealth University

Richmond

May, 2020

Andrew J. Eckert, PhD. Chair

Amy V. Whipple, PhD

Brian C. Verrelli, PhD

Maria C. Rivera, PhD

Rodney J. Dyer, PhD

Acknowledgments

I would like to begin by thanking my dissertation advisor, Dr. Andrew Eckert for his continuous support, guidance and for always pushing the limits of the science I did as a PhD student. I also want to thank Andrew along with Dr. Brian Verrelli for easing my transition into this PhD program. I want to thank my committee members, Dr. Maria Rivera and Dr. Rodney Dyer for helping me troubleshoot, refine experiments and analyses approach. Finally, I want to thank Dr. Amy Whipple for helping me understand the southwestern white pine system and facilitating the completion of my final dissertation chapter. This dissertation would not have been possible without the tremendous amount of work put in by several collaborators and volunteers across US and Mexico. I am grateful for their hard work, dedication as well as their friendship. To my fellow graduate students, specifically members of the Verrelli lab, Dyer lab & Eckert lab, you all were integral to my success and perseverance throughout this journey. Having a strong support system and a great group of friends both within VCU and outside, to share good and bad times was integral to the completion of this dissertation. Thank you for letting me fret when I needed to and for dancing it out with me when I needed to. Last but not the least my family, both here and back home in India, I want to thank you from the bottom of my heart for everything you have done for me, for I would not be here if not for you. To my husband, William, you have stood by me through several long and hard days, even sampling in the field and helping me cut needles. Scar and William, thank you for the crazy fun times through this journey. I can not imagine having gone through this journey without the two of you by my side.

Table of Contents

List of Tables	i
List of Figures	iii
Abstract	viii
Introduction.....	1
Chapter 1: The role of hybridization during ecological divergence of southwestern white pine (<i>Pinus strobiformis</i>) and limber pine (<i>P. flexilis</i>)	21
Chapter 2: Adaptive evolution in a conifer hybrid zone is driven by a mosaic of introgressed and standing genetic variants.....	70
Chapter 3: Intrinsic and extrinsic drivers of among population gene expression differentiation across the <i>P. strobiformis</i> - <i>P. flexilis</i> hybrid zone.....	135

LIST OF TABLES

Table 1.1 Ecological niche model performance and variable importance at 30 arc-second resolution

Table 1.2. Estimates of genetic diversity and divergence within and across the three groups, compared to a genome-wide $F_{ST-species}$ of 0.02 (95% CI: 0.008–0.03) and $F_{ST-strobiformis}$ of 0.009 (95% CI: 0.007–0.014).

Table 1.3. Model composite likelihoods and AIC model selection results for 11 alternative demographic models of *P. strobiformis* (Core and Periphery)–*P. flexilis* divergence. Results for the best-supported model are underlined, and the two best models are shown in boldface.

Table 1.S1. List of variables used for building ENMs and their correlation coefficients.

Table 1.S2. Sampling locations used in this study, their classification into Core and Periphery groups (*P. strobiformis* only), and mean admixture proportions.

Table 1.S3. Raw and converted parameter estimates from the $\partial A \partial I$ model that was best supported by AIC model selection.

Table 2.1: Variance partitioning, model R^2 and significance of multivariate models fitted using redundancy analyses.

Table 2.2: Top five environmental gradients (R^2 in parentheses) for three multivariate regression models partitioning the effect of geography and environment on median among and within population component of LD (D_{sr} & D_{is}), along with superscripts indicating sign of the regression coefficient. Environmental gradients are defined in Table 2.S1.

Table 2.S1: Explanation of all environmental gradients used in the present study and their mean value across the sampled range of pure *Pinus strobiformis* and *P. flexilis*. Seasons including winter, spring, summer and autumn encompass the following months: Winter = Jan, Feb & Dec; Spring = March to May; Summer = June to August; Autumn = September to November.

Table 2.S2: Summary statistics for Bayenv outlier SNPs associated with each of the 88 environmental gradients. Values in bold indicate significance at $p < 0.05$.

Table 2.S3: Proportion of times sets of three, four, five and six Bayenv outliers were represented in the same OCs.

Table 2.S4: R^2 estimates and the corresponding p -values from the multiple matrix regression for D_{is} and D_{st} across the 88 environmental gradients. For gradients with 10 or fewer unique response values, only the R^2 is reported.

Table 2.S5: Fold enrichment (FE) estimates as obtained from genomic cline analyses for all 88 environmental gradients.

Table 2.S6: Loadings of all 88 environmental gradients on the top seven PC axes that was used as the environmental matrix in RDA.

Table 3.1: Summary statistics for transcripts classified into various categories of adaptive differentiation based on Q_{ST} - F_{ST} comparisons. The multilocus estimate of F_{ST} across the 30 maternal trees was 0.0147 (95% CI = 0.01-0.02).

LIST OF FIGURES

Fig. 1.1 Map of sampling localities (black dots) overlaid on polygons showing geographical ranges for *Pinus strobiformis* (green) and *P. flexilis* (blue). Peripheral populations (squares) represent the putative hybrid zone. The corresponding locality information is available in Table S2.

Fig. 1.2. Results of niche divergence tests (Schoener's D) for **A**) Core versus Periphery **B**) Core versus *P. flexilis* and **C**) *P. flexilis* versus Periphery. Histograms indicate the background levels of niche divergence and arrows indicate the observed value of Schoener's D for each pair compared.

Fig. 1.3.A) Results of population genetic structure analysis using PCA on 51 633 SNPs. **B)** Results of assignment analyses for each tree in FASTSTRUCTURE for $K = 2$ clusters (right panel) plotted onto a topographic map of the study area (left panel). Each pie chart represents the average ancestry of a population from *P. strobiformis* and *P. flexilis*.

Fig. 1.4.A) Genomic distribution of F_{CT} , **B)** frequency distribution of hybrid index, **C)** variation in genomic ancestry as a function of hybrid index, **D)** correlation between genomic cline parameters, and **E)** 3D correlation plot of genomic cline parameters and F_{CT}

Fig. 1.5. The best-supported model from $\partial A \partial I$ analysis. This figure shows the parameter estimates for divergence times (T_i) in units of millions of years ago (Ma), reference effective population size (θ ; or after conversion, N_{ref}), lineage population sizes (N_i), and rates of gene flow (M_{ij}) for the optimal model determined by AIC model selection (see Table 1.3).

Fig. 1.S1. Schematics and parameter details for each of the 11 demographic models for the divergence of Core and Periphery groups within *P. strobiformis* and *P. flexilis* run in our $\partial A \partial I$ analysis. Parameters include divergence times (T_i), population sizes (N_i), homogeneous rates of gene flow (M_{ij} , gene flow from lineage j to i) and genomically heterogeneous rates of gene flow (M_{ijh}).

Fig. 1.S2. Ecological niche model projections for Core, Periphery, and *P. flexilis*, under present and past climate

Fig. 1.S3.A) Climate PCA with variables used in the ENMs, **B)** Distribution of precipitation seasonality (Bio15) at presence locations of Core & Periphery, and **C)** Difference between Core and Periphery for scaled and centered bioclimatic variables used in the final ENM

Figure 2.1: **(a)** Geographical distribution of *Pinus flexilis* and *P. strobiformis* with sampled populations indicated in yellow. The background colour palette represents a raster map of growing degree days (DD5), highlighting one of the environmental gradients of adaptive introgression. **(b)** Finer scale representation of the 98 hybrid populations used in this study. **(c)** Location of *P. strobiformis* and *P. flexilis* on the map of North-America, with the study region highlighted by a rectangular box.

Figure 2.2: Distribution of median Bayes factor (BF) values for **(a)** environmental gradients that strongly differentiate *P. flexilis* and *P. strobiformis* and **(b)** environmental gradients that least differentiate the two species. Distribution of multilocus F -statistics (F_{ST} and F_{CT}) for **(c)** environmental gradients that strongly distinguish *P. flexilis* and *P. strobiformis* and **(d)** environmental gradients that are least different between the two species. Red dots in the violin plots of BF indicates outlier SNPs. Red dots in the F -statistics boxplots indicate the observed

multilocus estimates for Bayenv outlier SNPs. Details of the environmental gradients are represented in Table 2.S1.

Figure 2.3: Proportion of SNPs across 100 replicates of LDna that were shared within an outlier cluster (OC) for the **(a)** strongly divergent and **(b)** least divergent environmental gradients between *P. strobiformis* and *P. flexilis* **(c)** Representation of one of the replicated sets from LDna, with red points indicating SNPs in the OC. The proportion of environmentally associated SNPs per OC, median LD for the cluster and the environmental gradients with which the SNPs in each OC are associated are indicated above each network. Regression coefficients for within population component of LD (D_{is}) for environmental gradients that were **(d)** most divergent and **(e)** least divergent between the parental species. Regression coefficients for among population component of LD and (D_{st}) for environmental gradients that were **(f)** most divergent and **(g)** least divergent between the parental species. Details of the environmental gradients represented on x-axis of panel a,b & d:g are represented in Table 2.S1.

Figure 2.4: Bootstrap distribution of *P. flexilis* ancestry fold enrichment (FE) with red points indicating the FE for outlier sets of SNPs associated with environmental gradients that were **(a)** strongly divergent between the two species and **(b)** least divergent between the two species. Details of the environmental gradients are represented in Table 2.S1.

Figure 2.S1: Schematic representation of the sequential variance partitioning approach used in RDA to estimate pure, joint and confounded effects.

Figure 2.S2: Ancestry proportions (Q -score) for each individual tree as obtained from NGSAdmix.

Figure 2.S3: (A) Trace plots of the determinant of the variance-covariance matrix for all 500,000 iterations across three independent Markov chains. (B) Trace plots after the burn-in, starting at iteration 20,000.

Figure 3. 1: Geographic location of sampled populations (red dots) and the two common gardens, Bear Springs (BS) & White Pockets (WP). Bear Springs is the high elevation common garden indicated in turquoise and White Pockets is the low elevation garden indicated in orange.

Figure 3.2: Distribution of transcript Q_{ST} for Bear Spring (BS) and White Pocket (WP).

Figure 3.3: (A) Boxplots for transcripts exhibiting steep Q_{ST} reaction norms. There were 5,244 transcripts which had high Q_{ST} only in BS (teal), while 8,481 had high Q_{ST} only in WP (salmon).

(B) Conceptual representation of transcript classification using the Q_{ST} - F_{ST} comparison approach with the number of transcripts classified under each category shown on the right.

Reaction norms are only provided for transcripts, multilocus F_{ST} is indicated in black without a reaction norm. Pink represents transcripts exhibiting adaptive differentiation in both gardens but no plasticity, yellow represents transcripts exhibiting adaptive plasticity, blue represents transcripts that are conditionally adaptive in WP and green represents transcripts that are conditionally adaptive in BS.

Figure 3.4: Distribution of adjusted R^2 for transcripts that are not highly differentiated between population against the adjusted R^2 (vertical line) for transcripts that were highly differentiated for (a) BS using the drought model, (b) BS using the freeze model, (c) WP using the drought model and (d) WP using the freeze model.

Figure 3.5: Maternal reaction norms for log normalised expression values (log-norm-Expr) of transcripts that are **(A)** conditionally adaptive only in BS, **(B)** Conditionally adaptive only in WP **(C)** Adaptive in neither gardens but displaying strong family variance in WP and **(D)** Adaptive in both gardens.

ABSTRACT

Interspecific gene flow potentiates adaptive evolution in a hybrid zone formed between *Pinus strobiformis* and *Pinus flexilis*

By Mitra Menon

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University. Virginia Commonwealth University, 2020.

Major Director: Dr. Andrew J. Eckert, Department of Biology

Species range margins are often characterised by high degrees of habitat fragmentation resulting in low genetic diversity and higher gene flow from populations at the core of the species range. Interspecific gene flow from a closely related species with abutting range margins can increase standing genetic diversity and generate novel allelic combinations thereby alleviating limits to adaptive evolution in range margin populations. Hybridization driven interspecific gene flow has played a key role in the demographic history of several conifer due to their life history characteristics such as weak crossability barriers and long generation times. Nevertheless, demonstrating whether introgression is adaptive and whether it helps overcome perils associated with high degrees of landscape fragmentation remains challenging in conifers due to limited among species differentiation and the lack of well developed genomic resources.

My dissertation addresses this challenge by first investigating the divergence history and the maintenance of species boundaries between two North American species of white pines: *Pinus strobiformis* and *P. flexilis*. By combining demographic modeling with ecological niche modeling and genomic cline analyses, I illustrate a divergence history of ecological speciation with gene flow and the absence of strong genomic incompatibilities. By combining genotyping-

by-sequencing datasets along with a transcriptomic dataset through a series of novel as well as established multifaceted approaches, I unravel the genetic architecture of adaptive evolution in fragmented range margin populations encompassing the *P. strobiformis*-*P. flexilis* hybrid zones.

Here, both introgressed and background genetic variants are shown to facilitate adaptive evolution along freeze and water-availability related environmental gradients, respectively. I also highlight the adaptive potential of novel allelic combinations formed by the interaction between introgressed and background genetic variants, that is unique to hybrid zone populations and will likely be crucial in responding to novel selective regimes imposed by climate change. Finally, by assaying transcriptional changes between hybrid zone populations through a common garden design, I reveal strong signatures of adaptive trait differentiation and of genotype-by-environment effects that is driven by variation in hybrid ancestry among populations.

This dissertation adds to the growing body of literature demonstrating the importance of introgression in assisting species response to changing climatic conditions via range shifts and through adaptive evolution. Contrary to the notion that extant conifers will be susceptible to rapid environmental change owing to their long generation times, I posit that the mosaics of allelic variants available within conifer hybrid zones will confer upon them greater resilience to ongoing and future environmental change and can be a key resource for conservation efforts.

INTRODUCTION

A key question in evolutionary biology is how neutral and selective processes shape the diversity of life forms. Early studies examined evolution from a dichotomous lens. One group led by Ronald Fisher, John Burdon Sanderson Haldane & Sewall Wright considered Darwinian selection as a primary driver of phenotypic change, while the other led by Motoo Kimura & Tomoko Ohta emphasised the random fixation of selectively neutral or nearly neutral mutations. The advent of molecular biology techniques in the 1960s and specifically of nucleotide sequence level data across several non-model organisms in the past decades have made the amalgamation of these two key concepts mainstream in evolutionary biology. This has shifted the focus towards assessing the relative importance of selective and neutral processes in shaping the observed patterns of genetic diversity.

Evolutionary dynamics of range margin populations

Species with broad geographic distributions provide an ideal system to understand the interplay between neutral and selective processes. Owing to their wide distribution, populations within such species are often separated by geographical and ecological barriers. Geographical barriers cause populations to undergo non-random mating, often independent of fitness differences, and become genetically structured (Wright, 1949). In species without obvious geographical barriers, the increased likelihood of mating among physically proximate individuals generates a clinal pattern of genome-wide population differentiation. Ecological barriers cause the type and intensity of selection to be spatially variable, generating an array of genomic architectures (McKay, 2001; Hansen, 2006). The genotype-phenotype-environment map is widely used to characterise the genetic architecture underlying adaptive traits (Sork et al. 2013).

Characterisation of the genomic architecture is ideally done through a multi-tier process involving the number and identity of causative variants, their location in the genome, their mutation rates, effect sizes, patterns of gene expression, pleiotropic effects, environmental influences, epistasis and additivity (McKay, 2001; Lind et al. 2018). While the utilization of this multifaceted approach is needed to accurately characterise the architecture underlying most quantitative traits, it remains a challenge for non-model organisms due to the paucity of genomic resources, limited sample sizes and the interest underlying the identification large effect causative variants. Small sample sizes further restrict these studies to identify only the low hanging fruits that represent loci of large effect sizes covering only a small fraction of the underlying polygenic architecture that is common to most quantitative traits (Rockman, 2012). Quantitative traits with polygenic architectures are characterised by a large number of alleles with very small effect sizes and a few with large effects (Gagnaire & Gaggiotti, 2006; reviewed in Lind et al. 2018).

The study of adaptive evolution has been centered primarily around detecting locally elevated signals of population differentiation beyond the background genomic level. This approach can be specifically problematic under complex demographic histories that are common to range margin populations. Populations occurring at the periphery of species' geographical ranges are considered range margin populations (Antonovics, 1976; Bridle and Vines, 2007). These populations are often at demographic non-equilibrium because they are at the epicentre of processes such as range expansions, bottlenecks and hybridization with a closely related sister taxon that has abutting range margins. Thus, for range margin populations, the interaction between gene flow and selection often extends beyond the intraspecific level, such that signatures of elevated differentiation could encompass regions associated with the maintenance

of species cohesion, which may or may not be associated with ecological barriers driving adaptive differences (Noor & Bennett, 2009; Feder & Nosil, 2010; Bierne et al. 2011; Han et al. 2017; Christe et al. 2017). It is now well characterised that speciation occurs along a spatial and genomic continuum of divergence. Range margin populations encompassing hybrid zones are at the midpoint of this spatial and genomic continuum. Hybrid zones are geographic areas where divergent lineages interbreed to produce individuals of mixed ancestry and often facilitate exchange of genetic variants through backcrossing into the gene pool of the divergent lineages (Barton & Hewitt, 1985). As such, they have been used as windows into the process of speciation and understanding whether barriers to gene flow are strong enough to maintain species integrity (Harrison, 1990).

The ability of lineages to hybridize is often held at the core of species definitions (Dobzhansky, 1937; Mayr, 1963; Harrison & Larson, 2014). While useful from a taxonomic standpoint to classify lineages exhibiting few morphological differences, reproductive isolation encompasses only one of the several facets of speciation. For instance, isolating mechanisms may be mediated by differences in environmental selective regimes between lineages, even though genetic incompatibilities are absent or are in their infancy (Agrawal et al. 2011). Buildup of genetic incompatibilities among loci can be environmentally dependent (extrinsic barriers) or independent (intrinsic barriers). The former is a result of ecological speciation (Schluter & Conte, 2009) where disruptive selection generates locally adapted lineages, while the latter is usually a result of negative epistasis occurring among lineage specific allelic variants (Orr, 1996). Regardless of the initial process of divergence, in most cases intrinsic and extrinsic barriers will eventually be coupled, such that genomic regions involved in intrinsic barriers to gene flow coincide with loci exhibiting ecological gradients in allele frequency (Bierne *et al.* 2011;

Cushman & Landguth 2016), ensuring the maintenance of species barriers despite the homogenizing effect of gene flow (Kulmuni & Westram 2017). Thus, even under ecological speciation, the interaction between intrinsic and extrinsic barriers can cause reduced hybrid fitness, unless the hybrids occur in an environment where the novel allelic combinations or breakdown of co-adapted gene complexes are favourable (Moore, 1977; Gompert et al. 2012; Schneemann et al. 2020). The surge in genomic datasets for non-model organisms provides us a unique opportunity to characterise the past and ongoing process of speciation, which is needed to accurately identify signatures of selection in range margin populations. Documenting the presence and the strength of isolating barriers as well as ongoing rates of interspecific gene flow (if any) is necessary to elucidate the evolutionary trajectory of hybridizing species.

Range margin populations often also occur in areas where suitable habitats are fragmented and small (Bridle & Vines, 2007). Small populations are prone to genetic drift which not only makes them depauperate in genetic diversity, but also reduces the efficacy of selection to remove deleterious variants, thereby increasing genetic load (Willi et al. 2018). Further, asymmetric gene flow into range margin populations from the core of the species range has been suggested to limit adaptive potential (Kirkpatrick & Barton, 1997). In contrast, several theoretical and empirical studies (Wright, 1982; Eckert et al. 2008; Bontrager & Angert, 2019) demonstrate that adaptive evolution proceeds faster with population structure and variable rates of gene flow. This interaction between gene flow and selection can be extended to include interspecific gene flow and intrinsic selection pressures arising due to the breakdown of co-adapted gene complexes in hybrid zones. On one hand, these incompatible gene complexes can reduce hybrid fitness thereby stagnating parental population growth which can often lead to the extinction of rare taxa (Wolf et al. 2001). On the other hand, when hybrids don't exhibit lower fitness relative

to their parents, a breakdown of co-adapted gene complexes can increase the standing levels of genetic diversity in fragmented range margin populations (Mesgaran et al. 2016; Taylor & Larson, 2019). Thus, the long-term consequences of hybridization have remained contentious and are often dependent on the demographic history of the two interacting species as well as the environmental conditions experienced by hybrid populations (Currat et al. 2008; Schneemann et al. 2019). Nevertheless, the availability of population-level, genome-wide datasets and methodological advances has enabled us to characterise the demographic history of several species which has revealed a genomic mosaic of differentiation between pairs of sister taxa (Payseur & Rieseberg, 2016). Such studies shed light on the potential of interspecific variants to introgress into novel genomic backgrounds by overcoming intrinsic selection pressures. Once introgressed, these variants have the potential to facilitate adaptive evolution and inhabit new niche spaces (i.e Hutchinson's niche: Hutchinson 1957), beyond that of the parental species alone (Pfennig et al. 2016; Pierce et al. 2017; Cronk & Suarez-Gonzalez, 2018; Taylor & Larson, 2019). By providing a segue for the transfer of adaptive variants, hybridization is an important player in the projected shifts of both the fundamental and realised niches of several species beyond their current range (Parmesan, 2006, Aitken et al. 2008). Predictive niche modeling for future or past environmental conditions assumes niche conservatism and lacks the ability to incorporate hybridization making such modeling efforts overly simplistic and even pessimistic. The latter is specifically true for plants, with nearly 25% of them experiencing natural hybridization (Mallet, 2005).

Hybridization can act as a conduit for species to track or expand their fundamental niche breadth (Pfennig et al. 2016; Bolte & Eckert, 2020). This expansion of niche breadth is influenced by the generation of novel allelic complexes and the input of adaptive variants via

hybridization (Pardo-diaz et al. 2012; Ma et al. 2019). The strongest evidence for the importance of hybridization in the evolutionary history of plants comes from studies in sunflowers, poplars and oaks (Gross & Rieseberg, 2005; Suarez-Gonzalez et al. 2016, 2018; Leroy et al. 2019). For instance, the combination of mesic-clay like soil adapted alleles from *Helianthus annuus* with drier-sandier soil adapted alleles from *H. petiolaris* enabled *H. anomalus* to inhabit sand dunes and eventually undergo homoploid hybrid speciation (Gross & Rieseberg, 2005). Studies in the *Populus trichocarpa* x *P. balsamifera* hybrid zone in western North America suggest that directional introgression of cold adapted alleles from *P. balsamifera* into the *P. trichocarpa* genomic background has enabled the latter to occupy colder habitats than is otherwise not typical of the species range (Suarez-Gonzalez et al. 2016, 2018). Further, they also demonstrate introgression of disease resistance alleles from *P. trichocarpa* into the hybrid genomic background, thereby making hybrid populations resilient to novel challenges such as changes in pathogen pressures imposed by changing climatic conditions. Well-developed hybrid zones dominated by backcrosses and advanced generation hybrids are becoming the focus of studies addressing positive evolutionary consequences of hybridization. For example, Mesarange et al. (2016) showed that when compared to symmetrical gene flow, hybrid zones experiencing asymmetrical gene flow, often typical of natural advanced generation hybrid zones, have the potential to overcome perils associated with small population sizes. Individuals within advanced generation hybrid zones have likely overcome incompatibilities associated with the generation of F1 hybrids and can contain novel variants or allelic combinations at a higher frequencies as compared to populations where de novo mutations are the dominant source of novel variation. Given that the rate of adaptation depends to a large degree on the initial frequencies of the

fitness-related genetic variants, hybrid zone populations are better equipped to respond to changing environmental conditions.

Evolutionary importance of hybridization in conifers

Artificial crossing in the hopes of generating progenies displaying hybrid vigour has a long history in the agricultural, animal husbandry, and forestry industries (Robison et al. 1981; Pearson, 1983; Knezick et al. 1984). Forests cover 30% of the earth's terrestrial surface (FAO 2015) and provide enormous ecosystem services such as food and shelter for wildlife, soil replenishment, sequestration of greenhouse gases as well as a key source of renewable natural resources (Whitaker, 1975). Many forest tree species have wide geographical distributions across heterogeneous landscapes. They harbour high levels of genetic diversity and display phenotypic plasticity that may have equipped them with the ability to withstand several decades of environmental fluctuations (Isabell et al. 2019). Environmental fluctuations, specifically the expansion and retreat of glaciers in the Northern Hemisphere, also caused dynamic changes in species' ranges, bringing about periods of contact and hybridization between closely related sister taxa which has often fueled post-glacial recolonisation (Klein et al. 2017). Thus, many long lived forest tree species have likely experienced episodes of interspecific gene flow thereby making them strong candidates to understand the potential of introgression in facilitating species persistence through neutral and adaptive evolutionary mechanisms. In addition to episodes of contact, the long lifespan of trees could delay the buildup of intrinsic barriers (Petit and Hampe, 2006; Stacy et al. 2014) thereby making hybridization a natural part of their evolutionary trajectory. Widespread evidence of hybridization within taxonomic families such as Cupressaceae and Pinaceae indicate the presence of weak intrinsic isolating barriers (Critchfield,

1986; Neale & Wheeler, 2019). Yet, several studies have demonstrated strong extrinsic barriers restricting, but not obliterating, interspecific gene flow among species within these families (Rehfeldt, 1999; Hamilton et al. 2013; De La Torre et al. 2014; De La Torre et al. 2015). For instance, *Pinus contorta* and *P. banksiana* hybridize in north-central Alberta, but the parental species remain differentiated along an edaphic gradient (Cullingham et al. 2012). Similarly, the homoploid hybrid species *Pinus densata* exhibits weak intrinsic barriers with its parental species (Zhao et al. 2014), yet demonstrates strong niche partitioning by occurring in high elevation environments where neither parent species are found. Spruce (genus *Picea*) provides one of the best examples of a complex and widespread hybrid zone involving nearly 6 different species, yet not leading to the collapse or loss of either parental species (Haselhorst et al. 2019). While not directly assessed in these studies, the maintenance of species barriers even under widespread hybridization could be attributed to high fecundity and strong selection against certain hybrid classes containing non-compatible allelic combinations during early life stages (Lindtke et al. 2014; Zhao et al. 2014).

Signatures of adaptive evolution across conifers

Conifers are one of the most widely distributed types of trees, with biodiversity hotspots centered in North America and Eurasia (Neale & Wheeler, 2019). Across western North America, they inhabit highly heterogeneous landscapes characterised by extremes of temperature and precipitation. Decades of provenance trials and recent molecular assays have demonstrated complex genetic, physiological and developmental processes aiding adaptation to inter- as well as intra-annual fluctuations in climatic conditions (Hermann & Lavender, 1968; Howe et al. 2003; Holliday, et al. 2008; Eckert et al. 2012; Eckert et al. 2015). Despite high levels of gene

flow, most conifers display strong signals of local adaptation that is guided by subtle and coordinated shifts in allele frequencies across populations rather than through localised fixation of alleles (Le Corre & Kremer, 2012; Hornoy et al. 2015; Lind et al. 2018; De La Torre et al. 2019). Conifers, like other organisms, are faced with the threat of global climate change, but are thought to be more susceptible given their sessile nature and long generation time. On the contrary, high levels of genetic diversity and co-ordinated shifts in allele frequencies of several small effect variants exhibiting redundant phenotypic effects typical of the polygenic architecture underlying quantitative traits in conifers could aid rapid adaptation to shifted fitness optima (Pritchard & Di Rienzo, 2010; De La Torre et al. 2019; Bitter et al. 2019). Additionally, genomic scans and theoretical models demonstrate that hybridisation could provide novel variants and increase standing levels of genetic diversity, making conifers more resilient to rapid environmental change than previously thought (Hamilton & Miller, 2016).

While long-term provenance trials in trees have provided evidence for local adaptation (Savolainen et al. 2007; Eckert et al. 2009), very few have accounted for the influence of introgression on adaptive evolution (De La Torre et al. 2014; Ma et al. 2019). Further, most of these studies do not evaluate components of fitness beyond current ranges for focal species, which is needed to predict responses to changing climatic conditions. Short-term seedling common gardens and genome-scale approaches encompassing a hybridizing species complex can provide space-for-time substitution (*sensu* Pickett, 1989). Broadly, space-for-time substitutions encompass analyses and experiments in which contemporary patterns of spatial environmental differences are used to predict and understand future or past unobservable events. Given the projected upslope and northward shift in the distribution of various cold adapted species such as conifers (Rehfeldt, 2004; Ledig et al. 2010), planting individuals further north of the focal

species' range or assaying individuals with some genomic ancestry from a northern sister species will be useful to forecast adaptive traits and their underlying architectures. The large intergenic spaces in conifers, signatures of selection residing on regulatory regions in species with larger genomes (Mei et al. 2019) and accumulating evidence for genotype-by-environment (G x E) effects even at the level of gene expression (Roberge et al. 2007; Leder et al. 2015) highlight the need for gathering transcriptome datasets from space-for-time substitution experiments. One of the biggest advantages of these datasets is the absence of ascertainment bias, since one does not have to a-priori pick traits that could facilitate adaptation to novel selective pressures imposed by anthropogenic climate change. When combined with genome-wide datasets and linkage maps, several of the subtle aspects of the genetic architecture underlying adaptive evolution in trees can be disentangled.

Hybridization & adaptive evolution in Pinus strobiformis

Forest tree species inhabiting high altitude landscapes in the southwestern North America are likely to experience increased drought intensity and seasonal fluctuations in climatic events such as early warming in the spring causing several trees to initiate active growth before the last date of frost has passed. Understanding the adaptive potential of tree species inhabiting semi-arid ecosystems, such as the ones in the southwest of the United States (US), to future climate scenarios will be critical for refining conservation frameworks (Schoettle & Sniezko, 2007; Aitken et al. 2008). For hybridizing species, introgression of variants from a species adapted to cooler climatic conditions could enable populations of a southern species to utilize the longer growing season while being resilient to sporadic frost events in early spring (*cf.* Suárez-González et al. 2016).

In this dissertation, I examine the interaction between hybridization and various environmental factors in shaping the evolutionary trajectory of the range margin populations of southwestern white pine (*Pinus strobiformis*). *P. strobiformis* is an important component of the mixed-conifer forests in southwestern North America, yet remains one of the most understudied species of soft pines (*Pinus* subgenus *Strobus*; Looney & Waring, 2013). It has a wide geographic distribution ranging from southern Colorado in the north to Jalisco, Mexico in the south. The range margin populations in the southwestern portions of the US occur on fragmented sky-islands, while in Mexico it has a more continuous distribution along the ridges of the Sierra Madre Occidental (Little, 1971). Despite the small sizes and isolated nature of the range margin populations, recent studies assaying physiological traits associated with drought and heat stress provide evidence for adaptive evolution within and across these populations (Goodrich et al. 2017; DaBell, 2018). These range margin populations of *P. strobiformis* also exhibit morphological characteristics that are intermediate between those noted in Mexico and the closely related northern sister species, *Pinus flexilis* (Frankis, 2009; Bisbee, 2014). *P. flexilis* inhabits montane ecosystems and ranges from northern Arizona to Alberta, Canada. Within the *P. strobiformis*-*P. flexilis* species complex, niche modeling efforts (Moreno-Letelier et al. 2013; Aguirre-Gutiérrez et al. 2015), phylogenies built using limited chloroplast and nuclear regions (Syring et al. 2007), as well as pattern of among species divergence at candidate loci associated with drought stress (Moreno-Letelier & Barraclough, 2015), all highlight the importance of divergent environmental selection in building and maintaining species boundaries. Overall, these corroborate the pattern of species differences noted across the genus *Pinus*. Despite strong niche differentiation between *P. flexilis* and *P. strobiformis*, the presence of extensive shared polymorphisms and their ability to undergo natural hybridization indicates porous species

boundaries and the potential for interspecific gene flow to facilitate adaptive evolution in range margin populations of the southwestern US. This overarching hypothesis is addressed through three chapters in my dissertation.

In **chapter 1**, I characterise the divergence history of *P. strobiformis* and *P. flexilis* and quantify the relative influence of extrinsic and intrinsic barriers to the maintenance of species boundaries. The primary finding from this chapter demonstrates a history of ecological speciation with gene flow mediated by divergence along drought and freezing temperatures. The second major finding of ongoing asymmetrical gene flow from *P. flexilis* into the range margin populations encompassing the hybrid zone lays the foundation for the second chapter. In **chapter 2**, I quantify the relative importance of introgressed variants from *P. flexilis* and locally available background genetic variants in facilitating adaptation to marginal habitats typical of range margin populations, which also encompasses the *P. strobiformis*-*P. flexilis* hybrid zone. This chapter reveals a complex architecture of adaptive evolution in the hybrid zone, with both introgressed and background genetic variants facilitating adaptive evolution, albeit along different environmental axes. Given the interaction between genomic ancestry and environmental conditions in shaping patterns of local adaptation within the hybrid zone, in **chapter 3** I assess patterns of adaptive trait differentiation and of plasticity (G x E effects) using transcriptome datasets generated from hybrid zone populations in response to environmental conditions predicted under climate change scenarios in the desert southwest of US. The primary finding from this chapter is of strong adaptive differentiation at several transcripts and conditional adaptation of transcripts in an environment dependent manner. I also demonstrate that hybridization contributes significantly towards both among population transcript differentiation as well as towards plasticity.

Overall, this dissertation highlights signals of adaptive evolution and of G x E effects in a hybrid zone that occurs as small isolated populations on a fragmented landscape. I show that hybridization has played a critical role in driving these signatures of adaptation and of G x E effects. I conclude by stating that hybrid zones within the genus *Pinus* are more likely to respond to the rapidly changing environmental conditions due to hybridization facilitated increase in standing levels of genetic diversity and the availability of novel allelic variants.

REFERENCES

1. Wright S. (1949) The genetical structure of populations. *Annals of Human Genetics*. 15(1):323–54.
2. McKay TF (2001). The genetic architecture of quantitative traits. *Annual Reviews of Genetics*. 35:303-39.
3. Hansen TF (2006) The evolution of genetic architecture. *Annual Reviews of Ecology Evolution and Systematics*. 37:123–157.
4. Sork VL, Aitken SN, Dyer RJ, Eckert AJ, Legendre P, Neale DB (2013) Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics and Genomes* 9:901–911.
5. Gagnaire PA, Gaggiotti OE (2016) Detecting polygenic selection in marine populations by combining population genomics and quantitative genetics approaches. *Current Zoology*. 62:603–616.
6. 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 & genomes* 14 (2), 29.
7. Rockman MV (2012). The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution*. 66(1):1-17
8. Antonovics J. (1976). The Nature of Limits to Natural Selection. *Annals of the Missouri Botanical Garden*. 63(2): 224-247
9. Bridle, J. and Vines, T. (2007). Limits to evolution at range margins: when and why does adaptation fail? – *Trends in Ecology and Evolution*. 22: 140–147.
10. Noor MAF, Bennett SM (2009) Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*, 103, 439–444.
11. Feder JL, Nosil P (2010) The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution*, 64, 1729–1747.
12. Bierne N, Welch J, Loire E, Bonhomme F & David P. (2011). The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology*. 20:2044-2072.
13. Orr HA(1996). Dobzhansky, Bateson and the genetics of speciation. *Genetics*. 144 (4): 1331-1335.
14. Han F, Lamichhaney S, Grant BR et al. (2017) Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. *Genome Research*. 27(6):1004-1015

15. Christe C, Stölting KN, Paris M et al. (2017). Adaptive evolution and segregating load contribute to the genomic landscape of divergence in two tree species connected by episodic gene flow. *Molecular Ecology*. 26(1): 59-76.
16. Barton, NH & Hewitt, GM. (1985). Analysis of hybrid zones. *Annual Reviews in Ecology and Evolution*. 16:113–148.
17. Harrison, RG. (1990) Hybrid zones: windows on evolutionary process. In: Futuyma, D. and Antonovics, J. (eds) *Oxford Surveys in Evolutionary Biology*, vol. 7, pp.69–128. Oxford University Press, Oxford.
18. Dobzhansky, T. (1937). Genetics and the origin of species. Columbia University Press, New York.
19. Mayr E (1963). Animal Species and Evolution. Belknap: Cambridge, MA.
20. Harrison, RG., & Larson, EL. (2014). Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, 105, 795–809.
21. Agrawal AF, Feder JL, Nosil P (2011). Ecological Divergence and the Origins of Intrinsic Postmating Isolation with Gene Flow. *International Journal of Ecology*, 2011, e435357.
22. Stankowski S, Sobel JM, Streisfeld MA (2015). The geography of divergence with gene flow facilitates multitrait adaptation and the evolution of pollinator isolation in *Mimulus aurantiacus*. *Evolution*. 69, 3054–3068.
23. Schluter D, Conte GL (2009). Genetics and ecological speciation. *Proceedings of the National Academy of Sciences*, 106, 9955–9962.
24. Kulmuni, J., and A. M. Westram. (2017). Intrinsic incompatibilities evolving as a by-product of divergent ecological selection: Considering them in empirical studies on divergence with gene flow. *Molecular Ecology*. 26:3093-3103.
25. Moore WS (1977) An evaluation of narrow hybrid zones in vertebrates. *Quarterly Review of Biology*, 52, 263–277.
26. Gompert Z, Lauren LK, Chris NC et al. (2012). Genomic regions with a history of divergent selection affect fitness of hybrids between two Butterfly Species. *Evolution*, 66, 2167-2181.
27. Currat M, Ruedi M, Petit RJ, Excoffier L (2008). The hidden side of invasions, massive introgression by local genes. *Evolution*. 62: 1908–1920.
28. Schneemann H, Sanctis BD, Roze D, Bierne N and Welch J (2020). The geometry and genetics of hybridization. *BioRxiv*.
29. Bridle, J. and Vines, T. (2007). Limits to evolution at range margins: when and why does adaptation fail? – *Trends in Ecology Evolution*. 22: 140–147.

30. Willi Y, Fracassetti M, Zoller S and Van Buskirk J (2018). Accumulation of mutational load at the edges of a species range. *Molecular Biology and Evolution*. 35(4): 781-791.
31. Kirkpatrick, M. & Barton, NH.(1997) Evolution of a species' range. *American Naturalist*. 150, 1–23.
32. Wright, S.W. (1982)The shifting balance theory and macroevolution. *Annual Review in Genetics* 16: 1-19.
33. Eckert, CG, Samis, KE & Loughheed, SC. (2008) Genetic variation across species geographical ranges: the central-marginal hypothesis and beyond. *Molecular Ecology* 17, 1170-1188.
34. Bontrager M & Angert AL. (2019). Gene flow improves fitness at a range edge under climate change. *Evolution Letters*. 3(1):55-58.
35. Wolf, D. E., Takebayashi, N., and Rieseberg, L. H. (2001). Predicting the risk of extinction through hybridization. *Conservation Biology*. 15, 1039–1053
36. Mesgarana MB, Lewis MA, Adesc PK et al. (2016). Hybridization can facilitate species invasions, even without enhancing local adaptation. *Proceeding of National Academy of Sciences*. 113(36): 10210-10214.
37. Taylor, SA., Larson, EL. (2019) Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology and Evolution*. 170–177.
38. Payseur BA, Rieseberg LH. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology*. 25:2337–2360.
39. Hutchinson GE (1957). Concluding remarks. *Cold Spring Harbor Symp* 22:415–427.
40. Pfennig KS, Kelly AL & Pierce A(2016). Hybridization as a facilitator of species range expansion. *Proceedings of Royal Society of Biological Sciences*. 283(1839): 20161329.
41. Pierce A, Gutierrez R, Rice A, Pfennig KS (2017). Genetic variation during range expansion: effects of habitat novelty and hybridization. *Proceedings of Royal Society of Biological sciences*. 284: 20170007
42. Cronk QC & Suarez-Gonzalez A (2018). The role of interspecific hybridization in adaptive potential at range margins. *Molecular Ecology*. 27(23): 4653-4656
43. Parmesan C (2006). Ecological and Evolutionary Responses to Recent Climate Change. *Annual Review of Ecology, Evolution, and Systematics*. 37: 637-669
44. Aitken SA, Yeaman S, Holliday JA, Wang T, Curtis-McLane S (2008).. Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications*. 1: 95-111
45. Mallet J (2005). Hybridization as an invasion of the genome. *Trends in ecology & evolution*. 20(5):229– 37

46. Bolte CE & Eckert AJ (2020). Determining the when, where and how of conifer speciation: a challenge arising from the study 'Evolutionary history of a relict conifer *Pseudotsaxus chienii*'. *Annals of Botany*. 124(1): v–vii
47. Pardo-Diaz, C., Salazar, C., Baxter, SW et al. (2012). Adaptive introgression across species boundaries in heliconius butterflies. *PLoS Genetics*. 8:e1002752
48. Ma Y, Wang J, Hu Q et al. (2019) Ancient introgression drives adaptation to cooler and drier mountain habitats in a cypress species complex. *Communications Biology*. 18: 210-213
49. Gross BL & Rieseberg LH. (2005). The Ecological Genetics of Homoploid Hybrid Speciation. *Journal of Heredity*. 96, 3, 241-252.
50. Suarez-Gonzalez A, Hefer, CA., Christe C et al. (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.
51. Suarez-Gonzalez, A., Hefer, CA., Lexer, C., Cronk, QC., and Douglas, CJ. (2018). Scale and direction of adaptive introgression between black cottonwood (*Populus trichocarpa*) and balsam poplar (*P. balsamifera*). *Molecular Ecology*. 27: 1667–1680.
52. Leroy T, Louvet JM, Lalanne C et al. (2019). Adaptive introgression as a driver of local adaptation to climate in European white oaks. *New Phytologist*.
53. Robison OW, McDaniel BT, Rincon EJ. (1981). Estimation of direct and maternal additive and heterotic effects from crossbreeding experiments in animals. *Journal of Animal Sciences*. 52(1):44-50.
54. Pearson OH, (1983). Heterosis in Vegetable Crops. In: Heterosis. 138-188.
55. Knezick, Kuser, and Sacalis (1984) Single clone orchard production of pitch x loblolly hybrids. In Proceedings- Northeastern Forest Tree Improvement Conference (USA)
56. Whitaker (1975) *Communities and Ecosystems*. Macmillian, London.
57. Isabell N, Holliday J & Aitken SN (2019). Forest genomics: Advancing climate adaptation, forest health, productivity, and conservation. *Evolutionary Applications*. 13:3-10.
58. Klein, EK, Lagache-Navarro, L & Petit, R.J (2017). Demographic and spatial determinants of hybridization rate. *Journal of Ecology*. 105, 29–38
59. Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. *Annual Reviews in Ecology*. 37:187–214
60. Stacy EA, Paritosh B, Johnson MA, Price DK (2017). Incipient ecological speciation between successional varieties of a dominant tree involves intrinsic postzygotic isolating barriers. *Ecology and Evolution*, 7, 2501–2512.

61. Critchfield WB (1986). Hybridization and classification of the white pines (Pinus section Strobilus). *Taxon*, 35, 647-656.
62. Neale D & Wheeler N. (2019). *The Conifers: Genomes, variation and evolution*. Springer.
63. Rehfeldt GE (1999). Systematics and genetic structure of Ponderosae taxa (Pinaceae) inhabiting the mountain islands of the Southwest. *American Journal of Botany*. 86, 741-752.
64. Hamilton, J.A, Lexer, C, Aitken, S.N. Genomic and phenotypic architecture of a spruce hybrid zone (*Picea sitchensis* × *P. glauca*). (2013). *Molecular Ecology*. 22, 827–841
65. De La Torre AR, Wang T, Jaquish B, Aitken SN (2014). Adaptation and exogenous selection in a *Picea glauca* × *Picea engelmannii* hybrid zone: implications for forest management under climate change. *New Phytologist*, 201, 687–699.
66. De La Torre AR, Ingvarsson PK, & Aitken SN (2015). Genetic architecture and genomic patterns of gene flow between hybridizing species of *Picea*. *Heredity*, 115, 153–164.
67. Cullingham C, James P, Cooke JEK & Coltman DW (2012). Characterizing the physical and genetic structure of the lodgepole pine × jack pine hybrid zone: mosaic structure and differential introgression. *Evolutionary Applications*. 5(8): 879-891.
68. Zhao W, Meng J, Wang B et al. (2014). Weak Crossability Barrier but Strong Juvenile Selection Supports Ecological Speciation of the Hybrid Pine *Pinus densata* on the Tibetan Plateau. *Evolution*. 68(11): 3120-3133.
69. Haselhorst M, Parchman T, Buerkle A (2019). Genetic evidence for species cohesion, substructure and hybrids in spruce. *Molecular Ecology*. 28(8): 2029-2045.
70. Lindtke D, Gompert Z, Lexer C, Buerkle CA (2014). Unexpected ancestry of *Populus* seedlings from a hybrid zone implies a large role of postzygotic selection in the maintenance of species. *Molecular Ecology*. 23: 4316-4330.
71. Hermann, R. K., & Lavender, D. P. (1968). Early growth of Douglas-fir from various altitudes and aspects in southern Oregon. *Silvae Genetica*, 17(4), 143–151.
72. Howe GT, Aitken SN, Neale DB et al. (2003). From genotype to phenotype: unraveling the complexities of cold adaptation in forest trees. *Canadian Journal of Botany* 81: 1247–1266.
73. Holliday JA, Ralph S, White R et al. (2008). Global monitoring of autumn gene expression within and among phenotypically divergent populations of Sitka spruce (*Picea sitchensis*). *New Phytologist* . 178: 103–122
74. Eckert AJ, Wegrzyn JL, Cumbie WP et al (2012) Association genetics of the loblolly pine (*Pinus taeda*, Pinaceae) metabolome. *New Phytologist*. 193:890–902.
75. Eckert AJ, Maloney PE, Vogler DR et al (2015) Local adaptation at fine spatial scales: an example from sugar pine (*Pinus lambertiana*, Pinaceae). *Tree Genetics and Genomes* 11:1–17.

76. Le Corre V, Kremer A (2012) The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology*. 21:1548–1566.
77. Hornoy B, Pavy N, Gérardi S, Beaulieu J, Bousquet J (2015) Genetic adaptation to climate in white spruce involves small to moderate allele frequency shifts in functionally diverse genes. *Genome Biology and Evolution*. 7:3269–3285
78. De La Torre A, White B, Neale D (2019). Environmental Genome-Wide Association Reveals Climate Adaptation Is Shaped by Subtle to Moderate Allele Frequency Shifts in Loblolly Pine. *Genome Biology and Evolution*. 11(10):2976–2989.
79. Pritchard JK, Di Rienzo A (2010). Adaptation—not by sweeps alone. *Nature Reviews Genetics*. 11(10):665–667.
80. Bitter, M.C., Kapsenberg, L., Gattuso, J. *et al.* (2019). Standing genetic variation fuels rapid adaptation to ocean acidification. *Nature Communications*. 10, 5821.
81. Hamilton, J., Miller, J. (2016). Adaptive introgression as a resource for management and genetic conservation in a changing climate. *Conservation Biology*. 30: 33-41.
82. Savolainen O, Pyhäjärvi T, Knürr T (2007) Gene flow and local adaptation in trees. *Annual Reviews in Ecology and Evolutionary Systematics*. 38:595–619.
83. Eckert AJ, Bower AB, Wegrzyn JL *et al.* (2009). Association Genetics of Coastal Douglas Fir (*Pseudotsuga menziesii* var. *menziesii*, Pinaceae). I. Cold-Hardiness Related Traits. *Genetics*. 182(4): 1289-1302
84. Pickett STA. (1989). Space-for-time substitution as an alternative to long-term studies. In: Long-term studies in ecology: approaches and alternatives. Springer-Verlag, New York, New York, USA. 110–135
85. Ledig, F. T., Rehfeldt, G. E., Saenz-Romero, C., & Flores-Lopez, C. (2010). Projections of suitable habitat for rare species under global warming scenarios. *American Journal of Botany*, 97(6), 970–987.
86. Mei W, Stetter MG, Stitzer MC. (2019). Adaptation in plant genomes : Bigger is different. *American Journal of Botany*, 105: 16-19
87. Schoettle A & Sniezko R. (2007). Proactive interventions to sustain high elevation Pine ecosystems ecosystems sustained by white pine blister rust. *Journal of Forest Research*. 121(5): 327-336.
88. Aitken SN, Yeaman S, Holliday JA *et al.* (2008). Adaptation, migration and extirpation: climate change outcomes for tree populations. *Evolutionary Applications*. 95-111.

89. Suarez-Gonzalez A, Hefer CA, Christie C, et al. (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.
90. Looney CE, Waring KM (2013). *Pinus strobiformis* (southwestern white pine) stand dynamics, regeneration, and disturbance ecology: *Annual reviews in Forest Ecology and Management*, 287, 90–102.
91. Little EL (1971). Atlas of United States trees. Volume 1. Conifers and important hardwoods. Misc. Publ. 1146. Washington, DC: U.S. Department of Agriculture, Forest Service. 320 p.
92. Goodrich BA, Waring KM & Kolb TE. (2016). Genetic variation in *Pinus strobiformis* growth and drought tolerance from southwestern US populations. *Tree Physiology*. 36: 1219-1235.
93. DaBell J. (2018). *Pinus strobiformis* response to an elevational gradient and correlation with source climate. (Master's thesis). Northern Arizona University.
94. Frankis MP. (2009). The high altitude white pines (*Pinus* L. subgenus *Strobus* Lemmon, Pinaceae) of Mexico and the adjacent SW USA. *International Dendrology Society Yearbook 2008*: 64-72.
95. Bisbee J (2014). Cone morphology of the *Pinus ayacahuite-flexilis* complex of the southwestern United States and Mexico. *Bulletin of the Cupressus conservation project*, 3, 3–33.
96. Moreno-Letelier A, Ortíz-Medrano A, Piñero D (2013). Niche Divergence versus Neutral Processes: Combined Environmental and Genetic Analyses Identify Contrasting Patterns of Differentiation in Recently Diverged Pine Species. *PLoS ONE*, 8, e78228.
97. Aguirre-Gutiérrez J, Serna-Chavez HM, Villalobos-Arambula A et al (2015). Similar but not equivalent: ecological niche comparison across closely-related Mexican white pines. *Diversity and Distributions*, 21, 245-257.
98. Syring J, Farrell K, Businský R, Cronn R, Liston A (2007). Widespread genealogical nonmonophyly in species of *Pinus* subgenus *Strobus*. *Systematic Biology*. 56(2):163-181.
99. Moreno-Letelier A, Barraclough TG (2015). Mosaic genetic differentiation along environmental and geographic gradients indicate divergent selection in a white pine species complex. *Evolutionary Ecology*, 29, 733–748.

CHAPTER 1

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

Introduction

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

Several studies of speciation have used hybrid zones as windows into the process of divergence between species (reviewed by Petit & Excoffier 2009). Studies conducted across the entire geographical range of hybridizing species have helped reveal not only the demographic context of speciation, but also the relative importance of intrinsic and extrinsic processes

(Schield *et al.* 2017; Ryan *et al.* 2017). Specifically, the maintenance of species boundaries has been shown to occur through tension zones (intrinsic incompatibilities *sensu* Barton & Hewitt 1985; Via *et al.* 2000; Barton 2001; Rundle 2002) and bounded hybrid superiority (extrinsic incompatibilities *sensu* Moore 1977; Milne *et al.* 2003; Hamilton *et al.* 2013). The former facilitates divergence through a buildup of genetic incompatibilities among loci causing environmentally independent reduction in hybrid fitness, whereas the latter involves increased hybrid fitness only in an intermediate environment to which the divergent parental allelic combinations confer a putative advantage. These two processes can be coupled, such that genomic regions involved in intrinsic incompatibility coincide with loci exhibiting ecological gradients in allele frequency (Bierne *et al.* 2011; Cushman & Landguth 2016), ensuring the maintenance of species barriers despite the homogenizing effect of gene flow (Kulmuni & Westram 2017). Thus, the interaction between intrinsic and extrinsic barriers to gene flow generates a genomic mosaic of introgression and differentiation that depends in part upon the demographic context and life history traits of the diverging lineages.

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

reproductive isolation can progress and be associated with several of these processes, such as with ecological niche partitioning (Agrawal *et al.* 2011).

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

In this study, we use an integrative approach to investigate processes leading to the divergence of two North American pine species—*Pinus strobiformis* Engelm. (southwestern white pine), and *P. flexilis* E. James. (limber pine). Our focal species inhabit a wide latitudinal range in the western part of North America, but display limited differences in morphological and reproductive traits (Benkman *et al.* 1984; Tomback *et al.* 2011; Bisbee 2014). Within a putative area of sympatry, located in the southern Rocky Mountains and Colorado Plateau, morphological evidence points towards the occurrence of hybridization (Steinhoff & Andresen 1971; Tomback & Achuff 2010). To examine the processes influencing species boundaries between these two conifers, we asked three questions: (1) Does the hybrid zone between *P. strobiformis* and *P. flexilis* occupy a niche ecologically divergent from either parent species? (2) Did the divergence of *P. strobiformis* and *P. flexilis* occur with continual gene flow? (3) Does a genome-wide mosaic of differentiation characterize divergence between *P. strobiformis* and *P. flexilis*, and is

this pattern attributed to extrinsic, intrinsic, or an interaction of both factors? Our results are consistent with ecological divergence occurring with continual gene flow between the focal species, with several lines of evidence supporting a strong influence of extrinsic factors in reinforcing species boundaries.

Materials and Methods

Focal taxa and field sampling

Pinus strobiformis and *P. flexilis* are closely related species of white pines that occur in the mountainous areas of western North America. The native range of *P. strobiformis* includes Mexico and the southwestern United States, and its distribution exhibits disjunctions across dry and wet boreal mixed forest ecosystems (Looney & Waring 2013; Fig. 1.1). *Pinus flexilis* inhabits areas across northern Arizona and northern New Mexico to Alberta, Canada, with a region of putative sympatry with *P. strobiformis* in the southern Rocky Mountains and Colorado Plateau (Fig. 1.1). Across this zone of putative sympatry, cone morphology and dispersal syndromes fall along a continuum of divergence, blending into the characteristics of populations in the allopatric zones of either species (Bisbee 2014).

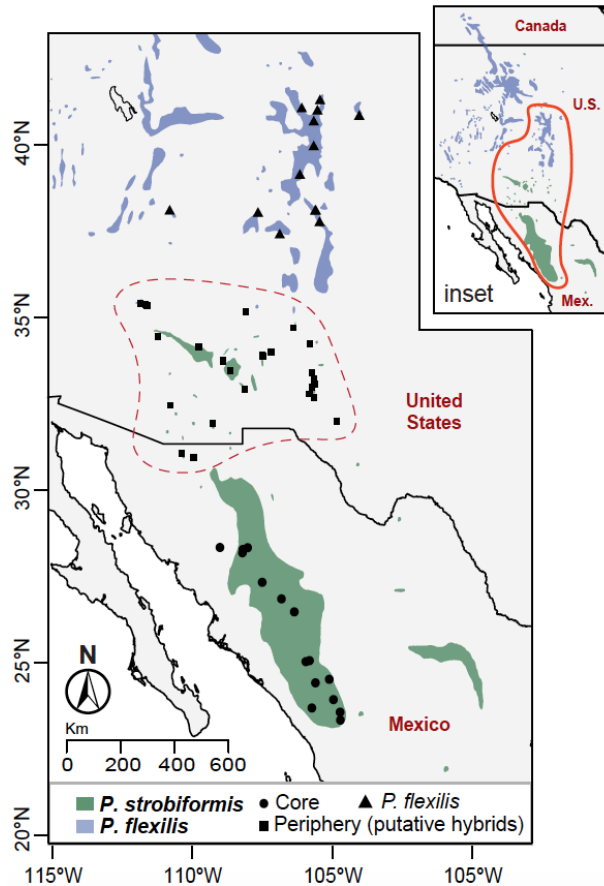


Fig 1.1: Map of sampling localities (black dots) overlaid on polygons showing geographical ranges for *Pinus strobiformis* (green) and *P. flexilis* (blue). Peripheral populations (squares) represent the putative hybrid zone.

We sampled 42 *P. strobiformis* populations encompassing a total of 376 trees (5-13 trees/population) from its entire geographical range. We avoided sampling the southeastern populations of *P. strobiformis*, as this region has been identified as putative hybrid zone with *P. ayacahuite* and trees here have been classified as *P. strobiformis* subspecies *veitchii* (Frankis 2009). Populations within *P. strobiformis* were classified into ‘Core’ (latitudinal range: 19–30.5 °N) and ‘Periphery’ (latitudinal range: 31–33 °N), such that Periphery represents the putative hybrid zone between *P. strobiformis* and *P. flexilis*. For *P. flexilis*, a total of 13 populations were sampled, with eight populations sampled from the southern range margin and five sampled closer to the range center (Fig. 1.1). Across these thirteen populations, we sampled a total of 69 trees

(4–10 trees/population). To help minimize relatedness, trees within the same site were sampled with a minimum spacing of 50 m (*P. strobiformis*) and 200 m (*P. flexilis*) from each other.

Data generation

Occurrence data

We assembled a comprehensive dataset of occurrences for ecological niche modeling (ENM) by supplementing our field site data with occurrence records downloaded from the Global Biodiversity Information Facility (GBIF), using functions from the DISMO package (Hijmans *et al.* 2017) available in the R environment (R Core Team 2017). GBIF records were filtered to be constrained within the known geographical distribution of both species as provided by Little (1971). These data were processed to remove duplicates, sub-species and records with geo-referencing errors, yielded 600 occurrence records for *P. strobiformis* and 420 for *P. flexilis*. Datasets available through GBIF are prone to several sampling biases, such as easily accessible locations are probably more intensely sampled regardless of the actual population density, thereby biasing the environmental space sampled (Boria *et al.* 2014). We observed this bias in our occurrence records for *P. strobiformis* where peripheral populations in the US were sampled more densely than the populations at the core of the species range in central Mexico. Another common bias is that occurrence records and predictor variables are not at identical spatial resolution; thus generating biased suitability scores for cells with higher numbers of individuals thereby increasing the rate of false positives. We addressed the biases discussed above by a) spatially projecting occurrence points and raster layers to the azimuthal equal area projection system under the WGS84 datum, and b) conducting spatial thinning to retain only one occurrence record within a 5 km radius of each point. These procedures yielded a final dataset containing 254 *P. strobiformis* occurrences and 336 *P. flexilis* occurrences.

Incorporating intraspecific genetic variation into ENMs can improve model fit and provide more accurate predictions when projecting across time and space (Knowles *et al.* 2007; Ikeda *et al.* 2017). Thus, we divided presence locations within *P. strobiformis* into the same Core and Periphery groups mentioned above in the ‘Focal taxa and field sampling’ section. These groups likely represent different genetic clusters given the geographically restricted phenotypic evidence of hybridization between *P. flexilis* and *P. strobiformis* (Steinhoff & Andresen 1971; Tomback & Achuff 2010; Bisbee 2014). We defined three groups that were the focus of our enquiries – (1) populations of *P. flexilis*, (2) populations of *P. strobiformis* from the northern range periphery (Periphery hereafter), and (3) populations of *P. strobiformis* from the range core (Core hereafter). Nineteen bioclimatic variables and altitude were used as predictors in the ENMs for all three groups. Present day geospatial data layers at 30 arc-second resolutions and at 2.5 arc-minute resolutions for the Last Glacial Maximum (LGM) were downloaded from WorldClim v.1.4 (Hijmans *et al.* 2005). For each of the twenty layers, data were extracted using the RASTER package (Hijmans *et al.* 2016) available in R.

DNA sequence data

We extracted total genomic DNA from 445 individuals sampled across 55 populations of both species using DNeasy Plant Kits (Qiagen). Five ddRADseq libraries (Peterson *et al.* 2012), each containing up to 96 multiplexed samples each, were prepared using the procedure detailed in Parchman *et al.* (2012). All libraries were digested using the *EcoRI* and *MseI* restriction enzymes followed by ligation of adaptors, barcodes, and primers. Post PCR, we selected DNA fragments in the 300–400 bp size range using agarose gel electrophoresis and isolated the pooled DNA using QIAquick Gel Extraction Kit (Qiagen). Single-end sequencing, with one multiplexed library per lane, was used to obtain 105 bp reads, with all sequencing conducted with Illumina

HiSeq 2500 at the Nucleic Acids Research Facility located at Virginia Commonwealth University. The resulting FASTQ files were processed using the DDOCENT bioinformatics pipeline (Puritz *et al.* 2014) and a series of downstream custom scripts to filter single nucleotide polymorphisms (SNPs) based on minor allele frequency cutoff, amount of missing data, PHRED quality score, read depth and F_{IS} values. The entire process yielded a total of 51 633 single nucleotide polymorphisms (SNPs), which were used as the starting dataset for all subsequent analyses.

Data analysis

Ecological niche modeling and niche divergence

We developed ENMs for each of the following groups: Core, Periphery, and *P. flexilis*, using algorithms available in the maximum entropy software program, MAXENT (Phillips *et al.* 2006). Since MAXENT was specifically developed for presence-only data, we drew a one-degree rectangular buffer around the known distribution of both species and obtained 100 000 background points at random without duplicates. Data processing, model fitting, and model evaluation using 5,000 iterations within MAXENT were conducted using the DISMO, RASTER, RGDAL (Bivand *et al.* 2017), and SPThin (Aiello-Lammens *et al.* 2015) packages available in R. ENMs were constructed from climate variables with an absolute correlation coefficient (r) less than 0.85 (Table 1.S1) to minimize collinearity that can inflate the effect of predictor variables (Braunisch *et al.* 2013). Two indices were used to assess model performance for each group: overall regularized training gain (RTG) and area under the curve (AUC). Since LGM data were not available at 30 arc-seconds resolution, we built two ENMs for each group (2.5 arc-minutes and 30 arc-seconds), but only used the 2.5 arc-minutes models for hindcasting to infer historical patterns of sympatry between species that could facilitate gene flow. We followed an average

projection ensemble approach across three LGM scenarios (CCSM4, MIROC, & MPI) to obtain a hindcasted suitability map. Changes in habitat suitability (stability) were assessed by adding MAXENT-predicted suitability maps across the LGM and present (as in Ortego *et al.* 2015). For these maps, values closer to 2 in a gridded cell are associated with the stability of highly suitable habitat for a given group across time points. In contrast, values closer to 0 are associated with the stability of highly unsuitable habitat for a given group across time points. Suitability scores across the full geographical extent for present conditions at 30 arc-seconds were obtained for all three groups delineated in our study. To investigate patterns of niche evolution, we conducted pairwise comparisons of these suitability scores. We accounted for potential biases towards niche divergence introduced by latitudinally-associated environmental variation in the present range of each pair, by performing asymmetric background randomization test, based on Schoener's D , in the R package ENMTOOLS (Warren *et al.* 2008). The two resulting null distributions obtained through this test correspond to the background level of niche divergence for each pair. An observed value of Schoener's D much smaller than expected after accounting for background differences could indicate niche divergence, whereas a value much larger than expected indicates niche conservatism (Warren *et al.* 2008).

Population structure and demographic modeling

We assessed the pattern and extent of genetic divergence between *P. strobiformis* and *P. flexilis* using multiple methods. First, we grouped the 42 *P. strobiformis* populations into the same Core and Periphery groups described above (see Data Generation & Fig. 1.1). We conducted a principal components analysis (PCA) to visualize grouping of sampled trees into the three groups delineated in our methods (Patterson *et al.* 2006; McVean 2009). To complement the PCA, we also conducted an individual-based assignment test using FASTSTRUCTURE (Raj

et al. 2014). We set the number of clusters (K) to 2, representing the two parental species investigated here, as we were interested in admixture between two defined species and not the potential number of genetic groups. Lastly, we utilized hierarchical fixation indices (F -statistics) to assess the extent of differentiation between species by nesting trees into populations and populations into species. There are two levels within the hierarchy, with F_{CT} describing differentiation among groups at the highest level of the hierarchy and F_{ST} describing differentiation among groups across all levels of the hierarchy (Yang 1998). A similar nested model with the highest level of hierarchy being groups within *P. strobiformis* (Core and Periphery) was used to assess intraspecific differentiation. For the former, F -statistics are denoted using the term ‘species’ in the subscripts, whereas the latter uses the term ‘groups’ in the subscripts. We used a similar hierarchical model with variance partitioning to estimate group specific and pairwise F -statistics for the three groups delineated in this study. We denote pairwise values of F_{ST} using one-letter abbreviations for the groups being compared (e.g. F_{ST-CP} indicates F_{ST} between Core and Periphery), and group specific values of F_{ST} with the name of the group in subscripts. We constructed 95% confidence intervals of multilocus F -statistics using bootstrap resampling ($n = 100$ replicates) in the HIERFSTAT package (Goudet 2005) available in R. Along with estimation of F -statistics, we also assessed overall levels of genetic diversity using multilocus estimates (i.e. means across SNPs) of observed and expected heterozygosities (H_o and H_e , respectively) per population.

Presence of individuals with mixed ancestry, as identified using FASTSTRUCTURE, can be a result of secondary contact, incomplete lineage sorting, or the presence of gene flow throughout the divergence history. Disentangling these explanations is important, because it directly influences our understanding of the relative importance of intrinsic and extrinsic factors

in facilitating speciation. For instance, when speciation is recent or has occurred with gene flow, we expect to see islands of divergence around regions experiencing strong intrinsic or extrinsic selection (Wu 2001; Feder *et al.* 2012). However, if hybrids are formed in areas with novel habitats, introgression might be selectively advantageous causing the absence of such islands. To infer the timing and influence of various demographic processes shaping the divergence history of our focal groups, we conducted demographic modeling using Diffusion Approximation for Demographic Inference ($\partial A \partial I$ v.1.7; Gutenkunst *et al.* 2009). We down-sampled the total SNP dataset for computational simplicity by creating bivariate 0.05-interval bins based on F_{ST} and heterozygosity, and then subsampling each bin such that the proportion of SNPs retained per bin represented each bin's contribution to the full dataset. We performed graphical checks using PCA to ensure that overall patterns of diversity and population genetic structure were preserved in the down-sampled data. To avoid demographic inference from being biased due to patterns of linkage disequilibrium we randomly sampled one SNP per assembled contig to obtain a final dataset of 6330 SNPs. To obtain the input for $\partial A \partial I$ runs we converted our vcf file containing the sub-sampled set of SNPs into a folded site frequency spectrum using mostly the default settings, but specifying appropriate projection values that maximised the number of segregating sites per group (Core, Periphery, and *P. flexilis*) using the '--project' flag in EASYSFS (Overcast 2017).

We compared a model of pure divergence with no gene flow (M_1) against a set of 10 alternative demographic models (M_2 – M_7) representing different speciation scenarios including varying timing and directionality of ancient or contemporary gene flow (Fig. 1.S1). Complexity was added to the models with gene flow by incorporating heterogeneity in the gene flow parameter across loci (Tine *et al.* 2014, models M_8 – M_{11} , Fig. 1.S1), which served as a test for

islands of divergence. We ran 10 replicate runs of each model in $\partial A \partial I$, using a $200 \times 220 \times 240$ grid space and the nonlinear Broyden-Fletcher-Goldfarb-Shannon (BFGS) optimization routine. Following Carstens *et al.* (2013), we conducted model selection in an information-theoretic framework using Akaike information criterion (AIC; Akaike 1974) and ΔAIC ($AIC_{\text{model } i} - AIC_{\text{best model}}$) scores (Burnham & Anderson 2002), calculated using results from the best replicate run (highest composite likelihood) for each model. We performed Fisher Information Matrix (FIM)-based uncertainty analysis on the best-supported model by setting the *eps* parameter to 10% in order to obtain upper and lower 95% confidence intervals (CIs) for all parameters. The *eps* parameter represents the step size for numerical derivatives within $\partial A \partial I$ that is used to calculate the curvature of the likelihood surface near the maximum likelihood estimates (MLEs) of our parameters. Unscaled parameter estimates, and their 95% CIs, were obtained using a per-lineage substitution rate of 7.28×10^{-10} substitutions/site/year rate estimated for Pinaceae by De La Torre *et al.* (2017) and a generation time of 50 years.

Genomics of interspecific introgression

Analyses of clines across hybrid zones are widely used to identify loci exhibiting exceptional patterns of introgression relative to the average genomic background (Fitzpatrick 2013; Gompert *et al.* 2012a; Gompert & Buerkle 2011; Stankowski *et al.* 2015). We classified our sampled trees into categories corresponding to admixed ($n_A = 111$) and parental species ($P. strobiformis = 277$, $P. flexilis = 54$) based on the Q -values from FASTSTRUCTURE. Trees with Q -values of 0.9 or higher were classified as pure $P. strobiformis$, those with Q of 0.10 or lower were classified as pure $P. flexilis$, and those with intermediate Q -values were classified as admixed (e.g. Ortego *et al.* 2014). As most loci exhibited little to no differentiation between parental species, we retained only loci with a minor allele frequency (MAF) difference of at least

10% between parental species ($n = 4,857$ SNPs). This allowed us to avoid false correlations between cline parameters and fixation indices (Parchman *et al.* 2013). We used this subset of 4,857 SNPs to perform a Bayesian genomic cline analysis in BGC v1.0 (Gompert & Buerkle 2012; Gompert & Buerkle 2011). Using Markov chain Monte Carlo (MCMC) sampling, BGC estimates the posterior distribution of ancestry for each locus as a function of the genome-wide admixture coefficient. The BGC model includes two genomic cline parameters, α (genomic cline center) and β (genomic cline rate, i.e. slope), determining the probability of *P. flexilis* ancestry, and the rate of transition from *P. flexilis* to *P. strobiformis* given a level of genomic admixture described by the hybrid index, h , respectively (Gompert & Buerkle 2012; Gompert *et al.* 2012a). A tree with $h = 0$ was classified as having solely *P. strobiformis* ancestry, whereas a tree with $h = 1$ was classified as having solely *P. flexilis* ancestry. We ran BGC for five replicate runs, each 45 000 steps in length, and, after discarding the first 25 000 steps as burn-in, we thinned the posterior distribution every 20 steps, thus yielding 1,000 samples which were used for inference of model parameters. We used TRACER v1.6 (Rambaut *et al.* 2013) to test for convergence among replicated runs, as well as appropriate mixing along MCMC chains. We identified excess ancestry loci (relative to the genome-wide average) as those with posterior α or β credible intervals (CrI; 95% equal-tail intervals) not containing zero. We identified outlier loci as those with posterior mean point estimates of α ($\hat{\alpha}$) or β ($\hat{\beta}$) significantly different from the rest of the genome, as judged by comparison to posterior quantiles of random-effect priors for α and β (Gompert *et al.* 2012a). Besides categorizing loci as excess ancestry or outlier, we also tested for correlations among locus-specific $F_{CT-species}$, α , and β , with and without absolute values for α and β . The sign of the cline parameters (specifically β) have direct implications for inferring the processes maintaining species boundaries and hence were incorporated in correlation tests.

Specifically, extremely positive values of β reflect strong selection against hybrids or population structure in the hybrid zone (Gompert *et al.* 2012b), while extremely negative values of β indicate a wide cline representing easy dispersal across species boundaries (Janoušek *et al.* 2012).

Although the hybrid index (h) obtained from BGC provides information about the age and stability of a hybrid zone, such inferences are limited to only one generation of admixture (Fitzpatrick 2012). We estimated h and interspecific heterozygosity using INTROGRESS (Gompert & Buerkle 2010), in order to extend our interpretations to a historical hybrid zone and categorize individuals into recent (F1s), advanced generation (FNs), and backcrossed hybrids (BCs). This was done using a modified classification from Hamilton *et al.* (2013). Both BGC and INTROGRESS yielded very similar estimates of h (Pearson's $r = 0.70, p < 0.001$), thus we used estimates from INTROGRESS due to the availability of inter-specific heterozygosity estimates from this software. To test for the influence of extrinsic factors in the maintenance of species boundaries, we performed linear regression analyses with backward variable selection using h against climate and geography as predictor variables. This was done using the reduced set of climate variables from the final ENMs (see Table 1.S1).

Results

Ecological niche modeling and niche divergence

ENMs for each of the three groups used in this study (Fig. 1.1) had high predictive ability, as indicated by AUC and RTG values (Table 1.1). For Core and Periphery, several covariates were important, with precipitation seasonality (Bio15) shared between Core and Periphery. For *P. flexilis*, altitude was consistently the most important variable across different measures of variable importance (Table 1.1).

Table 1.1 Ecological niche model performance and variable importance at 30 arc-second resolution

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

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

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

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

€: Variables with the highest permutation or percentage importance

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

Hindcasting the 2.5 arc-minute model onto LGM data layers supported a recent, post-LGM niche fragmentation and northward expansion in Periphery (Fig. 1.S2). A similar post-LGM northward expansion of suitable niche space was observed for *P. flexilis*. Furthermore, there was extensive range overlap between the two species during the LGM, which was greater than what is currently observed (Fig. 1.S2). Values of niche similarity based on Schoener's *D* ranged from 0.05 (*P. flexilis* – Core) to 0.17 (Periphery – Core). Background randomization tests revealed statistically significant niche divergence for two of the three comparisons (Fig. 1.2). For the third comparison, however, niche divergence was asymmetrical between Core and Periphery, with the niche of Periphery being conserved relative to the background of Core (Fig. 1.2A). A similar pattern was noted using only the presence points, where each group formed a distinct cluster within the multivariate climate space defined by the top two principal components (PCs) derived from PCA on the climate variables used for construction of the ENMs (Fig. 1.S3A).

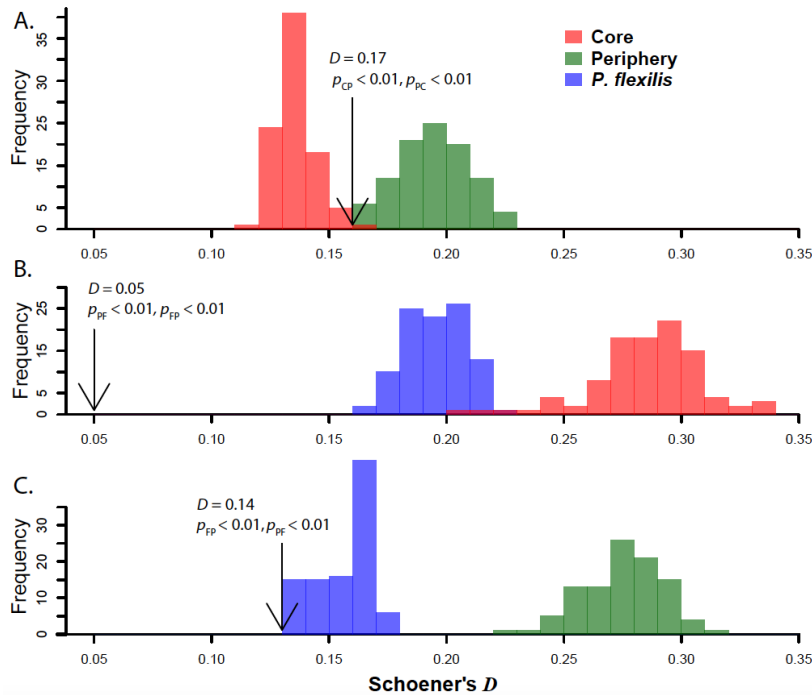


Fig. 1.2. Results of niche divergence tests (Schoener's D) for **A)** Core versus Periphery **B)** Core versus *P. flexilis* and **C)** *P. flexilis* versus Periphery. Histograms indicate the background levels of niche divergence and arrows indicate the observed value of Schoener's D for each pair compared.

Population structure and divergence history

The PCA using 51 633 SNPs was consistent with trees sampled from Core being differentiated from those of *P. flexilis*, which was most marked along PC1 (Fig. 1.3A). This PC explained 0.90% of the total genetic variance, which was in line with the overall level of differentiation estimated using hierarchical F -statistics ($F_{ST\text{-species}} = 0.021$, 95% CI: 0.008–0.031). Trees sampled from Periphery were located between those sampled from Core and *P. flexilis* (Fig. 1.3A), in line with Periphery containing hybrids. There was also a latitudinal gradient in the mean population Q -values, as estimated using FASTSTRUCTURE, with Core populations

exhibiting little to no ancestry from *P. flexilis* and Periphery being a mixture of *P. flexilis* and Core (Fig. 1.3B).

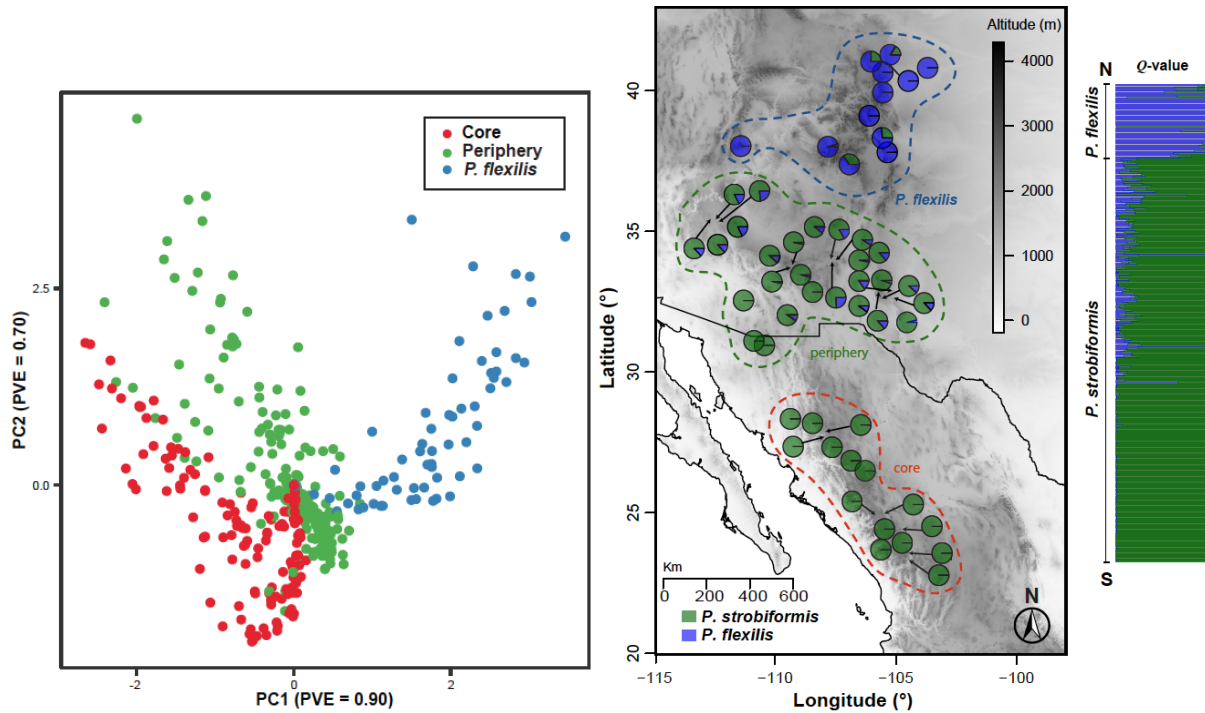


Fig. 1.3.A) Results of population genetic structure analysis using PCA on 51 633 SNPs. **B)** Results of assignment analyses for each tree in FASTSTRUCTURE for $K = 2$ clusters (right panel) plotted onto a topographic map of the study area (left panel). Each pie chart represents the average ancestry of a population from *P. strobiformis* and *P. flexilis*.

At the individual tree level, we observed a strong negative correlation (Pearson's $r = -0.69, p < 0.001$) between Q -values of putative hybrids and latitude, which is consistent with a geographical gradient of genomic introgression, such that trees geographically proximal to either parental species contain more ancestry from that parental species. Multilocus estimation of

differentiation between species ($F_{CT\text{-species}}$) was 0.01 (95% CI: 0.005–0.018, Fig. 1.4A), while that between groups within *P. strobiformis* ($F_{CT\text{-groups}}$) was 0.003 (95% CI: 0.0007–0.006). Group specific multilocus F_{ST} , pairwise F_{ST} , and heterozygosities differed little among the three groups, with the Core–*P. flexilis* comparison having the highest pairwise $F_{ST\text{-CF}} = 0.019$ (Table 1.2). Although populations of Periphery exhibited slightly higher heterozygosities and F_{ST} values ($F_{ST\text{-periphery}}$), this pattern was mainly driven by few populations, as indicated by the wider confidence interval around these estimates (Table 1.2).

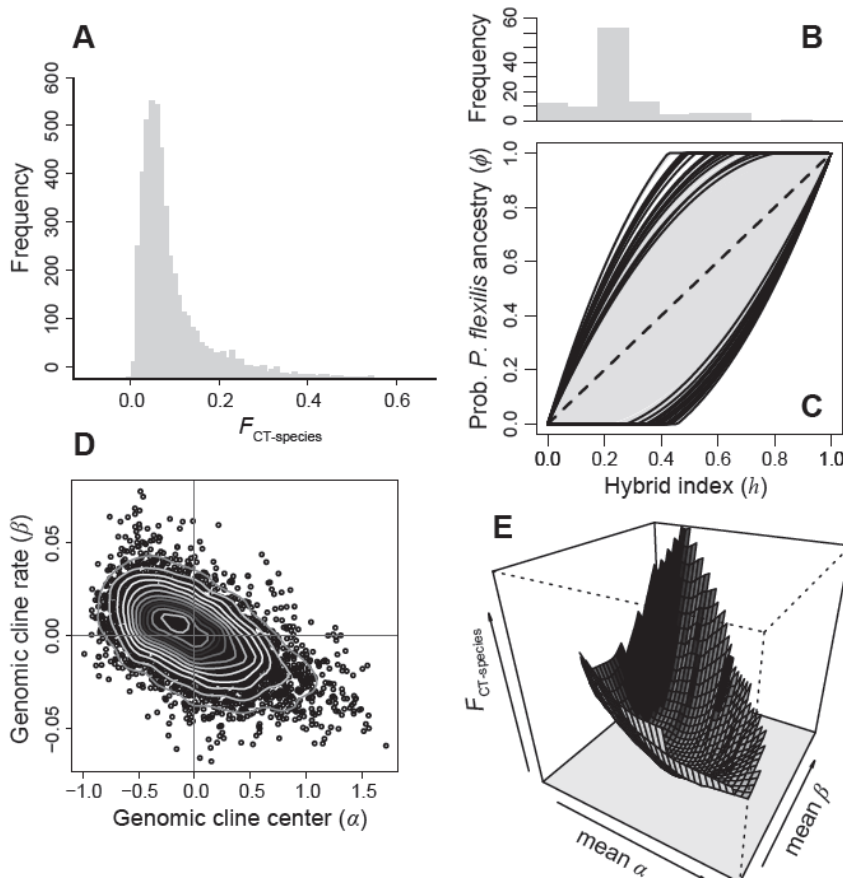


Fig. 1.4.A) Genomic distribution of F_{CT} , **B)** frequency distribution of hybrid index, **C)** variation in genomic ancestry as a function of hybrid index, **D)** correlation between genomic cline parameters, and **E)** 3D correlation plot of genomic cline parameters and F_{CT}

Table 1.2. Estimates of genetic diversity and divergence within and across the three groups, compared to a genome-wide $F_{ST-species}$ of 0.02 (95% CI: 0.008–0.03) and $F_{ST-strobiformis}$ of 0.009 (95% CI: 0.007–0.014).

Group	Multilocus F_{ST} (95% CI)	Pairwise F_{ST} (95% CI)	Mean $H_e \pm$ s.d.	Mean $H_o \pm$ s.d.
Core	0.003 (0.0025–0.0034)	Periphery: 0.009 (0.001–0.023) <i>P. flexilis</i> : 0.019 (0.006–0.032)	0.135 \pm 0.01	0.111 \pm 0.01
Periphery	0.007 (0.0071–0.0073)	<i>P. flexilis</i> : 0.015 (0.005–0.024) Core: 0.009 (0.001–0.023)	0.133 \pm 0.02	0.105 \pm 0.03
<i>P. flexilis</i>	0.003 (0.0025 – 0.0041)	Core: 0.019 (0.006–0.032) Periphery: 0.015 (0.005–0.024)	0.130 \pm 0.01	0.111 \pm 0.01

The best-supported demographic model was M_A , which is a model of symmetric ancient gene flow between the ancestral *P. strobiformis* and *P. flexilis* lineages, followed by contemporary gene flow between Periphery and *P. flexilis* (Table 1.3; Fig 1.4). This model was supported by a large AIC margin of 44.8 information units ($\Delta AIC_i \geq 44.8$). Converted parameter estimates indicated that the species diverged 18.04 million years ago (Ma) in the Miocene (95% CI: 26.29–9.79 Ma), but that the two groups within *P. strobiformis* diverged 3.63 Ma during the Pliocene (95% CI: 4.44–2.83 Ma) (Fig. 1.5; Table 1.S3). Overall rates of gene flow between species were substantial for both historical and contemporary periods; however, contemporary gene flow between species was geographically restricted to Periphery and *P. flexilis* (Table 1.S3). In addition, *P. flexilis* and Periphery experienced asymmetrical gene flow for which point estimates were larger in the direction of Periphery to *P. flexilis* ($M_{FP} = 11.53$ migrants/generation with a 95% CI: 0–57.94 versus $M_{PF} = 8.80$ with a 95% CI: 0–12.84). Periphery had the largest

population size estimate, while *P. flexilis* was inferred to have experienced an approximately 60% reduction in population size through time.

Table 1.3. Model composite likelihoods and AIC model selection results for 11 alternative demographic models of *P. strobiformis* (Core and Periphery)–*P. flexilis* divergence. Results for the best-supported model are underlined, and the two best models are shown in boldface.

Model	Model description	\ln Composite likelihood	k	AIC	ΔAIC_i
M_1	Strict isolation, no gene flow	–883.143	6	1778.29	65.44
hM_2	Secondary contact (Periphery– <i>P. flexilis</i>)	–886.227	7	1786.45	73.60
M_3	Ancient gene flow (speciation with gene flow)	–888.003	7	1790.01	77.16
M_4	Ancient gene flow, plus Periphery–<i>P. flexilis</i> gene flow	–847.424	9	1712.85	0.00
M_5	Ancient gene flow, plus Core–periphery gene flow	–885.428	9	1788.86	76.01
M_6	Secondary contact (Periphery– <i>P. flexilis</i>) and Core–Periphery gene flow	–883.949	10	1787.90	75.05
M_7	Ancient gene flow, followed by Periphery– <i>P. flexilis</i> gene flow, and Core–Periphery gene flow	–892.210	9	1806.42	93.57
M_8	Heterogeneous ancient gene flow	–869.824	14	1757.65	44.80
M_9	Heterogeneous ancient gene flow, plus Core–Periphery gene flow	–884.511	11	1791.02	78.17
M_{10}	Heterogeneous gene flow during secondary contact (Periphery– <i>P. flexilis</i>), and Core-Periphery gene flow	–902.279	9	1828.56	115.71
M_{11}	Heterogeneous ancient gene flow, followed by heterogeneous gene flow between Periphery– <i>P. flexilis</i> , and between Core–Periphery	–922.814	11	1873.63	160.78

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

Genomics of interspecific introgression

Hybrid index (h) values ranged from near zero to 0.80, with values around 0.20 being the most common thus suggesting overrepresentation of *P. strobiformis* ancestry (Fig. 1.4B).

Estimates of interspecific heterozygosity had a narrow range from 0.45 to 0.64, indicating weak reproductive barriers (Hamilton *et al.* 2013) and a long history of recombination within the

hybrid zone (Gompert *et al.* 2014). Classification of trees into genotypic classes based on h and interspecific heterozygosity revealed a dominance of advanced-generation hybrids (54%), with some trees being backcrossed into *P. strobiformis* (22%). No recent hybrids (F1s) were apparent. Stepwise linear regression analysis revealed a significant effect of geography and climate on h across the putative hybrid zone. Latitude (Pearson's $r = 0.41$, $p < 0.001$), precipitation seasonality (Pearson's $r = -0.32$, $p < 0.01$), and mean temperature of the warmest quarter (Pearson's $r = -0.18$, $p < 0.01$) had a strong influence on h , in line with the latter two being important predictor variables for Periphery in our ENM.

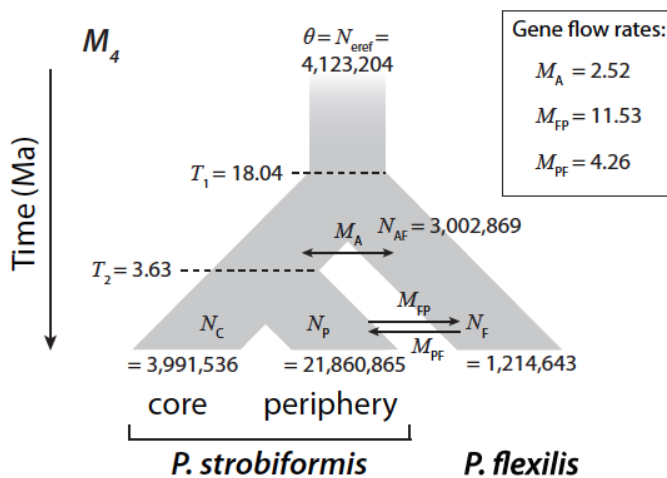


Fig. 1.5. The best-supported model from $\partial A \partial I$ analysis. This figure shows the parameter estimates for divergence times (T_i) in units of millions of years ago (Ma), reference effective population size (θ ; or after conversion, N_{eref}), lineage population sizes (N_i), and rates of gene flow (M_{ij}) for the optimal model determined by AIC model selection (see Table 1.3).

Substantial variation was found in estimates of genomic cline parameters (Fig. 1.4C,D), especially for α , with its range (-0.99 to 1.72) being 18.5-fold as wide as that of β (-0.068 to

0.078). Similar to the patterns observed in the distribution of h , an asymmetry towards *P. strobiformis* ancestry was noted in the genomic cline estimates. From the posterior distribution of α , we found 3,193 outlier loci, of which 570 (17.9%) had elevated probabilities of *P. flexilis* ancestry (positive $\hat{\alpha}$) and 2,623 (82.1%) had elevated probabilities of *P. strobiformis* ancestry (negative $\hat{\alpha}$). We identified fewer loci with excess ancestry, but in contrast to the pattern for outlier loci those with excess ancestry favored *P. flexilis* over *P. strobiformis* ancestry. Among the 287 loci with excess ancestry, 204 (71.1%) had excess *P. flexilis* ancestry (i.e. lower 95% CrI of $\alpha > 0$) and 83 (28.9%) had excess *P. strobiformis* ancestry (i.e. upper 95% CrI of $\alpha < 0$). The multilocus $F_{CT\text{-species}}$ estimate for loci with excess ancestry was 0.12 (95% CrI: 0.09–0.13) while for outlier loci it was 0.058 (95% CrI: 0.05–0.09). We did not identify any loci that were β outliers or had excess ancestry indicated by β . Hierarchical $F_{CT\text{-species}}$ was negatively correlated with raw values of α (Pearson's $r = -0.036$, $p = 0.01$), positively with raw values of β (Pearson's $r = 0.048$, $p < 0.001$) and positively with absolute values of both α (Pearson's $r = 0.14$, $p < 0.001$) and β (Pearson's $r = 0.26$, $p < 0.001$) (Fig. 1.4E).

Discussion

We identified strong evidence supporting ecological divergence with gene flow between *P. strobiformis* and *P. flexilis*. Our findings are generally consistent with previous reports on the species examined here; however, in contrast to the recent divergence time estimated by Moreno-Letelier *et al.* (2013), our demographic modeling is consistent with deeper divergence, as well as ongoing speciation with gene flow, that is driven and maintained primarily by extrinsic factors.

The latter was made possible by explicitly accounting for hybridization as a confounding and contributing factor to local adaptation and speciation.

Niche evolution and ecological divergence

Our results indicate that climatic factors have played a major role in driving niche divergence between *P. strobiformis* and *P. flexilis*. Populations within Periphery coincide with the known phenotypic hybrid zone between *P. strobiformis* and *P. flexilis* (Steinhoff & Andresen 1971; Tomback & Achuff 2010; Bisbee 2014) and formed a distinct group characterized by niche divergence from *P. flexilis* and asymmetrical niche divergence from Core. The asymmetrical pattern of niche divergence between Core and Periphery is likely the result of recent divergence. Under this scenario, we expect niche differentiation to occur primarily along a few environmental variables that strongly influence fitness in the transitional environmental conditions, with little to no differentiation among groups along other environmental axes (Fig. 1.S3C). In support of this expectation, precipitation seasonality was an important niche predictor for both Core and Periphery, but they were differentiated along this environmental axis (Fig. 1.S3B). While several other bioclimatic variables exhibited as large a difference as precipitation seasonality, they did not significantly contribute towards the niche of both Core and Periphery (Table 1.1; Fig. 1.S3C). These patterns reiterate the presence of hybrid populations in transitional environmental conditions, experiencing early stages of niche divergence.

In line with these results, precipitation seasonality and mean temperature of the warmest quarter had a strong negative association with genomic ancestry and contributed to the niche divergence of Periphery. These two climatic variables influence plant evapotranspiration and affect drought responses (Mishra & Singh 2010). Drought stress during the active growing season is widely recognized as a limiting factor to plant growth in the western parts of North

America (Williams *et al.* 2010; Restaino *et al.* 2016), and our results are indicative of adaptive divergence along a drought gradient between the three groups (Gitlin *et al.* 2006; Allen & Breshears 1998). Further, our study broadly agrees with other reports in *P. strobiformis* that indicate precipitation and altitude as important niche predictors (Aguirre-Gutiérrez *et al.* 2015; Shirk *et al.* 2017). Soil and vegetation variables used in previous ENMs, however, were not included in our analyses due to a lack of comparable data for *P. flexilis* and its unclear relationship to divergence history.

Despite fluctuations in suitable range size (Fig. 1.S2) and previous studies indicating reduction in genetic diversity at range margins using chloroplast markers (Moreno-Letelier & Piñero 2009), we find no evidence for this in our study. This could be explained by the asymmetry in gene flow between Periphery and *P. flexilis*, as inferred from the demographic modeling results (Bridle & Vines 2007; Ortego *et al.* 2014). Evidence of directional introgression from *P. flexilis* (positive α outliers), moreover, might also have facilitated adaptation to transitional environmental conditions. Such novel allelic combinations have often contributed to the ability of populations to colonize new niches that are intermediate but beyond the climatic conditions experienced by the parental species (De Carvalho *et al.* 2010; Hamilton *et al.* 2013; De La Torre *et al.* 2014b; Geraldés *et al.* 2014). Presence of a locally adapted and historical hybrid zone is supported by the absence of β outliers in our genomic cline results (Kamdem *et al.* 2016), as well as by a recent study identifying high Q_{ST} values associated with physiological traits primarily linked to drought tolerance within Periphery (Goodrich *et al.* 2016). The geographic cline in h , asymmetry in excess ancestry loci towards *P. flexilis*, and elevated estimates of $F_{ST\text{-periphery}}$, however, indicate the potential for geographically driven neutral introgression to generate biased signals of local adaptation within the peripheral

populations (Geraldès *et al.* 2014). Ongoing investigations using replicate populations in the hybrid zone across gradients of geographic proximity and climate similarity will be able to address this issue in further detail (Lotterhos & Whitlock 2015; Riquet *et al.* 2017).

Speciation with gene flow without islands of divergence

Demographic modeling indicated that divergence of *P. strobiformis* and *P. flexilis* is not recent (~18 Ma) on an absolute time scale and has occurred with continuous gene flow. The presence of continual gene flow and absence of a period of allopatry, moreover, is also supported by the L-shaped distribution of $F_{CT-species}$ values (Fig. 1.4A. Nosil & Feder 2012). Reduction in overlapping niche suitability from LGM to present, between *P. strobiformis* and *P. flexilis*, agrees with the best-supported demographic model indicating continuous but geographically restricted contemporary gene flow. Contemporary reduction in N_e for *P. flexilis* from our demographic modeling is contrary to the predicted post-LGM expansion of suitable habitat. This is likely due to the limited geographical sampling within *P. flexilis* for our genomic analyses or a nonlinear relationship between habitat suitability and realized population sizes. Specifically, due to the geographical bias in the sampling scheme, we were unable to account for further population structure within *P. flexilis*. This may also have biased our inference of gene flow, such that contemporary gene flow between the two species is restricted to geographically proximal genetic groups. However, the primary focus of our study was estimating whether or not divergence occurred with gene flow, which is unlikely to be influenced by sampling biases of this form. Further, based on results from the hindcasted niche models, the extensively sampled southeastern region of *P. flexilis* forms a putative refugium likely representing much of the diversity in southern *P. flexilis* that then expanded northward after the LGM. Thus, regardless of

the geographical bias in our sampling scheme, we are likely to have captured a sizeable fraction of the segregating variation within *P. flexilis*.

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

Genomic mosaic of introgression

The spatial context of loci within genomes, as well as the temporal scale of divergence between lineages, can influence patterns of introgression and are often depicted by a mosaic

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

A positive correlation between the steepness of genomic clines (β) and F_{CT} points towards coincidence of loci involved in disruptive selection and those involved in reproductive isolation. Such a positive association has been demonstrated across several taxa (*cf.* Janoušek *et al.* 2012; Parchman *et al.* 2013; Gompert *et al.* 2014; Ryan *et al.* 2017) and we suggest it to be indicative of disruptive selection driving the evolution of intrinsic barriers and its coupling with extrinsic processes. Several empirical and simulation based studies have demonstrated that both

α and β can reflect patterns of selection in the hybrid zone (Gompert et al. 2012b; Janoušek *et al.* 2012), but the interpretation of these values is influenced by the underlying demographic scenario (Gompert & Buerkle 2012; Gompert *et al.* 2012a, 2012b). Under the observed demographic scenario of ongoing gene flow, signatures of selection against hybrids (i.e., underdominance) would be reflected by steep genomic clines (positive β), while selection for hybrids (i.e., overdominance) would be reflected by wide genomic clines (negative β ; Gompert & Buerkle 2011; Janoušek *et al.* 2012). The observed absence of positive β outliers and of islands of divergence in our demographic analysis indicates that despite some evidence of coupling between intrinsic and extrinsic barriers, widespread intrinsic incompatibilities are absent in this system, at least for the loci examined in this study. This is consistent with studies demonstrating weak reproductive isolation examined through forced crosses among these and other white pine species (Critchfield 1986). Shared life history strategies among conifers, such as long generation time and high dispersal capacity, are likely to restrict the evolution of post- and pre-zygotic isolating mechanism (Stacy *et al.* 2017). The limited evidence of intrinsic incompatibilities noted in our study supports the above claim, and we suggest that this pattern could be generalized across conifers with similar divergence history. Absence of negative β outliers and of recent hybrids indicates widespread recombination within the hybrid zone and an intermediate stage of divergence between our focal species (Nosil *et al.* 2009). The intermediate stage of divergence between our focal species, despite a long period of divergence in absolute time (i.e. years), is not surprising given the long generation times and large N_e estimates for conifers, which would have reduced the realized period of divergence when measured in coalescent units. Overall, the total absence of β outliers indicates a viable hybrid zone maintained largely through extrinsic factors (Kamdem *et al.* 2016), which may be the first stage of coupling

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

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

Evidence of higher number of outliers with *P. strobiformis* ancestry and a negative association between our cline parameters (α and β) could be explained by three processes: (i) intrinsic incompatibilities resulting from Dobzhansky–Muller effects or complex epistatic effects disproportionately favoring allelic combinations from *P. strobiformis* in the hybrids relative to *P. flexilis* parental background, (ii) widespread directional selection on alleles from *P. strobiformis*

in the hybrid zone leading to the formation of co-adapted gene complexes, and (iii) incomplete lineage sorting resulting from recent divergence between Core and Periphery. In contrast to inferences from the Engelmann–white spruce hybrid zone (De La Torre *et al.* 2014b), the asymmetry of outlier loci is not due to high rates of gene flow from Core into Periphery, as the best demographic model excluded gene flow between these groups (see Figure 1.5B). A higher number of outlier loci with introgression favoring *P. strobiformis* is consistent with the strong influence of selection favoring alleles with *P. strobiformis* ancestry in the hybrid zone. Even without a linkage map, the cline results, along with asymmetrical niche divergence between Core and Periphery, points towards widespread directional introgression from *P. strobiformis* into the hybrid zone, which is consistent with local adaptation driving the evolution of co-adapted gene complexes from *P. strobiformis* and of emerging intrinsic incompatibilities (Gompert *et al.* 2012b). The geographic clines of h , despite the absence of current gene flow between the Core and Periphery, also points towards an effect of incomplete lineage sorting. However, higher directional introgression from *P. strobiformis* even after accounting for the skewed pattern of genomic ancestry in the hybrid individuals emphasizes the role of selection over incomplete lineage sorting.

Our results are in accordance with studies in other coniferous species demonstrating that speciation is likely initiated through ecological barriers, and several generations of hybridization might occur before the evolution of intrinsic barriers to gene flow (Hamilton *et al.* 2013; Zhou *et al.* 2014; Stacy *et al.* 2017). Integrating the existing genomic dataset with ongoing planting experiments involving climate treatments and measurements of fitness related traits should also help resolve the joint influence of extrinsic and intrinsic isolating mechanisms. Specifically, coincidence between the steepness of genomic, geographic, and trait specific clines would indicate

a dominant role of extrinsic factors in facilitating divergence and speciation (Holliday *et al.* 2010; De La Torre *et al.* 2015; Stankowski *et al.* 2015; Ryan *et al.* 2017). Alternatively, the presence of several loci showing steep clines, but lacking climatic or functional associations would indicate a dominance of intrinsic barriers (Ryan *et al.* 2017). Although the genomic cline analysis used in this study provided key insights into the complexity of species isolation, it lacks sufficient power to account for complex epistatic effects (Gompert & Buerkle 2011). These have likely played a key role in ecological speciation and in initiating the evolution of reproductive isolation (Lindtke *et al.* 2012; Flaxman *et al.* 2014). Ultimately, additional genomic resources will help test whether absolute measures of divergence are correlated with recombination rate. This study, however, provides concrete evidence of ecological speciation with gene flow, the presence of a historical hybrid zone maintained by extrinsic factors, and early stages of coupling between extrinsic and intrinsic barriers contributing towards diversification. Whether these patterns hold generally for speciation within conifers, given their life history characteristics as well as their complex and large genomes, is thus a worthwhile area of future research.

Appendix

Table 1.S1. List of variables retained in the final ENMs and their correlation coefficients

	Alt	Bio10	Bio13	Bio14	Bio15	Bio18	Bio19	Bio2	Bio3	Bio4	Bio6	Bio8	Bio9
Alt		-0.48	-0.25	0.38	-0.18	-0.16	0.06	0.44	0.68	-0.34	-0.01	-0.41	0.12
Bio10			0.11	-0.34	0.39	0.07	-0.26	0.32	0.41	-0.33	0.71	0.72	0.51
Bio13				0.21	0.46	0.73	0.42	-0.25	0.34	-0.42	0.37	0.18	0.2
Bio14					-0.61	0.21	0.47	-0.49	-0.32	0.09	-0.19	-0.35	-0.15
Bio15						0.39	-0.18	0.21	0.51	-0.37	0.42	0.53	0.21
Bio18							-0.06	-0.28	0.09	-0.09	0.08	0.38	-0.16
Bio19								-0.31	0	-0.21	0.09	-0.48	0.25
Bio2									0.59	-0.38	0.36	0.18	0.42
Bio3										-0.85	0.83	0.23	0.77
Bio4											-0.84	-0.08	-0.82
Bio6												0.36	0.8
Bio8													0.04
Bio9													

Table 1.S2. Sampling information and mean admixture proportion per population. Column headers “Cluster1” and “Cluster2” refer to ancestry from *P. strobiformis* and *P. flexilis*, respectively.

Species	Longitude	Latitude	cluster 1	cluster 2	N
<i>P. flexilis</i>	-107.655	38.005	0.047	0.953	5
<i>P. flexilis</i>	-106.155	39.106	0.000	1.000	5
<i>P. flexilis</i>	-105.526	40.969	0.000	1.000	5
<i>P. flexilis</i>	-105.658	40.652	0.000	1.000	5
<i>P. flexilis</i>	-104.030	40.813	0.007	0.993	4
<i>P. flexilis</i>	-106.876	37.380	0.347	0.653	5
<i>P. flexilis</i>	-105.588	38.071	0.006	0.994	5
<i>P. flexilis</i>	-105.660	39.933	0.000	1.000	5
<i>P. flexilis</i>	-106.088	41.032	0.251	0.749	5
<i>P. flexilis</i>	-110.817	38.068	0.000	1.000	5
<i>P. flexilis</i>	-105.449	37.736	0.037	0.963	5
<i>P. flexilis</i>	-105.604	38.074	0.262	0.738	5
<i>P. flexilis</i>	-105.434	41.267	0.164	0.836	10
<i>P. strobiformis</i>	-111.674	35.371	0.826	0.174	10

<i>P. strobiformis</i>	-109.964	30.951	0.999	0.001	9
<i>P. strobiformis</i>	-105.944	25.029	0.999	0.001	10
<i>P. strobiformis</i>	-108.901	33.717	0.964	0.036	9
<i>P. strobiformis</i>	-111.674	35.384	0.827	0.173	10
<i>P. strobiformis</i>	-106.815	26.846	1.000	0.000	9
<i>P. strobiformis</i>	-108.213	28.181	1.000	0.000	10
<i>P. strobiformis</i>	-104.717	23.565	1.000	0.000	10
<i>P. strobiformis</i>	-108.643	33.454	0.954	0.046	7
<i>P. strobiformis</i>	-107.510	33.899	0.758	0.242	6
<i>P. strobiformis</i>	-105.711	32.978	0.901	0.099	8
<i>P. strobiformis</i>	-109.772	34.132	0.883	0.117	5
<i>P. strobiformis</i>	-106.401	34.699	0.911	0.089	8
<i>P. strobiformis</i>	-104.954	23.928	0.999	0.001	10
<i>P. strobiformis</i>	-105.635	33.059	0.888	0.112	10
<i>P. strobiformis</i>	-104.830	31.980	0.936	0.064	10
<i>P. strobiformis</i>	-105.733	33.392	0.847	0.153	10
<i>P. strobiformis</i>	-105.103	24.517	1.000	0.000	9
<i>P. strobiformis</i>	-105.727	23.686	1.000	0.000	10

<i>P. strobiformis</i>	-105.641	33.232	0.948	0.052	7
<i>P. strobiformis</i>	-105.792	34.243	0.865	0.135	6
<i>P. strobiformis</i>	-104.706	23.335	1.000	0.000	10
<i>P. strobiformis</i>	-105.598	24.415	1.000	0.000	10
<i>P. strobiformis</i>	-111.856	35.397	0.870	0.130	7
<i>P. strobiformis</i>	-108.896	33.765	0.973	0.027	7
<i>P. strobiformis</i>	-105.638	32.682	0.870	0.130	9
<i>P. strobiformis</i>	-110.386	31.053	0.996	0.004	10
<i>P. strobiformis</i>	-111.243	34.450	0.871	0.129	9
<i>P. strobiformis</i>	-107.181	33.990	0.866	0.134	10
<i>P. strobiformis</i>	-108.143	32.922	0.971	0.029	9
<i>P. strobiformis</i>	-109.275	31.935	0.906	0.094	8
<i>P. strobiformis</i>	-111.624	35.348	0.740	0.260	10
<i>P. strobiformis</i>	-107.505	27.326	1.000	0.000	13
<i>P. strobiformis</i>	-110.783	32.451	1.000	0.000	7
<i>P. strobiformis</i>	-108.096	35.164	0.837	0.163	10
<i>P. strobiformis</i>	-108.196	28.277	0.997	0.003	5
<i>P. strobiformis</i>	-105.813	32.780	0.766	0.234	10

<i>P. strobiformis</i>	-108.026	28.327	0.994	0.006	10
<i>P. strobiformis</i>	-106.356	26.469	0.999	0.001	10
<i>P. strobiformis</i>	-107.486	33.880	0.896	0.104	9
<i>P. strobiformis</i>	-105.811	25.056	0.999	0.001	10
<i>P. strobiformis</i>	-109.018	28.335	1.000	0.000	10

Table 1.S3. Raw and converted parameter estimates, and their 95% confidence intervals (CIs), for the $\partial a \partial i$ model that was best supported by AIC model selection. Results are shown for model M_4 , and parameter estimates are given to three significant digits. Gene flow parameters are given of the form M_{ij} , denoting the per generation number of individuals in population i that originated in population j .

	Raw		Converted		
Parameter	Mean		Mean	Lower 95% CI bound	Upper 95% CI bound
N_{eref}	76.00285		4,123,203.89	3,456,361.429	4,790,046.36
N_{AF}	0.728		3,002,868.87	2,145,236.016	3,860,501.72
N_F	0.295		1,214,643.67	733,450.171	1,695,837.16

\underline{N}_C	0.968		3,991,536.02	2,066,050.429	5,917,021.60
N_P	5.302		21,860,864.99	16,661,087.379	27,060,642.59
M_A	6.908		2.52	0.697	5.214
M_{FP}	4.351		11.53	0.000	57.9479
M_{PF}	8.809		4.26	0.000	12.843
T_1	0.0437		18,037,965.62	9,789,789.953	26,286,141.28
T_2	0.00882		3,634,954.70	2,834,326.862	4,435,582.55

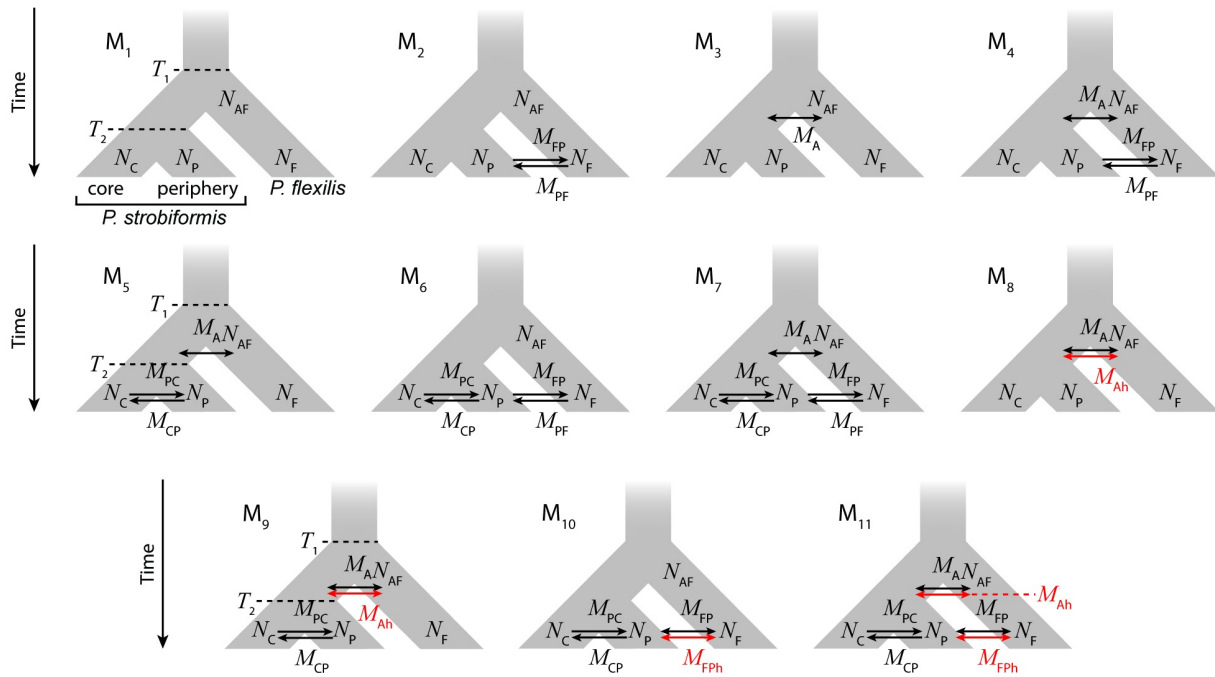


Fig. 1.S1. Schematics and parameter details for each of the 11 demographic models for the divergence of Core and Periphery groups within *P. strobiformis* and *P. flexilis* run in our $\partial A \partial I$ analysis. Parameters include divergence times (T_i), population sizes (N_i), homogeneous rates of gene flow (M_{ij} , gene flow from lineage j to i) and genomically heterogeneous rates of gene flow (M_{ijh}).

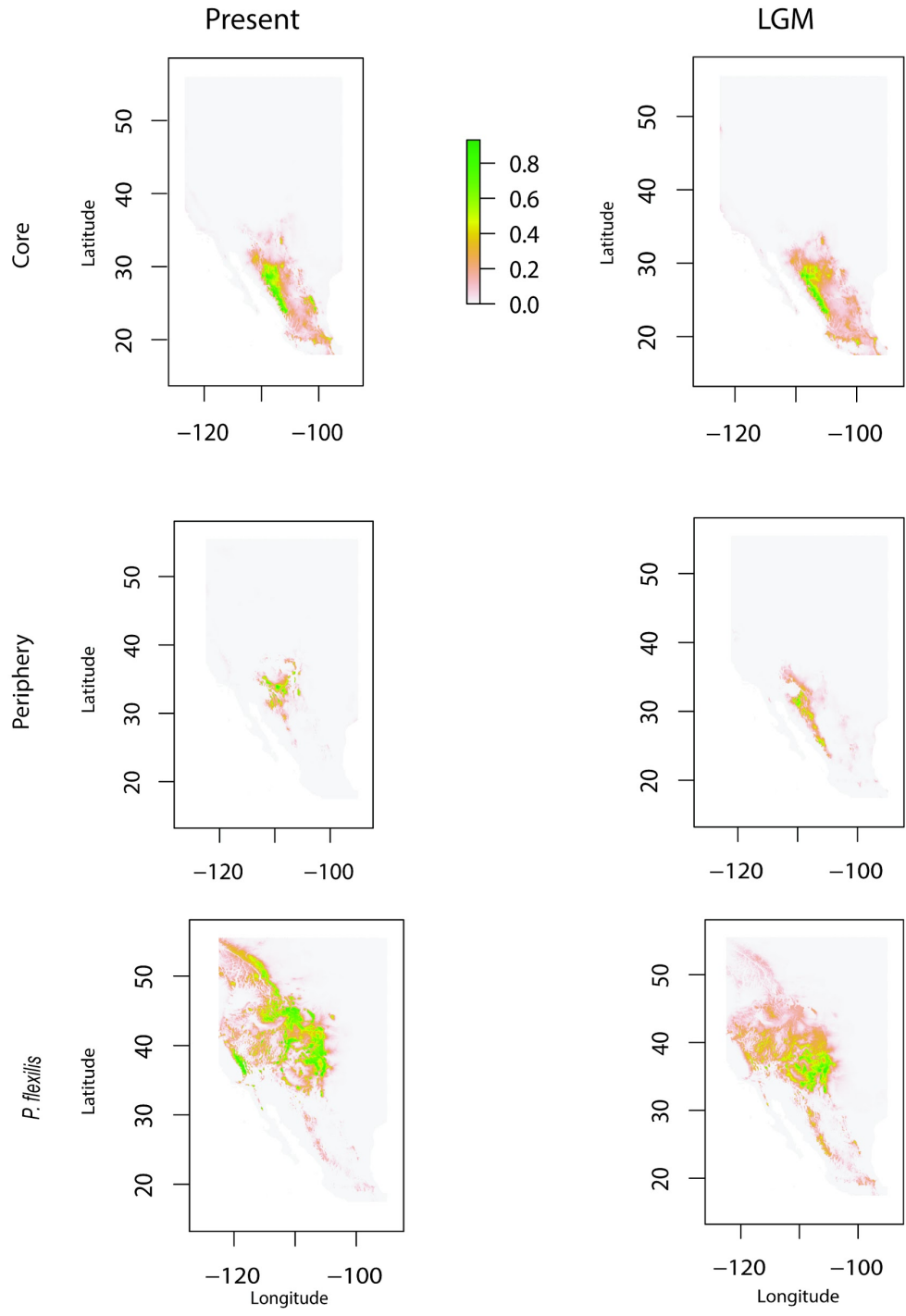


Fig. 1.S2. Ecological niche model projections for Core, Periphery, and *P. flexilis*, under present and past climate

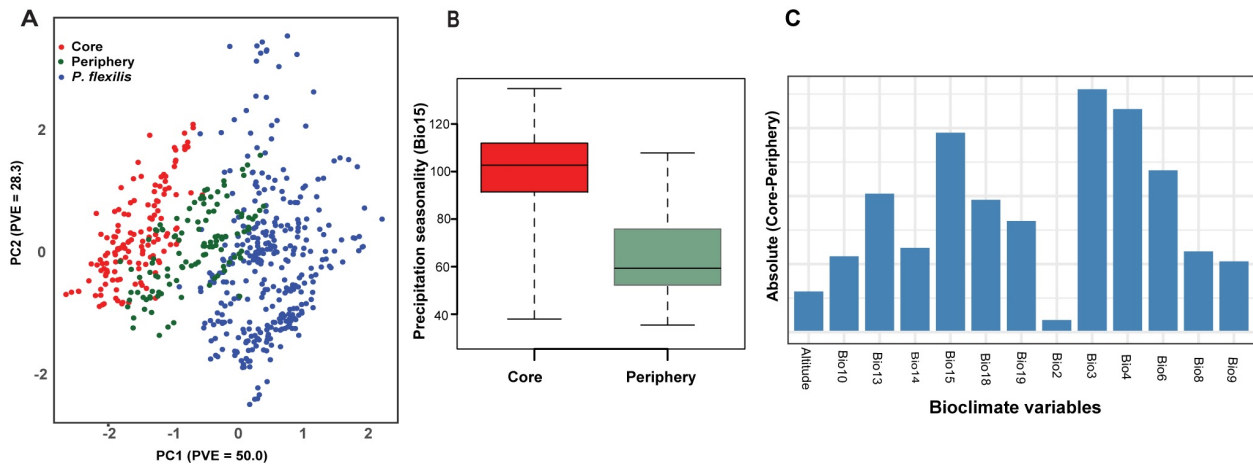


Fig. 1.S3. **A)** Climate PCA with variables used in the ENMs, **B)** Distribution of precipitation seasonality (Bio15) at presence locations of Core & Periphery, and **C)** Difference between Core and Periphery for scaled and centered bioclimatic variables used in the final ENM

REFERENCES

- Abott RJ & Brennan AC (2014). Altitudinal gradients, plant hybrid zones and evolutionary novelty. *Phil. Trans. R. Society B*, **369**, 20130346-20130346.
- Agrawal AF, Feder JL, Nosil P (2011). Ecological Divergence and the Origins of Intrinsic Postmating Isolation with Gene Flow. *International Journal of Ecology*, **2011**, e435357.
- Aguirre-Gutiérrez J, Serna-Chavez HM, Villalobos-Arambula A et al (2015). Similar but not equivalent: ecological niche comparison across closely-related Mexican white pines. *Diversity and Distributions*, **21**, 245-257.
- Aiello-Lammens ME, Boria RA, Radosavljevic A, Vilela B, Anderson RP (2015). spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, **38**, 541–545.
- Allen CD & Breshears DD (1998). Drought-induced shift of a forest–woodland ecotone: Rapid landscape response to climate variation. *Proc. Nat. Acad. Sci.* **95**, 14839–14842.
- Alberto FJ, Aitken SN, Alía R *et al.* (2013). Potential for evolutionary responses to climate change – evidence from tree populations. *Global Change Biology*, **19**, 1645–1661.
- Andrew RL & Rieseberg LH (2013). Divergence is focused on few genomic regions early in speciation: incipient speciation of sunflower ecotypes. *Evolution*, **67**, 2468–2482.
- Barton NH (2001). The role of hybridization in evolution. *Molecular Ecology*, **10**, 551–568.
- Barton NH, De Cara MAR (2009). The Evolution of Strong Reproductive Isolation. *Evolution*, **63**, 1171–1190.
- Barton NH, Hewitt GM (1985). Analysis of Hybrid Zones. *Annual Review of Ecology and Systematics*, **16**, 113–148.
- Benkman CW, Balda RP, Smith CC (1984). Adaptations for Seed Dispersal and the Compromises Due to Seed Predation in Limber Pine. *Ecology*, **65**, 632–642.
- Bierne N, Welch J, Loire E, Bonhomme F, David P (2011). The coupling hypothesis: why genome scans may fail to map local adaptation genes: The coupling hypothesis. *Molecular Ecology*, **20**, 2044–2072.
- Bisbee J (2014). Cone morphology of the *Pinus ayacahuite-flexilis* complex of the southwestern United States and Mexico. *Bulletin of the Cupressus conservation project*, **3**, 3–33.
- Bivand R, Keitt T, Rowlingson B *et al.* (2017). *rgdal: Bindings for the Geospatial Data Abstraction Library*.

- Braunisch V, Coppes J, Arlettaz R, Suchant R, Schmid H, et al. (2013) Selecting from correlated climate variables: a major source of uncertainty for predicting species distributions under climate change. *Ecography* 36:971–983.
- Bridle JR, Vines TH (2007). Limits to evolution at range margins: when and why does adaptation fail? *Trends in Ecology & Evolution*, **22**, 140–147.
- Burnham KP, Anderson DR (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer, New York.
- Buschiazzo E, Ritland C, Bohlmann J, Ritland K (2012). Slow but not low: genomic comparisons reveal slower evolutionary rate and higher dN/dS in conifers compared to angiosperms. *BMC Evolutionary Biology*, **12**, 8.
- Carstens BC, Brennan RS, Chua V *et al.* (2013). Model selection as a tool for phylogeographic inference: an example from the willow *Salix melanopsis*. *Molecular Ecology*, **22**, 4014–4028.
- Christe C, Stölting KN, Paris M *et al.* (2017). Adaptive evolution and segregating load contribute to the genomic landscape of divergence in two tree species connected by episodic gene flow. *Molecular Ecology*, **26**, 59–76.
- Coyne J, Orr HA (1989). Patterns of speciation in *Drosophilla*. *Evolution*, **43**, 362–381.
- Critchfield WB (1986). Hybridization and classification of the white pines (*Pinus* section *Strobus*). *Taxon*, **35**, 647-656.
- Cruickshank TE, Hahn MW (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, **23**, 3133–3157.
- Cunningham CI, Cooke JEK, Coltman DW (2014). Cross-species outlier detection reveals different evolutionary pressures between sister species. *The New Phytologist*, **204**, 215–229.
- Cushman, SA & Landguth, EL (2016). Spatially Heterogeneous Environmental Selection Strengthens Evolution of Reproductively Isolated Populations in a Dobzhansky–Muller System of Hybrid Incompatibility. *Front. Genet.* **7**, 209.
- De Carvalho D, Ingvarsson PK, Joseph J *et al.* (2010). Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula* L.), a widespread forest tree. *Molecular Ecology*, **19**, 1638–1650.
- De La Torre AR, Birol I, Bousquet J *et al.* (2014a). Insights into Conifer Giga-Genomes. *Plant Physiology*, **166**, 1724-1732.
- De La Torre AR, Wang T, Jaquish B, Aitken SN (2014b). Adaptation and exogenous selection in a *Picea glauca* × *Picea engelmannii* hybrid zone: implications for forest management under climate change. *The New Phytologist*, **201**, 687–699.

- De La Torre AR, Ingvarsson PK, & Aitken SN (2015). Genetic architecture and genomic patterns of gene flow between hybridizing species of *Picea*. *Heredity*, **115**, 153–164.
- De La Torre, A. R., Li, Z., Van de Peer, Y., & Ingvarsson, P. K. (2017). Contrasting Rates of Molecular Evolution and Patterns of Selection among Gymnosperms and Flowering Plants. *Molecular Biology and Evolution*, *34*(6), 1363-1377. Chicago
- de Lafontaine G, Prunier J, Gérardi S & Bousquet J. (2015). Tracking the progression of speciation: variable patterns of introgression across the genome provide insights on the species delimitation between progenitor–derivative spruces (*Picea mariana* × *P. rubens*). *Molecular Ecology*, **24**, 5229–5247.
- Dodd RS, Afzal-Rafii Z (2004). Selection and dispersal in a multispecies oak hybrid zone. *Evolution*, **58**, 261–269.
- Feder JL, Egan SP, Nosil P (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics*, **28**, 342–350.
- Fitzpatrick BM (2012). Estimating ancestry and heterozygosity of hybrids using molecular markers. *BMC Evolutionary Biology*, **12**, 131.
- Fitzpatrick BM (2013). Alternative forms for genomic clines. *Ecology and Evolution*, **3**, 1951–1966.
- Flaxman SM, Wacholder AC, Feder JL, Nosil P (2014). Theoretical models of the influence of genomic architecture on the dynamics of speciation. *Molecular Ecology*, **23**, 4074–4088.
- Frankis, M (2009). The high altitude white pines of Mexico and the adjacent SW USA (*Pinus* L. subgenus *Strobus* Lemmon, Pinaceae). International Dendrology Society. Yearbook 63-72.
- Geraldes A, Farzaneh N, Grassa CJ *et al.* (2014). The Landscape genomics of *Populus trichocarpa*: the role of hybridization, limited gene flow, and natural selection in shaping patterns of population structure: population differentiation in *Populus trichocarpa*. *Evolution*, **68**, 3260–3280.
- Gitlin AR, Sthultz CM, Bowker MA *et al* (2006). Mortality Gradients within and among Dominant Plant Populations as Barometers of Ecosystem Change During Extreme Drought. *Conservation Biology*, **20**, 1477–1486.
- Gompert Z, Buerkle CA (2010). introgress: a software package for mapping components of isolation in hybrids. *Molecular Ecology Resources*, **10**, 378–384.
- Gompert Z, Buerkle CA (2011). Bayesian estimation of genomic clines. *Molecular Ecology*, **20**, 2111–2127.
- Gompert Z, Buerkle CA (2012). bgc: Software for Bayesian estimation of genomic clines. *Molecular Ecology Resources*, **12**, 1168–1176.

- Gompert Z, Lucas LK, Buerkle CA *et al.* (2014). Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. *Molecular Ecology*, **23**, 4555–4573.
- Gompert Z, Parchman TL, Buerkle CA (2012a). Genomics of isolation in hybrids. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 439–450.
- Gompert Z, Lauren LK, Chris NC *et al.* (2012b). Genomic regions with a history of divergent selection affect fitness of hybrids between two Butterfly Species. *Evolution*, **66**, 2167–2181.
- Goodrich BA, Waring KM, Kolb TE (2016). Genetic variation in *Pinus strobiformis* growth and drought tolerance from southwestern US populations. *Tree Physiology*, **36**, 1219–1235.
- Goudet J (2005). hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, **5**, 184–186.
- Guerrero RF & Hahn MW (2017). Speciation as a sieve for ancestral polymorphism. *Molecular Ecology*, **26**, 5362–5368.
- 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**, e1000695.
- Hamilton JA, Lexer C, Aitken SN (2013). Genomic and phenotypic architecture of a spruce hybrid zone (*Picea sitchensis* × *P. glauca*). *Molecular Ecology*, **22**, 827–841.
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.
- Hijmans RJ, Eten J van, Cheng J *et al.* (2016). *raster: Geographic Data Analysis and Modeling*.
- Hijmans RJ, Phillips S, Elith JL and J (2017). *dismo: Species Distribution Modeling*.
- Holliday JA, Ritland K, & Aitken SN (2010). Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (*Picea sitchensis*). *The New Phytologist*, **188**, 501–514
- Ikeda DH, Max TL, Allan GJ *et al.* (2017). Genetically informed ecological niche models improve climate change predictions. *Global Change Biology*, **23**, 164–176.
- Janoušek V, Wang L, Luzynski K *et al.* (2012). Genome-wide architecture of reproductive isolation in a naturally occurring hybrid zone between *Mus musculus musculus* and *M. m. domesticus*. *Molecular Ecology*, **21**, 3032–3047.
- Kamdern C, Fouet C, Gamez S, White BJ (2016). Genomic signatures of introgression at late stages of speciation in a malaria mosquito. *bioRxiv*, 068239.

- Kane NC, King MG, Barker MS *et al.* (2009). Compative genomic and population genetic analyses indicate highly porous genomes and high levels of gene flow between divergent *Helianthus* species. *Evolution; international journal of organic evolution*, **63**, 2061–2075.
- Kimball S, Campbell DR, Lessin C (2008). Differential performance of reciprocal hybrids in multiple environments. *Journal of Ecology*, **96**, 1306-1318.
- Knowles LL, Carstens BC, Keat ML (2007). Coupling Genetic and Ecological-Niche Models to Examine How Past Population Distributions Contribute to Divergence. *Current Biology*, **17**, 940–946.
- Kulmuni J, Westram AM (2017). Intrinsic incompatibilities evolving as a by-product of divergent ecological selection: Considering them in empirical studies on divergence with gene flow. *Molecular Ecology*, **26**, 3093–3103.
- Lackey ACR, Boughman JW (2016). Evolution of reproductive isolation in stickleback fish: Evolution of isolation in stickelback fish. *Evolution*, **71**, 357–372.
- Lexer C, Joseph JA, Loo MV *et al.* (2010). Genomic Admixture Analysis in European *Populus* spp. Reveals Unexpected Patterns of Reproductive Isolation and Mating. *Genetics*, **186**, 699–712.
- Li X, Zhu C, Yeh CT *et al.* (2012). Genic and nongenetic contributions to natural variation of quantitative traits in maize. *Genome Research*, **22**, 2436–2444.
- Lindtke D, Buerkle CA, Barbará T, *et al.* (2012). Recombinant hybrids retain heterozygosity at many loci: new insights into the genomics of reproductive isolation in *Populus*. *Molecular Ecology*, **21**, 5042–5058.
- Lindtke D, Gompert Z, Lexer C, Buerkle CA (2014). Unexpected ancestry of *Populus* seedlings from a hybrid zone implies a large role for postzygotic selection in the maintenance of species. *Molecular Ecology*, **23**, 4316–4330.
- Lind BM, Freidline CJ, Wegrzyn JL *et al.* (2017). Water availability drives signatures of local adaptation in whitebark pine (*Pinus albicaulis* Englm.) across fine spatial scales of the Lake Tahoe Basin, USA. *Molecular Ecology*, **26**, 3168-3185.
- Looney CE, Waring KM (2013). *Pinus strobiformis* (southwestern white pine) stand dynamics, regeneration, and disturbance ecology: A review. *Forest Ecology and Management*, **287**, 90–102.
- Looney CE, Waring KM, Fairweather ML (2015). Monitoring the health of *Pinus strobiformis*: early impacts of white pine blister rust invasion. In: Forest Health Monitoring: National Status, Trends, and Analysis. USDA Forest service, pp:167-175.
- Losos JB, Arnold SJ, Bejerano G *et al.* (2013). Evolutionary Biology for the 21st Century. *PLoS Biology*, **11**.
- Lotterhos KE & Whitlock MC (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology*, **24**, 1031-1046.

- Marques DA, Lucek K, Haesler MP *et al.* (2017). Genomic landscape of early ecological speciation initiated by selection on nuptial colour. *Molecular Ecology*, **26**, 7–24.
- Mastretta-Yanes A, Moreno-Letelier A, Piñero D, Jorgensen TH, Emerson BC (2015). Biodiversity in the Mexican highlands and the interaction of geology, geography and climate within the Trans-Mexican Volcanic Belt. *Journal of Biogeography*, **42**, 1586–1600.
- McVean G (2009). A Genealogical Interpretation of Principal Components Analysis. *PLOS Genetics*, **5**, e100068
- Milne RI, Terzioglu S, Abbott RJ (2003). A hybrid zone dominated by fertile F1s: maintenance of species barriers in *Rhododendron*. *Molecular Ecology*, **12**, 2719–2729.
- Mishra AK, Singh VP (2010). A review of drought concepts. *Journal of Hydrology*, **391**, 202–216.
- Moore WS (1977) An Evaluation of Narrow Hybrid Zones in Vertebrates. *The Quarterly Review of Biology*, **52**, 263–277.
- Moreno-Letelier A & Piñero D (2009). Phylogeographic structure of *Pinus strobiformis* Engelm. across the Chihuahuan Desert filter-barrier. *Journal of Biogeography*, **36**, 121–131.
- Moreno-Letelier A, Barraclough TG (2015). Mosaic genetic differentiation along environmental and geographic gradients indicate divergent selection in a white pine species complex. *Evolutionary Ecology*, **29**, 733–748.
- Moreno-Letelier A, Ortíz-Medrano A, Piñero D (2013). Niche Divergence versus Neutral Processes: Combined Environmental and Genetic Analyses Identify Contrasting Patterns of Differentiation in Recently Diverged Pine Species. *PLoS ONE*, **8**, e78228.
- Noor MF, Bennett SM (2009). Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*, **103**, 439–444.
- Nosil P (2008). Speciation with gene flow could be common. *Molecular Ecology*, **17**, 2103–2106.
- Nosil P, Feder JL (2012). Genomic divergence during speciation: causes and consequences. *Phil. Trans. R. Soc. B*, **367**, 332–342.
- Nosil P, Harmon LJ, Seehausen O (2009). Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*, **24**, 145–156.
- Ortego J, Gugger PF, Riordan EC, Sork VL (2014). Influence of climatic niche suitability and geographical overlap on hybridization patterns among southern Californian oaks (B Emerson, Ed.). *Journal of Biogeography*, **41**, 1895–1908.
- Ortego J, Noguerales V, Gugger PF, Sork VL (2015). Evolutionary and demographic history of the Californian scrub white oak species complex: an integrative approach. *Molecular Ecology*, **24**, 6188–6208.

- Parchman TL, Gompert Z, Braun MJ *et al.* (2013). The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. *Molecular Ecology*, **22**, 3304–3317.
- Parchman TL, Gompert Z, Mudge J *et al.* (2012). Genome-wide association genetics of an adaptive trait in lodgepole pine: Association mapping of serotiny. *Molecular Ecology*, **21**, 2991–3005.
- Patterson N, Price AL, Reich D (2006). Population Structure and Eigenanalysis. *PLOS Genetics*, **2**, e190
- Pavy N, Pelgas B, Laroche J, Rigault P, Isabel N, & Bousquet, J. (2012). A spruce gene map infers ancient plant genome reshuffling and subsequent slow evolution in the gymnosperm lineage leading to extant conifers. *BMC Biology*, **10**, 84.
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012). Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLOS ONE*, **7**, e37135.
- Petit RJ, Excoffier L (2009). Gene flow and species delimitation. *Trends in Ecology & Evolution*, **24**, 386–393.
- Petit RJ, Hampe A (2006). Some Evolutionary Consequences of Being a Tree. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 187–214.
- Phillips SJ, Anderson RP, Schapire RE (2006). Maximum entropy modeling of species geographic distributions. *MAXENT v3.4.1. Ecological Modelling*, **190**, 231–259.
- Pritchard JK, Rienzo AD (2010). Adaptation – not by sweeps alone. *Nature Reviews Genetics*, **11**, 665–667.
- Puritz JB, Hollenbeck CM, Gold JR (2014). *dDocent* : a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ*, **2**, e431.
- R Core Team (2017). R v.3.3.2: A Language and Environment for Statistical Computing. Vienna, Austria : the R Foundation for Statistical Computing. *R Foundation for Statistical Computing, Vienna, Austria*.
- Raj A, Stephens M, Pritchard JK (2014). fastSTRUCTURE v.1.0: Variational Inference of Population Structure in Large SNP Data Sets. *Genetics*, **197**.
- Rajora OP, Eckert AJ, Zinck JWR (2016). Single-Locus versus Multilocus Patterns of Local Adaptation to Climate in Eastern White Pine (*Pinus strobus*, Pinaceae). *PLOS ONE*, **11**, e0158691.
- Rambaut, Suchard, Marc, Drummond, Alexei (2013). Tracer, Version 1.6.
- Rehfeldt GE (1999). Systematics and genetic structure of Ponderosae taxa (Pinaceae) inhabiting the mountain islands of the Southwest. *Am J Bot*, **86**, 741-752.
- Restaino CM, Peterson DL, Littell J (2016). Increased water deficit decreases Douglas fir growth throughout western US forests. *Proceedings of the National academy of Sciences*, **113**, 9557-9562.

- Riquet F, Haag-Liautard C, Woodall L *et al.* (2017). Parallel use of a shared genomic island of speciation in clinal and mosaic hybrid zones between cryptic seahorse lineages. *bioRxiv*, 161786
- Roesti M, Hendry AP, Salzburger W, Berner D (2012). Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Molecular Ecology*, **21**, 2852–2862.
- Rundle HD (2002). A test of ecologically dependent postmating isolation between sympatric sticklebacks. *Evolution*, **56**, 322–329.
- Ryan S, Fontaine MC, Scriber JM *et al.* (2017). Patterns of divergence across the geographic and genomic landscape of a butterfly hybrid zone associated with a climatic gradient. *bioRxiv*, 149179.
- Savolainen O, Lascoux M, Merilä J (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, **14**, 807–820.
- Schild DR, Adams RH, Card DC *et al.* (2017). Insight into the roles of selection in speciation from genomic patterns of divergence and introgression in secondary contact in venomous rattlesnakes. *Ecology and Evolution*, **7**, 3951–3966.
- Schluter D, Conte GL (2009). Genetics and ecological speciation. *Proceedings of the National Academy of Sciences*, **106**, 9955–9962.
- Shirk AJ, Cushman SA, Wehenkel CA *et al.* (2017). Southwestern white pine (*Pinus strobiformis*) species distribution models predict large range shift and contraction due to climate change. *Forest Ecology & Management (in review)*.
- Sork VL, Aitken SN, Dyer RJ *et al.* (2013). Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics & Genomes*, **9**, 901–911.
- Stacy EA, Paritosh B, Johnson MA, Price DK (2017). Incipient ecological speciation between successional varieties of a dominant tree involves intrinsic postzygotic isolating barriers. *Ecology and Evolution*, **7**, 2501–2512.
- Stankowski S, Sobel JM, Streisfeld MA (2015). The geography of divergence with gene flow facilitates multitrait adaptation and the evolution of pollinator isolation in *Mimulus aurantiacus*. *Evolution*, **69**, 3054–3068.
- Steinhoff RJ, Andresen JW (1971). Geographic variation in *Pinus flexilis* and *Pinus strobiformis* and its bearing on their taxonomic status. *Silvae Genet*, **20**, 159–167.
- Tine M, Kuhl H, Gagnaire PA *et al.* (2014). European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature Communications*, **5**.

- Tomback DF, Samano S, Pruett EL, Schoettle AW *et al.* (2011). Seed dispersal in limber and southwestern white pine: comparing core and peripheral populations. In: *The Future of High-Elevation, Five-Needle White Pines in Western North America: Proceedings of the High Five Symposium. Proceedings RMRS-P-63. Fort Collins, CO: US Department of Agriculture, Forest Service, Rocky Mountain Research Station*, pp. 69–71.
- Via S, Bouck AC, Skillman S (2000). Reproductive isolation between divergent races of pea aphids on two hosts. II. Selection against migrants and hybrids in the parental environments. *Evolution*, **54**, 1626–1637.
- Warren DL, Glor RE, Turelli M (2008). Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution*, **62**, 2868–2883.
- Williams AP, Allen CD, Millar CI, Swetnam TW *et al.* (2010). Forest responses to increasing aridity and warmth in the southwestern United States. *Proceedings of the National Academy of Sciences*. **107**, 21289-21294.
- Wu CI (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–865.
- Yang RC (1998). Estimating hierarchical *F*-statistics. *Evolution*, **52**, 950-956.
- Yeaman S, Otto SP (2011). Establishment and Maintenance of Adaptive Genetic Divergence Under Migration, Selection, and Drift. *Evolution*, **65**, 2123–2129.
- Yeaman S, Aeschbacher S, Bürger R (2016). The evolution of genomic islands by increased establishment probability of linked alleles. *Molecular Ecology*, **25**, 2542–2558.
- Zhou Y, Zhang L, Liu J, Wu G, Savolainen O (2014). Climatic adaptation and ecological divergence between two closely related pine species in Southeast China. *Molecular Ecology*, **23**, 3504–3522.

CHAPTER 2

Adaptive evolution in a conifer hybrid zone is driven by a mosaic of introgressed and standing genetic variants

INTRODUCTION

Despite growing evidence for hybridization across The Tree of Life, the evolutionary outcome of hybridization and introgression remains contentious due to the complex interplay between intrinsic and extrinsic selection pressures (Abbott, 2013; de Lafontaine & Bousquet, 2017; Todesco et al. 2016; Anderson & Stebbins, 1954). Given that conifers exhibit high levels of evolutionary conservation (de La Torre et al. 2017), weak reproductive isolating barriers (Critchfield 1986), large effective population sizes (N_e) (Charlesworth B, 2009; Bouille & Bousquet, 2005) and high fecundities; hybridization among conifers may less likely be maladaptive and could offer a mosaic of genomic variants to aid rapid evolution (Hamilton et al. 2013; Bresadola et al. 2019; Suarez-Gonzalez et al. 2016). Although investigations into adaptive introgression have been conducted in non-conifer tree taxa such as *Populus* and *Quercus* (Suarez-Gonzalez et al. 2016, 2018; Leroy et al. 2019), information on species with much larger genomes, such as conifers, is rare (Hufford et al. 2013; Takuno et al. 2015; Ma et al. 2019). Larger genome size is hypothesized to influence the genetic architecture of adaptive traits by minimizing hard sweeps and limiting genic enrichment of adaptive loci (Pyhäjärvi et al. 2013; Mei et al. 2018). These architectures thus typically evolve via subtle and coordinated allele frequency shifts of a large number of loci rather than drastic allele frequency changes at a handful of loci. Introgression can mediate the evolution of these architectures by providing a

mosaic of allelic variants to generate novel allelic combinations, typically not available within the range of either parental species. Therefore, conifers present ideal systems to investigate how large genome sizes and their aforementioned life-history characteristics interact to influence the genetic architecture of adaptive introgression within natural hybrid zones.

Our study focuses on two recently diverged conifer species (Chapter 1 & Menon et al. 2018). *Pinus strobiformis* is a key component of the montane mixed conifer ecosystems ranging from Jalisco in southern Mexico to southern Colorado in the US (Looney & Waring, 2013). Populations at the northern range primarily inhabit fragmented sky-islands and contain *Pinus flexilis*-*P. strobiformis* hybrids (Frankis 2009; Tomback et al. 2011; Bisbee 2014). Despite its fragmented and disjunct distribution, *P. strobiformis* exhibits overall weak population structure (Chapter 1 & Menon et al. 2018), corroborating findings across several gymnosperms with broad geographical distributions (Neal & Kremer, 2011). *Pinus flexilis* also inhabits montane ecosystems but is often seen dominating subalpine and tree line habitats (Schoettle & Rochelle, 2000). *P. flexilis* therefore has a broader ecological amplitude and is differentiated from *P. strobiformis* by occurring in cooler environments (Moreno-Letelier et al. 2013).

For long-lived sessile species such as trees, documenting the architecture of introgressed variants is crucial to assess the rate and mode of response to climate change (Hamilton & Miller, 2016). We intensively sampled the *P. strobiformis*-*P. flexilis* hybrid zone (Fig. 2.1) to address two hypotheses. First, projected increases in seasonality of temperature and precipitation in sky-island ecosystems harboring the hybrid zone populations (Hayhoe et al. 2004; Bell et al. 2004) and ongoing asymmetric gene flow from *P. flexilis* will favour the retention of cold tolerance-related variants from *P. flexilis*. Second, the relative contribution of standing genetic variants towards adaptive evolution will be greater along environmental gradients that are least divergent

between the two parental species and exert stronger selective pressures within the hybrid zone relative to that in the range of pure *P. flexilis*. We addressed these hypotheses using a multifaceted approach hinging on differences in expected genetic architectures of adaptive evolution from standing genetic variants and introgressed variants along several environmental gradients (Table 2.S1), while accounting for genetic drift and neutral introgression. We found strong signals of adaptive introgression along freeze-related environmental gradients, while water availability-related gradients are associated with adaptive evolution from standing genetic variants. Our work adds to a growing literature demonstrating the importance of introgression in assisting species responses to changing climatic conditions via range shifts and adaptive evolution.

METHODS

Sampling and generation of genetic data

We sampled 22 populations (3–8 trees per population) from the Mexican range and parts of New Mexico containing pure *Pinus strobiformis*, 12 populations (4–10 trees per population) from pure *P. flexilis* distributed from northern New Mexico to southern Wyoming and 98 populations (6–10 trees per population) from the *P. strobiformis*–*P. flexilis* hybrid zone (Fig. 2.1). Classification of populations into pure parentals and hybrids was based on findings from chapter 1 (Menon et al. 2018) and further refined here using NGSAdmix (Skotte et al. 2013) (Fig 2.S1). To assess patterns of fine-scale local adaptation within the hybrid zone, the 98 populations were sampled across a gridded design of latitude and longitude with paired high-low elevation sites (*cf.* Lotterhos & Whitlock, 2015).

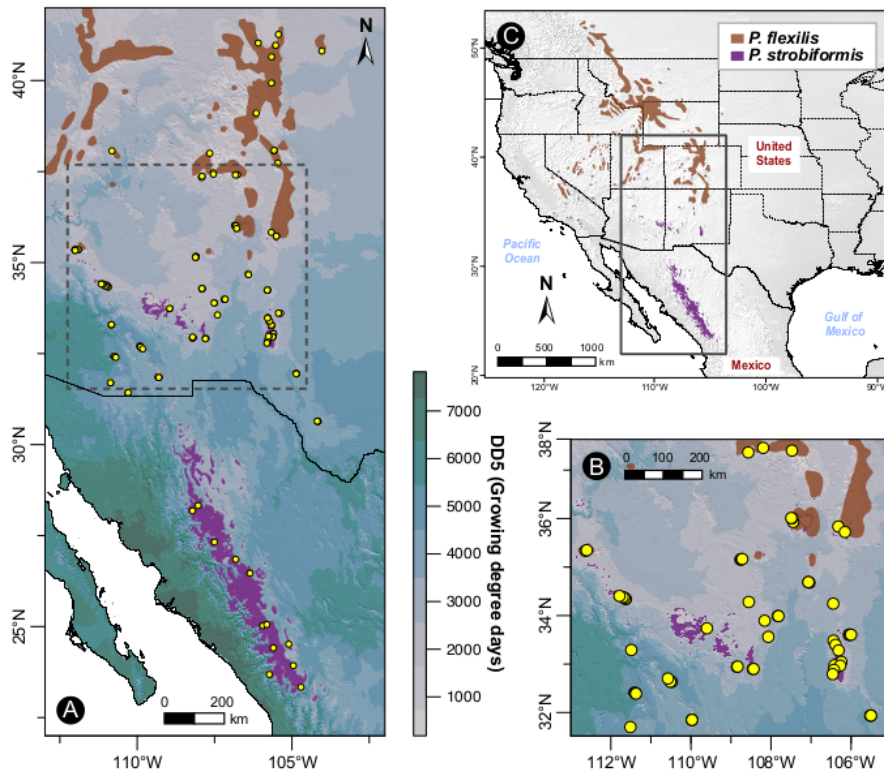


Figure 2.1: (a) Geographical distribution of *Pinus flexilis* and *P. strobiformis* with sampled populations indicated in yellow. The background colour palette represents a raster map of growing degree days (DD5), highlighting one of the environmental gradients of adaptive introgression. (b) Finer scale representation of the 98 hybrid populations used in this study. (c) Location of *P. strobiformis* and *P. flexilis* on the map of North-America, with the study region highlighted by a rectangular box.

We extracted genomic DNA from 1122 trees sampled across 132 populations using the DNeasy Plant Kit (Qiagen). Multiplexed ddRADseq libraries were prepared by pooling 96 trees per library, following the procedure detailed in Parchman et al (2012). Following size selection and isolation of pooled DNA from each library, we performed single-end sequencing of one library per lane (150 bp reads). All sequencing was conducted at Novogene using the Illumina HiSeq 4000 platform. The resulting FASTQ files, one per lane, were processed using dDocent (Puritz et al. 2014) and a series of custom post-filtering steps as conducted in Chapter 1 (Menon

et al. 2018). This process yielded a total of 73,243 SNPs, which were used as the starting dataset for all subsequent analyses.

Given the gridded design used for sampling the hybrid populations, we utilised latitude, longitude and elevation to obtain annual and seasonal climatic variables at 1 -km resolution from ClimateWNA v5.6 (Wang et al. 2016) for the 1981–2010 normal. We also added ten 1-km resolution soil variables from SoilGrids v.0.5.3 (Hengl et al. 2014). Whereas the analyses listed below used all 88 environmental gradients (Table 2.S1), the results presented herein focus only on 12 gradients that were amongst the most and least divergent between the two parental species. The most divergent environmental gradients included beginning of frost free period (bFFP), autumn degree days below zero °C (DD_0_at), spring degree days below 18 °C (DD_18_sp), spring degree days below 5 °C (DD5_sp), frost free period (FFP) and winter precipitation as snow (PAS_wt). The least divergent environmental gradients included annual heat moisture index (AHM), soil cation exchange capacity (CECSOL_s1), autumn degree days above 18 °C (DD18_at), extreme maximum temperature (EXT), spring relative humidity (RH_sp) and autumn maximum temperature (Tmax_at).

Identifying focal loci

To identify focal sets of loci displaying environmental associations within the hybrid zone, we utilised the Bayesian genotype-environment association (GEA) approach implemented in Bayenv2 (Coop et al. 2010; Gunther & Coop, 2013). We pruned our dataset to 72,889 biallelic SNPs from across the hybrid zone. Prior to conducting associations, we accounted for population history by estimating the variance-covariance matrix using 500,000 iterations across three independent Markov chains. Mixing and convergence across Markov chains were visually inspected using trace plots of the determinant of the variance-covariance matrix at every 500

steps (Brooks & Gelman, 1998). After verifying convergence (Fig. 2.S2), we randomly picked a covariance matrix at 250,000th iteration near the plateau to conduct single SNP based association analyses on the scaled and centered 88 environmental gradients. To ensure convergence during the association phase of the analyses, we ran three independent Markov chains each with 100,000 iterations. We utilised two levels of intersection to identify the most stringent set of outlier SNPs per environmental gradient. First, for each Markov chain, SNPs were classified as outliers if they fell outside the 99th percentile of both Bayes factor (BF) and Spearman's correlation coefficient (ρ). Next, outlier SNPs identified across all three chains per gradient were intersected to obtain the final set of strongly associated SNPs. For these SNPs, we estimated the median BF (\widetilde{BF}) across all three chains. To further understand genetic architectures of these outlier SNPs, we estimated multilocus F_{ST} , multilocus F_{CT} and median LD. We determined whether the focal sets fell outside the 95th percentile of the bootstrapped distribution generated using equal numbers of putatively neutral SNPs per environmental gradient. Bootstrapped sets were matched in two-dimensional bins based on the observed values of minor allele frequencies and proportions of missing data for each observed focal set. Differentiation measures were obtained through the hierarchical model implemented in the HIERFSTAT package v.0.04-22 (Goudet J, 2005) and LD was measured as the squared pairwise correlation coefficient (r^2) obtained through the genetics package v.1.3.8.1 (Warnes et al. 2013) in R v.3.3.2 (R core team, 2017).

Potential confounding influences of introgression

Given ongoing gene flow between *P. flexilis* and hybrid zone populations as identified in Chapter 1 (Menon et al. 2018), it is likely that the focal SNP sets detected above were products of both neutral and adaptive introgression from *P. flexilis*. To test this expectation, we used

Redundancy analysis (RDA) as implemented in the *vegan* package v.2.5.2 (Oksanen et al. 2013) in R. When compared with *Bayenv2*, RDA is able to control for multiple confounding factors. Hellinger-transformed allele frequency estimates for each of the 98 populations were used as the response matrix, while the predictor matrices included an environmental matrix, ancestry matrix, population structure matrix and a geographical matrix. For the environmental matrix, we conducted a principal components analysis on the scaled and centered environmental gradients and retained the top seven gradients as they explained >90% of variance in the dataset (Table 2.S2). The mean *Q*-score per population as estimated through *NGSAdmix* was used in the ancestry matrix. The first eigenvector of the covariance matrix obtained in *Bayenv2* (see above) was used to account for population structure. For the geographical matrix, we used scaled and centered spatial transformations of latitude, longitude and elevation (Liu Q, 1997). Overall, transformed SNP allele frequencies were modelled as a linear function of the predictor matrices and the significance of each fitted model was assessed using 9999 permutations. The *varpart* function in *vegan* was used to estimate proportions of variance in the genetic dataset explained by various combinations of the predictor matrices (Fig. 2.S3). We utilised the approach listed in Liu (1997) to estimate pure and confounded effects of the predictors on the response matrix. Since the primary objective of our study was to disentangle signatures of adaptive evolution from introgressed vs. standing genetic variants, we compared two models within RDA to assess the extent to which genetic variation was confounded between ancestry and environment. Model 1 contained the joint effect and the interaction effect of environment and ancestry, while model 2 only contained joint effects. Both models were conditioned on population structure and geography. If model 1 provided a significantly better fit to the data, it would indicate a confounding influence of environment and ancestry on outliers identified through *Bayenv*.

Identifying environmental gradients of adaptive introgression and local adaptation

Given the prevalence of polygenic architectures underlying local adaptation across non-domesticated species, selection should facilitate a buildup of covariance in allele frequencies across loci, manifested as linkage disequilibrium (LD), contributing to fitness differences (Kremer & Le Corre, 2012; Lind et al. 2017). Besides selection on standing genetic variants, elevated LD within hybrid zones could result from recent introgression and selection on introgressing variants. In the present study, *P. flexilis* inhabits areas experiencing cooler temperatures relative to *P. strobiformis* (Table 2.S1); thus, we expect the freeze-related Bayenv SNPs to exhibit elevated LD if they were adaptively introgressed from *P. flexilis*. This should hold true even after accounting for the expected on-average larger background LD among putatively neutral loci experiencing ongoing introgression. We utilised two LD-based approaches to test this hypothesis.

First, we conducted linkage disequilibrium based network analysis (LDna *sensu* Kempainen et al. 2015) to identify distinct clusters of SNPs exhibiting strong associations amongst themselves. Within the *P. strobiformis*–*P. flexilis* hybrid zone, LD clusters could arise due to positive selection on standing variants or due to positive selection on introgressed variants. To enhance our ability to detect these clusters, and to account for false positives due to co-variation in genomic ancestry with environmental gradients, we generated 100 matrices of pairwise LD values. For each matrix, we used all outliers from Bayenv and randomly generated an equal number of putatively neutral SNPs that were matched in minor allele frequency bins. Within LDna, the stringency of outlier cluster (OC) cutoff depends on Φ and E_{min} (Kempainen et al. 2015). Using a hierarchical tree constructed with Φ and E_{min} , the change in median LD among SNPs within a cluster before and after merger is given by λ and OCs are identified by

large λ values above the stringency cutoff. To determine the appropriate value of E_{min} and Φ we used a subset of the 100 matrices to generate a series of trees with values of E_{min} ranging from 4 to 10 but holding Φ constant at 2. We were consistently able to recover similar OCs across varying values of E_{min} , hence we decided to utilize an E_{min} of 9. Our choice of parameters is justified given the large genome sizes of conifer species, low coverage obtained through ddRAD-seq and on average rapid decay of LD. Across all matrices, we determined the proportion of times that Bayenv outliers were present in an OC and the environmental gradient that they were associated with, as detected in Bayenv. Specifically, we assessed how often sets of three to six Bayenv outliers were shared in an OC across replicate runs, the environmental gradient these sets were associated with, and the median LD across OCs.

While the LDna approach could detect evolutionary processes driving sets of SNPs to covary in their allele frequencies, it does not explicitly account for spatial heterogeneity in selection pressures. Variance partitioning of LD implemented using Ohta's D fills this gap by dissecting the geographical and environmental bases associated with LD among loci. Partitioning LD into within- and among-population components (D_{IS} & D_{ST} ; *sensu* Ohta 1982) can help understand the relative importance of selection and drift (Lind et al. 2017; Csillery et al. 2014). We extended this approach by utilising components of LD in a multiple matrix regression approach to distinguish selection on standing variants from selection on introgressed variants. First, we utilised the 88 outlier sets identified from Bayenv to partition LD among pairs of 98 populations using OhtaDstat v.2.0 (Beissinger et al. 2015) package in R. For each environmental gradient and each pairwise comparison, we estimated median D_{IS} & D_{ST} , which were then treated as the response matrices. Next, we obtained two predictor matrices: (a) pairwise geographical distances using the Vincenty ellipsoid formula implemented in geosphere v.1.5.7 (Hijmans,

2017) package in R and (b) pairwise absolute differences in the respective environmental gradients. Overall, for each set of outliers (88 sets, one per gradient) and both estimates of LD components (D_{ST} & D_{IS}), we conducted three matrix regressions using the `lgrMMRR` function within PopGenReport v.3.0.4 (Adamack & Gruber, 2014) package in R. These regressions included the total effect of environment and geography, the pure effect of environment and the pure effect of geography. The pure and confounded effects were calculated as in RDA. Using this approach, we were able to assess the contribution of environment, geography and the confounded effects towards the spatial partitioning of LD. Since the direction of introgression correlates with freezing temperatures, we expect the pure effect of environment on freeze-related Bayenv outliers to have a high predictive ability for D_{IS} . For environmental gradients that did not differentiate *P. strobiformis* and *P. flexilis* (Table 2.S1), but still likely impart strong selection within the hybrid zone, we expected the environment to have a high predictive ability for D_{ST} .

Genomic cline analyses and candidates for adaptive introgression

We used the genomic cline approach implemented in INTROGRESS v.1.2.3 (Gompert & Buerkle, 2009) to predict the parental genotypic probability of a marker in a hybrid individual as a function of genome-wide ancestry. As parental populations did not exhibit fixed differences at assayed SNPs, we utilised the parametric approach to identify SNPs exhibiting exceptional patterns of introgression (Gompert & Buerkle, 2009). To account for high false positives associated with the parametric approach, we only used SNPs that passed the Bonferroni corrected p -value threshold for displaying exceptional patterns of introgression (Janousek et al. 2012). These were subjected to two further filtering steps to declare a SNP as being significantly introgressed from *P. flexilis*: (a) the fitted estimate for the *P. flexilis*-like genotype for a tree should lie outside the upper 95% confidence interval obtained from neutral simulations and (b)

the SNP should be significantly introgressed across at least 20% of the trees. The choice of 20% was based on prior analyses showing that patterns of introgression were not sensitive to individual based cutoffs, that were evaluated using cutoffs in the range of 10% to 50%.

To identify candidates for adaptive introgression, we conducted an enrichment analysis for each of the 88 outlier sets. Specifically, we asked whether the Bayenv outlier SNPs for each environmental gradient (i) were overrepresented in the set of SNPs exhibiting significant introgression from *P. flexilis* using the following equation (following Hancock et al. 2011):

$$FE_{\text{env}(i)} = \frac{B_{pf}}{B} \bigg/ \frac{S_{pf}}{S}$$

where B_{pf} indicates the number of outliers identified through Bayenv that are also significantly introgressed from *P. flexilis*, B indicates the total number of Bayenv outliers, S_{pf} is the number of SNPs that are significantly introgressed from *P. flexilis* and S is the total number of SNPs used in INTROGRESS. Statistical significance of the observed enrichment for each of the 88 outlier sets was determined by running 10,000 null permutations of association between a representative number of randomly sampled SNPs classified as Bayenv outliers and exceptionally introgressed. This approach helped avoid false signals of adaptive introgression due to the latitudinal gradient of ancestry (Fig. 2.S1) covarying with several environmental gradients.

RESULTS

Environmental differences between parental species structure adaptive genetic variation within the hybrid zone

We identified strong association of environmental gradients with allele frequencies across the hybrid zone, with freeze and water availability driving most of the noted adaptive genetic

differentiation. Using Bayenv, 500 unique SNPs were associated with 88 environmental gradients (Table 2.S3). Strongly correlated gradients exhibited considerable, but incomplete overlap (max = 49%) in sets of associated SNPs. The number of outlier SNPs and the strength of association varied across environmental gradients, with a general trend of freeze-related and water availability-related gradients dominating sets of outliers. Using the intersection of SNPs outside the 99th percentile of both BF and $|\rho|$ across three independent runs of Bayenv, we identified a maximum of 45 (PAS_wt) and a minimum of 5 (PAS_sm) outlier SNPs. The \widetilde{BF} values for the outliers ranged from $1.48e+08$ (CMD_sm) to 0.84 (RAD_sm) (Fig. 2.2a & b and Table 2.S3). Overall, 11 SNPs had \widetilde{BF} values at or below 1 but none of them were associated with freeze-related gradients. Estimates of multilocus F_{ST} and F_{CT} were significantly greater than sets of random SNPs ($p < 0.05$) for 72 and 60 environmental gradients (Table 2.S3) with the average being 3 and 4 times larger than the global estimates. Median LD (measured as r^2) was significantly greater ($p < 0.05$) for 83 environmental gradients (Table 2.S3). Freeze and water availability-related environmental gradients had the highest point estimates for median LD, whereas only freeze-related gradients were ranked the highest for multilocus F_{ST} and F_{CT} values (Table 2.S3). In general, estimates of F_{ST} , F_{CT} and r^2 were larger for loci associated with environmental gradients most divergent between *P. flexilis* and *P. strobiformis* relative to those that were least divergent (Fig. 2.2c & d).

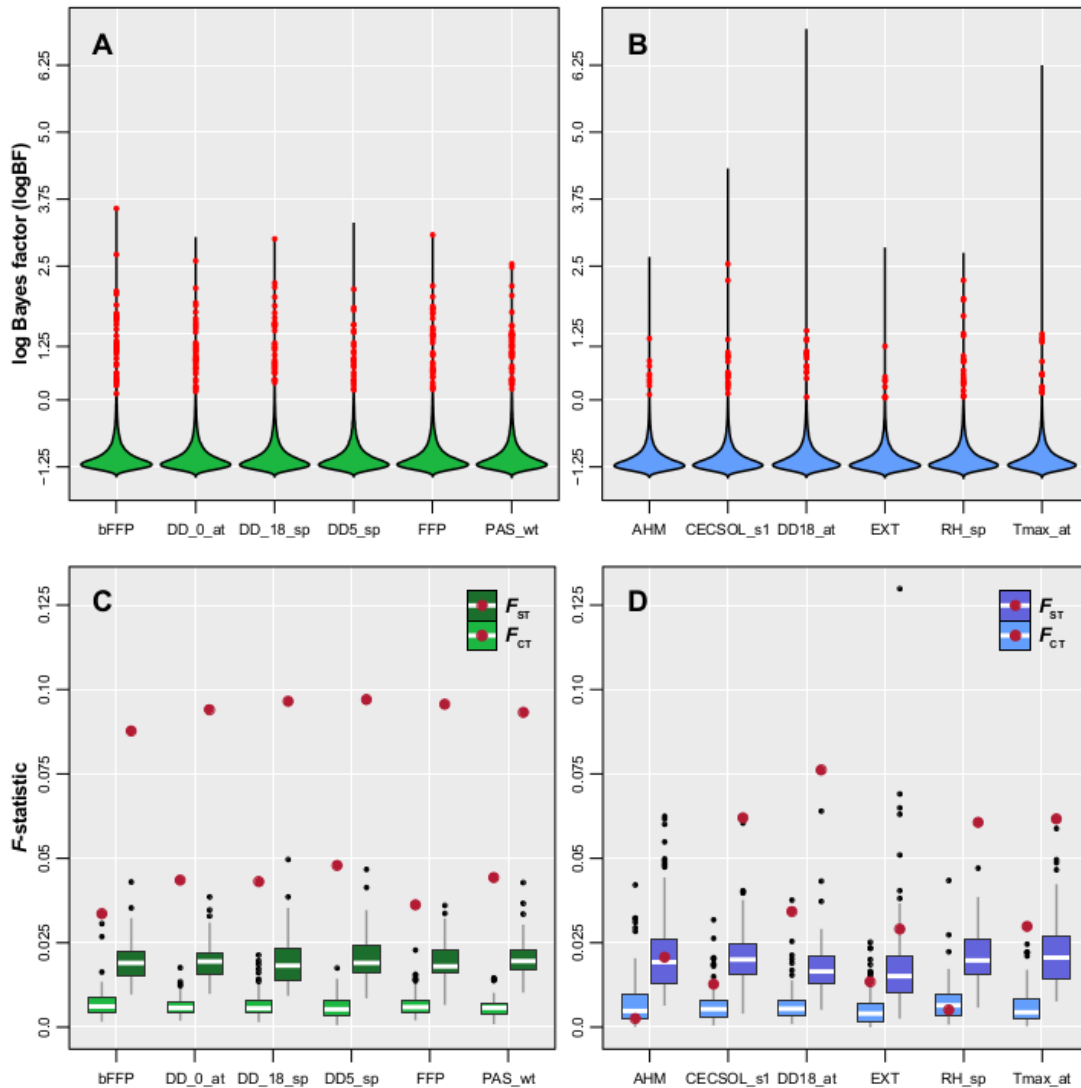


Figure 2.2: Distribution of median Bayes factor ($\log BF$) values for (a) environmental gradients that strongly differentiate *P. flexilis* and *P. strobiformis* and (b) environmental gradients that least differentiate the two species. Distribution of multilocus F -statistics (F_{ST} and F_{CT}) for (c) environmental gradients that strongly distinguish *P. flexilis* and *P. strobiformis* and (d) environmental gradients that are least different between the two species. Red dots in the violin plots of (a) and (b) indicate outlier SNPs. Red dots in the F -statistics boxplots indicate the observed multilocus estimates for Bayenv outlier SNPs. Details of the environmental gradients are represented in Table 2.S1.

Confounding influence of ancestry on signals of adaptive evolution

Estimates of ancestry were confounded with environment, as well as indirectly with other predictors through their relationship with environment, in explaining the observed genotypic variance. The full model within RDA which included environment, ancestry, population structure and geography explained a small ($R^2_{\text{adj}} = 0.027$, $F_{19,78} = 1.14$) yet significant ($p < 0.001$) amount of overall genotypic variance. Low R^2_{adj} could have resulted from weak population structure and the use of several putatively neutral genome wide SNPs (Harrison et al. 2017). Of the series of partial models fitted within RDA, most were significant and the variance explained by the model conditioned on population structure (2.5%) closely followed the variance explained by the full model (Table 2.1), reiterating the prevalence of weak population structure in conifers. By implementing a sequential variance partitioning approach (Fig. 2.S3), we quantified the independent and joint ability of various predictors to explain the genotypic variance. The pure effects together accounted for 56% of the total explained variance from the full model, while 43% was confounded in some way among them. Of the pure effects, geography had the largest contribution (33%) to the total variance while ancestry had the smallest (1%). Total confounded variance due to the interaction with all other predictors was highest for environment (37%) and lowest for population structure (1%). The amount of variance confounded between two predictors was highest for environment and geography (17%) and lowest for any combination including population structure. A formal model comparison through RDA reiterated that environment interacts with ancestry to best explain overall genotypic variance (Model 1: Environment + Ancestry + Environment * Ancestry, Model 2: Environment + Ancestry, $F_{71,78} = 1.04$, $p = 0.05$). The interaction effect model had an R^2_{adj} 1.6 fold larger than the model without it (Model 1: $R^2_{\text{adj}} = 0.014$, Model 2: $R^2_{\text{adj}} = 0.009$), indicating that outlier SNPs identified through

Bayenv were likely confounded with spatial variation in ancestry due to hybridization and ecological differentiation between the two species.

Table 2.1: Variance partitioning, model R^2 and significance of multivariate models fitted using redundancy analyses.

Predictors	R^2	R^2_{adj}	p -value
All	0.217	0.027	0.0001
Env	0.086	0.014	0.0001
Geo	0.12	0.020	0.0001
Ancestry	0.018	0.007	0.0001
PopStr	0.012	0.002	0.0001
Env X	0.07	0.004	0.01
Geo X	0.108	0.009	0.0002
Ancestry X	0.01	0.0004	0.366
PopStr X	0.011	0.002	0.01
Env+Geo X	0.116	0.018	0.0001
Env+Ancestry X	0.089	0.005	0.001
Env+PopStr X	0.085	0.006	0.0007
Env+Geo+Ancestry PopStr	0.133	0.025	0.0001
Env+Geo+PopStr Ancestry	0.126	0.019	0.0001
Env+Ancestry+PopStr Geo	0.104	0.007	0.0004

**X indicates all other matrices are partitioned out. Env stands for Environment, PopStr stands for population structure and Geo stands for geography.*

Adaptive variants are due to selection on introgressed and standing variants

LDna and LD variance partitioning revealed that Bayenv outliers were a product of selection on introgressed and standing genetic variants. SNPs associated with environmental gradients differentiating *P. flexilis* and *P. strobiformis* were overrepresented in the OCs identified through LDna (Table 2.S4). These OCs also tended to have a larger median LD relative to OCs containing SNPs associated with environmental gradients least divergent between the two species. Across replicate sets of LD matrices that were used to account for random SNP associations due to neutral introgression, the percentage of time sets of 3 to 6 Bayenv SNPs were present in an OC ranged from 74% to 65. While SNPs associated with freezing temperatures or with environmental gradients most divergent between the two species were dominant in these sets (Fig. 2.3a), very rarely were SNPs associated with any of the least divergent environmental gradients present together in an OC (Fig. 2.3b). Since our approach utilised neutral sets of SNPs in combination with outlier SNPs for each replicate, the OCs containing high median LD are less likely to be false positives due to spatially varying patterns of neutral introgression. We noted a few OCs containing a combination of freeze and water availability-related Bayenv SNPs; however, OCs containing a majority of SNPs associated with freeze-related gradients had the highest median LD across all replicate runs (Fig. 2.3c).

Partitioning of LD into among- and within-population components (D_{ST} & D_{IS}) and their association with climatic and geographical distance revealed the granularity of selection pressure across the hybrid zone and demonstrated a strong confounding between geography and freeze-related environmental gradients. Adaptive evolution should increase D_{ST} for loci associated with the environmental gradients differentiating hybrid zone populations due to locally divergent selection pressures.

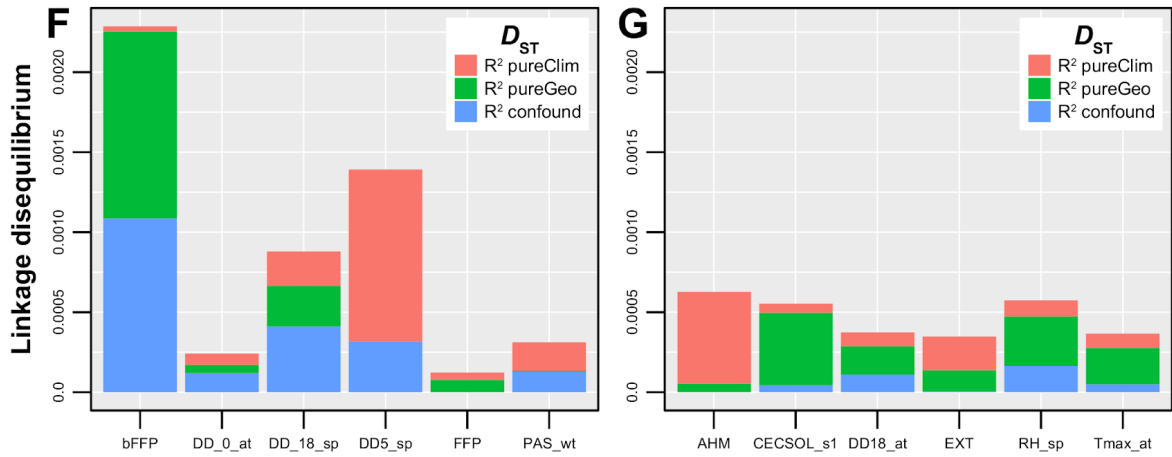
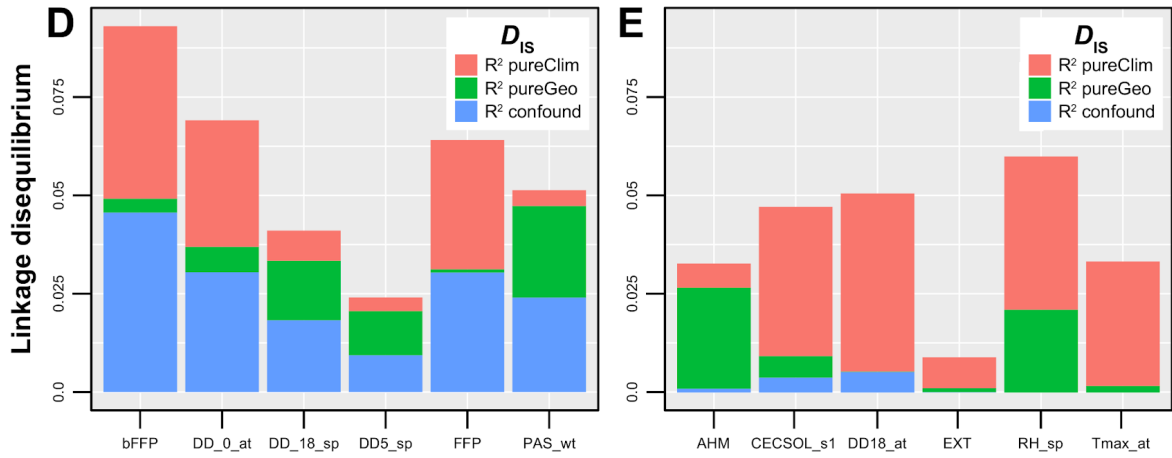
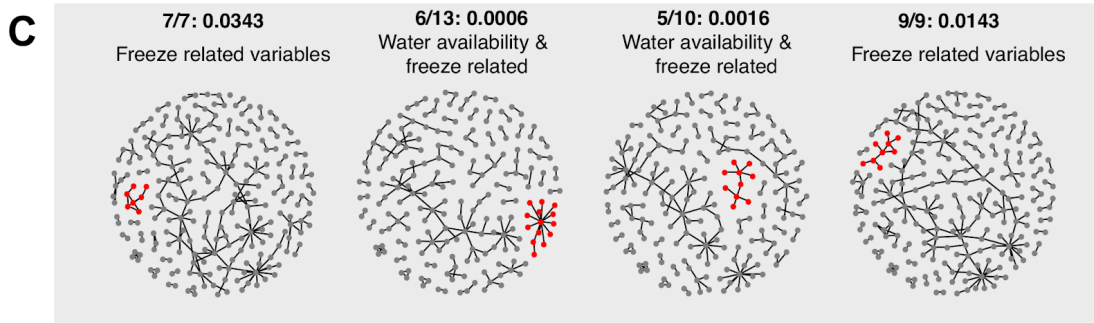
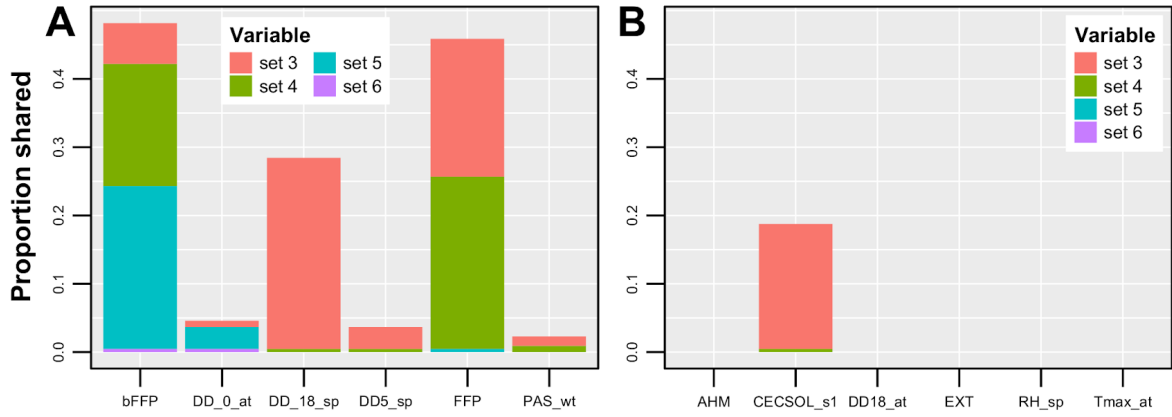


Figure 2.3: Proportion of SNPs across 100 replicates of LDna that were shared within an outlier cluster (OC) for the **(a)** strongly divergent and **(b)** east divergent environmental gradients between *P. strobiformis* and *P. flexilis* **(c)** Representation of one of the replicated sets from LDna, with red points indicating SNPs in the OC. The proportion of environmentally associated SNPs per OC, median LD for the cluster and the environmental gradients with which the SNPs in each OC are associated are indicated above each network. Regression coefficients for within population component of LD (D_{IS}) for environmental gradients that were **(d)** most divergent and **(e)** least divergent between the parental species. Regression coefficients for among population component of LD and (D_{ST}) for environmental gradients that were **(f)** most divergent and **(g)** least divergent between the parental species. Details of the environmental gradients represented on x-axis of panel a,b & d:g are represented in Table 2.S1.

In contrast, adaptive evolution within the hybrid zone should increase D_{IS} for environmentally associated loci that covary with the latitudinal axis of introgression and differentiate the two parental species. Using the full multivariate regression model containing the total effect of environment and geography, we identified solar radiation and water availability-related gradients to be the strongest predictors of D_{ST} , while freeze-related gradients had the highest R^2 for D_{IS} (Table 2.2). However, elevated LD (specifically D_{IS}) could arise in the absence of selection when populations experience recent and ongoing introgression (Schumer & Brandvain, 2016; Menon et al. 2019). To account for this, we implemented partial regression models to identify the primary components driving the elevated R^2 values for D_{ST} & D_{IS} along each environmental gradient. Environmental gradients listed below are described in Table 2.S1. First, for the pure effect of environment on D_{ST} , the highest R^2 was noted for SNPs associated with summer radiation (RAD_sm), whereas for D_{IS} , DD_0_wt associated SNPs had the highest value. Second, for the pure effect of geography on D_{ST} , SNPs associated with CMD_at and CECSOL_s1 had the highest R^2 , while PAS_wt associated SNPs had the highest R^2 when D_{IS} was the response

variable (Table 2.2, Fig 2.3 d:g). Overall, the total R^2 for D_{ST} was inversely related to the magnitude of difference in a given environmental gradients between the two parental species, while total R^2 for D_{IS} was directly related (Table 2.S5). We also note that, for most divergent environmental gradients, the confounding effect had a larger contribution to the total R^2 for D_{IS} relative to the least divergent gradients (Fig. 2.3 d & e). Specifically, even though SNPs associated with freeze-related gradients had the highest total R^2 for D_{IS} , this was rarely driven by the predictive ability of environment alone. In most cases, the total R^2 was partitioned equally between the pure effect of environment and the confounded effect (Fig. 2.3d, Table 2.S5). This is expected given (a) adaptive introgression is likely occurring along environmental gradients that also drive ecological speciation between *P. strobiformis* and *P. flexilis*, (b) latitudinal variation in ancestry and (c) environmental gradients related to freezing events covary with latitude.

Table 2.2: Top five environmental gradients (R^2 in parentheses) for three multivariate regression models partitioning the effect of geography and environment on median among and within population component of LD (D_{ST} & D_{IS}), along with superscripts indicating sign of the regression coefficient. Environmental gradients are defined in Table 2.S1.

Response	Full model	pureEnv	pureGeo	Confounded
D_{IS}	DD_0_wt (0.13)	DD_0_wt (0.079)	PAS_wt (0.023)	EMT (0.059)
	EMT (0.113)	DD_0 (0.054)	PAS (0.020)	eFFP (0.050)
	DD_0 (0.10)	EMT (0.053)	Tave_sp (0.019)	DD_0_wt (0.049)
	eFFP (0.099)	Tmin_wt (0.047)	DD5_at (0.018)	Tmin_wt (0.048)
	Tmin_wt (0.096)	bFFP (0.044)	PAS_at (0.018)	bFFP (0.045)

<i>D_{ST}</i>	RAD_sm (0.012)	RAD_sm (0.009)	CMD_at (0.004)	RH_sm (0.002)
	CMD_at (0.007)	AHM (0.006)	CECSOL_s1 (0.004)	RH_sp (0.002)
	BLD_s6 (0.007)	ORC_s1 (0.005)	BLD_s6 (0.004)	CMD_sm (0.001)
	AHM (0.006)	CMD (0.004)	MAP (0.003)	BLD (0.001)
	RH_sp (0.006)	RAD_at (0.03)	RH_sp (0.003)	RH (0.001)

Environmental gradients in bold were significant at $p < 0.01$

Abbreviations: *DD_0_wt*: winter degree days below zero, *PAS_wt*: winter precipitation as snow, *EMT*: extreme minimum temperature, *DD_0*: degree days below zero, *PAS*: precipitation as snow, *eFFP*: end of frost free period, *Tave_sp*: spring average temperature, *Tmin_wt*: winter minimum temperature, *bFFP*: beginning of frost free period, *PAS_at*: autumn precipitation as snow, *RAD_sm*: summer radiation, *CMD_at*: autumn Hargreaves climate moisture deficit, *RH_sm*: summer relative humidity, *AHM*: annual heat moisture index, *CMD*: Hargreaves climate moisture deficit, *MAP*: mean annual precipitation, *BLD_s6*: bulk density at 1m depth, *RH_sp*: spring relative humidity, *RH*: relative humidity, *CECSOL_s1*: cation exchange capacity at 0m depth, *RAD_at*: autumn radiation, *ORC_s1*: organic matter content at 0m depth.

Genomic cline analyses and drivers of adaptive introgression

Introgressed variants from *P. flexilis* facilitated adaptive evolution along freeze-related environmental gradients, while standing genetic variants drove adaptive evolution along water availability-related gradients. Of the 62,992 SNPs that were biallelic across the hybrid zone and biallelic across parental populations, 28,763 were significantly introgressed from *P. flexilis* in at least 20% of the individuals. Since our assessment of adaptive introgression was not sensitive to the individual based cutoff, we present results only for SNPs that were significantly introgressed across at-least 20% of the individuals. We note a higher than expected fold enrichment (FE) of *P. flexilis* ancestry for SNPs associated with freeze-related gradients and others that were strongly

divergent between the two parental species (Fig 2.4a, Table 2.S6, median FE: 1.5). Out of the 31 freeze-related environmental gradients (Table 2.S1), 27 were above the 95th percentile of null distribution of FE and 19 were above the 99th percentile. Conversely, none of the SNPs associated with water availability-related gradients or those least divergent between the two parental species, exhibited an enrichment of *P. flexilis* ancestry (Fig. 2.4b, median FE: 1.1). These findings were robust to potential confounding effects of geography and ancestry co-varying with several environmental gradients.

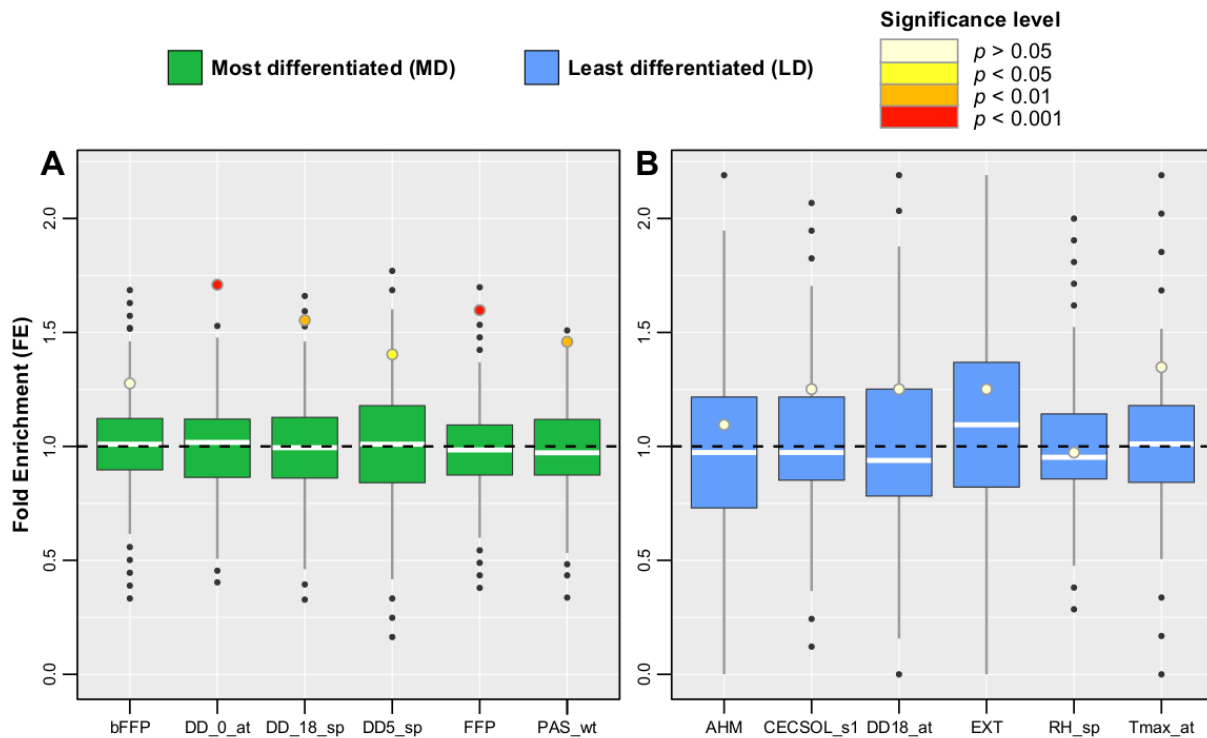


Figure 2.4: Bootstrap distribution of *P. flexilis* ancestry fold enrichment (FE) with red points indicating the FE for outlier sets of SNPs associated with environmental gradients that were (a) strongly divergent between the two species and (b) least divergent between the two species. Details of the environmental gradients are represented in Table 2.S1.

DISCUSSION

In line with several recent genome wide studies (Leroy et al. 2019; Whitney et al. 2013; Chhatre et al. 2018), the present work illustrates the importance of introgression from *P. flexilis* in facilitating adaptive evolution as well as characterizes the genetic architecture of putatively adaptive loci within the *P. flexilis*-*P. strobiformis* hybrid zone. The observed enrichment of *P. flexilis* ancestry among SNPs associated with freeze-related gradients supported our first hypothesis of the retention of freeze tolerance associated variants in a hybrid genomic background. Standing variants, moreover, were associated with adaptive evolution along water availability-related gradients. This supported our second hypothesis that the relative contribution of standing genetic variants is inversely related to the extent of divergence between the two parental species along an environmental gradient, as well as on selection pressures unique to the hybrid zone.

Adapting to rapidly changing climatic conditions is a major challenge for populations of long-lived species such as trees (Aitken et al. 2008; Alberto et al. 2013). As hybridization often occurs at species range margins that are characterised by low population density, a shift in fitness optima due to novel selective pressures imposed by climate change will purge non-adaptive alleles and increase genetic load (Hamilton & Miller, 2016; Kirkpatrick & Barton, 1997). The importance of introgression in alleviating genetic load and facilitating adaptive evolution (Stebbins, 1959) is likely to be amplified in fragmented range margin populations, where the geographical ranges of hybridizing species overlap (Petit & Excoffier, 2009). Conifer hybrid zones may be poised to overcome the challenges imposed by rapidly changing climatic conditions. This may be facilitated by their large effective population size, higher mutational input owing to reactivation of transposable elements (TEs) as noted in maize and sunflower hybrids (McClintock, 1984; Kawakami et al. 2011) and immediate adaptive evolution or

increases in standing genetic diversity from introgressed variants. By identifying the source of allelic variants and genetic architectures associated with adaptive evolution along several environmental gradients, we emphasise on the need for holistic conservation approaches that considers hybridization driven introgression, when possible.

Genetic architecture plays a key role in determining the fate of introgressed variants (Barton & Hewitt, 1985). Within advanced generation hybrid zones, the retention of introgressed variants depends on environmental conditions and the genomic background. This study, like several others (Mimura et al. 2014; De La Torre et al. 2014; Hamilton et l. 2015; Fraisse et al. 2016; Wu et al. 2018), demonstrated that for recently diverged species, or those with weak intrinsic isolating barriers, the retention of introgressed variants in hybrid zones can be favoured when fitness optima of hybrid populations align with those of the contributing sister species. In our study, this was evident when the environmental conditions of a population within the hybrid zone overlapped with those present in the range of *P. flexilis*, causing introgressing loci to likely experience positive selection. However, for populations that were environmentally distinct from *P. flexilis* and primarily dominated by *P. strobiformis* genomic background, the introgressing loci likely experience purifying selection. Specifically, when the environmental conditions of a hybrid population were similar to that in the range of *P. flexilis*, we noted greater proportion of loci with *P. flexilis* ancestry as well as less divergence in allele frequency (results not presented). This interaction between genomic ancestry and degree of environmental similarity with either parental species was also supported by variance partitioning approaches implemented in RDA and matrix regression of Ohta's *D*-statistics. Both approaches indicated a strong contribution of the confounding effect between environment and ancestry towards the granular spatial variation in genetic diversity across the hybrid zone (Table 2.1 & Fig 2.3. d:g).

Ongoing introgression causes localised increases in LD mimicking patterns expected under strong selection. Thus, in addition to the already expected polygenic architecture of local adaptation in species experiencing high gene flow (Kremer & Le Corre, 2012), outlier scans dependent on elevated patterns of differentiation alone will be underpowered and prone to high false positive rates in populations experiencing introgression. By intensively sampling the hybrid zone through a gridded sampling design (see methods) and assessing spatial patterns of LD, we have unravelled aspects of these subtle architectures and demonstrated that different environmental gradients are associated with adaptive evolution from introgressed and standing genetic variants. Specifically, introgressed variants drove adaptive evolution along freeze-related gradients that were most divergent between the two species, while standing genetic variants were predominantly associated with water availability-related gradients or those that were least divergent between the species. Among our results, greater magnitudes of LD for several outlier sets, dominance of freeze associated SNPs in the OCs and a significant association of LD variance components (D_{ST} & D_{IS}) with environmental gradients all highlight the importance of covariance in allele frequencies facilitating adaptive evolution in species with large genome sizes and large N_e (Eckert et al. 2015; Hornoy et al. 2015; Lind et al. 2017). The co-occurrence of freeze and water availability-related SNPs in some of the OCs identified through LDna echoes the role of hybridization in generating novel, putatively adaptive gene complexes unique to hybrid zones (Rieseberg et al. 2007; Lewontin & Birch, 1966). These novel gene complexes likely resulted from ongoing asymmetric introgression from *P. flexilis* providing new allelic variants into a genomic background dominated by *P. strobiformis* and may confer a competitive advantage to hybrid populations under new selective regimes generated by rapidly changing climatic conditions (Hamilton & Miller, 2016). Together, by utilising random sets of SNPs

within LDna and the permutation approach in enrichment analysis grants us confidence that despite the possibility of false positives due to strong confounding between ancestry and environmental selection gradients, several of the SNPs in the OCs are either physically linked to the true candidates or themselves the target of adaptive introgression. Further, none of the SNPs shared in an OC were from the same contig, indicating that they are less likely to be outliers due to physical linkage.

Since our study was conducted in a hybrid zone, signatures of local adaptation and clinal change in allele frequencies detected here could be confounded with spatial variation in the strength of purifying selection (Kim et al. 2018) or with the coupling of intrinsic and extrinsic barrier loci (Bierne et al. 2011). Purifying selection can generate weakly deleterious mutations, known to contribute significantly towards standing genetic diversity (Eyre-Walker et al. 2006; Christe et al. 2017). Relative to studies conducted in non-hybrid populations, those in hybrid zones are more likely to experience heterogeneity in the strength of purifying selection due to the interaction of introgressed alleles with different genomic backgrounds. Identifying the proportion of loci under spatially varying purifying selection (dependent or independent of environment) will have to await further development of population genomic resources in conifers (but see Lu et al. 2019). Coupling between intrinsic and extrinsic barriers is more likely to generate signatures that parallel local adaptation in tension zones formed by secondary contact (Bierne et al. 2011). In the absence of secondary contact in our study system, such coupling might occur in patchy hybrid zones where certain populations are predominantly differentiated along environmental gradients facilitating ecological speciation between *P. strobiformis* and *P. flexilis*. This could be driving elevated F_{ST} values of some SNPs associated with freezing temperatures and water availability (Table 2.S3), or influencing the noted spatial variation in D_{IS} as a function

of freeze-related environmental gradients (Fig. 2.3d). Approaches that scan the genome for signals of environmental associations and elevated differentiation are likely to pick up longer term processes (Whitlock MC, 1992) which could have identified some of the coupled loci mentioned above, if they exist. However, by combining genomic cline analyses with environmental associations we were able to capture the contribution of shorter term processes such as ongoing introgression (Lexer et al. 2010) towards adaptive evolution within the hybrid zone.

Utilisation of ddRADseq datasets for detecting signals of local adaptation has been criticised due to low genomic coverage, specifically of genic regions, and high error rates (Lowry et al. 2017). However, the uniform genome sampling provided through RADseq like approaches reduces ascertainment bias and false positives (Parchman et al. 2018). Adding to the influence of genome size on the architecture of adaptive evolution (Mei et al. 2019) and regulatory regions being enriched for candidate loci (Pyhäjärvi et al. 2013), the inverse relationship between N_e and the frequency of neutral mutations (Gossmann et al. 2012) indicates that selection could be more prevalent in conifers due to their large N_e . Thus, compared to ddRADseq based association studies conducted in organisms with smaller genomes and lower N_e , those in conifers may more likely pick up true signals of adaptive evolution, even if they are only partially characterized due to low genomic coverage. Given the reactivation of TEs in hybrids with complex genomes (Kawakami et al. 2011; Liu & Wendel, 200), and the dominance of TEs in conifer genomes (Nystedt et al. 2013; Stevens et al. 2016), we propose that some of the candidate loci in this study likely mapped to TEs, which would be harder to detect in association studies utilising non-hybrid populations.

CONCLUSION

Hybridization and introgression are pervasive across the Tree of Life. Several studies utilising genomic data have revealed a key role of introgressed variants in facilitating adaptation either through immediate advantage in the hybrid genomic background or by increasing standing genetic diversity (Colosimo et al. 2005; Jagoda et al. 2018). Lack of an assembled genome prohibited us from disentangling the aforementioned two paths to adaptive evolution. Nevertheless, our work identifies the environmental gradients associated with adaptive evolution from standing genetic and introgressed variants, and quantifies their relative importance. The combination of freeze and water availability-related SNPs in *P. strobiformis*–*P. flexilis* hybrid populations potentially make them an ideal seed source for conservation efforts focused on climate change mitigation. Preliminary results from ongoing common garden assays of *P. strobiformis* (*pers. comm.* E. Bucholz) dovetail with our inference of higher drought and freeze tolerance in the hybrid zone populations likely generated through the presence of *P. flexilis* alleles in a genomic background dominated by *P. strobiformis*. Beyond the hybrid zone literature, we corroborate theoretical and empirical studies demonstrating that gene flow between ecologically differentiated populations or species can buffer population decline by increasing genetic diversity and providing novel allelic combinations.

Appendix

Table 2.S1: Explanation of all environmental gradients used in the present study and their mean value across the sampled range of pure *Pinus strobiformis* and *P. flexilis*. Seasons including winter, spring, summer and autumn encompass the following months: Winter = Jan, Feb & Dec; Spring = March to May; Summer = June to August; Autumn = September to November.

Abbreviation	Variable	Parent Diff	<i>P. strobiformis</i>	<i>P. flexilis</i>	Freeze related	Water availability related
AHM	Annual heat-moisture index	7.95	30.35	22.4		Y
bFFP	Begining of frost free period	79.5	99.5	179	Y	
BLD_s1	Bulk density at 0m depth	218.783	1310.090	1091.307		
BLD_s6	Bulk density at 1m depth	62.080	1521.227	1583.307		
CECSOL_s1	Cation exchange capacity at 0m depth	1.734	25.727	27.461		Y
CECSOL_s6	Cation exchange capacity at 1m depth	5.8	22.954	17.153		Y
CMD	Hargreaves climatic moisture deficit	310	626	316		Y
CMD_at	Hargreaves climatic moisture deficit autumn	75.5	118.5	43		Y
CMD_sm	Hargreaves climatic moisture deficit summer	112	115	227		Y

CMD_sp	Hargreaves climatic moisture deficit spring	291.5	344.5	53		Y
CMD_wt	Hargreaves climatic moisture deficit autumn	88	88	0		Y
DD_0	Degree days below 0C	839.5	39.5	879	Y	
DD_0_at	Degree days below 0C autumn	105.5	2.5	108	Y	
DD_0_sp	Degree days below 0C spring	173	4	177	Y	
DD_0_wt	Degree days below 0C winter	549	32	581	Y	
DD_18	Degree days below 18C	3351.5	2026.5	5378	Y	
DD_18_at	Degree days below 18C autumn	834.5	435.5	1270	Y	
DD_18_sm	Degree days below 18C summer	393	98	491	Y	
DD_18_sp	Degree days below 18C spring	1066.5	459.5	1526	Y	
DD_18_wt	Degree days below 18C winter	1159	979	2138	Y	
DD18	Degree days above 18C	236	252	16		
DD18_at	Degree days above 18C autumn	50	53	3		
DD18_sm	Degree days above 18C summer	142	154	12		
DD18_sp	Degree days above 18C spring	44	45	1		

DD18_wt	Degree days above 18C winter	1.5	1.5	0		
DD5	Degree days above 5C	2185	3207	1022		
DD5_at	Degree days below 5C autumn	626.5	819.5	193	Y	
DD5_sm	Degree days below 5C summer	535.5	1249.5	714	Y	
DD5_sp	Degree days below 5C spring	704.5	799.5	95	Y	
DD5_wt	Degree days below 5C winter	232.5	238.5	6	Y	
eFFP	End of frost free period	67	312	245	Y	
EMT	Extreme minimum temperature in C	21.95	-18.45	-40.4	Y	
Eref	Reference evapotranspiration	647.5	1355.5	708		Y
Eref_at	Reference evapotranspiration autumn	156.5	293.5	137		Y
Eref_sm	Reference evapotranspiration summer	58	463	405		Y
Eref_sp	Reference evapotranspiration spring	239	408	169		Y
Eref_wt	Reference evapotranspiration winter	213	213	0		Y
EXT	Extreme maximum temperature in C	4.5	36.7	32.2		Y

FFP	Frost free period	145.5	212.5	67	Y	
MAP	Mean annual precipitation (mm)	136	783	647		Y
MAR	Mean annual solar radiation	3.55	20.75	17.2		
MAT	Mean annual temp (C)	10.3	13.5	3.2		
MCMT	Mean coldest month temperature (C)	12.55	6.55	-6	Y	
MSP	Mean annual summer precipitation (mm)	259.5	539.5	280		Y
MWMT	Mean warmest month temperature (C)	4.15	18.55	14.4		
NFFD	Number of frost free days	176.5	296.5	120	Y	
NFFD_at	Number of frost free days autumn	54.5	80.5	26	Y	
NFFD_sm	Number of frost free days summer	11	92	81	Y	
NFFD_sp	Number of frost free days spring	64	77	13	Y	
NFFD_wt	Number of frost free days winter	41	42	1	Y	
ORC_s1	Organic matter content at 0m depth	16.769	89	105.769		
ORC_s6	Organic matter content at 1m depth	1.479	5.136	6.6155		
PAS	Precipitation as snow (mm)	232.5	3.5	236	Y	
PAS_at	Precipitation as snow	26	1	27	Y	

	autumn (mm)					
PAS_sm	Precipitation as snow summer (mm)	1	1	2	Y	
PAS_sp	Precipitation as snow spring (mm)	101	0	101	Y	
PAS_wt	Precipitation as snow winter (mm)	119.5	1.5	121	Y	
PHIHOX_s1	pH in H2O at 0m depth	1.832	61.909	60.076		
PHIHOX_s6	pH in H2O at 1m depth	1.206	63.409	64.615		
PHIKCL_s1	pH in KCl at 0m depth	3.227	54.227	51		
PHIKCL_s6	pH in KCl at 1m depth	0.304	54.227	53.923		
PPT_at	Precipitation autumn (mm)	23	178	155		Y
PPT_sm	Summer precipitation (mm)	222	403	181		Y
PPT_sp	Precipitation spring (mm)	113.5	69.5	183		Y
PPT_wt	Precipitation winter (mm)	32	112	144		Y
Rad_at	Radiation autumn	4.65	18.65	14		
Rad_sm	Radiation summer	1.6	25.7	24.1		
Rad_sp	Radiation spring	3.4	24.1	20.7		
Rad_wt	Radiation winter	4.25	14.45	10.2		
RH	Relative humidity	3.5	54.5	51		Y
RH_at	Relative humidity autumn	4.5	56.5	52		Y

RH_sm	Relative humidity summer	7.5	58.5	51		Y
RH_sp	Relative humidity spring	0.5	50.5	50		Y
RH_wt	Relative humidity winter	2.5	53.5	51		Y
SHM	Summer heat-moisture index	16.5	35.7	52.2		Y
Tave_at	Average temp autumn (C)	9.85	13.95	4.1		
Tave_sm	Average temp summer (C)	5.9	18.6	12.7		
Tave_sp	Average temp spring (C)	12.2	13.6	1.4		
Tave_wt	Average temp winter (C)	12.9	7.1	-5.8		
TD	Continentality (C)	8.65	12.15	20.8		
Tmax_at	Max temp autumn (C)	9.3	20.4	11.1		
Tmax_sm	Max. temp summer (C)	4.2	25.6	21.4		
Tmax_sp	Maximum temperature spring (C)	12.4	21.5	9.1		
Tmax_wt	Maximum temperature winter (C)	14.3	15.4	1.1		
Tmin_at	Minimum temp autumn (C)	10.5	6.8	-3.7	Y	
Tmin_sm	Min temp summer (C)	6.95	11.45	4.5	Y	
Tmin_sp	Minimum temp	10.6	4.7	-5.9	Y	

	spring (C)					
Tmin_wt	Minimum temp winter (C)	13.1	0.1	-13	Y	

Table 2.S2: Loadings of all 88 environmental gradients on the top seven PC axes that was used as the environmental matrix in RDA.

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Tmax_wt	-0.124	-0.08	-0.034	0.154	0.1	-0.038	-0.025
Tmax_sp	-0.123	-0.123	-0.01	0.087	0.096	-0.023	-0.037
Tmax_sm	-0.076	-0.204	-0.116	-0.039	0.032	-0.022	-0.014
Tmax_at	-0.115	-0.143	-0.081	0.074	0.073	-0.013	0.01
Tmin_wt	-0.139	0.087	-0.005	-0.004	-0.033	0.051	0.022
Tmin_sp	-0.139	0.072	0.024	-0.084	0.001	0.053	-0.034
Tmin_sm	-0.138	0.034	-0.038	-0.124	-0.045	0.022	-0.002
Tmin_at	-0.138	0.08	-0.026	-0.071	-0.03	0.045	-0.007
Tave_wt	-0.145	0.015	-0.02	0.073	0.029	0.013	0.002
Tave_sp	-0.147	-0.021	0.01	-0.006	0.05	0.02	-0.04
Tave_sm	-0.13	-0.087	-0.086	-0.101	-0.013	0.004	-0.008
Tave_at	-0.146	-0.017	-0.056	-0.011	0.017	0.023	0.001
PPT_wt	-0.004	0.048	-0.327	-0.027	0.045	-0.063	0.213
PPT_sp	0.06	0.045	-0.278	-0.125	0.077	-0.063	0.107
PPT_sm	-0.067	0.159	0.05	0.201	0.172	-0.039	0.015
PPT_at	-0.013	0.166	-0.2	-0.058	0.187	-0.144	0.037
Rad_wt	-0.055	0.078	-0.112	0.201	-0.247	-0.225	-0.077
Rad_sp	-0.088	0.078	-0.099	0.193	-0.216	-0.13	-0.056
Rad_sm	-0.033	0.122	-0.089	0.081	-0.303	-0.234	-0.087
Rad_at	-0.06	0.075	-0.141	0.182	-0.264	-0.191	-0.084

DD_0_wt	0.142	-0.022	-0.007	-0.083	0.012	-0.053	-0.093
DD_0_sp	0.143	0.013	-0.029	-0.026	0.021	-0.078	-0.121
DD_0_at	0.141	-0.009	-0.008	-0.055	0.02	-0.068	-0.144
DD5_wt	-0.13	-0.006	-0.058	0.046	0.087	-0.059	-0.15
DD5_sp	-0.144	-0.023	0.004	-0.017	0.073	-0.003	-0.099
DD5_sm	-0.13	-0.087	-0.087	-0.102	-0.013	0.004	-0.009
DD5_at	-0.143	-0.027	-0.071	-0.03	0.021	0.014	-0.033
DD_18_wt	0.145	-0.016	0.02	-0.073	-0.028	-0.013	-0.002
DD_18_sp	0.147	0.02	-0.01	0.004	-0.047	-0.021	0.034
DD_18_sm	0.13	0.086	0.064	0.103	0.036	-0.014	-0.005
DD_18_at	0.146	0.015	0.051	0.009	-0.016	-0.027	-0.006
DD18_wt	-0.057	0.065	0.144	-0.168	0.189	-0.118	-0.229
DD18_sp	-0.107	0.025	0.124	-0.047	0.229	-0.086	-0.166
DD18_sm	-0.124	-0.084	-0.116	-0.098	0.025	-0.015	-0.029
DD18_at	-0.133	-0.043	-0.099	-0.045	0.047	-0.015	-0.065
NFFD_wt	-0.128	0.092	-0.015	-0.066	0.05	-0.041	-0.117
NFFD_sp	-0.14	0.069	0.01	-0.074	-0.007	0.061	-0.024
NFFD_sm	-0.114	0.056	0.018	-0.03	-0.182	0.093	0.12
NFFD_at	-0.137	0.085	-0.026	-0.057	-0.039	0.053	0.005
PAS_wt	0.12	0.014	-0.168	-0.097	0.071	-0.097	-0.079
PAS_sp	0.122	0.021	-0.121	-0.066	0.077	-0.107	-0.236
PAS_sm	0.068	0.132	0.114	0.208	0.012	0.013	0.036
PAS_at	0.125	0.029	-0.109	-0.069	0.078	-0.12	-0.203
Eref_wt	-0.127	0.019	-0.074	0.124	0.116	-0.011	-0.047
Eref_sp	-0.106	-0.137	0.003	0.165	0.047	0.003	0.079
Eref_sm	-0.024	-0.24	-0.105	0.066	0.043	-0.023	0.014
Eref_at	-0.093	-0.16	-0.052	0.17	0.064	-0.013	0.076
CMD_wt	-0.067	-0.046	0.191	-0.019	0.065	0.016	-0.322

CMD_sp	-0.102	-0.113	0.14	0.153	-0.006	0.037	-0.079
CMD_sm	0.048	-0.207	-0.098	-0.099	-0.151	0.003	0.012
CMD_at	-0.031	-0.225	0.062	0.078	-0.107	0.09	-0.041
RH_wt	-0.082	0.181	0.023	-0.137	-0.118	0.095	0.046
RH_sp	-0.062	0.198	0.04	-0.173	-0.087	0.088	-0.008
RH_sm	-0.087	0.193	0.054	-0.104	-0.066	0.038	0.004
RH_at	-0.076	0.199	0.031	-0.137	-0.088	0.06	-0.02
MAT	-0.147	-0.026	-0.037	-0.007	0.023	0.016	-0.012
MWMT	-0.114	-0.111	-0.126	-0.131	-0.027	-0.002	0.023
MCMT	-0.143	0.021	-0.022	0.081	0.02	0.007	0.004
TD	0.074	-0.137	-0.098	-0.236	-0.052	-0.01	0.019
MAP	-0.02	0.146	-0.254	0.037	0.164	-0.096	0.147
MSP	-0.06	0.176	0.062	0.158	0.199	-0.052	0.004
AHM	-0.04	-0.144	0.22	-0.051	-0.209	0.13	-0.142
SHM	0.023	-0.188	-0.071	-0.173	-0.2	0.054	0.021
DD_0	0.142	-0.015	-0.011	-0.071	0.014	-0.059	-0.104
DD5	-0.142	-0.048	-0.057	-0.05	0.028	0.002	-0.051
DD_18	0.148	0.018	0.026	-0.005	-0.022	-0.02	0.008
DD18	-0.13	-0.072	-0.097	-0.09	0.047	-0.022	-0.048
NFFD	-0.139	0.081	-0.007	-0.066	-0.014	0.04	-0.027
bFFP	0.132	-0.088	-0.034	0.103	0.013	-0.046	0.038
eFFP	-0.137	0.079	-0.021	-0.074	-0.036	0.048	-0.016
FFP	-0.136	0.085	0.011	-0.092	-0.024	0.048	-0.026
PAS	0.123	0.019	-0.148	-0.084	0.075	-0.104	-0.142
EMT	-0.138	0.09	-0.016	-0.011	-0.034	0.034	-0.002
EXT	-0.078	-0.192	-0.093	-0.057	0.084	-0.074	-0.04
Eref	-0.11	-0.127	-0.071	0.151	0.088	-0.013	0.02
CMD	-0.026	-0.244	0.031	0.02	-0.122	0.044	-0.067

MAR	-0.065	0.086	-0.12	0.186	-0.261	-0.201	-0.08
RH	-0.079	0.195	0.038	-0.138	-0.089	0.071	0.007
BLD_s1	-0.075	-0.101	0.2	-0.041	-0.058	-0.123	0.033
BLD_s6	0.032	-0.038	-0.168	-0.201	-0.13	0.032	-0.167
CECSOL_s1	0.051	0.067	-0.031	0.066	0.141	0.103	-0.325
CECSOL_s6	-0.067	-0.058	-0.01	-0.122	0.157	-0.039	0.065
ORC_s1	0.071	0.067	-0.105	0.067	0.137	0.275	-0.01
ORC_s6	0.042	-0.006	0.062	-0.052	0.112	0.123	0.375
PHIHOX_s1	-0.074	-0.042	0.215	-0.08	0.055	-0.309	0.047
PHIHOX_s6	-0.046	-0.049	0.241	-0.112	0.022	-0.327	-0.025
PHIKCL_s1	-0.076	-0.02	0.137	-0.111	0.027	-0.312	0.26
PHIKCL_s6	-0.032	-0.027	0.153	-0.166	0.019	-0.39	0.216

Table 2.S3: Summary statistics for Bayenv outlier SNPs associated with each of the 88 environmental gradients. Values in bold indicate significance at $p < 0.05$.

Variable	minBF	maxBF	medianBF	multi.Fct	multi.Fst	LDmedian	Noutliers
AHM	1.265	14.066	1.265	0.002	0.020	0.0007	9
bFFP	1.313	3892.487	1.313	0.033	0.087	0.0006	39
BLD_s1	1.026	77.088	1.0261	0.014	0.051	0.0004	17
BLD_s6	1.765	63.378	1.765	0.029	0.062	0.0003	9
CECSOL_s1	1.317	345.225	1.317	0.012	0.063	0.0006	18
CECSOL_s6	0.996	21017.445	0.996	0.014	0.047	0.0005	11
CMD	1.092	80.953	1.092	0.003	0.052	0.0006	10
CMD_at	1.0457	61.564	1.045	0.008	0.032	0.0005	8
CMD_sm	1.340	148260043	1.3405	0.036	0.089	0.0004	31
CMD_sp	1.286	38.068	1.28	0.035	0.083	0.0004	21

CMD_wt	1.232	43.329	1.232	0.001	0.049	0.0006	9
DD_0	1.256	514.810	1.256	0.042	0.092	0.0005	43
DD_0_at	1.446	403.949	1.446	0.043	0.093	0.0005	43
DD_0_sp	1.934	322.791	1.934	0.042	0.096	0.0005	39
DD_0_wt	1.487	511.698	1.487	0.042	0.092	0.0006	44
DD_18	1.51	185.070	1.513	0.0413	0.090	0.0004	35
DD_18_at	1.31	128.456	1.311	0.042	0.092	0.0004	32
DD_18_sm	1.234	15.272	1.234	0.024	0.050	0.0006	18
DD_18_sp	2.117	1046.740	2.117	0.042	0.096	0.0005	33
DD_18_wt	1.443	301.280	1.443	0.051	0.103	0.0005	37
DD18	1.054	28.622	1.054	0.026	0.061	0.0004	17
DD18_at	1.124	19.673	1.124	0.034	0.076	0.0005	14
DD18_sm	1.060	24.525	1.060	0.031	0.063	0.0003	14
DD18_sp	1.605	175256.911	1.605	0.0196	0.0716	0.0005	24
DD18_wt	1.574	2845.489	1.573	0.0143	0.073	0.001	26
DD5	1.275	29.782	1.2752	0.028	0.058	0.0004	24
DD5_at	1.556	50.003	1.556	0.039	0.075	0.0004	28
DD5_sm	1.060	14.843	1.060	0.023	0.055	0.0004	18
DD5_sp	1.547	119.051	1.547	0.047	0.096	0.0004	26
DD5_wt	1.742	85.734	1.74	0.025	0.068	0.0005	18
eFFP	1.551	573.833	1.55	0.036	0.090	0.0005	44
EMT	1.880	458.259	1.880	0.041	0.091	0.0005	36
Eref	1.108	14.057	1.108	0.018	0.043	0.00054	10
Eref_at	0.898	14.036	0.898	0.020	0.040	0.0003	10

Eref_sm	0.912	57.167	0.912	0.009	0.053	0.0003	14
Eref_sp	1.284	16.778	1.284	0.017	0.035	0.0005	11
Eref_wt	1.259	76.576	1.259	0.032	0.081	0.0004	27
EXT	1.09	10.046	1.099	0.013	0.029	0.0001	8
FFP	1.580	1235.937	1.580	0.035	0.095	0.0006	40
MAP	0.953	17.616	0.953	0.004	0.039	0.0003	15
MAR	2.061	291.054	2.061	0.012	0.053	0.0002	11
MAT	1.193	151.464	1.193	0.041	0.089	0.0005	33
MCMT	1.361	290.476	1.361	0.057	0.107	0.0006	33
MSP	1.356	65.128	1.356	0.042	0.091	0.0004	33
MWMT	1.165	29.331	1.165	0.013	0.038	0.0003	15
NFFD	1.457	270.417	1.457	0.036	0.092	0.0007	41
NFFD_at	1.414	331.937	1.414	0.036	0.091	0.0006	44
NFFD_sm	2.071	2852.013	2.071	0.038	0.1004	0.001	33
NFFD_sp	1.734	1752.234	1.734	0.037	0.094	0.0006	42
NFFD_wt	1.202	178.775	1.202	0.022	0.063	0.0006	27
ORC_s1	1.636	4.743	1.636	0.003	0.029	0.0004	6
ORC_s6	0.931	16.072	0.931	0.011	0.039	0.0003	7
PAS	1.349	377.340	1.349	0.049	0.101	0.0005	41
PAS_at	1.249	187.726	1.249	0.053	0.105	0.0007	35
PAS_sm	1.186	3.813	1.186	0.0003	0.003	0.0003	5
PAS_sp	1.725	217.210	1.725	0.0526	0.108	0.0006	37
PAS_wt	1.611	352.031	1.611	0.043	0.092	0.0006	45
PHIHOX_s1	1.826	3110.343	1.826	0.002	0.0288	0.0006	14

PHIHOX_s6	1.082	1433.299	1.082	0.002	0.018	0.0005	17
PHIKCL_s1	1.934	18.863	1.934	0.003	0.016	0.0004	9
PHIKCL_s6	1.38	123.311	1.38	0.003	0.014	0.0003	9
PPT_at	0.927	14.634	0.927	0.011	0.029	0.0002	16
PPT_sm	1.362	42.295	1.362	0.044	0.094	0.0005	36
PPT_sp	1.358	63.551	1.358	0.037	0.093	0.0004	14
PPT_wt	1.217	163.590	1.217	0.008	0.046	0.0004	17
Rad_at	2.134	43.503	2.134	0.008	0.048	0.0002	12
Rad_sm	0.849	4.710	0.849	0.004	0.033	0.0003	9
Rad_sp	1.831	532.119	1.831	0.043	0.087	0.0004	20
Rad_wt	3.957	12.145	3.957	0.006	0.026	0.0005	5
RH	1.251	174.281	1.251	0.029	0.086	0.0004	23
RH_at	1.434	92.504	1.434	0.027	0.082	0.0004	23
RH_sm	1.945	158.298	1.945	0.032	0.083	0.0004	23
RH_sp	1.158	171.993	1.158	0.005	0.060	0.0003	23
RH_wt	1.179	124.388	1.179	0.034	0.092	0.0004	24
SHM	1.231	4086680.91	1.231	0.010	0.055	0.0005	26
Tave_at	1.381	114.238	1.381	0.044	0.092	0.0005	30
Tave_sm	1.075	15.197	1.075	0.024	0.056	0.0005	18
Tave_sp	1.866	1021.718	1.866	0.041	0.095	0.0004	35
Tave_wt	1.450	296.517	1.450	0.052	0.103	0.0005	38
TD	1.234	548.570	1.234	0.058	0.103	0.0005	28
Tmax_at	1.363	17.012	1.363	0.029	0.061	0.0003	13
Tmax_sm	0.905	4.970	0.905	0.002	0.015	0.0003	13

Tmax_sp	1.413	24.555	1.413	0.031	0.068	0.0003	15
Tmax_wt	1.589	32.724	1.589	0.049	0.101	0.0004	15
Tmin_at	1.479	496.463	1.479	0.038	0.091	0.0006	40
Tmin_sm	1.780	230.117	1.780	0.031	0.078	0.0005	34
Tmin_sp	1.587	596.087	1.587	0.040	0.097	0.0006	36
Tmin_wt	1.316	559.018	1.316	0.039	0.090	0.0005	41

Table 2.S4: Proportion of times sets of three, four, five and six Bayenv outliers were represented in the same OCs.

Variable	set3	set4	set5	set6
bFFP	0.482	0.423	0.244	0.005
FFP	0.459	0.257	0.005	0
NFFD_sm	0.445	0.441	0.276	0.244
eFFP	0.299	0.257	0	0
NFFD_at	0.299	0.285	0.01	0
NFFD_sp	0.294	0.285	0.005	0
NFFD	0.294	0.253	0	0
DD_0_sp	0.289	0.005	0.005	0
Tmin_sm	0.289	0.01	0	0
DD_18_at	0.285	0.01	0	0
DD_18_sp	0.285	0.005	0	0
Tave_sp	0.285	0.01	0	0
CECSOL_s1	0.189	0.005	0	0
DD18_wt	0.083	0.037	0.019	0.005
DD_0_wt	0.051	0.037	0.037	0.005

Tmin_at	0.051	0.014	0	0
DD_0_at	0.046	0.037	0.037	0.005
DD_0	0.046	0.037	0.037	0.005
DD_18_wt	0.042	0.037	0.005	0
DD_18	0.042	0.005	0	0
EMT	0.042	0.005	0	0
PAS_at	0.042	0.005	0.005	0
PAS_sp	0.042	0.005	0.005	0
Tave_at	0.042	0.005	0	0
Tave_wt	0.042	0.037	0.005	0
Tmin_sp	0.042	0.005	0	0
Tmin_wt	0.042	0.037	0.005	0
DD5_sp	0.037	0.005	0	0
MAT	0.037	0.005	0	0
MCMT	0.037	0.037	0.005	0
PPT_sm	0.037	0.005	0	0
Rad_sp	0.037	0	0	0
TD	0.037	0.005	0	0
Eref_wt	0.023	0	0	0
PAS_wt	0.023	0.01	0	0
PAS	0.01	0.005	0	0
BLD_s1	0.01	0.005	0	0
CMD_sm	0.005	0	0	0
DD18	0.005	0	0	0

DD5_at	0.005	0	0	0
DD5	0.005	0	0	0
MSP	0.005	0	0	0
NFFD_wt	0.005	0	0	0
RH_wt	0.005	0	0	0
SHM	0.005	0	0	0
PHIHOX_s6	0.005	0	0	0
AHM	0	0	0	0
CMD_at	0	0	0	0
CMD_sp	0	0	0	0
CMD_wt	0	0	0	0
CMD	0	0	0	0
DD_18_sm	0	0	0	0
DD18_at	0	0	0	0
DD18_sm	0	0	0	0
DD18_sp	0	0	0	0
DD5_sm	0	0	0	0
DD5_wt	0	0	0	0
Eref_at	0	0	0	0
Eref_sm	0	0	0	0
Eref_sp	0	0	0	0
Eref	0	0	0	0
EXT	0	0	0	0
MAP	0	0	0	0

MAR	0	0	0	0
MWMT	0	0	0	0
PAS_sm	0	0	0	0
PPT_at	0	0	0	0
PPT_sp	0	0	0	0
PPT_wt	0	0	0	0
Rad_at	0	0	0	0
Rad_sm	0	0	0	0
Rad_wt	0	0	0	0
RH_at	0	0	0	0
RH_sm	0	0	0	0
RH_sp	0	0	0	0
RH	0	0	0	0
Tave_sm	0	0	0	0
Tmax_at	0	0	0	0
Tmax_sm	0	0	0	0
Tmax_sp	0	0	0	0
Tmax_wt	0	0	0	0
BLD_s6	0	0	0	0
CECSOL_s6	0	0	0	0
ORCDRC_s1	0	0	0	0
ORCDRC_s6	0	0	0	0
PHIHOX_s1	0	0	0	0
PHIKCL_s1	0	0	0	0

PHIKCL_s6	0	0	0	0
-----------	---	---	---	---

Table 2.S5: R^2 estimates and the corresponding p -values from the multiple matrix regression for D_{IS} and D_{ST} across the 88 environmental gradients. For gradients with 10 or fewer unique response values, only the R^2 is reported.

Variable	LD estimate	R2Full	p-Tot	R2pureEnv	pEnv	R2pureGeo	pGeo	Confound
AHM	Dis	0.00327	0.037	0.00062	0.356	0.00257	0.035	0.0001
bFFP	Dis	0.09302	0.001	0.04386	0.001	0.00352	0.039	0.04565
BLD_s1	Dis	0.00732	0.001	0.00703	0.002	0.00027	0.452	0.00002
BLD_s6	Dis	0.00805	0.029	0.00771	0.033	0.00012	0.682	0.00022
CECSOL_s1	Dis	0.00471	0.027	0.00379	0.078	0.00055	0.36	0.00038
CECSOL_s6	Dis	0.00005	0.505	0.00004	0.846	0.00001	0.907	-0.00001
CMD	Dis	0.00053	0.201	0.00049	0.314	0.00004	0.819	0.00001
CMD_at	Dis	0.01065	0.02	0.01039	0.022	0.00031	0.544	-0.00004
CMD_sm	Dis	0.0077	0.002	0.00002	0.875	0.00569	0.006	0.002
CMD_sp	Dis	0.00288	0.068	0.00189	0.087	0.00146	0.061	-0.00048
CMD_wt	Dis	0.00913	0.007	0.0041	0.079	0.00553	0.008	-0.0005
DD_0	Dis	0.10068	0.001	0.05412	0.001	0.00355	0.031	0.04302
DD_0_at	Dis	0.0691	0.001	0.03219	0.001	0.00648	0.005	0.03045
DD_0_sp	Dis	0.05959	0.001	0.03208	0.001	0.00367	0.022	0.02385
DD_0_wt	Dis	0.13034	0.001	0.07974	0.001	0.00082	0.294	0.04979
DD_18	Dis	0.0405	0.001	0.00983	0.004	0.01211	0.001	0.01857
DD_18_at	Dis	0.02428	0.001	0.00393	0.056	0.00946	0.003	0.0109

DD_18_sm	Dis	0.00852	0.001	0.00798	0.001	0.00003	0.808	0.00051
DD_18_sp	Dis	0.04105	0.001	0.00766	0.009	0.01512	0.001	0.01828
DD_18_wt	Dis	0.03531	0.001	0.01301	0.001	0.0039	0.014	0.01841
DD18	Dis	0.00777	0.002	0.00701	0.006	0.00143	0.106	-0.00066
DD18_at	Dis	0.00505	0.011	0.00452	0.011	0.00002	0.881	0.00052
DD18_sm	Dis	0.0039	0.026	0.00352	0.024	0.00069	0.264	-0.00031
DD18_sp	Dis	0.00691	0.013	0.00443	0.018	0.00027	0.426	0.00222
DD18_wt	Dis	0.02087	0.007	0.01648	0.006	0.00132	0.159	0.00308
DD5	Dis	0.00179	0.172	0.00006	0.82	0.0013	0.197	0.00044
DD5_at	Dis	0.03783	0.001	0.00543	0.031	0.01815	0.001	0.01426
DD5_sm	Dis	0.00174	0.099	0.00008	0.709	0.00139	0.105	0.00027
DD5_sp	Dis	0.02406	0.002	0.00346	0.096	0.01124	0.001	0.00937
DD5_wt	Dis	0.00496	0.1	0.0045	0.108	0.00059	0.382	-0.00013
eFFP	Dis	0.09977	0.001	0.04304	0.001	0.00597	0.01	0.05077
EMT	Dis	0.11365	0.001	0.05319	0.001	0.00144	0.177	0.05903
Eref	Dis	0.00768	0.005	0.00737	0.009	0.00062	0.295	-0.0003
Eref_at	Dis	0.02097	0.001	0.02051	0.001	0.00067	0.34	-0.0002
Eref_sm	Dis	0.00652	0.003	0.00597	0.006	0.00024	0.501	0.00032
Eref_sp	Dis	0.00476	0.011	0.00461	0.014	0.00002	0.833	0.00014
Eref_wt	Dis	0.00561	0.001	0.00482	0.001	0.00136	0.099	-0.00057
EXT	Dis	0.0009	0.218	0.00079	0.356	0.0001	0.717	0.00001
FFP	Dis	0.06409	0.001	0.03287	0.001	0.00084	0.289	0.03039
MAP	Dis	0.00405	0.032	0.00324	0.025	0.0011	0.171	-0.00029
MAR	Dis	0.01101	0.009	0.00794	0.02	0.00464	0.006	-0.00157

MAT	Dis	0.04286	0.001	0.01074	0.002	0.01361	0.001	0.01852
MCMT	Dis	0.0124	0.001	0.00689	0.002	0.00012	0.607	0.0054
MSP	Dis	0.02137	0.001	0.00142	0.156	0.01299	0.001	0.00697
MWMT	Dis	0.00057	0.292	0.0005	0.332	0.00013	0.557	-0.00005
NFFD	Dis	0.07565	0.001	0.03604	0.001	0.00129	0.167	0.03833
NFFD_at	Dis	0.07372	0.001	0.03294	0.001	0.00239	0.063	0.03839
NFFD_sm	Dis	0.00501	0.021	0.0037	0.034	0.00001	0.979	0.00132
NFFD_sp	Dis	0.07049	0.001	0.0306	0.001	0.00298	0.05	0.03693
NFFD_wt	Dis	0.01514	0.007	0.00854	0.006	0.00028	0.512	0.00632
ORC_s1	Dis	0.00636	0.067	0.00416	0.065	0.00277	0.054	-0.00057
ORC_s6	Dis	0.00168	0.3	0.00111	0.315	0.00068	0.322	-0.0001
PAS	Dis	0.03211	0.001	0.00029	0.708	0.02091	0.001	0.01092
PAS_at	Dis	0.02792	0.001	0.00127	0.396	0.01777	0.001	0.00889
PAS_sm	Dis	0.00653	0.018	0.00553	0.063	0.00082	0.31	0.00019
PAS_sp	Dis	0.01658	0.001	0.0005	0.541	0.01408	0.001	0.00201
PAS_wt	Dis	0.05133	0.001	0.00402	0.122	0.02331	0.001	0.02402
PHIHOX_s1	Dis	0.01489	0.003	0.01388	0.002	0.00172	0.126	-0.0007
PHIHOX_s6	Dis	0.0139	0.005	0.01058	0.01	0.00534	0.006	-0.00202
PHIKCL_s1	Dis	0.01329	0.003	0.01235	0.001	0.00169	0.068	-0.00075
PHIKCL_s6	Dis	0.02918	0.001	0.02854	0.001	0.00107	0.186	-0.00043
PPT_at	Dis	0.00224	0.08	0.00119	0.254	0.00098	0.256	0.00008
PPT_sm	Dis	0.02157	0.001	0.00168	0.112	0.01193	0.001	0.00796
PPT_sp	Dis	0.01983	0.001	0.01653	0.001	0.00015	0.649	0.00316
PPT_wt	Dis	0.00569	0.018	0.00405	0.021	0.00259	0.044	-0.00094

Rad_at	Dis	0.01868	0.002	0.01797	0.001	0.00119	0.146	-0.00048
Rad_sm	Dis	0.01512	0.005	0.01489	0.009	0.00047	0.465	-0.00024
Rad_sp	Dis	0.00679	0.001	0.00014	0.641	0.00607	0.001	0.00059
Rad_wt	Dis	0.00316	0.151	0.00106	0.359	0.00145	0.162	0.00066
RH	Dis	0.01761	0.001	0.00106	0.372	0.01108	0.001	0.00548
RH_at	Dis	0.01823	0.001	0.00094	0.4	0.01215	0.001	0.00515
RH_sm	Dis	0.02158	0.005	0.0099	0.009	0.00291	0.044	0.00877
RH_sp	Dis	0.00517	0.032	0.00389	0.033	0.0021	0.056	-0.00082
RH_wt	Dis	0.02883	0.001	0.00281	0.114	0.01647	0.001	0.00956
SHM	Dis	0.01508	0.001	0.00136	0.211	0.01022	0.001	0.00351
Tave_at	Dis	0.02581	0.001	0.00452	0.062	0.00987	0.003	0.01144
Tave_sm	Dis	0.00047	0.518	0.00015	0.614	0.00022	0.539	0.00011
Tave_sp	Dis	0.04631	0.001	0.00764	0.007	0.01861	0.001	0.02007
Tave_wt	Dis	0.04314	0.001	0.01833	0.001	0.0033	0.036	0.02152
TD	Dis	0.01626	0.001	0.00014	0.664	0.01183	0.001	0.0043
Tmax_at	Dis	0.00325	0.116	0.00316	0.124	0.00017	0.665	-0.00009
Tmax_sm	Dis	0.00023	0.592	0.00011	0.688	0.00014	0.658	-0.00001
Tmax_sp	Dis	0.00322	0.17	0.00204	0.226	0.00158	0.175	-0.0004
Tmax_wt	Dis	0.0029	0.034	0.00274	0.029	0.00007	0.692	0.0001
Tmin_at	Dis	0.07033	0.001	0.0331	0.001	0.00167	0.112	0.03557
Tmin_sm	Dis	0.04259	0.001	0.01685	0.001	0.00448	0.009	0.02128
Tmin_sp	Dis	0.0864	0.001	0.04283	0.001	0.00179	0.152	0.0418
Tmin_wt	Dis	0.0966	0.001	0.04704	0.001	0.00066	0.334	0.04891
AHM	Dst	0.00615	0.001	0.00575	0.001	0.00053	0.149	-0.00012

bFFP	Dst	0.00229	0.001	0.00004	0.756	0.00117	0.027	0.00109
BLD_s1	Dst	0.00062	0.04	0.00013	0.444	0.00036	0.22	0.00014
BLD_s6	Dst	0.00693	0.001	0.00111	0.047	0.00448	0.001	0.00135
CECSOL_s1	Dst	0.00553	0.001	0.00057	0.161	0.00454	0.001	0.00043
CECSOL_s6	Dst	0.00002	0.533	0.00001	0.867	0.00001	0.891	-0.00001
CMD	Dst	0.0052	0.001	0.00369	0.002	0.00148	0.014	0.00004
CMD_at	Dst	0.00747	0.001	0.00301	0.001	0.00457	0.001	-0.0001
CMD_sm	Dst	0.00375	0.001	0.00033	0.237	0.00201	0.004	0.00142
CMD_sp	Dst	0.00008	0.417	0.00008	0.578	0.00001	0.938	0.00001
CMD_wt	Dst	0.0008	0.064	0.00063	0.117	0.00014	0.448	0.00004
DD_0	Dst	0.00089	NA	0.00005	NA	0.0007	NA	0.00015
DD_0_at	Dst	0.00025	NA	0.00007	NA	0.00006	NA	0.00013
DD_0_sp	Dst	0.00077	NA	0.00001	NA	0.00052	NA	0.00025
DD_0_wt	Dst	0.00051	NA	0.00001	NA	0.00033	NA	0.00018
DD_18	Dst	0.00164	0.004	0.00008	0.599	0.00092	0.038	0.00066
DD_18_at	Dst	0.0015	NA	0.00001	NA	0.00104	NA	0.00046
DD_18_sm	Dst	0.00297	0.001	0.00134	0.014	0.00094	0.04	0.0007
DD_18_sp	Dst	0.00088	0.026	0.00022	0.326	0.00026	0.306	0.00042
DD_18_wt	Dst	0.00207	0.002	0.00035	0.249	0.00059	0.098	0.00114
DD18	Dst	0.00309	0.002	0.00134	0.024	0.00108	0.036	0.00068
DD18_at	Dst	0.00374	0.001	0.00086	0.068	0.00179	0.01	0.0011
DD18_sm	Dst	0.00399	0.001	0.00161	0.009	0.00162	0.008	0.00077
DD18_sp	Dst	0.00106	0.035	0.00052	0.094	0.00014	0.455	0.00041
DD18_wt	Dst	0.00163	0.016	0.00041	0.182	0.00089	0.03	0.00034

DD5	Dst	0.00221	0.001	0.00049	0.128	0.00098	0.027	0.00076
DD5_at	Dst	0.00135	0.021	0.00107	0.041	0.00001	0.966	0.00028
DD5_sm	Dst	0.00315	0.001	0.00302	0.002	0.0001	0.552	0.00004
DD5_sp	Dst	0.0014	0.045	0.00108	0.064	0.00001	0.986	0.00032
DD5_wt	Dst	0.00143	0.02	0.0011	0.036	0.00002	0.805	0.00032
eFFP	Dst	0.0007	NA	0.00039	NA	0.00001	NA	0.00031
EMT	Dst	0.0008	0.088	0.00046	0.128	0.00001	0.936	0.00035
Eref	Dst	0.00362	0.001	0.00165	0.014	0.00148	0.015	0.00051
Eref_at	Dst	0.00323	0.001	0.00225	0.003	0.00054	0.115	0.00045
Eref_sm	Dst	0.00202	0.007	0.00163	0.013	0.00025	0.305	0.00015
Eref_sp	Dst	0.0013	0.008	0.00005	0.646	0.00113	0.032	0.00013
Eref_wt	Dst	0.00188	0.002	0.00047	0.155	0.00073	0.057	0.00069
EXT	Dst	0.00348	0.001	0.00211	0.008	0.00133	0.018	0.00005
FFP	Dst	0.00012	NA	0.00005	NA	0.00008	NA	-0.00002
MAP	Dst	0.00453	0.001	0.00087	0.076	0.00329	0.002	0.00039
MAR	Dst	0.00185	0.038	0.00139	0.049	0.00071	0.12	-0.00025
MAT	Dst	0.00077	0.072	0.00067	0.059	0.00007	0.637	0.00005
MCMT	Dst	0.0017	0.002	0.00008	0.608	0.00078	0.056	0.00085
MSP	Dst	0.00237	NA	0.00033	NA	0.00116	NA	0.00088
MWMT	Dst	0.00085	0.027	0.00025	0.299	0.00046	0.158	0.00015
NFFD	Dst	0.001	NA	0.00069	NA	0.00039	NA	-0.00009
NFFD_at	Dst	0.0015	NA	0.00072	NA	0.00096	NA	-0.00018
NFFD_sm	Dst	0.00023	0.265	0.00018	0.271	0.00008	0.61	-0.00003
NFFD_sp	Dst	0.00076	0.07	0.00057	0.107	0.00008	0.602	0.00012

NFFD_wt	Dst	0.00107	0.044	0.00087	0.067	0.00022	0.351	-0.00002
ORC_s1	Dst	0.00483	0.001	0.00478	0.004	0.00007	0.67	-0.00002
ORC_s6	Dst	0.00105	0.031	0.00004	0.768	0.00098	0.079	0.00004
PAS	Dst	0.00033	NA	0.00003	NA	0.00017	NA	0.00015
PAS_at	Dst	0.00057	NA	0.00004	NA	0.0005	NA	0.00004
PAS_sm	Dst	0.00028	0.324	0.00028	0.32	0.00001	0.951	-0.00001
PAS_sp	Dst	0.00136	NA	0.00062	NA	0.00097	NA	-0.00023
PAS_wt	Dst	0.00032	NA	0.00018	NA	0.00001	NA	0.00013
PHIHOX_s1	Dst	0.00137	0.026	0.00126	0.024	0.0002	0.334	-0.00009
PHIHOX_s6	Dst	0.00177	0.003	0.00078	0.072	0.00056	0.125	0.00043
PHIKCL_s1	Dst	0.00203	0.014	0.00199	0.02	0.00002	0.832	0.00003
PHIKCL_s6	Dst	0.00105	0.015	0.00049	0.157	0.00051	0.132	0.00005
PPT_at	Dst	0.00024	0.199	0.00021	0.352	0.00003	0.736	0.00001
PPT_sm	Dst	0.0016	NA	0.00137	NA	0.00034	NA	-0.00011
PPT_sp	Dst	0.00237	0.004	0.00179	0.005	0.00001	0.851	0.00058
PPT_wt	Dst	0.00047	0.157	0.00034	0.213	0.00001	0.998	0.00013
Rad_at	Dst	0.00539	0.001	0.00351	0.014	0.00092	0.087	0.00098
Rad_sm	Dst	0.01178	0.001	0.00991	0.001	0.00107	0.061	0.00081
Rad_sp	Dst	0.00002	0.495	0.00001	0.9	0.00001	0.911	0.00001
Rad_wt	Dst	0.00255	0.002	0.00136	0.041	0.00068	0.108	0.00052
RH	Dst	0.00345	0.002	0.00073	0.087	0.00142	0.02	0.00131
RH_at	Dst	0.00352	0.001	0.00104	0.053	0.00121	0.033	0.00129
RH_sm	Dst	0.00533	0.001	0.00081	0.091	0.0023	0.004	0.00223
RH_sp	Dst	0.00574	0.001	0.00101	0.052	0.0031	0.001	0.00164

RH_wt	Dst	0.00308	0.001	0.00075	0.066	0.00116	0.028	0.00118
SHM	Dst	0.00137	0.007	0.00033	0.229	0.00066	0.09	0.00039
Tave_at	Dst	0.00128	0.009	0.00009	0.548	0.00067	0.069	0.00052
Tave_sm	Dst	0.00304	0.002	0.00291	0.003	0.00017	0.44	-0.00004
Tave_sp	Dst	0.00075	0.054	0.00034	0.226	0.00009	0.536	0.00033
Tave_wt	Dst	0.00225	0.001	0.00034	0.224	0.00069	0.072	0.00122
TD	Dst	0.00007	0.515	0.00006	0.634	0.00002	0.845	-0.00001
Tmax_at	Dst	0.00366	0.001	0.00091	0.044	0.00227	0.002	0.0005
Tmax_sm	Dst	0.00166	0.006	0.00158	0.004	0.00005	0.653	0.00003
Tmax_sp	Dst	0.00297	0.001	0.0016	0.012	0.0009	0.043	0.00048
Tmax_wt	Dst	0.00088	0.014	0.00013	0.46	0.00055	0.128	0.00021
Tmin_at	Dst	0.0007	NA	0.00013	NA	0.00015	NA	0.00043
Tmin_sm	Dst	0.0009	NA	0.00009	NA	0.00036	NA	0.00046
Tmin_sp	Dst	0.0005	0.183	0.00028	0.286	0.00001	0.949	0.00023
Tmin_wt	Dst	0.00106	0.011	0.00006	0.622	0.00037	0.21	0.00065

Table 2.S6: Fold enrichment (FE) estimates as obtained from genomic cline analyses for all 88 environmental gradients.

Variable	FE	Permutation test significance
AHM	1.096	NS
bFFP	1.278	0.05
BLD_s1	1.461	0.05
BLD_s6	1.369	NS
CECSOL_s1	1.252	NS

CECSOL_s6	1.315	NS
CMD	0.822	NS
CMD_at	0.939	NS
CMD_sm	1.252	NS
CMD_sp	1.534	0.05
CMD_wt	0.626	NS
DD_0	1.71	0.001
DD_0_at	1.71	0.001
DD_0_sp	1.717	0.001
DD_0_wt	1.773	0.001
DD_18	1.527	0.05
DD_18_at	1.511	0.05
DD_18_sm	1.643	0.05
DD_18_sp	1.555	0.05
DD_18_wt	1.565	0.001
DD18	1.16	NS
DD18_at	1.252	NS
DD18_sm	1.096	NS
DD18_sp	0.986	NS
DD18_wt	0.674	NS
DD5	1.461	0.05
DD5_at	1.402	0.05
DD5_sm	1.339	NS
DD5_sp	1.402	0.05

DD5_wt	0.974	NS
eFFP	1.588	0.001
EMT	1.546	0.001
Eref	1.315	NS
Eref_at	0.877	NS
Eref_sm	0.996	NS
Eref_sp	0.598	NS
Eref_wt	1.753	0.001
EXT	1.252	NS
FFP	1.599	0.001
MAP	1.011	NS
MAR	1.315	NS
MAT	1.575	0.001
MCMT	1.575	0.001
MSP	1.242	NS
MWMT	1.252	NS
NFFD	1.729	0.001
NFFD_at	1.534	0.001
NFFD_sm	1.136	NS
NFFD_sp	1.573	0.001
NFFD_wt	1.055	NS
ORC_s1	1.315	NS
ORC_s6	0.939	NS
PAS	1.499	0.05

PAS_at	1.482	0.05
PAS_sm	0.439	NS
PAS_sp	1.611	0.001
PAS_wt	1.461	0.05
PHIHOX_s1	1.096	NS
PHIHOX_s6	1.096	NS
PHIKCL_s1	1.565	0.05
PHIKCL_s6	0.313	NS
PPT_at	1.096	NS
PPT_sm	1.611	0.05
PPT_sp	1.408	0.05
PPT_wt	0.685	NS
Rad_at	1.315	NS
Rad_sm	0.939	NS
Rad_sp	1.729	0.001
Rad_wt	1.315	NS
RH	1.153	NS
RH_at	1.038	NS
RH_sm	1.384	0.05
RH_sp	0.974	NS
RH_wt	1.148	NS
SHM	1.429	0.05
Tave_at	1.487	0.05
Tave_sm	1.339	NS

Tave_sp	1.575	0.05
Tave_wt	1.582	0.001
TD	1.623	0.001
Tmax_at	1.348	NS
Tmax_sm	0.996	NS
Tmax_sp	1.252	NS
Tmax_wt	1.461	0.05
Tmin_at	1.539	0.001
Tmin_sm	1.388	0.05
Tmin_sp	1.675	0.001
Tmin_wt	1.629	0.001

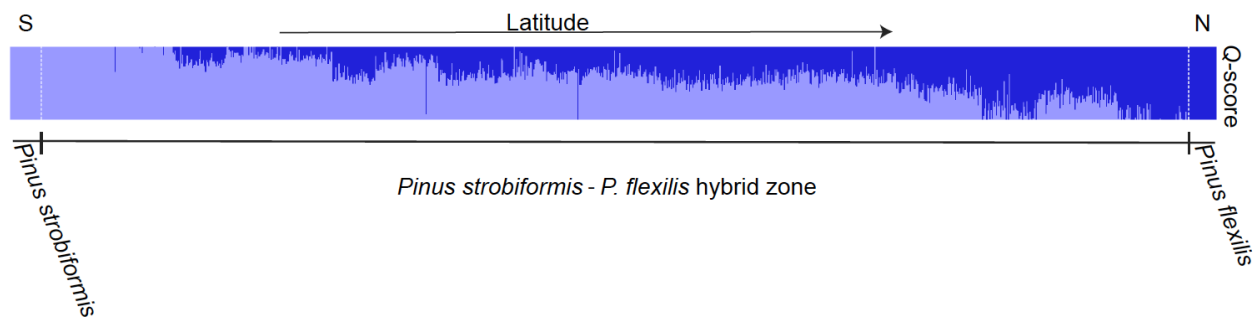


Figure 2.S1: Ancestry proportions (Q -score) for each individual tree as obtained from NGSAdmix.

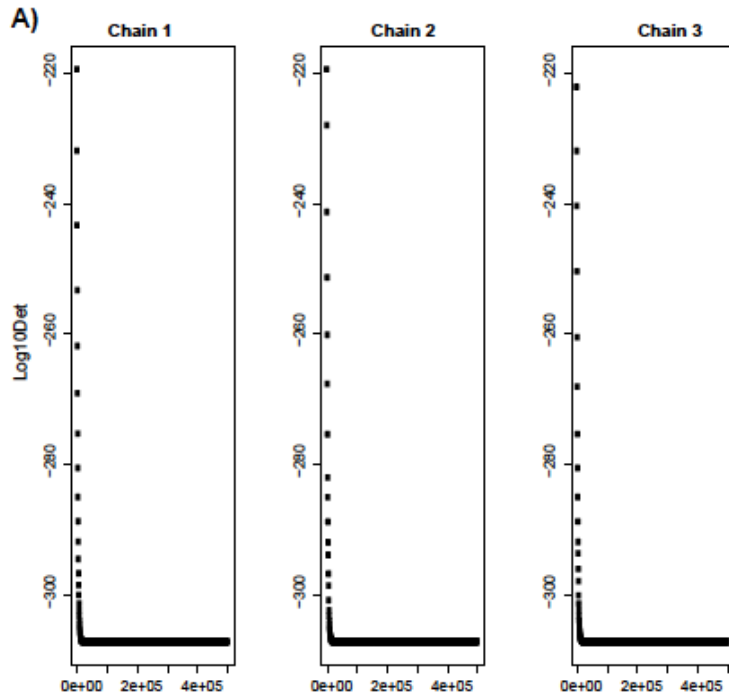
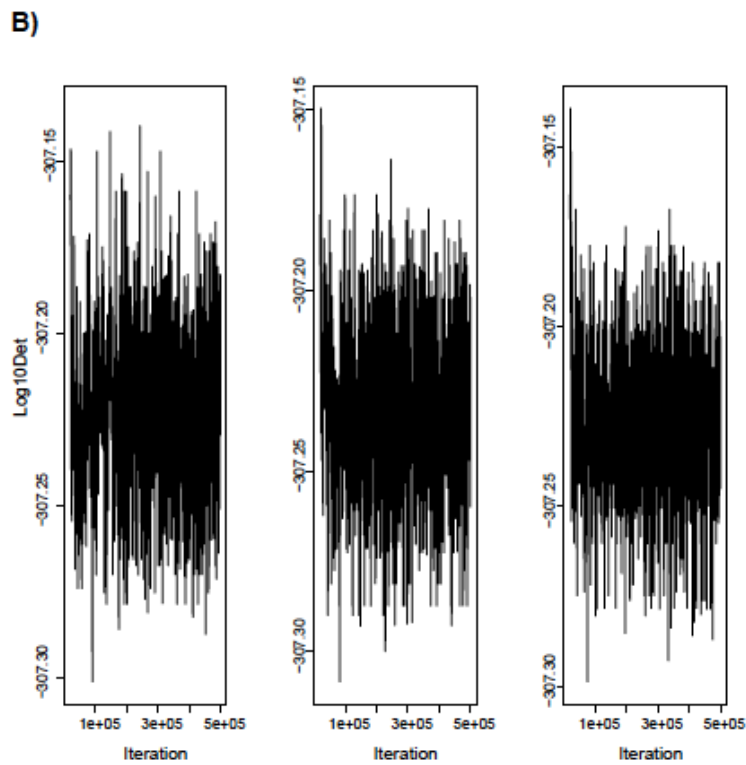


Figure 2.S2: (A) Trace plots of the determinant of the variance-covariance matrix for all 500,000 iterations across three independent Markov chains. (B) Trace plots after the burn-in, starting at iteration 20,000.



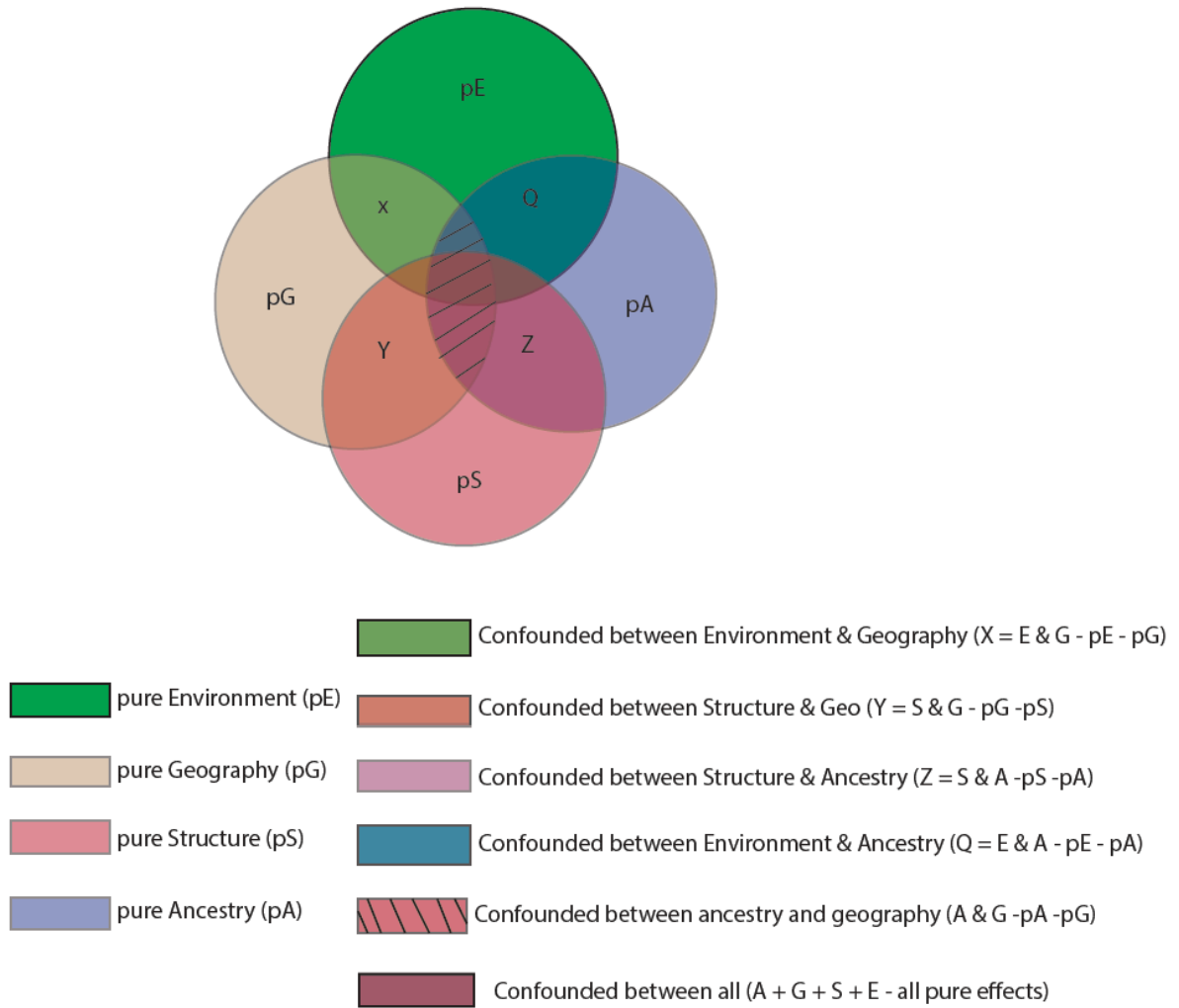


Figure 2. S3: Schematic representation of the sequential variance partitioning approach used in RDA to estimate pure, joint and confounded effects.

REFERENCES

1. Abbott, RD et al. Hybridization and speciation. *J. Evol. Bio.* 26:229–246. (2013)
2. de Lafontaine G & Bousquet Jean. Asymmetry matters: A genomic assessment of directional biases in gene flow between hybridizing spruces. *Ecol. Evol.* 7: 3883-3893. (2017)
3. Todesco, M et al. Hybridization and extinction. *Evol. Appl.* 9: 892–908. (2016)
4. Anderson & Stebbins. Hybridization as an evolutionary stimulus. *Evolution.* 8, 378-388 (1954).
5. De La Torre, A. R., Li, Z., Van de Peer, Y., & Ingvarsson, P. K. Contrasting rates of molecular evolution and patterns of selection among gymnosperms and flowering plants. *Mol. Bio. Evol.*, 34:1363– 1377 (2017).
6. Critchfield, W.B. Hybridization and classification of the white pines (*Pinus* section *Strobus*). *Taxon.* 35, 647-656 (1986).
7. Charlesworth B. Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* 10: 195–205 (2009).
8. Bouille M, Bousquet J. Trans-species shared polymorphisms at orthologous nuclear gene loci among distant species in the conifer *Picea* (Pinaceae): Implications for long term maintenance of genetic diversity in trees. *Am. J. Bot.* 92: 63-73 (2005).
9. Hamilton, J.A, Lexer, C, Aitken, S.N. Genomic and phenotypic architecture of a spruce hybrid zone (*Picea sitchensis* × *P. glauca*). *Mol. Ecol.*, 22, 827–841 (2013).
10. Hamilton, J., Miller, J. Adaptive introgression as a resource for management and genetic conservation in a changing climate. *Conserv. Biol.* 30, 33-41 (2016).
11. Bresadola L et al. Admixture mapping in interspecific *Populus* hybrids identifies classes of genomic architectures for phytochemical, morphological and growth traits. *New Phytol.* 223:2076-2089 (2019).
12. Suarez-Gonzalez, A. et al. Genomic and functional approaches reveal a case of adaptive introgression from *Populus balsamifera* (balsam poplar) in *P. trichocarpa* (black cottonwood). *Mol. Ecol.* 25: 2427–2442 (2016).
13. Suarez-Gonzalez, A., Hefer, CA., Lexer, C., Cronk, QC., and Douglas, CJ. Scale and direction of adaptive introgression between black cottonwood (*Populus trichocarpa*) and balsam poplar (*P. balsamifera*). *Mol. Ecol.* 27: 1667–1680 (2018).
14. Leroy et al. Adaptive introgression as a driver of local adaptation to climate in European white oaks. *New. Phy.* (2019).

15. Hufford et al. Genomic signature of crop-wild introgression in Maize. *PLoS Genet.* **9**: e100347. (2013).
16. Takuno et al. Independent Molecular Basis of Convergent Highland Adaptation in Maize. *Genetics.* **200**: 1297-1312 (2015).
17. Ma, Y et al. Ancient introgression drives adaptation to cooler and drier mountain habitats in a cypress species complex. *Communi. Biol.* **18**: 210-213 (2019)
18. Pyhäjärvi T, Hufford MB, Mezouk S, Ross-Ibarra J. Complex patterns of local adaptation in Teosinte. *Genome Biol. Evol.* **5**: 1594-1609 (2013).
19. Mei W, Stetter MG, Stitzer MC. Adaptation in plant genomes : Bigger is different. *Am. J. Bot.* **105**: 16-19 (2019).
20. Menon M et al. The role of hybridization during ecological divergence of southwestern white pine (*Pinus strobiformis*) and limber pine (*P. flexilis*). *Mol. Ecol.* **27**: 1245–1260. (2018).
21. Looney, CE., & Waring, KM. *Pinus strobiformis* (southwestern white pine) stand dynamics, regeneration, and disturbance ecology: A review. *Forest Ecol. Manag.* **287**, 90–102 (2013).
22. Frankis MP. The high altitude white pines (*Pinus* L. subgenus *Strobus* Lemmon, Pinaceae) of Mexico and the adjacent SW USA. *International Dendrology Society Yearbook 2008*: 64-72 (2009).
23. Tomback DF et al. Seed dispersal in limber and southwestern white pine: comparing core and peripheral populations. In: *The Future of High Elevation, Five-Needle White Pines in Western North America: Proceedings of the High Five Symposium. Proceedings RMRS-P- 63*. Fort Collins, CO: US Department of Agriculture, Forest Service, Rocky Mountain Research Station, pp. 69–71 (2011).
24. Bisbee J. Cone morphology of the *Pinus ayacahuite-flexilis* complex of the southwestern United States and Mexico. *Bulletin of the Cupressus conservation project.* **3**: 3–33 (2014).
25. Neale DB & Kremer A. Forest tree genomics: growing resources and applications. *Nat. Rev. Genet.* **12**: 111–122 (2011).
26. Schoettle, AW and SG Rochelle. Morphological variation of *Pinus flexilis* (Pinaceae), a bird-dispersed pine, across a range of elevations. *Am. J. Bot.* **87**:1797-1806 (2000).
27. Moreno-Letelier, A., Ortíz-Medrano, A. & Pinaero, D. Niche Divergence versus Neutral Processes: Combined Environmental and Genetic Analyses Identify Contrasting Patterns of Differentiation in Recently Diverged Pine Species. *PLoS ONE* **8**, e78228 (2013).
28. Hayhoe, KD et al. Emissions pathways, climate change, and impacts on California. *Proc. Natl Acad.Sci. USA* **101**:12422–12427 (2004).

29. Bell, JL., Sloan LC, and Snyder MA. Regional changes in extreme climatic events: a future climate scenario. *J. Climate* **17**:81-87 (2004).
30. Harrison KA et al. Signatures of polygenic adaptation associated with climate across the range of a threatened fish species with high genetic connectivity. *Mol. Ecol.* **26**: 6253-6269 (2017).
31. Schumer M & Brandvain Y. Determining epistatic selection in admixed populations. *Mol. Ecol.* **25**: 2577-2591 (2016)
32. Menon et al. Tracing the footprints of a moving hybrid zone under a demographic history of speciation with gene flow. *Evol. Appl.* (2019).
33. Whitney, KD et al. Quantitative trait locus mapping identifies candidate alleles involved in adaptive introgression and range expansion in a wild sunflower. *Mol. Ecol.* **24**: 2194-2211 (2015).
34. Chhatre VE, Evan LM, DiFazio SP & Keller SR. Adaptive introgression and maintenance of a trispecies hybrid complex in range-edge populations of *Populus*. *Mol. Ecol.* **27**: 4820-4838 (2018).
35. Aitken SA, Yeaman S, Holliday JA, Wang T, and Curtis-McLane S. Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evol. Appl.* **1**: 95-111 (2008).
36. Alberto FJ et al. Potential for evolutionary responses to climate change – evidence from tree populations. *Glob. Chn. Bio.* **19**: 1645-1661. (2013)
37. Kirkpatrick, M & Barton, NH. Evolution of a species' range. *Am. Nat.* **150**:1–23 (1997).
38. Stebbins, GL. The role of hybridization in evolution. *Proc. Am. Phil. Soc.*, **103**: 231–251 (1959).
39. Petit, R.J & Excoffier, L. Gene flow and species delimitation. *Trends Ecol. Evol.* **24**, 386–393 (2009).
40. McClintock, B. The significance of responses of the genome to challenge. *Science*, **226**: 792–801 (1984).
41. Kawakami T et al. Transposable Element Proliferation and Genome Expansion Are Rare in Contemporary Sunflower Hybrid Populations Despite Widespread Transcriptional Activity of LTR Retrotransposons. *Genome Biol. Evol.* **3**: 156–167. (2011).
42. Barton, NH & Hewitt, GM. Analysis of hybrid zones. *Annu. Rev. Ecol. Evol. S*, **16**:113–148 (1985).
43. Mimura M, Mishima M, Lascoux M & Yahara T. Range shift and introgression of the rear and leading populations in two ecologically distinct *Rubus* species. *BMC Evol. Biol.* **2014**: 209 (2014).

44. De La Torre, AR., Wang, T., Jaquish, B. & Aitken, SN. Adaptation and exogenous selection in a *Picea glauca* × *Picea engelmannii* hybrid zone: implications for forest management under climate change. *New Phytol.* **201**, 687–699 (2014).
45. Hamilton JR, De La Torre AR & Aitken SN. Fine-scale environmental variation contributes to introgression in a three-species spruce hybrid complex. *Tree Genet. Genomes* **11**: 1-14 (2015).
46. Fraïsse CK, Belkhir J, Welch J & Bierne N. Local Interspecies Introgression Is the Main Cause of Extreme Levels of Intraspecific Differentiation in Mussels. *Mol. Ecol.* **25**: 269–770 (2016).
47. Wu DD et al. Pervasive introgression facilitated domestication and adaptation in the Bos species complex. *Nature Ecol. Evol.* **2**: 1139-1145 (2018).
48. Kremer A & Le Corre V. Decoupling of differentiation between traits and their underlying genes in response to divergent selection. *Heredity*, **10**•, 375–385. (2012).
49. Eckert AJ et al. Local adaptation at fine spatial scales: an example from sugar pine (*Pinus lambertiana*, Pinaceae). *Tree Genet. Genomes* **11**: 42 (2015).
50. Hornoy B, Pavy N, Gérardi S, Beaulieu J & Bousquet J. Genetic adaptation to climate in white spruce involves small to moderate allele frequency shifts in functionally diverse genes. *Genome Biol. Evol.* **7**: 3269-3285 (2015).
51. Lind BM et al. Water availability drives signatures of local adaptation in whitebark pine (*Pinus albicaulis* Engelm.) across fine spatial scales of the Lake Tahoe Basin, USA. *Mol. Ecol.* **26**: 3168–3185. (2017).
52. Rieseberg LH et al. Hybridization and the colonization of novel habitats by annual sunflowers. *Genetica*, **129**: 149–165. (2007).
53. Lewontin RC & Birch LC. Hybridization as a source of variation for adaptation to new environments. *Evolution*, **20**: 315-336 (1966).
54. Kim BY, Huber CD & Lohmueller K. Deleterious variation shapes the genomic landscape of introgression. *PLoS Genet.* **14**: e1007741 (2018).
55. Bierne N, Welch J, Loire E, Bonhomme F & David P. The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Mol. Ecol.* **20**:2044-2072 (2011).
56. Eyre-Walker A, Woolfit M & Phelps T. The Distribution of Fitness Effects of New Deleterious Amino Acid Mutations in Humans. *Genetics*. **173**: 891-900 (2006).
57. Christie C et al. Adaptive evolution and segregating load contribute to the genomic landscape of divergence in two tree species connected by episodic gene flow. *Mol. Ecol.* **26**: 59-76 (2017).
58. Lu M, Hodgins KA, Degner JC & Yeaman S. Purifying selection does not drive signatures of convergent local adaptation of lodgepole pine and interior spruce. *BMC Evol. Biol.* **19**:110 (2019)

59. Whitlock MC. Temporal fluctuations in demographic parameters and the genetic variance among populations. *Evolution* **46**: 608–615 (1992)
60. Lexer C et al. Genomic Admixture Analysis in European *Populus* spp. Reveals Unexpected Patterns of Reproductive Isolation and Mating. *Genetics*. **186**: 699-712 (2010).
61. Lowry D et al. *Mol. Ecol. Resources*. **17**: 142–152 (2017).
62. Parchman TL et al. RADseq approaches and applications for forest tree genetics. *Tree Genet. Genomes*. **14**: 39 (2018).
63. Gossmann TI, Keightley PD and Eyre-Walker A. The effect of variation in the effective population size on the rate of adaptive molecular evolution in eukaryotes. *Genome Biol. Evol.*, **4**: 658-667. (2012)
64. Liu B & Wendel JF. Retrotransposon activation followed by rapid repression in introgressed rice plants. *Genome*. **43**:874-80. (2000)
65. Nystedt B et al. The Norway spruce genome sequence and conifer genome evolution. *Nature* **497**:579–584 (2013).
66. Stevens KA et al. Sequence of the sugarpine megagenome. *Genetics*. **204**:1613-1626 (2016).).
67. Colosimo, PF et al. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* **307**: 1928-1933 (2005).
68. Jagoda E et al. Disentangling Immediate Adaptive Introgression from Selection on Standing Introgressed Variation in Humans. *Mol. Biol. Evol.* **35**: 623–630. (2018)
69. Skotte L Korneliussen TS & Albrechtsen A. Estimating individual admixture proportions from next generation sequencing data. *Genetics* **195**: 693–702. (2013)
70. Lotterhos K and Whitlock M. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Mol. Ecol.* **24**: 1031-46. (2015)
71. Parchman TL et al. Genome -wide association genetics of an adaptive trait in lodgepole pine: Association mapping of serotiny. *Mol. Ecol.* **21**: 2991 –3005. (2012)
72. Puritz JB, Hollenbeck CM & Gold JR. dDocent : a RADseq, variant -calling pipeline designed for population genomics of non -model organisms. *PeerJ*, **2**: e431. (2014)
73. Wang T, Hamann A, Spittlehouse DL, & Carroll C. Locally downscaled and spatially customizable climate data for historical and future periods for North America. *PLoS One* **11**: e0156720. (2016)
74. Hengl T et al. SoilGrids1km — Global Soil Information Based on Automated Mapping. *PLoS ONE* **9**: e105992 (2014)
75. Coop G, Witonsky D, Di Rienzo A & Pritchard JK. Using environmental correlations to identify loci underlying local adaptation. *Genetics*, **185**: 1411–1423 (2010)

76. Günther T & Coop G. Robust identification of local adaptation from allele frequencies. *Genetics*. **195**: 205–220 (2013).
77. Brooks, S.P & Gelman, A. General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Statist.* **7**: 434–455 (1998).
78. Goudet J. hierfstat, a package for R to compute and test hierarchical F -statistics. *Mol. Ecol. Notes*. **5**: 184 –186 (2005).
79. G Warnes, G Gorjanc, F Leisch & M Man. *genetics: Population Genetics*. R package version 1.3.8.1. 2013)
80. R Core Team. R v.3.3.2: A Language and Environment for Statistical Computing. Vienna, Austria : the R Foundation for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2017) .
81. Oksanen J et al. *vegan: Community Ecology Package*. R package version 2.5-2. (2013)
82. Legendre P & Legendre L. *Numerical ecology*, 2nd English edn. Elsevier, Amsterdam (1998)
83. Liu Q. Variation partitioning by partial redundancy analysis (RDA). *Environmetrics* **8**: 75-85 (1997).
84. Kempainen, P. *et al.* Linkage disequilibrium network analysis (LDna) gives a global view of chromosomal inversions, local adaptation and geographic structure. *Mol. Ecol. Resour.* **15**: 1031–1045 (2015).
85. Ohta, T. Linkage disequilibrium with the island model. *Genetics*, **101**: 139– 155. (1982)
86. Csillery K et al. Detecting short spatial scale local adaptation and epistatic selection in climate-related candidate genes in European beech (*Fagus sylvatica*) populations. *Mol. Ecol.* **23**: 4696–4708. (2014)
87. Beissinger TM et al. Using the variability of linkage disequilibrium between subpopulations to infer sweeps and epistatic selection in a diverse panel of chickens. *Heredity*, **116**: 58– 166 (2015)
88. Hijmans RJ. *geosphere: Spherical trigonometry*. R package version 1.5-7. (2017)
89. Adamack, AT & Gruber, B. PopGenReport: simplifying basic population genetic analyses in R. *Methods Ecol. Evol.*, **5**: 384-387 (2014).
90. Gompert Z & Buerkle AC. *introgress: methods for analyzing introgression between divergent lineages*. R package version 1.2.3 (2012).
91. Gompert Z. & Buerkle CA. A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Mol. Ecol.* **18**: 1207-1224 (2009).
92. Janoušek V et al. Genome-wide architecture of reproductive isolation in a naturally occurring hybrid zone between *Mus musculus musculus* and *M. m. domesticus*. *Mol. Ecol.* **21**: 3032-3047. (2012).

93. Hancock AM et al. Adaptation to climate across the *Arabidopsis thaliana* genome, *Science*.
334:83-86 (2011).

CHAPTER 3

Intrinsic and extrinsic drivers of among population gene expression differentiation across the *P. strobiformis*-*P. flexilis* hybrid zone

Introduction

The ability of species to survive under rapidly changing climatic conditions will depend on the alignment of their multivariate trait values with the local fitness optima and on the availability of genetic variation to drive evolution towards the shifted fitness optima (Price et al. 2003; Jump et al. 2009). The former can be facilitated by environment dependent expression of trait values (i.e., phenotypic plasticity). Phenotypic plasticity is ubiquitous across organisms and is hypothesised as a major mechanism facilitating survival under changing climatic conditions (Bradshaw, 2006; Nicotra et al. 2015; Benito Garzon et al. 2019). The potential for phenotypic plasticity to facilitate long-term persistence of populations, however, relies on the level of genetic variation underlying reaction norms (West-Eberhard, 1989). Decades of common garden experiments in plants provide remarkable evidence for genetic variation in plasticity indicated by the differential performance of genotypes across gardens which generates genotype-by-environment effects (GxE, reviewed in Savolanein et al. 2013). While GxE effects provide raw material for populations to respond to selective pressures, it does not by itself provide evidence for adaptive evolution unless the environmentally induced phenotype is demonstrated to improve fitness either in one (i.e. conditionally adaptive) or across all measured environments (i.e. adaptive plasticity, Pigliucci & Schlichting, 1996; Baythavong & Stanton, 2010). Demonstrating a relationship between plasticity and fitness can be challenging in long-lived organisms such as

trees. By conceptualising plasticity as a quantitative trait some of these challenges can be alleviated as it facilitates the application of approaches uniting the influence of both selection and neutral processes on the evolution of phenotypic plasticity (Falconer & McKay, 1996; McKay & Latta, 2002). When the extent of among population variation in reaction norms is higher than that expected from neutral processes alone, it can be taken as evidence for adaptive plasticity (reviewed in Josephs, 2018).

While examples of variation in phenotypic plasticity among populations are plenty, only a few empirical studies have shed light on the evolutionary processes driving plasticity (Williams et al. 1996; Schmid et al. 2019). As populations move away from their optimal environmental conditions, especially at range margins, an increase in slope of the reaction norm is expected (Wright, 1982; Kirkpatrick & Barton, 1997). In empirical studies, this is often demonstrated as pronounced environmental (E) or GxE for populations originating from heterogeneous environments (Akman et al. 2018), which is typical for range margin populations. If these populations continue to experience predictable environmental heterogeneity, selection can facilitate the evolution of adaptive phenotypic plasticity (Via & Lande, 1985; Ghalambor et al. 2007). In addition to the influence of environmental conditions at the source population's location, neutral processes such as range expansions and patterns of gene flow could also drive plasticity (Schmid et al. 2019). For instance, in range margin populations higher landscape fragmentation is often accompanied with lower genetic diversity which can restrict the evolution of phenotypic plasticity as well as its potential to be adaptive (Eckert et al. 2008). Thus, disentangling how demographic and climatic processes vary among populations is important to determine the environmental context of adaptive trait differentiation.

A frequently unaccounted factor in the evolution of phenotypic plasticity is the influence of interspecific gene flow across hybrid zones defined by abutting ranges of sister taxa. Interspecific gene flow can facilitate range expansion and aid adaptation to novel selective regimes by increasing standing genetic variation through the generation of novel allelic combinations (Barton, 2001; Taylor & Larson, 2019). Within a hybrid zone, GxE effects can manifest as an interaction of the individual's genotype with the environment, the interaction of a specific allele in an individual with its genomic background or both (Dlugosch et al. 2015; Gould et al. 2018). A handful of studies utilizing natural hybrid zones have demonstrated hybrid class dependent expression of trait values, such that F1s are usually intermediate or exhibit transgressive phenotypes, but backcrossed individuals are more reflective of the parent with which they share a greater genomic composition (Silim et al. 2001; Schweitzer et al. 2002; Hegarthy et al. 2008; Favre & Karrenberg, 2011). Thus, despite the prevalence of hybridization and phenotypic plasticity across the Tree of Life, we lack clear empirical support for whether and how hybrid ancestry influences the evolution of phenotypic plasticity.

Common garden experiments for plants usually focus on fitness-related traits such as height, mortality and phenology to infer signatures of adaptive trait differentiation and of phenotypic plasticity. These traits are influenced by a cascade of molecular phenotypes that can be measured through gene expression profiling (Aubin-Horth & Renn, 2009; Richards et al. 2012). Gene expression variation in natural populations is known to be highly heritable, describes the joint influence of genetic and environmental divergence and reflects signatures of various selective pressures (Whitehead & Crawford, 2006). While earlier studies mostly focused on assessing the role of selection in driving gene expression divergence at candidate genes (Roberge et al. 2007; Reddy Palle et al. 2010; Keller et al. 2012; Menon et al. 2015), recent work

has utilised genome-wide transcriptome profiling (Lewis & Reed, 2018; Gould et al. 2018; Hamala et al. 2020). Transcriptome-wide studies help reduce ascertainment bias associated with targeted trait assays, uncover trait associations through co-expression networks and facilitate simultaneous estimation of polymorphism levels associated with traits. When combined with demographic inferences obtained from non-genic regions, gene expression profiling can also be very powerful to identify spatial variation in selection pressures. For instance, Akman et al. (2018) used source population environmental conditions to demonstrate that gene expression plasticity was adaptive only for certain populations of *Protea repens*. Using *Theobroma cacao*, Hamala et al. (2020) combined among population differentiation in co-expression networks with demographic modeling to demonstrate spatial variation in purifying selection and in adaptive trait differentiation. While within population, among populations and across environmental assessment of transcriptional changes are ample, studies combining multiple genotypes with environments to demonstrate GxE are only beginning to accumulate (see Huang et al. 2015; Akman et al. 2018). Although these studies have enhanced our understanding of differential transcriptional responses across populations and environments, the influence of interspecific gene flow on generating GxE effects due to differential interactions with the extrinsic and genomic environments are still in their infancy. As highlighted by Dlugosch et al. (2015), admixture can unmask cryptic genetic variation by converting epistatic variation into additive genetic variation, which is directly related to the ability of a trait to respond to selection (Falconer & McKay, 1996).

Our study focuses on the hybrid zone formed between two ecologically divergent species of pines, *Pinus strobiformis* and *P. flexilis*. Both species have a wide geographic distribution across the western part of North America and are known to hybridize in the sky-islands of New

Mexico, Arizona and southern Colorado (Bisbee, 2014; Menon et al. 2018). These populations likely continue to experience strong genetic drift due to the high degree of landscape fragmentation and ongoing asymmetric gene flow from *P. flexilis* which is also promoting northward range expansion (Menon et al. 2019). Being present on the western sky-islands, the hybrid zone is subjected to severe drought and seasonal frost, as well as diurnal fluctuations in temperature and solar radiation (Adams & Kolb, 2004). Current predictions by the Intergovernmental Panel on Climate change (IPCC) indicate an increase in the seasonality of drought and frost events in these sky-island landscapes (Seager and Vecchi, 2010). Theoretical predictions state that phenotypic plasticity could be a dominant mechanism facilitating response to future climatic conditions in organisms where the rate of environmental change is lower than the generation time of organisms (Levins, 1968). This will likely be the case for most conifers such as *P. strobiformis* given their long generation time and sessile nature, causing them to experience several episodes of environmental change within a lifetime. Further, if genetic variation exists for phenotypic plasticity (GxE), it can be acted upon by selection to drive the evolution of locally adapted populations. The genome-wide pattern of ancestry variation across the *P. flexilis*-*P. strobiformis* hybrid zone and locus specific enrichment of *P. flexilis* ancestry along freeze-related environmental gradients as demonstrated in Chapter 1 & 2 (Menon et al. 2018; 2020 *in review*) makes this study system ideal to address the complex interaction between climate and demographic processes in determining the evolution of phenotypic plasticity and ascertaining whether trait differentiation as well as GxE effects are driven by divergent selection.

The overarching goal of this work is to leverage hybrid ancestry information and source climatic conditions within a common garden framework to evaluate patterns of gene expression GxE. Preliminary physiological and phenotyping assays in this system have demonstrate strong

GxE component and hybrid individuals exhibiting higher survival across an array of environmental conditions (DaBell, 2017; Buchlotz, 2020). Gene expression profiling provides a way to complement standard trait measurements and helps assess a multitude of physiological responses simultaneously. Specifically, we aim to answer the following questions: 1) What are the environmental drivers of common garden-specific expression divergence among populations? 2) What proportion of transcripts exhibit signatures of adaptive plasticity and of conditional neutrality? 3) What is the contribution of hybrid ancestry towards population level transcriptome divergence?

Methods

Field sampling and climate dataset

Seeds from open-pollinated cones were collected across the full range of *P. strobiformis*, including the hybrid zone. Being open pollinated, we assume that the resulting progeny were a 50:50 mixture of half-sibs and full-sibs and henceforth refer to our design as mixed-sib (Squillace, 1974). Assuming a mixed-sib design would be conservative since it underestimates the additive genetic variance relative to that from a half-sib design. The mixed-sib progenies were planted across two common gardens established in the North Kaibab National Forest, AZ (Fig 3.1) that were a part of the Southwest Experimental Garden Array (SEGA). Bear Springs (BS) is a high elevation garden that experiences cool temperatures and precipitation in the form of snow. White Pockets (WP) is a low elevation garden that experiences warm and dry conditions. Within each garden, individual trees were planted in a randomized block design between the years of 2015 and 2017. In fall 2019, we sampled 180 trees representing 90 mixed-sibs (3 sibs/maternal tree) planted across the two common gardens that represented 30 maternal

trees originating from 10 populations (3 maternal trees/population) (Fig. 3.1). The sampling was done across two clear weather condition days in August between the hours of 12:00 and 17:00. For each tree, up to three current year's fascicles were sampled and quickly flash frozen in liquid nitrogen. We utilised latitude and longitude to obtain annual and seasonal climate variables at a resolution of 1 km from ClimateWNA v6.1 (Wang et al. 2016) for the normal 1981-2010 for each of the 30 maternal trees. Population level estimates were obtained by averaging the values across the three maternal trees representing a population.

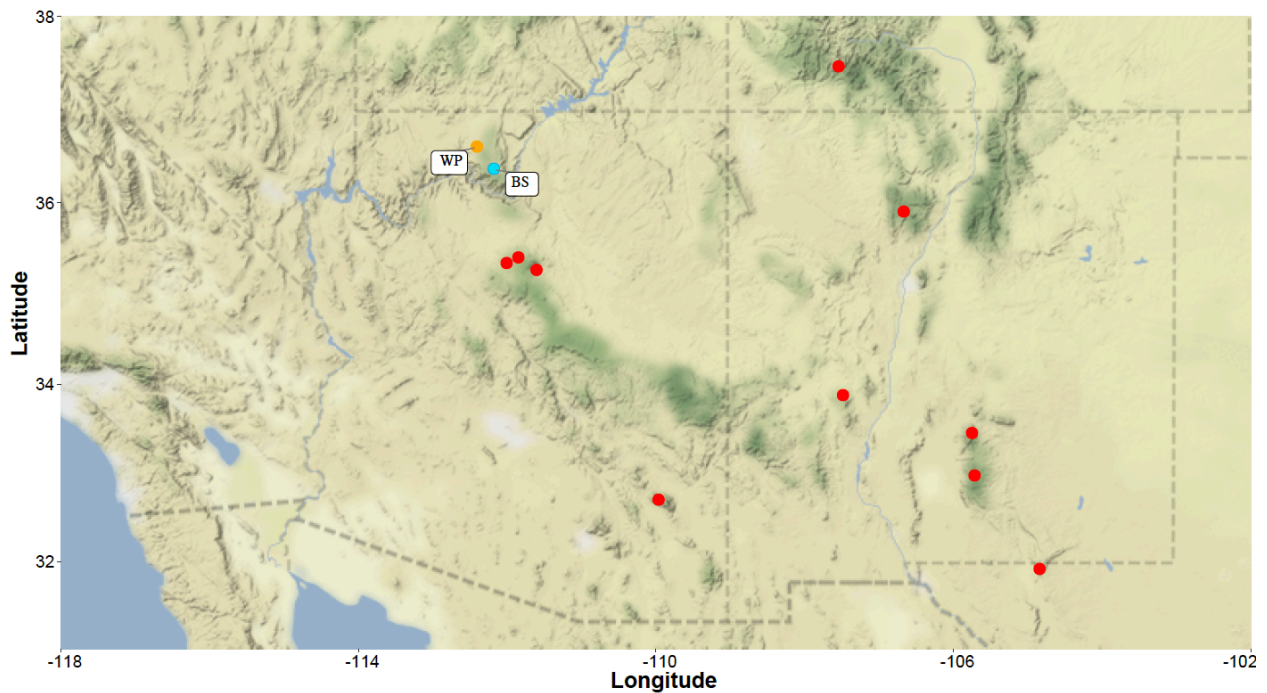


Figure 3.1: Geographic location of sampled populations (red dots) and the two common gardens, Bear Springs (BS) & White Pockets (WP). Bear Springs is the high elevation common garden indicated in turquoise and White Pockets is the low elevation garden indicated in orange.

RNA isolation and transcriptome assembly

A maximum of 100 mg of needle was ground into a fine powder using liquid nitrogen in a mortar and pestle with the addition of 40 mg of Polyvinylpyrrolidone (PVP-40). Total RNA was extracted using Spectrum Plant total RNA kit with modifications to the manufacturer's protocol. Genomic DNA contamination was removed using the On-column DNase1 kit from Sigma. RNA concentration and purity were assessed using NanoDrop 2000, gel electrophoresis and finally through RNA integrity number from Bioanalyser run on an Agilent 2100. Post polyA tail selection to enrich for mRNA, paired-end ($2 \times 150\text{bp}$) RNA libraries were prepared following standard protocols for NEBNext Ultra II RNA Library Prep by Illumina and sequenced on Novaseq6000. All steps starting from polyA tail selection to sequencing were conducted at Novogene (Sacramento, CA).

Raw fastq files were assessed for quality using fastqc and subsequently trimmed for quality and to remove adapters using TrimGalore v.0.6.4 (Krueger, 2015). For species harbouring high levels of genetic diversity and large intergenic spaces, transcriptomes built by concatenating several individual reads can generate a fragmented assembly and underestimate heterozygosity since multiple reads across individuals will often be mapped to different transcripts (*pers. comm.* J. Wegrzyn). To avoid this issue, we built 20 individual transcriptome assemblies using one representative from each of our 10 populations across the two common gardens. These 20 *de novo* transcriptome assemblies were built using Trinity v.2.8.5 (Grabherr et al. 2011; Haas et al. 2013) with a *k*-mer length of 24 and a minimum contig size of 300bp. Redundancy reduction of this hyper-assembly was conducted through the tr2aacds pipeline in EvidentialGene v.2018.06.18 (Gilbert, 2013). The pipeline implemented within EvidentialGene first predicts coding regions and then removes transcripts that (a) are completely redundant, (b) exhibit high clustering similarity and (c) are perfect fragments of larger transcripts. Transcripts

that passed these steps were queried against the UniprotKB/Swissprot (release 2019_11) Diamond database implemented within the EnTap pipeline v.0.9.1 (Hart et al. 2019) to filter out fungal and bacterial contamination from our reference transcriptome. This process retained a total of 192,987 transcripts and 186,117 genes that were the starting point of further analyses.

We utilised a series of steps to assess the quality and the completeness of our transcriptome. First, we estimated the median length of contigs in our assembly (N50) using only the longest isoform per gene as well as the expression level sensitive matrix that is considered suitable for RNAseq datasets called ExN50 (Haas et al. 2013). This process was done to evaluate the quality of our assembly and to explore the saturation of full length constructed transcripts as a function of the read depth. The N50 for our dataset was 654bp and Ex peak noted at E90 was 2kb corresponding to 13,872 genes out of the total 186,117 genes. Second, the completeness of our assembly was assessed by using *Arabidopsis thaliana* as the reference database in the orthology-based algorithm implemented in BUSCO v.2.0 (Seppey et al. 2019). This approach identified 83.8% of the transcripts as matching and complete and 13% as missing. To evaluate the extent of conifer specific transcriptome space covered in our assembly we queried our transcripts against the publicly available *P. lambertiana* v.1.0 and *P. flexilis* (Liu et al. 2016) transcriptomes using blastp with an e-score threshold of 10^{-50} and 10^{-100} , respectively. This process yielded 26,296 matches against the *P. lambertiana* transcriptome, covering 45% of the transcriptome space in *P. lambertiana*. For *P. flexilis* we identified 34,438 matches that covered 20% of the transcriptome space in *P. flexilis*. Third, we assessed mapping rates for individual fastq files to determine whether a sufficient amount of the transcriptome space was covered in the *de novo* assembled transcriptome. This process indicated an average mapping rate of 73% across 180 samples. Further, the mapping rate did not exhibit an upward bias towards the 20

samples that were used in the *de novo* assembly. The assembled transcriptome was annotated through the implementation of uniprotKB and PlantRef database (release 98) within the EnTAP pipeline that was specifically designed for non-model organisms. Read counts for each individual tree were extracted using RSEM with bowtie v.2.0 (Langmead & Salzberg, 2012) and ranged from 13 to 28 million reads per sample. Lowly expressed transcripts were filtered using the lowest level of hierarchy in our dataset, i.e. we removed any transcript that was not expressed across all three sibs of a maternal tree in either garden. This process resulted in a dataset of 70,250 transcripts.

Estimation of maternal trait value

Prior to conducting further analyses using the 70,250 transcripts, we normalised the read count data for varying library sizes using the Trimmed mean of M values (TMM) approach implemented within the `calcNormFactors` function from edgeR v.3.14.0 (Robinson et al. 2010) in R v.3.6.3 (R core team, 2020). We utilized the mixed-sib design to determine each maternal tree's normalised expression value for each of the 70,250 transcripts using the following equation:

$$Y_{ijklm} = \mu + Batch_i + SY_j + (Garden_k | Pop_l / Fam_m) + \epsilon \quad (1)$$

Here, Y represents the normalised expression of the transcript, μ represents the global mean for the transcript, $Batch$ represents the date of sequencing, SY is the year that the seedlings were planted, maternal family (Fam) is nested within the population (Pop) and its effect on Y is considered to vary by $Garden$. This mixed effect model was fitted using the `variancePartition` v.1.16.1 (Hoffman & Schadt, 2016) package in R, which utilises log-transformed normalised counts and precision weights to incorporate the mean-variance trend typical of gene expression

datasets. Using the fitted model in (1), we then obtained garden-specific maternal values and garden-specific population values with the following:

$$MatVal_{ijk} = \mu + Fam_i + Pop_j + Garden_k \quad (2)$$

$$PopVal_{jk} = \mu + Pop_j + Garden_k \quad (3)$$

Here, μ represents the global intercept, Fam_i represents the effect of maternal family i , Pop_j represents the population from which maternal tree i originates and $Garden_k$ represents the effect of the Garden k .

Estimating the effect of environment on among population differentiation

To determine how gene expression varies among populations and gardens, we utilised the variance components for family and population from equation (1) and estimated garden specific measures of gene expression differentiation. Treating expression as a quantitative trait enabled us to estimate the degree of among population divergence (Q_{ST}) attributed to heritable variance components (Spitze 1993). Given overall low levels of genome wide differentiation in our study system (multilocus $F_{ST} = 0.007$, Menon et al. 2018) we declared all transcripts with a $Q_{ST} \geq 0.5$ to be strongly differentiated. To identify climatic drivers of the among population differentiation in gene expression for the strongly differentiated transcripts, we utilised redundancy analysis (RDA) implemented in the vegan package v.2.5.6 (Oksanen et al. 2019) in R. Here, we used population-level estimates of gene expression from equation (3) as the response matrix along with drought-related variables, freeze-related variables and geography as the predictor matrices. Classification of climatic variables into freeze or drought related is the same as used in Chapter 2 and is presented in Table 2.S1. To reduce the dimensionality of each of these matrices, we performed principal component analyses (PCAs) and only retained PC axes that jointly explained up to 90% of the variance in each predictor dataset. For geography, we utilised the scaled and

centred estimates of latitude and longitude. We built three models in RDA for each of the two gardens. Our first model used drought-related, freeze-related and geography-related variables as the predictors. The second model used only drought-related and geography-related variables, while the third model used only freeze-related and geography-related variables as the predictors. Finally, we generated an empirical null distribution of 10,000 adjusted R^2 values from non-highly differentiated transcripts ($Q_{ST} < 0.5$) to test the statistical significance of models 2 and 3. For each garden, we compared our observed R^2 for the highly differentiated transcripts against these empirical null distributions. Observed R^2 values located in the upper tail of the empirical distribution for model 2 would indicate an important role of drought-related variables, while higher than expected observed R^2 for model 3 would indicate an important role of freeze-related variables in driving population differences in expression values for the highly differentiated transcripts ($Q_{ST} \geq 0.5$).

Assessing signatures of adaptive differentiation

Similar to Q_{ST} , F_{ST} describes the degree of divergence in allele frequency among populations. While Q_{ST} - F_{ST} comparisons are widely used to assess divergent selection on phenotypic and physiological traits, only a handful of studies have implemented them for gene expression data (Gibson & Weir, 2005; Roberge et al. 2007). To identify signatures of adaptive evolution and of adaptive plasticity across the gardens for each transcript, we made use of a formal comparison between F_{ST} and Q_{ST} implemented in the $Q_{ST}F_{ST}Comp$ package (Gilbert & Whitlock, 2015) in R. Given our assumption of a mixed-sib design, the family variance used here is $1/3^{\text{rd}}$ of the additive genetic variance. To incorporate the garden design in our estimation of Q_{ST} , we first regressed out the effect of SY and Batch and used the residual normalised log count expression values for each individual tree to assess signatures of selection. Estimates for F_{ST}

were obtained from putatively neutral single nucleotide polymorphisms (SNPs) generated through ddRADseq genotyping (see next section) for the same set of maternal trees as used in our RNAseq. Since F_{ST} describes a null expectation for the degree of divergence due to neutral processes such as drift and gene flow, transcripts with Q_{ST} estimates significantly higher than the multilocus F_{ST} (upper tail p-value < 0.05) were declared as being adaptive. Our goal here was not only to identify transcripts exhibiting signatures of adaptive evolution, but to categorise transcripts into those exhibiting (a) adaptive differentiation but no plasticity across gardens, where $Q_{ST} > F_{ST}$ for both gardens but the Q_{ST} reaction norm has a slope of zero, (b) adaptive plasticity, where $Q_{ST} > F_{ST}$ across both gardens and Q_{ST} reaction norm has a slope greater than 0 and (c) conditionally adaptive, where $Q_{ST} > F_{ST}$ only in one of the gardens. This then defines four categories of transcripts, namely: adaptive differentiation only, adaptive plasticity, conditionally adaptive in BS and conditionally adaptive in WP.

Hybrid index and differential expression analyses

Since we were interested in assessing whether hybridization influences changes in expression patterns across gardens, we obtained hybrid index (HI) estimates for each of our maternal trees from a larger dataset that sampled several populations across the range of *P. strobiformis*, *P. flexilis* and the hybrid zone. Specifically, we genotyped 282 individuals from 80 populations (including the 10 used for RNAseq) using ddRADseq and called SNPs using the dDocent pipeline (Puritz et al. 2014), as implemented in Chapters 1 and 2 (Menon et al. 2018; 2020 *in review*). The raw set of SNPs were further subjected to a series of filters based on biallelic SNPs, indels, missing data, minor allele frequency cutoff of 0.01, F_{IS} , depth and quality. For these we followed a similar procedure as listed in Chapters 1 and 2 (Menon et al. 2018; 2020 *in review*). We obtained estimates of hybrid index using the genotype likelihood based clustering

algorithm implemented in NGSAdmix (Skotte et al. 2013) with $K=2$ to represent the two focal species (as done in Menon et al. 2018).

To determine whether hybrid ancestry contributed towards strong population differentiation and towards adaptive evolution, we correlated mean population level estimate of *P. flexilis* ancestry with population level expression values for the highly differentiated transcripts at each garden. Similarly, we correlated ancestry estimates with the absolute difference in population level expression values between the two gardens to assess whether hybridization facilitates the evolution of phenotypic plasticity. Each set of observed correlation coefficients was compared against an empirical null distributions of correlation coefficients using the Kolmogorov-Smirnov test implemented in R. The empirical null was generated using all transcripts that did not fall into the category that it was being compared against.

RESULTS

Expression variation among populations and across gardens

Estimates of Q_{ST} for both gardens followed a U-shaped distribution and ranged from 0 to 1 (Fig. 3.2) with a mean of 0.29 ± 0.38 (sd) for BS and 0.35 ± 0.38 for WP, respectively. As expected, most transcripts exhibited little among population differentiation. However, several transcripts exhibited high among population differentiation ($Q_{ST} \geq 0.5$), with 17,440 (24.8%) in BS, 21,628 (30.7%) in WP and 10,724 (15%) transcripts being shared across both gardens. We used the absolute difference in Q_{ST} between gardens to identify transcripts exhibiting steep Q_{ST} reaction norms, which could be reflective of population-level phenotypic plasticity. The mean Q_{ST} difference across all 70,250 transcripts was 0.24 ± 0.29 . We identified 13,725 transcripts

(19.5%) exhibiting steep Q_{ST} reaction norms. Of these, 5,244 had higher Q_{ST} in BS relative to WP, while 8,481 had higher Q_{ST} in WP relative to that in BS (Fig. 3.3a).

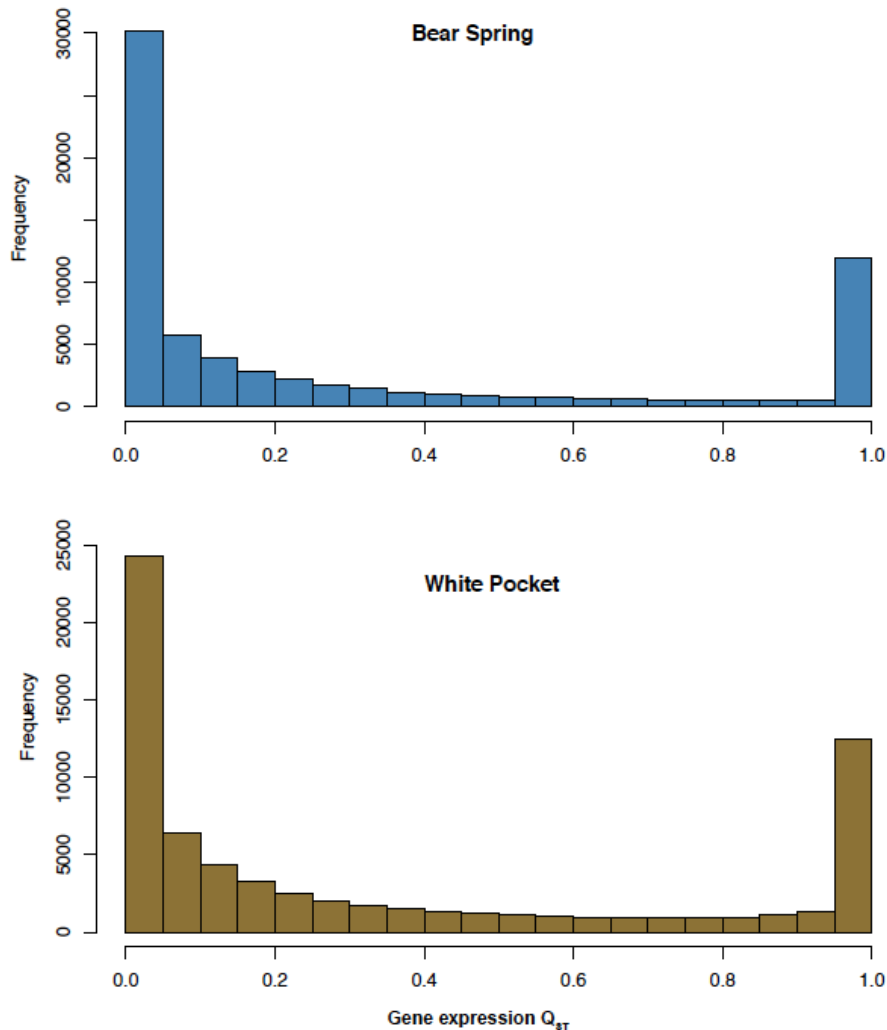


Figure 3.2: Distribution of transcript Q_{ST} for Bear Spring (BS) and White Pocket (WP).

Drivers of garden-specific population differentiation in expression

The full RDA model (model 1) using population-level expression values for the highly differentiated transcripts as the response matrix yielded an adjusted $R^2 = 0.27$ ($p = 0.03$) for BS and $R^2 = 0.034$ ($p = 0.4$) for WP, respectively. The RDA model used to assess the impact of drought-related variables on among population differences in transcript levels (model 2) yielded

an adjusted $R^2 = 0.28$ ($p = 0.003$) for BS and $R^2 = 0.13$ ($p = 0.19$) for WP, respectively. The observed R^2 for the model 2 was significantly higher than the empirical null distribution of R^2 values only in BS ($p < 0.001$) (Fig. 3.4), demonstrating that drought-related variables were more important in explaining the noted among population differentiation for the high Q_{ST} transcripts. Finally, the RDA model to assess the impact of freezing-related variables on among population differences in transcript levels (model 3) yielded an adjusted $R^2 = 0.026$ ($p = 0.37$) for BS and $R^2 = 0.083$ ($p = 0.14$) for WP, respectively. This observed R^2 was significantly lower than the empirical null distribution of R^2 values only in WP ($p < 0.001$), indicating that freeze-related variables were less important in explaining the noted among population differentiation for the high Q_{ST} transcripts (Fig. 3.4).

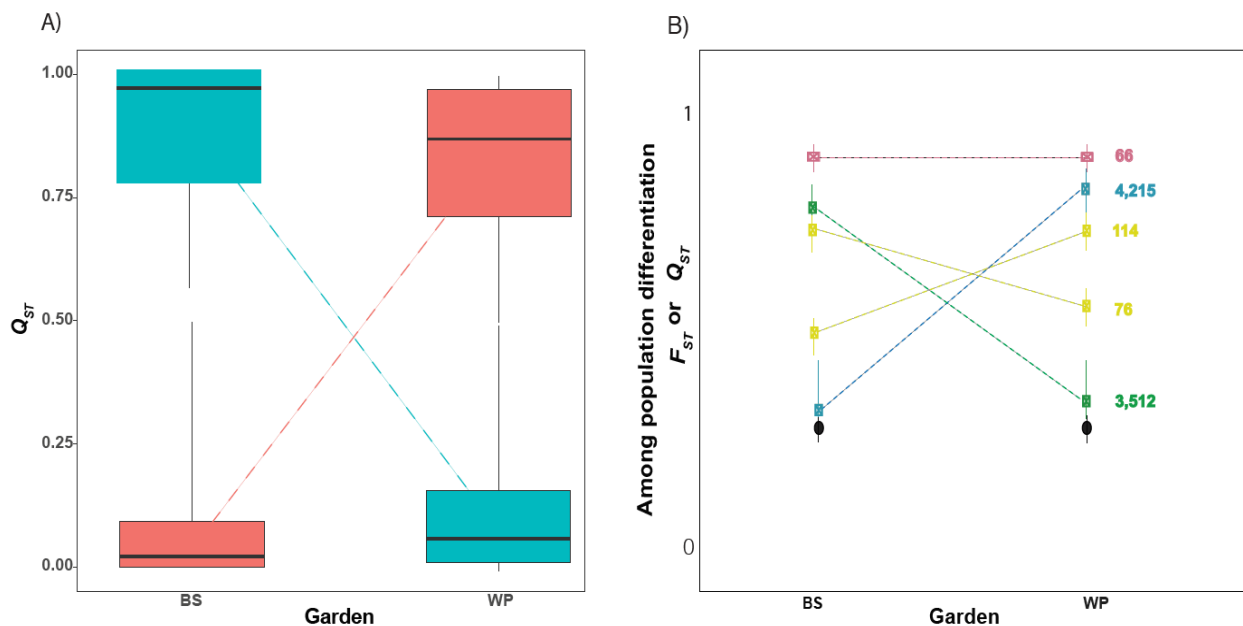


Figure 3.3: (A) Boxplots for transcripts exhibiting steep Q_{ST} reaction norms. There were 5,244 transcripts which had high Q_{ST} only in BS (teal), while 8,481 had high Q_{ST} only in WP (salmon). (B) Conceptual representation of transcript classification using the Q_{ST} - F_{ST} comparison

approach with the number of transcripts classified under each category shown on the right. Reaction norms are only provided for transcripts, multilocus F_{ST} is indicated in black without a reaction norm. Pink represents transcripts exhibiting adaptive differentiation in both gardens but no plasticity, yellow represents transcripts exhibiting adaptive plasticity, blue represents transcripts that are conditionally adaptive in WP and green represents transcripts that are conditionally adaptive in BS.

Transcript level signatures of selection

We filtered our ddRADseq SNP dataset down to the 30 maternal trees for which transcriptome data was available. This dataset was further filtered to retain only SNPs with a minor allele frequency of 0.01 and maximum missing data of 50%. Overall, this yielded a total of 11,431 SNPs that were used for estimating F_{ST} and to perform Q_{ST} - F_{ST} comparisons for assessing signatures of selection. The multilocus estimate of F_{ST} was 0.015 (95% confidence interval, CI= 0.01-0.02) and ranged from 0 to 1 across SNPs. By comparing the CI of F_{ST} with Q_{ST} for each transcript, we classified transcripts into several categories that are conceptualised in Fig. 3.3B. We identified 5% of the transcripts as exhibiting signatures of adaptive trait differentiation ($Q_{ST} > F_{ST}$) at BS and 6% at WP (Table 3.1). Of these, only 0.094% of the transcripts exhibited zero reaction norms slopes (adaptive differentiation in both gardens but no plasticity). We identified 6% of the transcripts as being significantly differentiated from F_{ST} only in WP (conditionally adaptive in WP), while 5% were significantly differentiated from F_{ST} only in BS (conditionally adaptive in BS). These transcripts also exhibited non-zero reaction norm slopes. We only noted 0.28% of the transcripts to exhibit adaptive plasticity, such that $Q_{ST} > F_{ST}$ at both gardens and they also had non-zero reaction norm slopes.

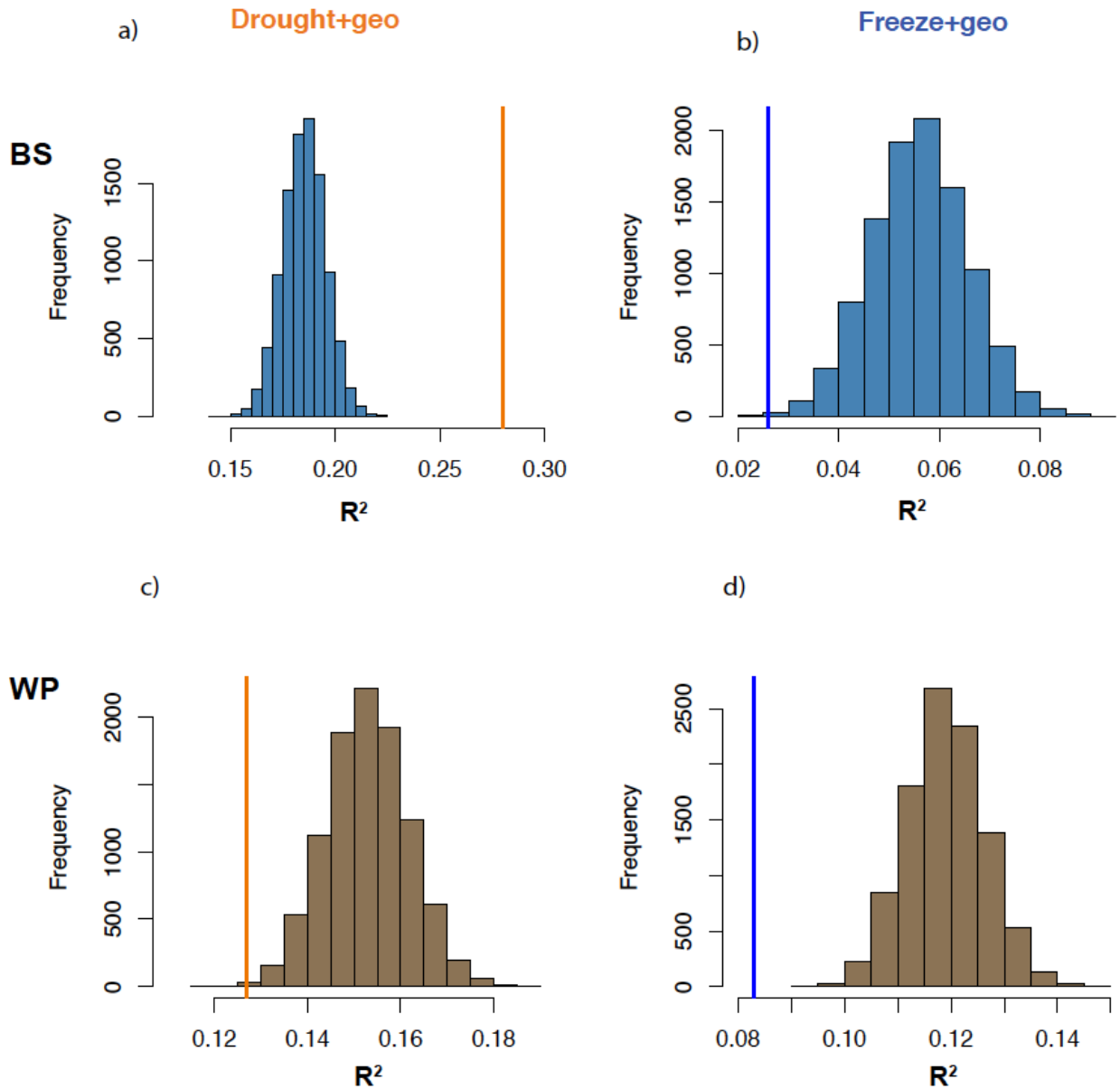


Figure 3.4: Distribution of adjusted R^2 for transcripts that are not highly differentiated between population against the adjusted R^2 (vertical line) for transcripts that were highly differentiated for (a) BS using the drought model, (b) BS using the freeze model, (c) WP using the drought model and (d) WP using the freeze model.

Table 3.1: Summary statistics for transcripts classified into various categories of adaptive differentiation based on Q_{ST} - F_{ST} comparisons. The multilocus estimate of F_{ST} across the 30 maternal trees was 0.0147 (95% CI = 0.01-0.02).

Category	Mean Q_{ST} (\pm sd)	Percent of transcripts	Number of transcripts
Adaptive differentiation but no plasticity	BS = 1 ± 0	0.094	66
	WP = 1 ± 0		
Conditionally adaptive in BS	BS = 0.84 ± 0.19	5	3,512
	WP = 0.05 ± 0.09		
Conditionally adaptive in WP	BS = 0.04 ± 0.08	6	4,215
	WP = 0.83 ± 0.19		
Adaptive plasticity	BS = 0.77 ± 0.19	0.28	190
	WP = 0.82 ± 0.2		

Influence of hybrid ancestry

Our estimates of hybrid ancestry for the 30 maternal trees ranged from 0.18 to 1, with 1 indicating 100% genomic ancestry from *Pinus flexilis*. The distribution of correlation coefficients demonstrating the relationship between mean population ancestry and population level transcript expression value ranged from -1 to 1 for most categories. A formal distribution comparison using the Kolmogorov-Smirnov test demonstrated significant differences between distributions.

Specifically, the distribution of correlation coefficients for highly differentiated transcripts in both gardens was shifted towards the left relative to the empirical null ($p_{BS} < 0.01$ and $p_{WP} < 0.01$). On the contrary, the distribution for transcripts with an absolute among population Q_{ST} difference above 0.5 was shifted to the right relative to the empirical null ($p < 0.01$). However, three out of the four categories of adaptive differentiation defined using the Q_{ST} - F_{ST} comparisons exhibited no deviation from the empirical null distribution of correlation coefficients ($p > 0.05$). The distribution of correlation coefficients for transcripts exhibiting adaptive differentiation across both gardens but no plasticity was slightly, yet significantly, shifted left relative to the empirical null distribution ($p = 0.04$).

DISCUSSION

Range margin populations are often considered depauperate in genetic diversity, which should restrict the ability of plasticity to aid long-term persistence and to facilitate adaptation to putatively novel habitats encountered at species' range boundaries. On the contrary, if range margin populations encompass hybrid zones, then an increase in genetic diversity is expected, which could facilitate the evolution of phenotypic plasticity. By utilising a transcriptome-wide dataset from a controlled common garden design containing the range margin populations encompassing the *P. strobiformis*-*P. flexilis* hybrid zone, we provide strong evidence for adaptive trait differentiation and for adaptive plasticity. We also demonstrate that hybridization likely facilitates the evolution of phenotypic plasticity through increases to standing genetic variation and the presence of novel allelic combinations.

The recent surge in the availability of various genome-wide datasets for model and non-model systems has revealed a strong role of regulatory and non-genic elements in driving

adaptive evolution (Gould et al. 2019; Bachtiar et al. 2019; Mei et al. 2019). Adaptive evolution is often studied through the assessment of population differentiation at fitness-related traits. Gene expression patterns, while reliant on environmental stimuli, are known to have strong heritable components and hence can be treated as fitness-related traits contributing towards adaptive evolution (Pyhäjärvi et al. 2013; Huang et al. 2019). By conceptualising gene expression as a quantitative trait, which is not novel to our approach, we revealed strong patterns of among population differentiation (Q_{ST}) that varied across two common garden sites representative of climatic variability within the natural range of *P. strobiformis*. The large values of Q_{ST} are unusual when compared to previous assessments in conifers or even in *P. strobiformis* that used an array of phenotypic and physiological traits (Goodrich et al. 2016; reviewed in Lind et al. 2018). A direct comparison to these studies can only be possible with matching sample sizes as our study was limited by the number of populations and families assayed. Nevertheless, Ogasawara & Okubo (2009) demonstrated that gene expression differences tend to be strongest during initial stages of species divergence and proceed towards an asymptote as species diverge further. Further, presence of strong selective pressures in the form of post-zygotic isolating barriers in early life stages across trees (Lindtke et al. 2014; Zhao et al. 2014) could also have contributed towards the noted strong population differentiation. We thus suggest that the large proportion of transcripts with $Q_{ST} \geq 0.5$ could be a result of sampling younger cohorts from a hybrid zone formed between two recently diverged species and hence would be a product of interaction between intrinsic and extrinsic selection pressures. With respect to gene expression differentiation specifically, our results are in-line with studies across other systems demonstrating strong heritable components to gene expression values, as well as strong GxE effects (Roberge et al. 2007; Leder et al. 2015).

While a vast number of transcripts (15%) were strongly differentiated in both BS and in WP, the RDA models examining the effect of drought-related variables and freezing-related variables on among population expression differences was only significant in BS for drought-related variables. These results are contrary to previous work in these common gardens demonstrating strong population differentiation and strong associations between source climatic conditions and fitness-related physiological traits (Goodrich et al. 2016; DaBell, 2017). Comparing the observed R^2 for each garden against the empirical null distribution revealed that drought-related variables were more important in explaining the noted among population differences at BS, while at WP both freeze-related and drought-related variables were deemed less important (Fig. 3.4). The noted differences across the two gardens could be a result of (a) among population variation driven by differences in hybrid ancestry or (b) differences in selective pressures between the two gardens. Specifically, trees in BS experience severe winters and aseasonal frost events which could cause freeze tolerance related transcripts to exhibit a baseline level of expression across all trees regardless of the source they originated from and thus cause among population difference in expression to be driven by other climatic gradients such as drought. On the contrary, WP is a dry and hot site which could remove any source climate related differences in transcript levels of drought associated genes and cause among population differences to be driven by climatic gradients such as freezing temperatures. Argument (b), however, only partially supports our results because freeze-related variables did not exhibit stronger association with highly differentiated transcripts at WP (Fig. 4). The stronger than expected associations between genomic ancestry and among population variation in gene expression levels noted here provides support for argument (a). We suggest that while hybrid ancestry was associated with strong differentiation of transcripts at both gardens, this effect

maybe exaggerated in BS which is a high elevation environment with conditions similar to that in the range of *P. flexilis*.

The availability of transcriptome-wide datasets from common garden studies has resulted in the identification of sets of genes that exhibit differences in expression patterns among populations and across environments. These approaches by themselves, however, do not clearly allow for further assessments involving the estimation of heritable components of gene expression variation that are needed to evaluate the role of selection. Similarly, for studies utilising multiple common gardens, inference of phenotypic plasticity is possible, but ascertaining the proportion of traits that are adaptively plastic is not possible without further assessing fitness components. Since an assessment of fitness is challenging in long-lived organisms such as trees, we utilised a formal test that compares levels of genetically-based among population expression differences (Q_{ST}) with genome-wide patterns of differentiation at putatively neutral loci (F_{ST}) to ascertain the relative contribution of selection in driving expression differences within and across multiple environments (reviewed in Leinonen et al. 2008; Josephs, 2018). Overall, 5% of the transcripts examined exhibited signatures consistent with divergent selection, with the number of adaptively differentiated transcripts being lower in BS relative to WP (Table 3.1).

We utilised Q_{ST} reaction norms to demonstrate population-level GxE effects, which are a form of plasticity, and to assess whether these were adaptive or conditionally neutral. Plasticity is often thought of at the level of individual genotypes (Schlichting, 1986), however, inference of adaptive plasticity can only be made at the population level and hence our approach made use of among population variation in reaction norms. To demonstrate that our assessment of plasticity based on population reaction norms is representative of maternal reaction norms to a large extent, we plotted the reaction norms for maternal tree values obtained in equation 2. Maternal

tree values should be reflective of the same genotype being planted in different environments and hence should comply with the traditional definition of plasticity. We noticed a wide array of maternal value reaction norms, with some transcripts exhibiting minimal among population or among family variation in a garden while some others demonstrating strong among population differences across both gardens (Fig. 3.5). All maternal trees demonstrated non-zero expression variation across gardens for the transcripts classified as exhibiting steep Q_{ST} reaction norms. Further, 72% of these transcripts had at least half the maternal trees exhibiting an absolute expression difference of 0.5 or higher across gardens. Our results also demonstrate strong signals of GxE effects, with approximately 5% of the transcripts examined being conditionally adaptive in one of the gardens and 0.28% exhibiting adaptive plasticity. The latter were primarily associated with response to heat stress and pollen development.

The distribution of correlation coefficients for transcripts exhibiting plasticity was shifted higher relative to those that did not exhibit plasticity, likely indicating that ancestry from *P. flexilis* contributed towards steeper reaction norms. Since we only used genome-wide estimates of ancestry and not transcript-specific estimates, we lack confidence in pinpointing specific transcripts that exhibit strong among population differentiation due to hybrid ancestry. Nevertheless, the higher than expected association noted between *P. flexilis* ancestry and the slope of reaction norms would be in-line with previous expectations (Ackerly et al. 2000; Schmid et al. 2019). To elaborate, we suggest that the presence of heterogeneous and harsh climatic conditions in the range margin populations and the likely increase in additive genetic variance resulting from admixture (Goodnight, 1995; Whitlock et al. 1993) could have contributed towards the noted strong relationship between plasticity and *P. flexilis* ancestry for the strongly differentiated transcripts. Future studies performing differential gene expression

analyses using hybrid ancestry information from maternal sibs rather than only the maternal trees and mapping the gene sequences identified here to the transcriptomes of parental *P. flexilis* and *P. strobiformis* would discover links between the ancestry of specific genomic regions and their contributions toward adaptive trait differentiation as well as towards the maintenance of species boundaries that are often expressed in early life stages in trees.

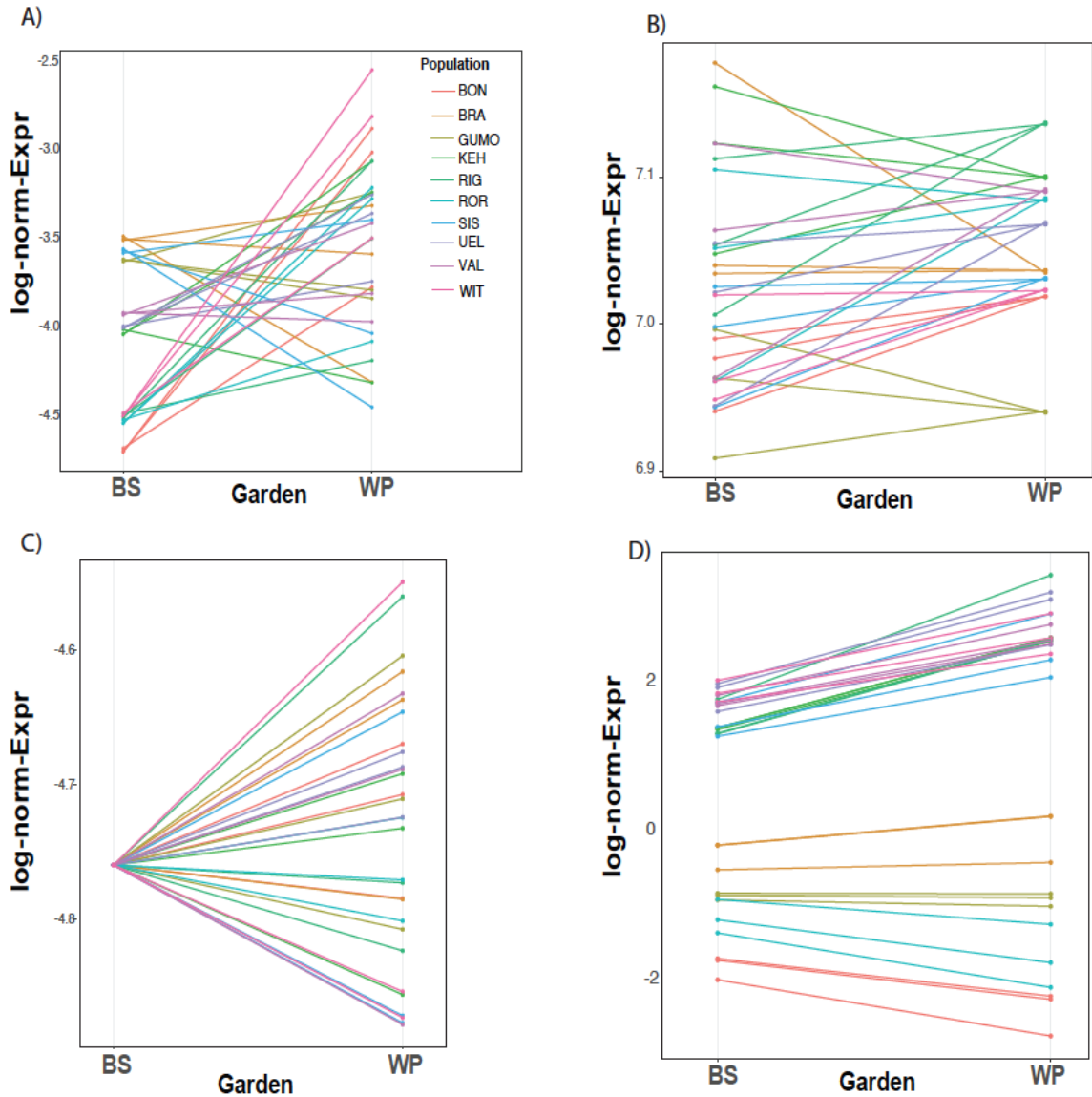


Figure 3.5: Maternal reaction norms for log normalised expression values (log-norm-Expr) of transcripts that are (A) conditionally adaptive only in BS, (B) Conditionally adaptive only in WP (C) Adaptive in neither gardens but displaying strong family variance in WP and (D) Adaptive in both gardens.

Conclusion

Adaptive evolution often involves changes in gene expression, with nearly 12-78% of transcriptional changes likely to influence the organismal phenotype (Greenbaum et al. 2003). While several studies consider expression changes to be transient and drastically influenced by environmental changes, others have shown these to also be heritable and exhibit strong GxE effects. Our study is one of the few to extend the well-developed field of quantitative genetics to gene expression and evaluate the proportion of transcripts exhibiting adaptive divergence. By utilising hybrid trees planted across two common gardens, we have identified transcripts exhibiting adaptive plasticity and have quantified the contribution of hybridization towards gene expression divergence. While this study is limited due to the moderate number of populations sampled ($n = 10$), it significantly advances the field of transcriptomic studies by co-opting an array of well-developed approaches to understand the evolution of genomic reaction norms. Through these we provide strong evidence of environment dependent adaptive differentiation and adaptive plasticity, reiterating the view that gene expression patterns experience similar evolutionary pressures as any quantitative trait and thus can be adaptive, maladaptive or neutral.

REFERENCES

1. Price, T.D., Qvarnstrom, A. & Irwin, D.E. The role of phenotypic plasticity in driving genetic evolution. *Proceedings of the Royal Society of London, Series B*. 270, 1433–1440. (2003).
2. Jump AS, Marchant R, Peñuelas J. Environmental change and the option value of genetic diversity. *Trends in plant Sciences*. 12(1): 51-58. (2009)
3. Bradshaw, A.D. Unravelling phenotypic plasticity—why should we bother? *New Phytologist* 170, 644–64, (2006).
4. Nicotra AB, Atkin OK *et al.* Plant phenotypic plasticity in a changing climate. *Trends in Plant Science*. 15,12,684–692 (2010).
5. Benito Garzón, Robson M & Hampe A. Δ TraitSDMs: species distribution models that account for local adaptation and phenotypic plasticity. *New Phytologist* . 222: 1757-1765 (2019).
6. West-Eberhard MJ. Phenotypic plasticity and the origin of diversity. *Annual Reviews in Ecology and Systematics*. 20:249-78 (1989).
7. Savolanein O, Lascoux M & Merilä J. Ecological genomics of local adaptation. *Nature Reviews Genetics*. 14, 807–820. (2013).
8. Pigliucci M & Schlichting CD. Reaction norms of Arabidopsis. IV. Relationships between plasticity and fitness. *Heredity*. 76(5): 427-436 (1996).
9. Baythavong BS & Stanton ML. Characterizing selection on phenotypic plasticity in response to natural environmental heterogeneity. *Evolution* 64: 2904–2920. (2010).
10. Falconer DS, MacKay TFC. *Introduction to Quantitative Genetics*. Longman Scientific and Technical, Essex. (1996)
11. McKay JK, Latta RG. Adaptive population divergence: markers, QTL and traits. *Trends in Ecology and Evolution*, 17, 285–291. (2002).
12. Josephs E. Determining the evolutionary forces shaping G x E. *New Phytologist* 219: 31–36. (2018)
13. Williams, G.C. *Adaptation and Natural Selection*. Princeton University Press, Princeton. (1966)
14. Schmid M, Dallo R & Guillaume F. Species' range dynamics affect the evolution of spatial variation in plasticity under environmental change. *The American Naturalist*. 193, : 798-813. (2019)
15. Akman M, Carlson JE & Latimer A. Climate gradients explain population-level divergence in drought-induced plasticity of functional traits and gene expression in a South African Protea. *BioRxiv*. (2018)
16. Wright, S.W. The shifting balance theory and macroevolution. *Annual Review of Genetics* 16: 1-19 (1982).

17. Kirkpatrick, M. & Barton, N. H. Evolution of a species' range. *American Naturalist*. 150, 1–23 (1997).
18. Eckert, CG, Samis, KE & Loughheed, SC. Genetic variation across species geographical ranges: the central-marginal hypothesis and beyond. *Molecular Ecology*. 17, 1170-1188 (2008).
19. Barton, N. H. The role of hybridization in evolution. *Molecular Ecology*. 10, 551–568 (2001).
20. Taylor, S.A., & Larson, EL. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology & Evolution*, 3(2), 170–177. (2019)
21. Dlugosch KM, Anderson SA, Braasch J, Cang A & Gillette HD. The devil is in the details: genetic variation in introduced populations and its contributions to invasion. *Molecular ecology* 24 (9), 2095-2111 (2015)
22. Gould BA, Chen Y & Lowry DB. Gene Regulatory Divergence Between Locally Adapted Ecotypes in Their Native Habitats. *Molecular Ecology*. 21: 4174-4188. (2018)
23. Silim S, Guy R *et al.* Plasticity in water-use efficiency of *Picea sitchensis*, *P. glauca* and their natural hybrids. *Oecologia*, 128, 3, 317-325. (2001).
24. Schweitzer J, Martinsen GD, Whitham TG. Cottonwood hybrids gain fitness traits of both parents: a mechanism for their long-term persistence? *American Journal of Botany*, 89, 6, 981-990 (2002).
25. Hegarthy MJ, Barker GL, Brennan AC, Edwards KJ, Abbott RJ, Hiscock SJ. Changes to gene expression associated with hybrid speciation in plants: further insights from transcriptomic studies in *Senecio*. *Philos Trans R Soc Lond B Biol Sci*. 363(1506):3055-69 (2008)
26. Favre A & Karrenberg S. Stress tolerance in closely related species and their first-generation hybrids: a case study of *Silene*. *Journal of Ecoogy*. 99, 6, 1415-1423 (2011).
27. Whitehead, A., and D. L. Crawford. Variation within and among species in gene expression: raw material for evolution. *Molecular Ecology*. 15: 1197–1211 (2006)
28. Roberge C, Guderley H and L. Bernatchez. Genome wide Identification of Genes Under Directional Selection: Gene Transcription Q_{ST} Scan in Diverging Atlantic Salmon Subpopulations. *Genetics* 177: 1011–1022 (2007)
29. Reddy Palle *et al.* Natural variation in expression of genes involved in xylem development in loblolly pine (*Pinus taeda* L.) *Tree Genetics & Genomes* 7, 193–206 (2011).
30. Keller SR, Levsen N, Olson MS, Tiffin P. Local adaptation in the flowering-time gene network of balsam poplar, *Populus balsamifera* L. *Molecular Biology and Evolution* 29: 3143–3152 (2012).

31. Menon M, Barnes W, Olson MS. Population genetics of freeze tolerance among natural populations of *Populus balsamifera* across the growing season. *New Phytologist*. 207 (3), 710-722. (2015)
32. Lewis J & Reed R. Genome-wide regulatory adaptation shapes population-level genomic landscapes in *Heliconius*. *Molecular Biology and Evolution*. 36:159–173 (2019).
33. Hamala et al. Gene Expression Modularity Reveals Footprints of Polygenic Adaptation in *Theobroma cacao*. *Molecular Biology and Evolution*. 37: 110–123. (2020)
34. Huang Y et al. Genome-Wide Genotype-Expression Relationships Reveal Both Copy Number and Single Nucleotide Differentiation Contribute to Differential Gene Expression between Stickleback Ecotypes. *Genome Biology and Evolution* 11(8):2344–2359. (2015)
35. Falconer DS & Mackay T. Introduction to Quantitative Genetics. Longman, New York (1996)
36. Adams HD, Kolb TE. Drought responses of conifers in ecotone forests of northern Arizona: tree ring growth and leaf $\delta^{13}C$. *Oecologia*, 140: 217– 225. (2004).
37. Seager, R & Vecchi, G. A. Greenhouse warming and the 21st century hydroclimate of southwestern North America. *Proceedings of National Academy of Sciences*, 107, 21277–21282. (2010).
38. Levins R. Evolution in changing environments: some theoretical explorations. 2. Princeton University Press; (1968).
39. DaBell J. *Pinus strobiformis* response to an elevational gradient and relationships with source climate. Master's thesis. December 2017. Northern Arizona University.
40. Buchlotz ER. Early growth, water relations and growth: common garden studies of *Pinus strobiformis* under climate change. May 2020. Northern Arizona University.
41. Squillace AE. Average genetic correlations among offspring from open-pollinated forest trees. *Silvae Genetica*. 23: 149- 156 (1974).
42. Wang, T., Hamann, A. Spittlehouse, D.L. and Carroll, C. Locally downscaled and spatially customizable climate data for historical and future periods for North America. *PLoS One* 11: e0156720 (2016).
43. Krueger F. Trim Galore!: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files.
http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ (2015).
44. Grabherr MG et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology*. 29(7): 644–652 (2013).
45. Haas BJ et al. *De novo* transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nature Protocols*. 8: 1-43 (2014)

46. Gilbert D. Gene-omes built from mRNA seq not genome DNA. 7th annual arthropod genomics symposium. Notre Dame. <http://arthropods.eugenes.org/EvidentialGene/trassembly.html> . (2013)
47. Hart AJ et al. EnTAP: Bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes. *Molecular Ecology Resources*. 20: 591-604 (2020).
48. Seppey M., Manni M., Zdobnov E.M. BUSCO: Assessing Genome Assembly and Annotation Completeness. In: Kollmar M. (eds) Gene Prediction. *Methods in Molecular Biology*. 1962. Humana, New York, NY. (2019)
49. Liu JJ, Schoettle AW, Sniezko RA et al. Genetic mapping of *Pinus flexilis* major gene (*Cr4*) for resistance to white pine blister rust using transcriptome-based SNP genotyping. *BMC Genomics*. 17:753. (2016).
50. Langmead B & Salzberg S. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. 9:357-359 (2012).
51. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139-140. (2010)
52. R Core Team. R v.3.6.3: A Language and Environment for Statistical Computing. Vienna, Austria : The R Foundation for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2020)
53. Spitze, K. Population-structure in *Daphnia-obtusa* - quantitative genetic and allozymic variation. *Genetics* 135: 367–374. (1993)
54. Oksanen J et al. vegan: Community Ecology Package. R package version 2.5-2. (2013).
55. Gibson G & Weir B. The quantitative genetics of transcription. *Trends in Genetics*. 21: 616–623 (2005).
56. Gilbert KJ & MC Whitlock. *QST FST* comparisons with unbalanced half-sib designs. *Molecular Ecology Resources*, 15: 262-267 (2015).
57. Puritz JB, Hollenbeck CM & Gold JR. dDocent : a RADseq, variant -calling pipeline designed for population genomics of non -model organisms. *PeerJ*, 2: e431. (2014)
58. Skotte L Korneliussen TS & Albrechtsen A. Estimating individual admixture proportions from next generation sequencing data. *Genetics* 195: 693–702. (2013).
59. Bachtar et al. Architecture of population-differentiated polymorphisms in the human genome. *PLoS ONE* 14(10): e0224089. (2019)
60. Mei W, Stetter MG, Stitzer MC. Adaptation in plant genomes : Bigger is different. *American Journal of Botany*. 105: 16-19 (2019).

61. Pyhäjärvi T, Hufford MB, Mezouk S, Ross-Ibarra J. Complex patterns of local adaptation in Teosinte. *Genome Biology and Evolution*. 5: 1594-1609 (2013).
62. Lind BM, Menon M, Bolte CE, Faske TM & Eckert AJ. The genomics of local adaptation in trees: Are we out of the woods yet?. *Tree genetics & genomes*. 14 (2), 29. (2018).
63. Ogasawara O & Okubo K. On theoretical models of gene expression evolution with random genetic drift and natural selection. *PLoS One* 4:e7943 (2009)
64. Leder EH et al. The Evolution and Adaptive Potential of Transcriptional Variation in Sticklebacks—Signatures of Selection and Widespread Heritability. *Molecular Biology and Evolution*. 32(3):674–689 (2015).
65. Goodrich BA, Waring KM & Kolb TE. Genetic variation in *Pinus strobiformis* growth and drought tolerance from southwestern US populations. *Tree Physiology*. 36: 1219-1235. (2016).
66. Leinonen, T., R. O'Hara, J. M. Cano and J. Merilä. Comparative studies of quantitative trait and neutral marker divergence: a meta-analysis. *Journal of Evolutionary Biology*. 21 1–17. (2008)
67. Schlichting CD. The evolution of phenotypic plasticity in plants. *Annual Reviews in Ecology in Systematics*. 17:667-93. (1986)
68. Ackerly DD, Dudley SA, Sultan SE et al. The evolution of plant ecophysiological traits: Recent advances and future directions: new research addresses natural selection, genetic constraints, and the adaptive evolution of plant ecophysiological traits. *Bioscience*. 50: 979–995 (2000).
69. Goodnight CJ. Epistasis and the increase in additive genetic variance: Implications for phase I of Wright's shifting-balance process. *Evolution* 49: 502–511. (1995)
70. Whitlock MC, Phillips PC, Wade MJ. Gene interaction affects the additive genetic variance in subdivided populations with migration and extinction. *Evolution* 47: 1758–1769. (1993)
71. Greenbaum D, Colangelo C, Williams K, Gerstein M. Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biology*. 4:117. (2003)