Implications of genetic heterogeneity for plant translocation during ecological restoration

Taylor M. Crow[1], Daniel E. Runcie[1], Kristina M. Hufford[2]

1 Department of Plant Sciences, University of California, Davis, CA, USA. [2] Ecosystem Science and Management, University of Wyoming, Laramie, WY 82071, USA

Keywords: *Cercocarpus montanus*, phylogeography, genetic diversity, genetic differentiation, genetic structure, niche

Abstract

Ecological restoration practices often require translocating plant material from distant sites. Yet published guidelines for seed transfer are available in very few species. Accurately predicting how plant populations will perform when transferred requires multi-year and multi-environment field trials and comprehensive follow-up census work. In this study, we analyzed the genetic structure of an important shrub used in ecological restorations in the Southern Rocky Mountains called alderleaf mountain mahogany (*Cercocarpus montanus*) by sequenced 1440 individual plant samples from 48 populations across a broad geographic range. We found that geographic and environmental covariates of genetic differentiation and diversity reflected the complex climate and topography across which the species is distributed. We identified several temperature and precipitation variables that were useful predictors of genetic differentiation in this species and can be used to generate seed transfer recommendations. These results will be useful for defining management and restoration practices for mountain mahogany and other widespread montane plant species.

Introduction

The restoration of vegetation after a disturbance event can improve many ecosystem services (Barral et al., 2015). For example, soil stabilization, pollinator and wildlife habitat, nutrient cycling, and carbon sequestration are all positively correlated with successful ecological restoration (Benayas et al., 2009). However, bringing foreign plant material to a restoration site can have unintended consequences. Importing maladaptated populations can result in large-scale plant mortality (Johnson et al., 2004), inbreeding depression of introduced material, or outbreeding depression within future local and foreign hybrid populations (Hufford and Mazer, 2003). Therefore, optimizing the fitness of imported plant material is of vital importance.

Seed transfer guidelines are intended to establish criteria to aid in the selection of plant material for restoration. However, traditional common garden experiments are expensive and time-consuming (Johnson et al., 2004), requiring multi-year and multi-environment field trials and comprehensive follow-up census work. Typically, the relationship between phenotypic variation and environment or geography are used to create categorical seed transfer zones (Campbell and Sorensen, 1978; Bower and Aitken, 2008), continuous seed transfer guidelines Parker and Niejenhuis (1996), or both (Hamann et al., 2000; Saenz-Romero and Tapia-Olivares, 2008). These experiments are limited by the number of populations, number of environments, and the amount of time it may take to quantify consequences of foreign plant material (Johnson et al., 2004).

Integrating estimates of genetic variation with metrics of geographic distribution, and environmental niche can improve how seed transfer guidelines are created. Species distributions are the result of complex interactions and historical contingencies (Fukami, 2015). The niche concept incorporates both the environmental and geographic distribution of a species, and seeks to explain potential factors governing range limits (Sexton et al., 2009). One conceptualization of a niche is summarized by Hutchison's n-dimensional hypervolume (Hutchison, 1957), described as a set of biologically relevant and independent variable axes

within which a species occurs. The multivariate environmental space represents conditions that accommodate population persistence and growth (Hutchinson, 1978). As habitat quality or availability decreases, population size and gene flow are likely to decrease (Brown, 1984; Eckert et al., 2008). Testing the effect of geographic and environmental centrality on genetic diversity can further differentiate whether genetic structure is caused by selection, and warrants evaluation for seed transfer guidelines.

Spatial (Bower et al., 2014; Crow et al., 2018) or genetic analyses (Krauss and He, 2006) modeling approaches for may be a useful tool for establishing plant population translocation recommendations without the financial or time investment required by a transplant experiment. For example, genetic structure analyses can estimate genetic connectivity between adjacent populations, which can be used to predict the likelihood of reduced fitness during a potential transfer (Ellstrand and Elam, 1993; Sexton et al., 2014). Preserving genetic structure in restoration is important for maintaining adapted combinations of genes in native populations. Further, gene flow between introduced and native populations may lead to outbreeding depression when locally adapted gene complexes are disrupted by immigrant alleles after admixture (Fenster and Galloway, 2000; Montalvo and Ellstrand, 2001). Identifying geographic and environmental patterns related to genetic differentiation and diversity can therefore provide useful guidelines for seed introductions in ecological restoration (Montalvo and Ellstrand, 2001).

In this study we investigated genetic variation relevant for restoration in a native perennial shrub, alderleaf mountain mahogany, *Cercocarpus montanus*. Alderleaf mountain mahogany is a habitat dominant species which forms dense stands where it occurs, and hosts a nitrogen-fixing actinobacteria in the genus *Frankia*. The species has a widespread spatial distribution throughout the Southern Rocky mountains, and occurs on both sides of the continental divide. Mountain mahogany is used in restoration projects because of its value as a forage plant for large ungulates, especially in the winter months. We collected and sequenced DNA from 1440 individual plant samples from 48 populations, estimated diversity, and generated

over 6,000 single nucleotide polymorphisms (SNPs) to describe genetic structure. We tested whether genetic diversity was a function of latitude, habitat quality, range centrality, or a combination. Our goal was to make recommendations for seed sourcing in ecological restoration.

Methods

Study species

Cercocarpus montanus Raf. is a deciduous, perennial shrub species in the rose family (Rosaceae) with a large geographic distribution in western North America (Dorn, 2001). The species occurs on both sides of the Continental Divide from northern Mexico to approximately 45 degrees in latitude near the Wyoming-Montana state borders in the United States (Fig. 1). Populations are generally distributed between 1200 and 3000 m in elevation and often grow in rocky, limestone soils (Williams et al., 2004). Fruits are achenes with an elongated style. Mountain mahogany shrubs serve as hosts for nitrogen-fixing actinomycete bacteria (genus *Frankia*) in root nodules, and this adaptation contributes to successional processes in arid regions dominated by unstable, low nitrogen soils (Klemmedson, 1979).

DNA extraction, sequencing, assembly and variant detection

Mountain mahogany populations were located along a north-south cline throughout the southern Rocky Mountains (Fig. 1). We collected leaf tissue from 30 individuals per population from 48 populations and extracted DNA using a modified cetyl-trimethyl ammonium bromide (CTAB) protocol (Doyle, 1987). DNA was quantified with a NanoDrop 2000 spectrophotometer (Thermo Fisher, Inc.), and additional extractions were conducted when necessary due to high levels of contaminants or low DNA concentrations. We prepared genomic libraries for genotype-by-sequencing (GBS) following protocols in Parchman et al. (2012). To summarize, we digested sample DNA with two restriction enzymes (MseI and EcoRI)

and ligated barcodes containing unique 8-10 bp sequences to the resulting DNA fragments for each sample to ensure that sequence reads could be assigned to individuals. We then PCR amplified the barcoded restriction-ligation products with standard Illumina primers (1, 5′ - AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC-CGATCT - 3′; 2, 5′ - CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT - 3′) (Illumina, Inc.).

Barcoded PCR products were selected randomly and combined into two multiplexed libraries of 720 individual samples and sequenced at the University of Texas Genomic Sequencing and Analysis Facility (GSAF; Austin, Texas, USA) on the Illumina HiSeq 2500 platform using single-end 100 bp reads. After filtering reads for oligonucleotides used in library synthesis and the PhiX genome, with subsequent demultiplexing and assignment of reads to individuals, we had 397 866 962 sequence reads for further analysis. We completed a de novo genome assembly with a randomly chosen subset of 2.4×10^7 reads using SEQMAN NGEN software (DNASTAR, Inc.). This step resulted in construction of an artificial, partial reference genome containing 111 967 contigs. We used `bwa` (Burrows-Wheeler Aligner; Li and Durbin 2009) to map reads from each individual to this partial reference genome. Once complete, 15 520 448 total reads (64.6%) assembled to the partial reference genome. Aligned reads were then indexed and sorted using samtools and bcftools (Li et al., 2009). We used the command 'mpileup -P ILLUMINA -u -g -I -f cemo.fasta sorted.bam | bcftools view -N -c -e -g -v -I -d 0.8 -p 0.01 -P full -t 0.001 -o variants.vcf' to calculate genotype likelihoods and filter variant sites. We then removed variant sites on the same contig with a mean allele frequency less than 0.05.

Population genetic analyses

We calculated genotypes as point estimates of genotype likelihoods and constructed a genetic covariance matrix for all individuals. We ran a principal components analysis (PCA) of the genotype matrix using the *prcomp* function in *R* to summarize genetic variation. We

tested for correlations between the individual scores on the first two principle component axes and potential drivers of genetic variation such as latitude, elevation, precipitation and temperature. Additionally, genotype data were used to calculate individual admixture coefficients using the sparse non-negative matrix factorization algorithm (sNMF) implemented in the LEA package (Frichot et al., 2014; Frichot and François, 2015) in R. This algorithm estimates ancestry coefficients in a computationally efficient manner. The sNMF algorithm is similar to the program STRUCTURE (Pritchard et al., 2000; Falush et al., 2003), which estimates ancestry independently for each individual, and does not require a-priori assumptions about population membership. To determine the best-supported number of genetic clusters (K) within our collections of mountain mahogany, we used a cross-entropy criterion from K=1 to K=10 from the SNMF function. This criterion uses a masked genotype testing set to determine the prediction accuracy of the model at each K value.

Point estimates of allele frequencies within each population were calculated from the genotype likelihoods, and allele frequencies were used to calculate F_{ST} and Nei's genetic distance (D_A) as measures of genetic differentiation (Nei et al., 1983; Takezaki and Nei, 1996; Weir and Hill, 2002). We used Bayesian linear models with Nei's D_A as the response variable, and pairwise geographic distance, environment distance, and a binary variable representing the Continental Divide as model predictors. Population pairs were assigned 0 if they originated from the same side of the Continental Divide, or assigned 1 if they were collected from opposite sides of the divide. Environmental distances were measured as the population pair difference for each environmental variable centered on the mean and divided by the standard deviation (z-score). Environmental variables included thirty-year normal temperature and precipitation estimates from thin plate spline surfaces (`http://forest.moscowfsl.wsu.edu/climate`). All predictor variables (Table S1) were standardized prior to modeling so that their magnitude could be compared. We fit the full and reduced models for genetic differentiation in R with the RJAGS package for MCMC models in JAGS (Plummer, 2003). We ran Markov Chain Monte Carlo (MCMC) simulations for 10 000 iterations with

the first 2000 runs discarded as burn-in. We thinned the MCMC chain every 5 steps for a total posterior sample of 1600 for each of 3 chains. The deviance information criterion (DIC) was used to select the model that best accounted for genetic distance, as well as to compare models with and without geographic distance, environmental distance, and topographic barriers as covariates.

Relative contribution of geographic and environmental distance to genetic differentiation

Geographic and environmental distances contribute to adaptive differentiation that can affect translocation outcomes of seed sources. We used a model that explicitly differentiates between the effects of environment and potential genetic barriers, relative to geographic distance. This model was developed by Bradburd et al. (2013) called Bayesian Estimation of Differentiation in Alleles by Spatial Structure and Local Ecology, implemented in the R package BEDASSLE. We tested the complete dataset, and used the beta-binomial Markov Chain Monte Carlo model to account for data overdispersion. We ran MCMC simulations for 3×10^6 iterations, thinned the chain every 20 iterations, and checked the trace plots for convergence and acceptance rates.

Genetic diversity in central and peripheral habitat

We modeled genetic diversity as a function of geographic and environmental centrality. We estimated genetic diversity for each population using the program ANGSD (Korneliussen et al., 2014). Sorted BAM files were used as input to calculate each population's site allele frequencies from genotype likelihoods. We filtered sites that had a minimum mapping quality of 10 and a minimum q-score of 20. The allele frequency likelihoods were used to calculate the maximum likelihood estimate (MLE) of the site frequency spectrum (SFS) using the EM algorithm. Estimates of nucleotide polymorphisms were calculated as θ_π (Tajima, 1983),

a measure of average pairwise differences, and Watterson θ_W (Watterson, 1975), which is based on the number of segregating sites. Theta estimates were calculated using the empirical Bayesian approach with the SFS as priors (following `http://popgen.dk/angsd/index.php/Thetas,Tajima,Neutrality_tests`).

To model geographic and environmental centrality of our collections in the context of the entire range of *C. montanus*, we used range-wide occurrence points from a previous study of mountain mahogany (Crow et al., 2018). Geographic centrality was calculated as the great circle geographic distance (van Etten, 2018) from each of our sampled populations to the mean latitude and longitude of the species' range, and the range of each individual genetic cluster separately (Fig. 1). We calculated geographic peripherality as the distance between each population and the shortest linear distance to the minimum convex polygon of the species' range. Environmental centrality was calculated as the multidimensional euclidean distance of each population to the species' environmental centroid, and the centroid of each genetic cluster (Blonder et al., 2014). We also used the probability of occurrence derived from a previously published species distribution model (SDM) of mountain mahogany (Crow et al., 2018) as an indicator of habitat quality, which served as a potential predictor of population genetic diversity. In summary, environmental variables were selected for the SDM using a model improvement ratio following (Murphy et al., 2010), and a Random Forests algorithm was used to generate the distribution model. We then used linear models to determine if habitat quality, geographic centrality, or environmental marginality were predictive of genetic diversity using the LM and ANOVA function from the STATS packages in R.

Niche similarity among genetic clusters

Niche overlap statistics were used to test if genetic clusters defined by the sNMF admixture analysis occupied distinct subsets of the overall environmental range. Broennimann et al. (2012) developed methods to get an unbiased estimate of niche overlap using kernel smoother functions applied to densities of occurrence points in environmental space, calibrated on the

available environmental space across the study area. We calculated kernel densities for each genetic cluster, and used D metrics (Schoener, 1970) to determine if there was significant overlap among niche space between genetic groups:

$$D = 1 - 0.5(\sum_{ij} |z_{1_{ij}} - z_{2_{ij}}|)$$

where $z_{1_{ij}}$ and $z_{2_{ij}}$ are the occupancy of the environment calculated from kernel density functions of entity one and two respectively. The D metric is 0 if there is no overlap between genetic groups and 1 if there is complete overlap. We used the ECOSPAT package (Broennimann et al., 2017) in R (R Core Team, 2018) to calculate niche similarity and overlap. Ecospat performs a randomization test where $z1_{ij}$ and $z2_{ij}$ are combined and randomly separated into two groups, and the D statistic is calculated 100 times to build a null distribution. The observed D statistics, using genetic clusters as entity designations, were calculated and compared to the distribution of simulated D values for each pair of genetic clusters separately. Presence points and environmental data for the distribution of mountain mahogany from Crow et al. (2018) were incorporated as background points.

Results

Sequence alignment and SNP discovery

We identified 12 022 single nucleotide variants using samtools and bcftools (Li and Durbin, 2009). For a variant site to be identified, we required that at least 50% of all individuals have a minimum of one read at that locus. After removing sites with a minor allele frequency of <0.05% and randomly selecting one variant per contig to ensure independence of loci, we retained 6352 single nucleotide polymorphisms (SNPs) for further analyses of population genetic structure. In sum, 1366 of the 1440 individual samples of *C. montanus* had sufficient sequencing coverage to be retained for further analysis, resulting in a range of 22–30

individuals per population. Remaining samples had an average of 8.5 reads per SNP.

Population genetic analyses

The first PC axis (PC1) accounted for 89.7% of the genetic variation among individuals of mountain mahogany, and reflected latitude of origin and the effect of the Continental Divide as a barrier (Fig. 2). PC2 accounted for 3.1% of genetic variation, and separated populations of *C. montanus* collected near Albuquerque, NM and Flagstaff, AZ, from those in the remainder of the range. The first PC axis shows that mountain mahogany has continuous genetic variation in the southern portion of its range, and two separate clusters in northern latitudes. Pearson's correlation coefficients (r) between each environment variable and PC1 were used to determine the likely drivers of population genetic structure. We found that two environmental variables: growing season precipitation (GSP) and the number of degree days less than zero °C (DD0), had the highest correlations (0.44 and 0.42 respectively) with PC1.

The mean Nei's D_A genetic distance between populations was 0.0346 (SD=0.017), with a range of 0.009–0.108, comparable to previous studies of plant species (Reynolds et al., 2013; Abraham et al., 2015). Pairwise F_{ST} had an overall mean of 0.161 (SD=0.0856). The mean F_{ST} among pairs of populations from opposite sides of the Continental Divide was 0.241 (SD=0.079), while the mean F_{ST} among populations on the same side of the divide was 0.135 (SD=0.07). Pairwise F_{ST} was positively correlated with geographic distance, and population pairs from opposite sides of the Continental Divide had elevated F_{ST} resulting from the large geographic barrier (Fig. 3). Growing season precipitation and degree days less than 0C were standardized and combined as a single mean Euclidean distance for each population pair, and served as environmental predictors in modeling. The Bayesian linear model with the lowest DIC included both geographic and environmental distance as predictors of genetic differentiation (Table 1). The best predictor in a univariate model of genetic differentiation was geographic distance, followed by environmental distance, while the binary distance matrix representing the Continental Divide was the worst predictor.

The best supported number of clusters for sNFM admixture analysis was K=4 (Fig. S3). Populations were assigned to a single cluster based on the predominant population admixture coefficient of individuals within each population. Although genetic variation is continuous through most of mountain mahogany's range, the map of admixture composition shows that the genetic clusters are partitioned in geographic space (Fig. 2, panel A), with more highly admixed zones between clusters. The genetic clusters occupied regions of the species environmental space with different multivariate centroids (Fig. 4 panel A). Cluster one and three had no detected overlap in their environmental niche, while clusters one and two and two and three had partial, but not significant overlap in environmental space (Table S3).

The BEDASSLE analysis calculated the ratio of environmental and geographic effect sizes on genetic differentiation (α_{E}:α_{D}). We used growing season precipitation and degree days less than zero as environmental variables, as well as a binary distance matrix representing the Continental Divide to quantify the effect of the environment on genetic distance. A difference of one degree days less than zero is comparable to approximately 8 kilometers, and a 1 cm change in growing season precipitation has the same effect on genetic differentiation as approximately 70 kilometers. The Continental Divide had the largest effect on genetic differentiation relative to geographic distance. Crossing the Continental Divide had the same effect on genetic differentiation in mountain mahogany as moving 1.7 x 10^7 km, a larger distance than our collection area.

We detected significant variation in genetic diversity across mountain mahogany's central range. Nucleotide diversity estimates were highly correlated (r>0.9, θ_π and θ_W), and we therefore arbitrarily chose θ_π for further modeling (Table S2). Genetic diversity was not correlated with latitude (P = 0.266, df = 43, R^2 = 0.028). We used two measures of precipitation (growing season precipitation and summer precipitation balance) and two temperature metrics (degree days less than zero and frost-free period) to model genetic diversity because of low colinearity between variables and high correlation with diversity estimates.

Genetic diversity was lower in populations further from the species' multidimensional environmental centroid. Geographic periphery however, was a poor predictor of θ_π. We found that the environmental distance to the centroid of each genetic cluster best described genetic diversity (Table 2). We also found significant variation among genetic clusters for the effect of environmental and geographic distance; namely genetic variation within the northern and southern genetic clusters (Cluster 1 and 3) both had a significant relationship to environmental marginality, whereas within the central genetic cluster (cluster 2) diversity was not correlated with environment (Fig. 4).

Discussion

Mountain mahogany is increasingly used in restoration programs where it occurs as a dominant shrub, particularly because it hosts nitrogen-fixing actinobacteria that allow establishment in nutrient-poor soils, and provides important overwintering forage for wildlife. Despite widespread occurrence in the Rocky Mountain West, no prior ecological genetic studies have characterized genetic diversity across mountain mahogany's central range. We sequenced 1440 individuals from six U.S. states in the Southern Rocky Mountains to investigate the factors that affect population genetic diversity. Results provide preliminary data for seed sourcing guidelines in mountain mahogany restoration.

Species-wide distributions of genetic variation are a critical component of ecosystem management, and long-term population viability depends on genetic diversity (Reed and Frankham, 2003). We found evidence that genetic differentiation is impacted by geographic distance and the environment, including large geographical barriers and climate gradients. These results have range-wide implications for mountain mahogany shrubland management, and lay the groundwork for critical decision-making under environmental change. The study area spanned a large portion of the Southern Rocky Mountains, where orographic effects disrupt the correlation between temperature, precipitation and latitude that is present in

the adjacent Great Plains of North America. Complex topography has been shown to increase genetic variation (Wang et al., 2014; Rodríguez et al., 2015), likely due to geographic isolation of populations.

While mountain mahogany varied continuously across the sampled geographic range, distinct clusters suggest that populations may be adapted to local environmental conditions. However, we cannot infer ecotypic variation in this study based on genetic variation alone. Field studies are needed to determine if individuals have higher fitness within geographic clusters relative to individuals grown at sites outside of their cluster of origin. Despite this limitation, model results suggest that the source of seeds for translocation may impact the viability of the resulting population. The Bayesian model with the best fit included both geographic and environmental distance as factors in population differentiation. Results of the BEDASSLE model, designed to disentangle the effects of geographic and environmental distance, showed that growing season precipitation (GSP) and the number of degree days less than zero (DD0) had large effects on genetic structure in this species. This outcome provides support for seed sourcing guidelines that limit collection to the genetic and the correlated environmental cluster represented by the restoration site.

Mountain mahogany populations had significantly elevated genetic differentiation across the Continental Divide, especially in central Colorado, where the Continental Divide is at its highest altitude. Several studies have shown that the Continental Divide is a strong barrier to gene flow (Schield et al., 2018; Machado et al., 2018). However, to date, no published study has documented this in plant species. Many studies have found significant effects of geographic barriers on genetic differentiation in plant species, including: seas (Jaros et al., 2017), lakes and terrain (Ju et al., 2018), rivers (Geng et al., 2015), mountains (Zhu et al., 2017; Reeves and Richards, 2014), and basins (Bontrager and Angert, 2018). Our data agree with these studies and indicate that populations from opposite sides of the Continental Divide are genetically isolated, despite what may appear to be close geographic proximity (Fig. 3). Populations from the western slopes of the Rocky Mountains had high among-

population genetic differentiation, especially populations 3 and 4 (Fig. 1 panel B and C). Population 3 and 4 may have been founded separately from other western slope populations, or may be hybrids with a closely related species, *Cercocarpus ledifolius*, that co-occurs in this region (Stutz, 1988). The two most genetically differentiated populations (47 and 48), in New Mexico and Arizona respectively (Fig. 1 panel B and C), inhabit isolated locations surrounded by desert regions with low habitat suitability (Crow et al., 2018). Populations 47 and 48 are likely adapted to high temperature and low precipitation conditions, and may warrant further investigation into their taxonomic status..

Despite the heterogeneity of climatic conditions in our study area, we found that genetic clusters were partitioned in distinct geographic groupings (Fig. 2). Further, the genetic clusters inhabited significantly different environmental space (Fig. 4A), which corroborates linear modeling results showing that geography and environment are both factors involved in shaping genetic variation. Given these results, we analyzed patterns of genetic diversity across both geographic and environmental gradients.

Model outcomes suggested that environmental centrality was a better predictor of genetic diversity than geographic distance. This analysis was completed for all sampled populations, as well as for individual genetic clusters. In both cases, genetic diversity was lower near the environmental niche periphery and not strongly correlated with geography. A previous study by Lee-Yaw et al. 2017 found similar results, where genetic diversity of *Arabidopsis lyrata* ssp. *lyrata* was lower at the edge of the environmental niche, but not the limits of the sampled geographic range. Several meta-analyses have shown that the geographic and environmental range limits do not necessarily coincide, and that the geographic range frequently does not explain patterns of genetic variation (Eckert et al., 2008; Pironon et al., 2017). Another review by Lira-Noriega and Manthey (2014) found that only about half of species ranges have any correlation between geographic and environmental marginality, and that environmental marginality was consistently associated with genetic diversity, while geographic marginality was not.

Reduced genetic variation associated with range limits does not distinguish whether populations occuring at range limits are demographic sinks maintained by immigration from more central habitat, or are important genetic resources adapted to marginal conditions by selection. However, the correlation of genetic diversity and environmental centrality bolsters our findings that the environment (namely, climate variation) is a selective force across the genome of mountain mahogany. Other studies of genetic variation near range limits have found contrasting results, even among populations of a species. For example, Hargreaves and Eckert (2019) found that while some populations of the annual plant *Rhinanthus minor* near the range margin had lower fitness, other edge populations were locally adapted. Aguirre-Liguori et al. (2017) found that genetic diversity was lower near the geographic range margin of teosinte, and candidate adaptive SNPs were positively correlated with distance to niche centroid, arguing that populations near the geographic range margins were isolated, while populations near the edges of the environmental niche were locally adapted. In *Picea sitchensis*, populations proximal to the range margin are more likely to carry rare alleles (Gapare et al., 2005), however, a second study of *P. sitchensis* determined that populations near the range limit were locally adapted (Mimura and Aitken, 2010).

The results of our study have implications for the conservation of western shrublands dominated by alderleaf mountain mahogany. The effect of the Continental Divide on genetic structure was significant. This suggests that transferring populations across the Continental Divide would increase the likelihood of maladaptation, and subsequent risks for outbreeding depression among progeny of local and introduced plants. Two climate variables, degree days less than zero and growing season precipitation, were significant drivers of genetic structure as well as reduced genetic diversity. These two variables could delimit collection sites when transferring seed sources during restoration. Choosing a commercial seed source or collection location that is most environmentally similar to the restoration site may increase chances of introducing adapted genotypes (Hufford and Mazer, 2003). In the case of mountain mahogany, preliminary seed collection zones could be delineated by the four common clusters

in genetic analysis. This is a practical approach given that the four clusters represent large geographic areas for collection despite considerable elevational and microhabitat variation. Whether populations near range margins are important resources for conservation in mountain mahogany remains unclear. Plants are subjected to biotic and abiotic stressors that influence population dynamics (Pagel and Schurr, 2012; Franklin et al., 2016), seed predators (Louda, 1982), pollinators (Biesmeijer et al., 2006), and dispersers (Merow et al., 2011). As a result, additional studies are needed to determine the adaptive value of mountain mahogany populations along range margins for ecological restoration, particularly in light of changing climate conditions.

Acknowledgments

We are grateful for the assistance of personnel at New Mexico State University, the USDA Manitou Experimental Forest, Boulder County Parks and Open Space and Mountain Cement Co. in Laramie, Wyoming. This material is based upon work that is supported by Wyoming Agricultural Experiment Station funding provided through the National Institute of Food and Agriculture, U.S. Department of Agriculture, McIntire-Stennis under #228001. Work was also supported by The Berry Biodiversity Center research grant, and Boulder County Parks and Open Space research grant. We would like to also thank P. McIlvenna and D. Bergman for help collecting leaf material, and L. Mandeville for assistance with DNA library preparation. Special thanks to A. Buerkle for guidance for analyses and writing of this manuscript.

References

Abraham, E. M., I. Ganopoulos, P. Giagourta, M. Osathanunkul, I. Bosmali, A. Tsaftaris, A. Papaioannou, and P. Madesis. 2015. Genetic diversity of *Lotus corniculatus* in relation

to habitat type, species composition and species diversity. Biochemical Systematics and Ecology 63:59–67.

Aguirre-Liguori, J., M. Tenaillon, A. Vázquez-Lobo, B. Gaut, J. Jaramillo-Correa, S. Montes-Hernandez, V. Souza, and L. Eguiarte. 2017. Connecting genomic patterns of local adaptation and niche suitability in teosintes. Molecular Ecology 16:4226–4240.

Barral, M. P., J. M. R. Benayas, P. Meli, and N. O. Maceira. 2015. Quantifying the impacts of ecological restoration on biodiversity and ecosystem services in agroecosystems: a global meta-analysis. Agriculture, Ecosystems & Environment 202:223–231.

Benayas, J. M. R., A. C. Newton, A. Diaz, and J. M. Bullock. 2009. Enhancement of biodiversity and ecosystem services by ecological restoration: a meta-analysis. science 325:1121–1124.

Biesmeijer, J. C., S. P. Roberts, M. Reemer, R. Ohlemüller, M. Edwards, T. Peeters, A. Schaffers, S. G. Potts, R. Kleukers, C. Thomas, et al. 2006. Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. Science 313:351–354.

Blonder, B., C. Lamanna, C. Violle, and B. J. Enquist. 2014. The n-dimensional hypervolume. Global Ecology and Biogeography 23:595–609.

Bontrager, M., and A. L. Angert. 2018. Genetic differentiation is determined by geographic distance in *Clarkia pulchella*. bioRxiv . `https://www.biorxiv.org/content/early/2018/07/23/374454`

Bower, A. D., and S. N. Aitken. 2008. Ecological genetics and seed transfer guidelines for *Pinus albicaulis* (Pinaceae). American Journal of Botany 95:66–76.

Bower, A. D., J. B. St. Clair, and V. Erickson. 2014. Generalized provisional seed zones for native plants. Ecological Applications 24:913–919.

Bradburd, G. S., P. L. Ralph, and G. M. Coop. 2013. Disentangling the effects of geographic and ecological isolation on genetic differentiation. Evolution 67:3258–3273.

Broennimann, O., V. Di Cola, and A. Guisan. 2017. ecospat: Spatial ecology miscellaneous methods. R package version 2.2.0. `https://CRAN.R-project.org/package=ecospat`

Broennimann, O., M. C. Fitzpatrick, P. B. Pearman, B. Petitpierre, L. Pellissier, N. G. Yoccoz, W. Thuiller, M.-J. Fortin, C. Randin, N. E. Zimmermann, et al. 2012. Measuring ecological niche overlap from occurrence and spatial environmental data. Global Ecology and Biogeography 21:481–497.

Brown, J. H. 1984. On the relationship between abundance and distribution of species. The American Naturalist 124:255–279.

Campbell, R., and F. Sorensen. 1978. Effect of test environment on expression of clines and on delimitation of seed zones in Douglas-fir. Theoretical and Applied Genetics 51:233–246.

Crow, T. M., S. E. Albeke, C. A. Buerkle, and K. M. Hufford. 2018. Provisional methods to guide species-specific seed transfer in ecological restoration. Ecosphere 9:e02059.

Dorn, R. D. 2001. Vascular plants of Wyoming, volume 3. Cheyenne, Wyoming: Mountain West Publishing 340p.-illus.. En Icones, Keys.

Doyle, J. J. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem bull 19:11–15.

Eckert, C., K. Samis, and S. Lougheed. 2008. Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. Molecular Ecology 17:1170–1188.

Ellstrand, N. C., and D. R. Elam. 1993. Population genetic consequences of small population size: implications for plant conservation. Annual review of Ecology and Systematics 24:217–242.

Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587.

Fenster, C. B., and L. F. Galloway. 2000. Inbreeding and outbreeding depression in natural populations of *Chamaecrista fasciculata* (Fabaceae). Conservation Biology 14:1406–1412.

Franklin, J., J. M. Serra-Diaz, A. D. Syphard, and H. M. Regan. 2016. Global change and terrestrial plant community dynamics. Proceedings of the National Academy of Sciences 113:3725–3734.

Frichot, E., and O. François. 2015. LEA: an R package for landscape and ecological association studies. Methods in Ecology and Evolution 6:925–929.

Frichot, E., F. Mathieu, T. Trouillon, G. Bouchard, and O. François. 2014. Fast and efficient estimation of individual ancestry coefficients. Genetics 196:973–983.

Fukami, T. 2015. Historical contingency in community assembly: integrating niches, species pools, and priority effects. Annual Review of Ecology, Evolution, and Systematics 46:1–23.

Gapare, W. J., S. N. Aitken, and C. E. Ritland. 2005. Genetic diversity of core and peripheral Sitka spruce (*Picea sitchensis* (Bong.) Carr) populations: implications for conservation of widespread species. Biological Conservation 123:113–123.

Geng, Q., Z. Yao, J. Yang, J. He, D. Wang, Z. Wang, and H. Liu. 2015. Effect of Yangtze River on population genetic structure of the relict plant *Parrotia subaequalis* in eastern China. Ecology and evolution 5:4617–4627.

Hamann, A., M. Koshy, G. Namkoong, and C. Ying. 2000. Genotype× environment interactions in alnus rubra: developing seed zones and seed-transfer guidelines with spatial statistics and gis. Forest Ecology and Management 136:107–119.

Hargreaves, A. L., and C. G. Eckert. 2019. Local adaptation primes cold-edge populations for range expansion but not warming-induced range shifts. Ecology letters 22:78–88.

Hufford, K. M., and S. J. Mazer. 2003. Plant ecotypes: genetic differentiation in the age of ecological restoration. Trends in Ecology & Evolution 18:147–155.

Hutchinson, G. E. 1978. An introduction to population ecology. Hew Haven Connecticut Yale University Press 1978.

Hutchison, G. 1957. Concluding remarks. Cold Spring Harbor Symposia on Quantitative Biology 22:415–427.

Jaros, U., A. Tribsch, and H. P. Comes. 2017. Diversification in continental island archipelagos: new evidence on the roles of fragmentation, colonization and gene flow on the genetic divergence of *Aegean Nigella* (Ranunculaceae). Annals of botany 121:241–254.

Johnson, G., F. C. Sorensen, J. B. St Clair, and R. C. Cronn. 2004. Pacific Northwest forest tree seed zones: a template for native plants? Native Plants Journal 5:131–140.

Ju, M.-M., Y. Fu, G.-F. Zhao, C.-Z. He, Z.-H. Li, and B. Tian. 2018. Effects of the tanaka line on the genetic structure of *Bombax ceiba* (Malvaceae) in dry-hot valley areas of southwest China. Ecology and evolution 8:3599–3608.

Klemmedson, J. 1979. Ecological importance of actinomycete-nodulated plants in the western United States. Botanical Gazette 140:S91–S96.

Korneliussen, T. S., A. Albrechtsen, and R. Nielsen. 2014. ANGSD: analysis of next generation sequencing data. BMC Bioinformatics 15:356.

Krauss, S. L., and T. H. He. 2006. Rapid genetic identification of local provenance seed collection zones for ecological restoration and biodiversity conservation. Journal for Nature Conservation 14:190–199.

Lee-Yaw, J. A., M. Fracassetti, and Y. Willi. 2017. Environmental marginality and geographic range limits: A case study with *Arabidopsis lyrata* ssp. *lyrata*. Ecography 41:622–634.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760.

Li, J., Q. Li, W. Hou, K. Han, Y. Li, S. Wu, Y. Li, and R. Wu. 2009. An algorithmic model for constructing a linkage and linkage disequilibrium map in outcrossing plant populations. Genetics research 91:9–21.

Lira-Noriega, A., and J. D. Manthey. 2014. Relationship of genetic diversity and niche centrality: a survey and analysis. Evolution 68:1082–1093.

Louda, S. M. 1982. Distribution ecology: variation in plant recruitment over a gradient in relation to insect seed predation. Ecological Monographs 52:25–41.

Machado, A. P., L. Clément, V. Uva, J. Goudet, and A. Roulin. 2018. The Rocky Mountains as a dispersal barrier between barn owl (*Tyto alba*) populations in North America. Journal of Biogeography .

Merow, C., N. LaFleur, J. A. Silander Jr, A. M. Wilson, and M. Rubega. 2011. Developing dynamic mechanistic species distribution models: predicting bird-mediated spread of invasive plants across northeastern North America. The American Naturalist 178:30–43.

Mimura, M., and S. Aitken. 2010. Local adaptation at the range peripheries of Sitka spruce. Journal of Evolutionary Biology 23:249–258.

Montalvo, A. M., and N. C. Ellstrand. 2001. Nonlocal transplantation and outbreeding depression in the subshrub *Lotus scoparius* (Fabaceae). American Journal of Botany 88:258–269.

Murphy, M. A., J. S. Evans, and A. Storfer. 2010. Quantifying *Bufo boreas* connectivity in Yellowstone National Park with landscape genetics. Ecology 91:252–261.

Nei, M., T. Maruyama, and C.-I. Wu. 1983. Models of evolution of reproductive isolation. Genetics 103:557–579.

Pagel, J., and F. M. Schurr. 2012. Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. Global Ecology and Biogeography 21:293–304.

Parchman, T. L., Z. Gompert, J. Mudge, F. D. Schilkey, C. W. Benkman, and C. Buerkle. 2012. Genome-wide association genetics of an adaptive trait in lodgepole pine. Molecular Ecology 21:2991–3005.

Parker, W. H., and A. v. Niejenhuis. 1996. Regression-based focal point seed zones for *Picea mariana* from northwestern Ontario. Canadian Journal of Botany 74:1227–1235.

Pironon, S., G. Papuga, J. Villellas, A. L. Angert, M. B. García, and J. D. Thompson. 2017. Geographic variation in genetic and demographic performance: new insights from an old biogeographical paradigm. Biological Reviews 92:1877–1909.

Plummer, M. 2003. Jags: A program for analysis of bayesian graphical models using gibbs sampling. in Proceedings of the 3rd international workshop on distributed statistical computing.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. `https://www.R-project.org/`

Reed, D. H., and R. Frankham. 2003. Correlation between fitness and genetic diversity. Conservation biology 17:230–237.

Reeves, P. A., and C. M. Richards. 2014. Effect of a geographic barrier on adaptation in the dwarf sunflower (*Helianthus pumilus* nutt.). International Journal of Plant Sciences 175:688–701.

Reynolds, L. K., M. Waycott, and K. J. McGlathery. 2013. Restoration recovers population structure and landscape genetic connectivity in a dispersal-limited ecosystem. Journal of Ecology 101:1288–1297.

Rodríguez, A., M. Börner, M. Pabijan, M. Gehara, C. F. Haddad, and M. Vences. 2015. Genetic divergence in tropical anurans: deeper phylogeographic structure in forest specialists and in topographically complex regions. Evolutionary ecology 29:765–785.

Saenz-Romero, C., and B. Tapia-Olivares. 2008. Genetic variation in frost damage and seed zone delineation within an altitudinal transect of *Pinus devoniana* (*P. michoacana*) in Mexico. Silvae Genetica 57:165.

Schield, D. R., R. H. Adams, D. C. Card, A. B. Corbin, T. Jezkova, N. R. Hales, J. M. Meik, B. W. Perry, C. L. Spencer, L. L. Smith, et al. 2018. Cryptic genetic diversity, population structure, and gene flow in the Mojave rattlesnake (*Crotalus scutulatus*). Molecular phylogenetics and evolution 127:669–681.

Schoener, T. W. 1970. Nonsynchronous spatial overlap of lizards in patchy habitats. Ecology 51:408–418.

Sexton, J. P., S. B. Hangartner, and A. A. Hoffmann. 2014. Genetic isolation by environment or distance: which pattern of gene flow is most common? Evolution 68:1–15.

Sexton, J. P., P. J. McIntyre, A. L. Angert, and K. J. Rice. 2009. Evolution and ecology of species range limits. Annual Review of Ecology, Evolution, and Systematics 40.

Stutz, H. C. 1988. Taxonomy and evolution of *Cercocarpus* in the western United States. in

K. L. Johnson, editor. 5th utah shrub ecology workshop: The genus cercocarpus, volume 16093, pages 15–25 . Utah Utah State University, Logan, Utah.

Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437–460.

Takezaki, N., and M. Nei. 1996. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. Genetics 144:389–399.

van Etten, J. 2018. gdistance: Distances and routes on geographical grids. R package version 1.2-2. `https://CRAN.R-project.org/package=gdistance`

Wang, Z.-F., J.-Y. Lian, W.-H. Ye, H.-L. Cao, and Z.-M. Wang. 2014. The spatial genetic pattern of castanopsis chinensis in a large forest plot with complex topography. Forest ecology and management 318:318–325.

Watterson, G. 1975. On the number of segregating sites in genetical models without recombination. Theoretical Population Biology 7:256–276.

Weir, B. S., and W. G. Hill. 2002. Estimating F-statistics. Annual review of genetics 36:721–750.

Williams, S., J. Johnson, L. Munn, and T. Nieder. 2004. Edaphic characteristics of nitrogen fixing nodulation (Actinorhizae) by *Cercocarpus montanus* Raf. and *Purshia tridentata* (Pursh) DC. in USDA Forest Service Proceedings RMRS-P, volume 31, pages 84–90 .

Zhu, Q., B.-Y. Liao, P. Li, J.-C. Li, X.-M. Deng, X.-S. Hu, and X.-Y. Chen. 2017. Phylogeographic pattern suggests a general northeastward dispersal in the distribution of *Machilus pauhoi* in South China. PloS one 12:e0184456.

Table 1: Bayesian linear regression models and coefficients. Predictor variables were standardized using a z-score prior to modeling. Genetic distance (GenDist) was calculated as Nei's D_A. Environmental distance is a multivariate distance matrix of degree days less than zero, and growing season precipitation. Geography is a pairwise geographic distance matrix. The smallest DIC indicates the best model.

Model	β_1 (SD)	β_2 (SD)	μ (SD)	DIC
GenDist ~ Environment + Geography	0.173 (0.013)	0.297 (0.0126)	-0.0028 (0.0457)	601.242
GenDist ~ Barrier + Geography	0.364 (0.0367)	0.301 (0.0132)	-9.63e-05 (0.0419)	679.119
GenDist ~ Barrier + Environment	0.465 (0.0408)	0.207 (0.0151)	0.000258 (0.0538)	921.633
GenDist ~ Geography	0.327 (0.0135)		-0.00132 (0.0494)	759.724
GenDist ~ Environment	0.229 (0.0156)		0.00428 (0.061)	1026.883
GenDist ~ Barrier	0.533 (0.0435)		0.000125 (0.0491)	1105.838

Table 2: Summary of linear regression models and model selection criterion for the effects of geographic and environmental centrality on genetic diversity. *P<0.05,**P<0.01.

Model	β_1 (CI)	β_2 (CI)	μ (CI)	R^2	adj. R^2	AIC
θ_π ~ Environment (Env)	-0.01** (-.03 - 0.001)	NA	0.36 (0.32-0.40)	.106	.085	-206
θ_π ~ Geography (Geo)	-0.01 (-0.01 - 0)	NA	0.31 (0.31-0.32)	0.033	0.010	-202
θ_π ~ Env + Geo	-.01 (-.03 - 0)	0 (-0.01 - 0)	0.36 (0.32-0.40)	0.109	.067	-204

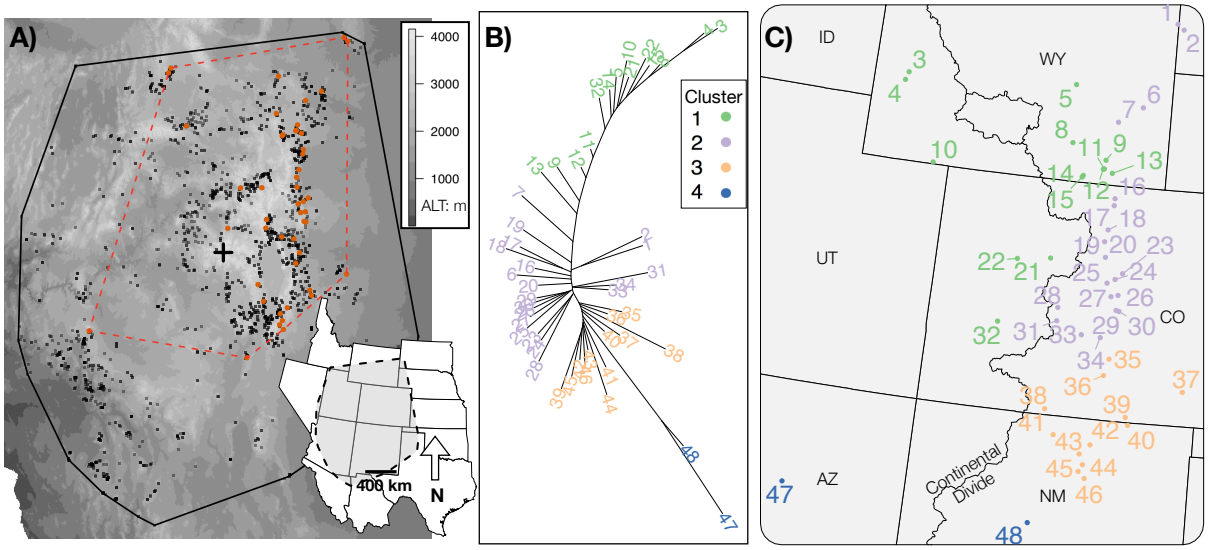

Figure 1: A) Minimum convex polygon (mcp) of species range (black line) around species occurrence points (black squares) and the dashed red line is a mcp around 48 sampled populations (red diamonds). The geographic center of the overall species distribution mcp is marked with a cross. B) Unrooted neighbor-joining tree of Nei's D_A, colors correspond to assigned genetic cluster. C) Map of sampled populations with numbers from 1 to 48 based on latitude for reference.

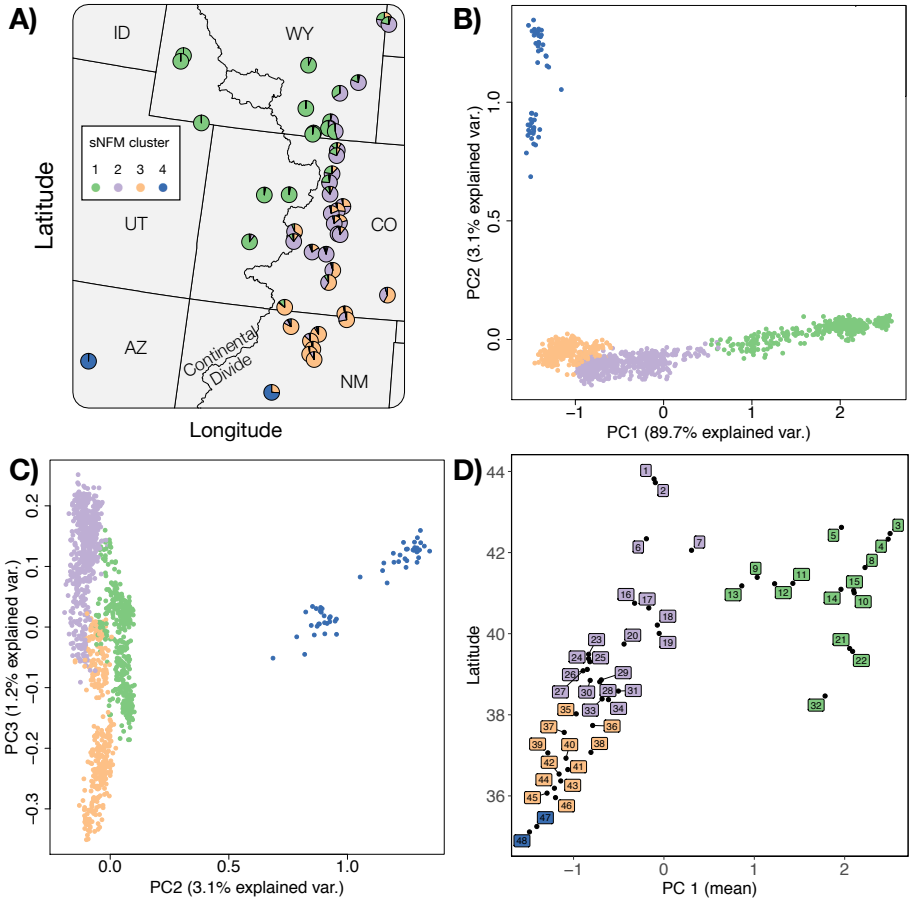

Figure 2: sNFM admixture and principal component analysis of *Cercocarpus montanus*. A) Pie chart of sNFM admixture proportions from k=4 ancestral gene pools (Figure S3) for each of the 48 populations collected in our study. B) PC axis one and two and C) two and three show continuous genetic variation across individuals within clusters. D) Scatter plot of the mean PC axis one score for each of the 48 populations plotted with latitude to visualize geographic structure. Points are colored based on the predominant population assignment from admixture analysis.

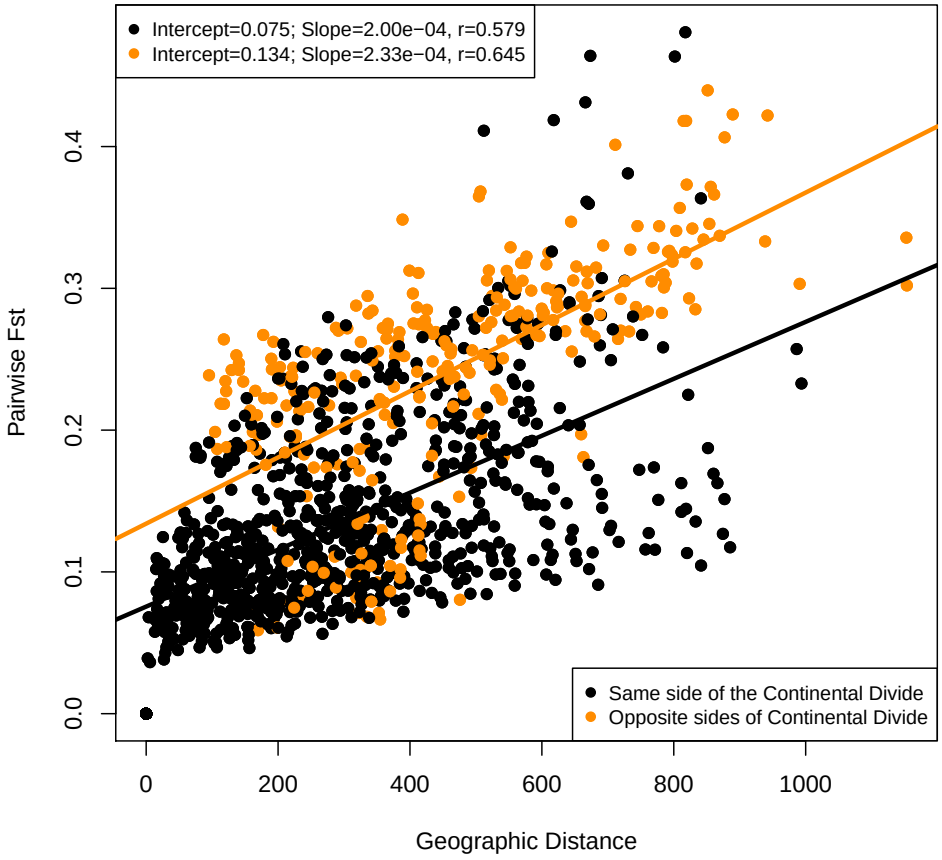

Figure 3: Matrix regression of pairwise genetic and geographic distances. Red points are point pairs from opposite sides of the Continental Divide, while black points are point pairs from the same side of the Continental Divide. Two separate linear models results are listed and model line and summaries correspond to point colors.

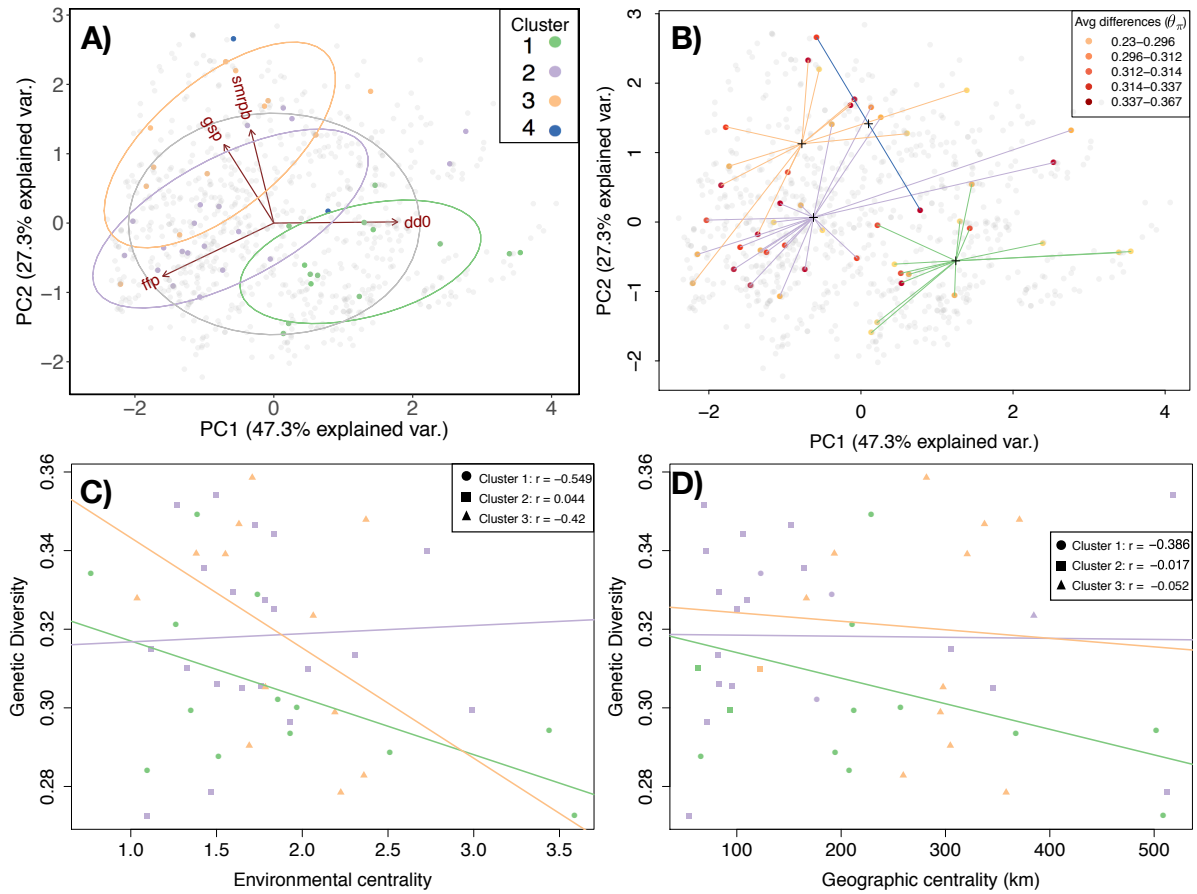

Figure 4: Principal component analysis of growing season precipitation (GSP), summer precipitation balance (smrpb), frost-free period (ffp) and degree days below zero Celsius (dd0). A) The genetic cluster assignment in environmental PCA space, B) shows genetic diversity for each population (point) and the distance (lines) of each population to the cluster-specific environmental centroid (crosses). Finally genetic diversity (θ_π) plotted over C) environmental centrality and D) geographic centrality. More central populations are closer to zero. Regression lines were modeled for each genetic cluster separately.