bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Implications of genetic heterogeneity for plant translocation during ecological restoration

Taylor M. Crow[1], C. Alex Buerkle[2], Daniel E. Runcie[1], Kristina M. Hufford[3]

1 Department of Plant Sciences, University of California, Davis, CA, USA. [2] Department of Botany, University of Wyoming, Laramie, WY, 82071, USA, [3] Ecosystem Science and Management, University of Wyoming, Laramie, WY 82071, USA

Keywords: *Cercocarpus montanus*, ecological restoration, genetic differentiation, genetic structure, niche, phylogeography, seed transfer, seed transfer zones

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 # Abstract

2 Ecological restoration often requires translocating plant material from distant sites. Yet
3 published guidelines for seed transfer are available for very few species. Accurately predicting
4 how plants will perform when transferred requires multi-year and multi-environment field
5 trials and comprehensive follow-up work. In this study, we analyzed the genetic structure
6 of an important shrub used in ecological restorations in the Southern Rocky Mountains
7 called alder-leaf mountain mahogany (*Cercocarpus montanus*). We sequenced DNA from
8 1440 plants in 48 populations across a broad geographic range. We found that genetic
9 heterogeneity among populations reflected the complex climate and topography across which
10 the species is distributed. We identified several temperature and precipitation variables that
11 were useful predictors of genetic differentiation and can be used to generate seed transfer
12 recommendations. These results will be valuable for defining management and restoration
13 practices for mountain mahogany and other widespread montane plant species.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

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 individuals 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 environmental and spatial distance 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 however are limited by the number of populations, number of environments, and the amount of time it may take to quantify consequences of importing foreign plant material (Johnson et al., 2004).

Models based on climate (Bower et al., 2014; Crow et al., 2018) or genetic data (Krauss and He, 2006), or a combination of both (Massatti et al., 2020) may be useful 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

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

41 combinations of alleles in native populations. Further, gene flow between introduced and
42 native populations may lead to outbreeding depression when locally adapted gene complexes
43 are disrupted by immigrant alleles after admixture (Fenster and Galloway, 2000; Montalvo
44 and Ellstrand, 2001). Identifying geographic and environmental patterns related to genetic
45 differentiation can therefore provide useful guidelines for seed introductions in ecological
46 restoration (Montalvo and Ellstrand, 2001).

47 In addition to field experiments, knowledge of genetic subdivision of species across space
48 and its association with dimensions of the environmental niche can also contribute to the
49 development of seed transfer guidelines. The niche concept incorporates both the environ-
50 mental and spatial distribution of a species, and can be used in understanding factors gov-
51 erning range limits (Sexton et al., 2009). One conceptualization of a niche is summarized by
52 Hutchison's n-dimensional hypervolume (Hutchison, 1957), described as a set of biologically
53 relevant and independent environmental axes within which a species occurs. The multi-
54 variate environmental space represents conditions that accommodate population persistence
55 and growth (Hutchinson, 1978). As habitat quality or availability decreases, population size
56 and gene flow are expected to decrease (Brown, 1984; Eckert et al., 2008). Understanding
57 the relationship between species' genetic structure and niche can lead to the identification
58 of evolved population differences and locally adapted ecotypes to inform guidelines for seed
59 transfer.

60 In this study we investigated genetic variation relevant for restoration of a native perennial
61 shrub, alder-leaf mountain mahogany, *Cercocarpus montanus*. Mountain mahogany is used in
62 restoration projects because of its value as a forage plant for large ungulates, especially in the
63 winter months. We collected and sequenced DNA from 1440 individual plant samples from 48
64 populations, estimated genetic diversity within populations, and measured variation at over
65 6,000 single nucleotide polymorphisms (SNPs) to describe genetic structure. We tested to
66 what extent genetic structure was a function of latitude, habitat quality, niche centrality, or a
67 combination thereof, with the goal of informing seed transfer recommendations for mountain

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

68 mahogany.

69 Methods

70 Study species

71 *Cercocarpus montanus* Raf. is a deciduous, perennial shrub species in the rose family (Rosaceae)
72 with a large spatial distribution in western North America (Dorn, 2001). The species occurs
73 on both sides of the Continental Divide and from northern Mexico to the Wyoming-Montana
74 state borders in the United States (Fig. 1). Populations are generally distributed between
75 1200 and 3000 meters in elevation and often grow in rocky, limestone soils (Williams et al.,
76 2004). Mountain mahogany are monoecious and have wind-pollinated flowers. Fruits are ach-
77 enes with an elongated style that twists in later development, and is covered in trichomes.
78 These structures are hypothesised to aid in wind and animal mediated dispersal (Gucker,
79 2006). Mountain mahogany shrubs serve as hosts for nitrogen-fixing actinomycete bacteria
80 (genus *Frankia*) in root nodules, and this adaptation contributes to successional processes in
81 arid regions dominated by unstable, low nitrogen soils (Klemmedson, 1979).

82 DNA extraction, sequencing, assembly and variant detection

83 Mountain mahogany populations were located along a north-south axis in the southern Rocky
84 Mountains (Fig. 1). We collected leaf tissue from 30 individuals in each of 48 populations
85 and extracted DNA using a modified cetyl-trimethyl ammonium bromide (CTAB) proto-
86 col (Doyle, 1987). DNA was quantified with a NanoDrop 2000 spectrophotometer (Thermo
87 Fisher, Inc.), and additional extractions were conducted when necessary due to high levels
88 of contaminants or low DNA concentrations. We prepared genomic libraries for genotype-
89 by-sequencing (GBS) following protocols in Parchman et al. (2012). To summarize, we
90 digested sample DNA with two restriction enzymes (MseI and EcoRI) and ligated barcodes
91 containing unique 8–10 bp sequences to the resulting DNA fragments for each sample to

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

92 ensure that sequence reads could be assigned to individuals. We then PCR amplified the
93 barcoded restriction-ligation products with standard Illumina primers (1, 5′ - AATGAT-
94 ACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT - 3′;
95 2, 5′ - CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT - 3′) (Illumina, Inc.).

96 Barcoded PCR products were combined into two multiplexed libraries of 720 individual
97 samples (with individuals allocated to the libraries randomly to avoid confounding library
98 effects) and sequenced at the University of Texas Genomic Sequencing and Analysis Facility
99 (Austin, Texas, USA) on the Illumina HiSeq 2500 platform using single-end 100 bp reads.
100 After filtering reads for oligonucleotides used in library synthesis and the PhiX genome, with
101 subsequent demultiplexing and assignment of reads to individuals, we had 24000000 sequence
102 reads for further analysis. We completed a de novo genome assembly with a randomly
103 chosen subset of 2.4×10^7 reads using SEQMAN NGEN software (DNASTAR, Inc.). This
104 step resulted in construction of an artificial, partial reference genome containing 111 967
105 contigs. We used `bwa` (Burrows-Wheeler Aligner; Li and Durbin 2009) to map reads from
106 each individual to this partial reference genome. Once complete, 15 520 448 total reads
107 (64.6%) assembled to the partial reference genome. Aligned reads were then indexed and
108 sorted using samtools and bcftools (Li et al., 2009). We used the command 'mpileup -P
109 ILLUMINA -u -g -I -f cemo.fasta sorted.bam | bcftools view -N -c -e -g -v -I -d 0.8 -p 0.01
110 -P full -t 0.001 -o variants.vcf' to calculate genotype likelihoods and filter variant sites. We
111 then retained a single SNP per contig and removed SNPs with an allele frequency less than
112 0.05.

113 Population genetic analyses

114 We estimated genotypes as the mean of the genotype likelihood distribution and constructed
115 a genetic covariance matrix for all individuals. We ran a principal components analysis
116 (PCA) of the genetic covariance matrix using the *prcomp* function in *R* to summarize genetic
117 variation. We tested for correlations between the individual scores on the first two principal

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

118 component axes and potential drivers of genetic variation such as latitude, elevation, pre-
119 cipitation and temperature. Additionally, genotype data were used to calculate individual
120 admixture coefficients using the sparse non-negative matrix factorization algorithm (sNMF)
121 implemented in the LEA package (Frichot et al., 2014; Frichot and François, 2015) in R. This
122 algorithm estimates ancestry coefficients in a computationally efficient manner. The sNMF
123 algorithm is similar to the program STRUCTURE (Pritchard et al., 2000; Falush et al.,
124 2003), which estimates ancestry independently for each individual, and does not require *a*
125 *priori* assumptions about population membership. To determine the best-supported number
126 of genetic clusters (K) within our collections of mountain mahogany, we used a cross-entropy
127 criterion from K=1 to K=10 from the SNMF function. This criterion uses a masked genotype
128 testing set to determine the prediction accuracy of the model at each K value.

129 Point estimates of allele frequencies within each population were calculated from the
130 genotype likelihoods, and allele frequencies were used to calculate the Weir moment estima-
131 tor of F_{ST} (Weir and Hill, 2002) and Nei's genetic distance (D_A) (Nei et al., 1983; Takezaki
132 and Nei, 1996) as measures of genetic differentiation. F_{ST} was calculated using the CALCU-
133 LATE.ALL.PAIRWISE.FST function in the BEDASSLE package in R, and D_A was calculated
134 using a custom R script.

135 We used Bayesian linear models with Nei's D_A as the response variable, and pairwise ge-
136 ographic distance, environment distance, and a binary variable representing the Continental
137 Divide as model predictors. Population pairs were assigned 0 if they originated from the
138 same side of the Continental Divide, or assigned 1 if they were collected from opposite sides
139 of the divide. Environmental distances were measured as the population pair difference for
140 each environmental variable centered on the mean and divided by the standard deviation (z-
141 score). Environmental variables included thirty-year normal temperature and precipitation
142 estimates from thin plate spline surfaces (`http://forest.moscowfsl.wsu.edu/climate`).
143 All predictor variables (Table S1) were standardized prior to modeling so that the magni-
144 tude of their estimated coefficients could be compared. We fit the full and reduced models for

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 steps 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 spatial distance, environmental distance, and topographic barriers as covariates.

Relative contribution of geographic and environmental distance to genetic differentiation

Geographic and environmental distances could 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 topographic barriers to gene flow, relative to spatial distance. This model was developed by Bradburd et al. (2013), is called Bayesian Estimation of Differentiation in Alleles by Spatial Structure and Local Ecology, and is implemented in the R package BEDASSLE. We tested the complete dataset, and used the beta-binomial Markov Chain Monte Carlo model. 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 spatial and environmental centrality. We estimated genetic diversity for each population using the program ANGSD (Korneliussen et al., 2014). Sequence alignments to the pseudo-reference (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

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

169 frequency likelihoods were used to calculate the maximum likelihood estimate (MLE) of the
170 site frequency spectrum (SFS) using the EM algorithm. Estimates of nucleotide polymor-
171 phisms were calculated as θ_π (Tajima, 1983), a measure of average pairwise differences, and
172 Watterson θ_W (Watterson, 1975), which is based on the number of segregating sites. Theta
173 estimates were calculated using the empirical Bayesian approach with the SFS as priors
174 (following `http://popgen.dk/angsd/index.php/Thetas,Tajima,Neutrality_tests`).

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

192 Niche similarity among genetic clusters

193 Niche overlap statistics were used to test if genetic clusters defined by the sNMF admixture
194 analysis occupied distinct subsets of the overall environmental range. Broennimann et al.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

195 (2012) developed methods to get an unbiased estimate of niche overlap using kernel smoother
196 functions applied to densities of occurrence points in environmental space, calibrated on the
197 available environmental space across the study area. We calculated kernel densities for
198 the environment occupied by each genetic cluster, and used D metrics (Schoener, 1970) to
199 determine if there was significant overlap of niche space between genetic groups:

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

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

210 Results

211 Sequence alignment and SNP discovery

212 We identified 12 022 single nucleotide variants using samtools and bcftools (Li and Durbin,
213 2009). For a variant site to be identified, we required that at least 50% of all individuals have
214 a minimum of one read at that locus. After removing sites with a minor allele frequency
215 of <5% and randomly selecting one variant per contig to ensure independence of loci, we
216 retained 6352 single nucleotide polymorphisms (SNPs) for further analyses of population

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

217 genetic structure. In sum, 1366 of the 1440 individual samples of *C. montanus* had suffi-
218 cient sequencing coverage to be retained for further analysis, resulting in a range of 22–30
219 individuals per population. Remaining samples each had an average of 8.5 reads per SNP.

220 Population genetic analyses

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

231 The mean Nei's D_A genetic distance between populations was 0.0346 (SD=0.017), with
232 a range of 0.009–0.108, comparable to previous studies of plant species (Reynolds et al.,
233 2013; Abraham et al., 2015). Pairwise F_{ST} had an overall mean of 0.161, and a SD of 0.0856
234 (Fig. S1). The mean F_{ST} among pairs of populations from opposite sides of the Continental
235 Divide was 0.241 (SD=0.079), while the mean F_{ST} among populations on the same side
236 of the divide was 0.135 (SD=0.07). Pairwise F_{ST} was positively correlated with spatial
237 distance, and population pairs from opposite sides of the Continental Divide had elevated
238 F_{ST} resulting from the effective topographic barrier (Fig. 3). Growing season precipitation
239 and degree days less than 0°C were standardized and combined as a single mean Euclidean
240 distance for each population pair, and served as environmental predictors in modeling. The
241 Bayesian linear model with the lowest DIC included both spatial and environmental distance
242 as predictors of genetic differentiation (Table 1). The best predictor in a univariate model

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

243 of genetic differentiation was geographic distance, followed by environmental distance, while
244 the binary design matrix representing the Continental Divide was the worst predictor.

245 The best supported number of clusters for sNFM admixture analysis was K=4 (Fig. S2).
246 Populations were assigned to a single cluster based on the predominant population admixture
247 coefficient of individuals within each population (Fig. S3). Although genetic variation is
248 continuous through most of mountain mahogany's range, the map of admixture composition
249 shows that the genetic clusters are partitioned in geographic space (Fig. 2, panel A), with
250 more highly admixed zones between clusters. The genetic clusters occupied regions of the
251 species environmental space with different multivariate centroids (Fig. 4 panel A). Cluster
252 one and three had no detected overlap in their environmental niche, while clusters one and
253 two and two and three had partial, but not significant overlap in environmental space (Table
254 S3).

255 The BEDASSLE analysis calculated the ratio of environmental and spatial distance effect
256 sizes on genetic differentiation (α_{E}:α_{D}). We used growing season precipitation and degree
257 days less than 0°C as environmental variables, as well as a binary design matrix representing
258 the Continental Divide to quantify the effect of the environment on genetic distance. A
259 difference of one degree days less than 0°C is comparable to approximately 8 kilometers, and
260 a 1 cm change in growing season precipitation has the same effect on genetic differentiation
261 as approximately 70 kilometers spatial distance. The Continental Divide had the largest
262 effect on genetic differentiation relative to spatial distance. Crossing the Continental Divide
263 had the same effect on genetic differentiation in mountain mahogany as moving 1.7×10^7
264 km, a larger distance than our collection area.

265 We detected significant variation in genetic diversity across mountain mahogany's cen-
266 tral range. Nucleotide diversity estimates were highly correlated (r>0.9, θ_π and θ_W), and
267 we therefore arbitrarily chose θ_π for further modeling (Table S2). Genetic diversity was not
268 correlated with latitude (P = 0.266, df = 43, $R^2 = 0.028$). We used two measures of precipi-
269 tation (growing season precipitation and summer precipitation balance) and two temperature

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

270 metrics (degree days less than zero and frost-free period) to model genetic diversity because
271 of low colinearity between variables and high correlation with diversity estimates. Genetic
272 diversity was lower in populations farther from the species' multidimensional environmental
273 centroid. Spatial centrality however, was a poor predictor of θ_π. Likewise, we found that
274 spatial centrality was a poor predictor of the probability of occurrence (Fig. S4). The envi-
275 ronmental distance to the centroid of each genetic cluster best described genetic diversity,
276 and had a negative correlation (Table 2). We also found significant variation among genetic
277 clusters for the effect of environmental and spatial distance; namely genetic variation within
278 the northern and southern genetic clusters (Cluster 1 and 3) both had a significant rela-
279 tionship to environmental marginality, whereas within the central genetic cluster (cluster 2)
280 diversity was not correlated with environment (Fig. 4).

281 Discussion

282 Mountain mahogany is increasingly used in restoration programs, particularly because it
283 hosts nitrogen-fixing actinobacteria that allow establishment in nutrient-poor soils, and pro-
284 vides important overwintering forage for wildlife. Despite widespread occurrence in the
285 Rocky Mountain West, no prior ecological genetics study has characterized genetic structure
286 across mountain mahogany's central range. We sequenced 1440 individuals from six U.S.
287 states in the Southern Rocky Mountains to learn the extent of genetic heterogeneity across
288 the geographic range and the environments occupied by the species.

289 We found evidence that genetic structure of mountain mahogany was affected by spatial
290 and environmental distance, as well as topographic barriers. The results provide preliminary
291 data for seed sourcing guidelines for mountain mahogany. Genetic variation is important
292 to consider for species management, especially in a restoration setting where hundred or
293 thousands of individual plants are transplanted to a new site (Reed and Frankham, 2003).
294 These results have range-wide implications for mountain mahogany shrubland management,

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

295 and lay the groundwork for critical decision-making under environmental change.

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

309 The Continental Divide is associated with greater genetic differences between Mountain
310 mahogany populations, especially in central Colorado, where the Continental Divide is at its
311 highest altitude. Several studies have shown that the Continental Divide is a strong barrier
312 to gene flow (Schield et al., 2018; Machado et al., 2018). However, to date, no published
313 study has documented this in plant species. Several studies have found significant effects
314 of topographic barriers on genetic differentiation in plant species, including: seas (Jaros
315 et al., 2017), lakes and terrain (Ju et al., 2018), rivers (Geng et al., 2015), mountains (Zhu
316 et al., 2017; Reeves and Richards, 2014), and basins (Bontrager and Angert, 2018). Our data
317 agree with these studies and indicate that populations from opposite sides of the Continental
318 Divide are genetically more isolated, despite what may appear to be close spatial proximity
319 (Fig. 3). Populations from the western slopes of the Rocky Mountains had high among-
320 population genetic differentiation, especially populations 3 and 4 (Fig. 1 panel B and C).
321 Population 3 and 4 may have been founded separately from other western slope populations,

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

322 or may contain hybrids with a closely related species, *Cercocarpus ledifolius*, that co-occurs
323 in this region (Stutz, 1988). The two most genetically differentiated populations (47 and 48),
324 in New Mexico and Arizona respectively (Fig. 1 panel B and C), inhabit isolated locations
325 surrounded by desert regions with low habitat suitability (Crow et al., 2018). Populations
326 47 and 48 are likely adapted to high temperature and low precipitation conditions, and may
327 warrant further investigation into their taxonomic status.

328 Despite the heterogeneity of climatic conditions in our study area, we found that the
329 best supported genetic clusters corresponded to plants in cohesive geographic regions (Fig.
330 2). Further, the genetic clusters were associated with significantly different environmental
331 space (Fig. 4A), which corroborates linear modeling results showing that spatial distance and
332 environment are both factors related to genetic variation. Given these results, we analyzed
333 patterns of genetic diversity across both spatial and environmental gradients.

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

347 Reduced genetic variation associated with range limits does not distinguish whether pop-
348 ulations occurring at range limits are demographic sinks maintained by immigration from

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

349 more central habitat, or are important genetic resources adapted to marginal conditions by
350 selection. However, the correlation of genetic diversity and environmental centrality bolsters
351 our findings of genetic structure covarying with the environment. The lack of genetic homo-
352 geneity in mountain mahogany indicates that populations are not equivalent, and caution
353 should be taken when planning transfer of plant propagules, particularly during restora-
354 tion. Other studies of genetic variation near range limits have found contrasting results,
355 even among populations of a species. For example, Hargreaves and Eckert (2018) found
356 that while some populations of the annual plant *Rhinanthus minor* near the range margin
357 had lower fitness, other edge populations were locally adapted. Aguirre-Liguori et al. (2017)
358 found that genetic diversity was lower near the geographic range margin of teosinte, and
359 candidate adaptive SNPs were positively correlated with distance to niche centroid, arguing
360 that populations near the geographic range margins were isolated, while populations near
361 the edges of the environmental niche were locally adapted. In *Picea sitchensis*, populations
362 proximal to the range margin are more likely to carry rare alleles (Gapare et al., 2005), how-
363 ever, a second study of *P. sitchensis* determined that populations near the range limit were
364 locally adapted (Mimura and Aitken, 2010). These studies illustrate that range margins can
365 harbor both source and sink genetic pools even within species, and that making predictions
366 about population fecundity near range margins is difficult.

367 The results of our study suggest that populations of mountain mahogany have genetic
368 structure across its range that is correlated with differences in the environment. The effect of
369 the Continental Divide on genetic structure was significant. This suggests that transferring
370 populations across the Continental Divide would increase the likelihood of maladaptation,
371 and subsequent risks for outbreeding depression among progeny of local and introduced
372 plants. Two climate variables, degree days less than zero and growing season precipitation,
373 were significantly related to population genetic structure as well as differences in genetic
374 diversity. These two variables could delimit collection sites when transferring seed sources
375 during restoration. Choosing a commercial seed source or collection location that is most

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

376 environmentally similar to the restoration site may increase chances of introducing adapted
377 genotypes (Hufford and Mazer, 2003). In the case of mountain mahogany, preliminary seed
378 collection zones could be delineated by the four common clusters in genetic analysis. This is
379 a practical approach given that the four clusters represent large spatial regions for collection
380 despite considerable altitudinal and microhabitat variation. Whether populations near range
381 margins are important resources for conservation in mountain mahogany remains unclear.
382 Plants are subjected to biotic and abiotic stressors that influence population dynamics (Pagel
383 and Schurr, 2012; Franklin et al., 2016), seed predators (Louda, 1982), pollinators (Biesmeijer
384 et al., 2006), and dispersers (Merow et al., 2011). As a result, additional studies are needed
385 to determine the adaptive value of mountain mahogany populations along range margins for
386 ecological restoration, particularly in light of changing climate conditions.

387 Acknowledgments

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

397 References

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

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

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

408 Benayas, J. M. R., A. C. Newton, A. Diaz, and J. M. Bullock. 2009. Enhancement of
409 biodiversity and ecosystem services by ecological restoration: a meta-analysis. science
410 325:1121–1124.

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

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

416 Bontrager, M., and A. L. Angert. 2018. Genetic differentiation is determined by geographic
417 distance in *Clarkia pulchella*. bioRxiv . `https://www.biorxiv.org/content/early/`
418 `2018/07/23/374454`

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

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

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

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

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

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

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

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

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

443 Ellstrand, N. C., and D. R. Elam. 1993. Population genetic consequences of small popula-
444 tion size: implications for plant conservation. Annual review of Ecology and Systematics
445 24:217–242.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

446 Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of population structure using
447 multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–
448 1587.

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

451 Franklin, J., J. M. Serra-Diaz, A. D. Syphard, and H. M. Regan. 2016. Global change and
452 terrestrial plant community dynamics. Proceedings of the National Academy of Sciences
453 113:3725–3734.

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

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

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

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

464 Gucker, C. L. 2006. Cercocarpus montanus. Fire Effects Information System,[Online]. US De-
465 partment of Agriculture, Forest Service, Rocky Mountain Research Station, Fire Sciences
466 Laboratory (Producer) .

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

470 Hargreaves, A. L., and C. G. Eckert. 2018. Local adaptation primes cold-edge populations
471 for range expansion but not warming-induced range shifts. bioRxiv page 259879 .

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

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

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

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

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

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

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

488 Korneliussen, T. S., A. Albrechtsen, and R. Nielsen. 2014. ANGSD: analysis of next gener-
489 ation sequencing data. BMC Bioinformatics 15:356.

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

493 Lee-Yaw, J. A., M. Fracassetti, and Y. Willi. 2017. Environmental marginality and geo-
494 graphic range limits: A case study with *Arabidopsis lyrata* ssp. *lyrata*. Ecography 41:622–
495 634.

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

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

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

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

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

508 Massatti, R., R. K. Shriver, D. E. Winkler, B. A. Richardson, and J. B. Bradford. 2020.
509 Assessment of population genetics and climatic variability can refine climate-informed
510 seed transfer guidelines. Restoration Ecology .

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

516 Montalvo, A. M., and N. C. Ellstrand. 2001. Nonlocal transplantation and outbreeding
517 depression in the subshrub *Lotus scoparius* (Fabaceae). American Journal of Botany
518 88:258–269.

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

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

523 Pagel, J., and F. M. Schurr. 2012. Forecasting species ranges by statistical estimation of
524 ecological niches and spatial population dynamics. Global Ecology and Biogeography
525 21:293–304.

526 Parchman, T. L., Z. Gompert, J. Mudge, F. D. Schilkey, C. W. Benkman, and C. Buerkle.
527 2012. Genome-wide association genetics of an adaptive trait in lodgepole pine. Molecular
528 Ecology 21:2991–3005.

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

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

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

543 Reeves, P. A., and C. M. Richards. 2014. Effect of a geographic barrier on adaptation in
544 the dwarf sunflower (*Helianthus pumilus* nutt.). International Journal of Plant Sciences
545 175:688–701.

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

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

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

556 Schoener, T. W. 1970. Nonsynchronous spatial overlap of lizards in patchy habitats. Ecology
557 51:408–418.

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

562 Stutz, H. C. 1988. Taxonomy and evolution of *Cercocarpus* in the western United States. in
563 K. L. Johnson, editor. 5th utah shrub ecology workshop: The genus cercocarpus, volume
564 16093, pages 15–25 . Utah Utah State University, Logan, Utah.

565 Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics
566 105:437–460.

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

569 van Etten, J. 2018. gdistance: Distances and routes on geographical grids. R package version
570 1.2-2. `https://CRAN.R-project.org/package=gdistance`

571 Watterson, G. 1975. On the number of segregating sites in genetical models without recom-
572 bination. Theoretical Population Biology 7:256–276.

573 Weir, B. S., and W. G. Hill. 2002. Estimating F-statistics. Annual review of genetics 36:721–
574 750.

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 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

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

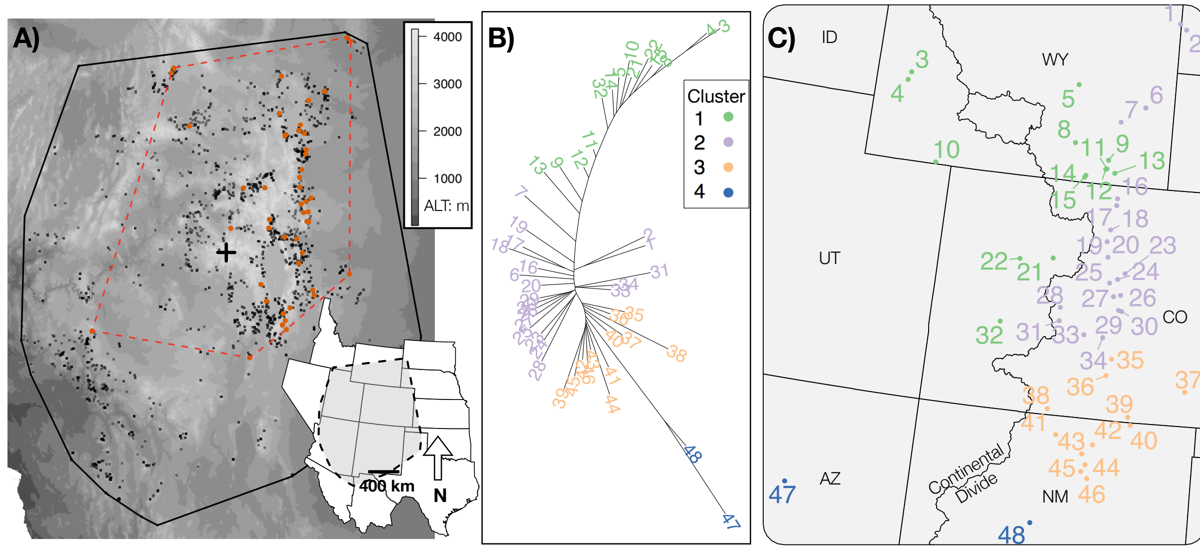

Figure 1: A) Minimum convex polygon (mcp) of species range (black line) around species occurrence points (black squares) and the dashed red line is a mcp around 48 sampled populations (red diamonds). The geographic center of the overall species distribution mcp is marked with a cross. B) Unrooted neighbor-joining 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.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

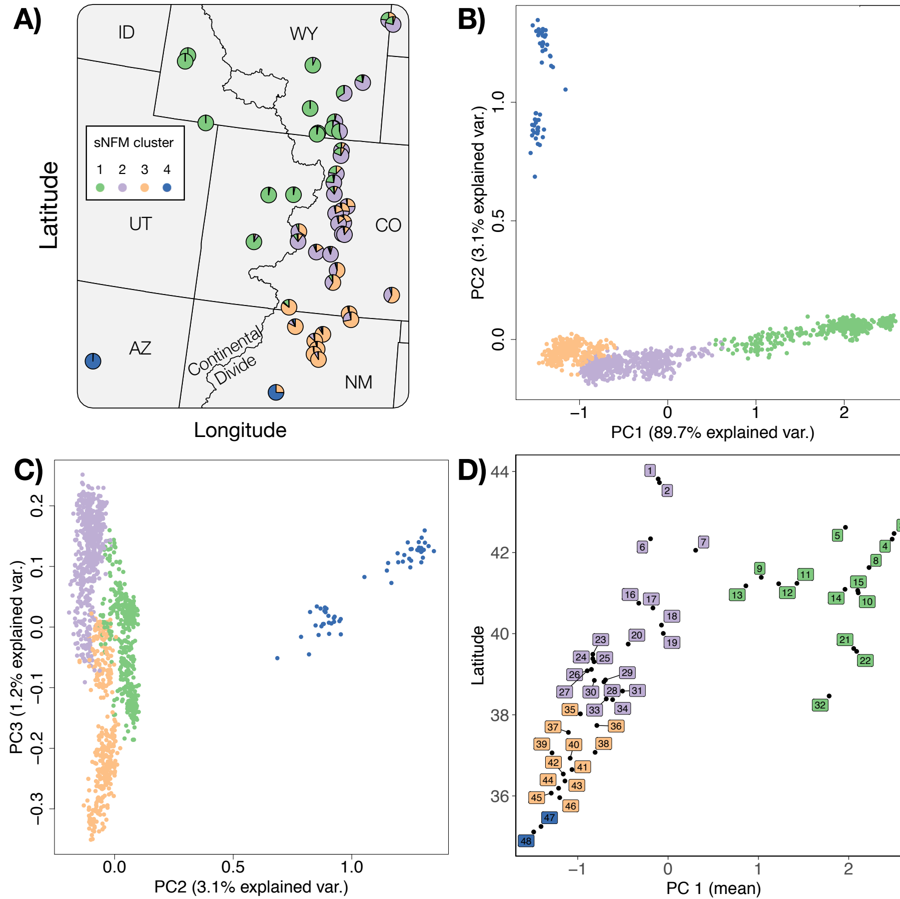

Figure 2: sNFM admixture and principal component analysis of *Cercocarpus montanus*. A) Pie chart of sNFM admixture proportions from k=4 ancestral gene pools (Figure S3) for each of the 48 populations collected in our study. B) PC axis one and two and C) two and three show continuous genetic variation 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.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

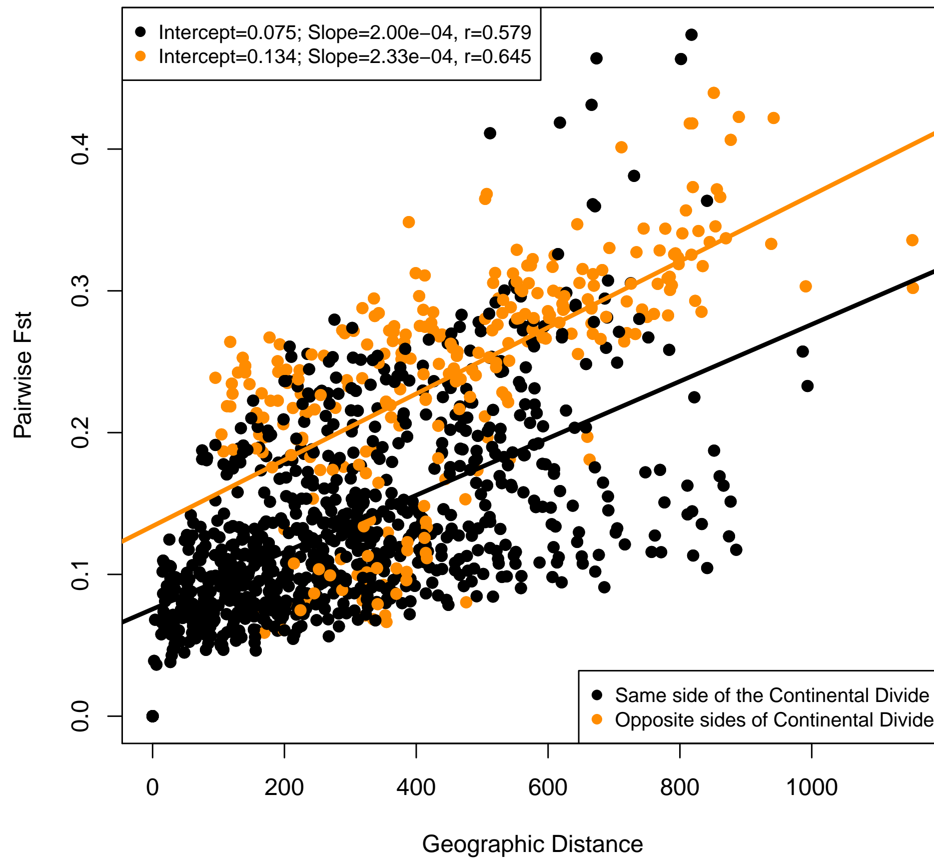

Figure 3: Matrix regression of pairwise genetic and geographic distances. Red points are point pairs from opposite sides of the Continental Divide, while black points are point pairs from the same side of the Continental Divide. Two separate linear models results are listed and model line and summaries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

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.923524; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

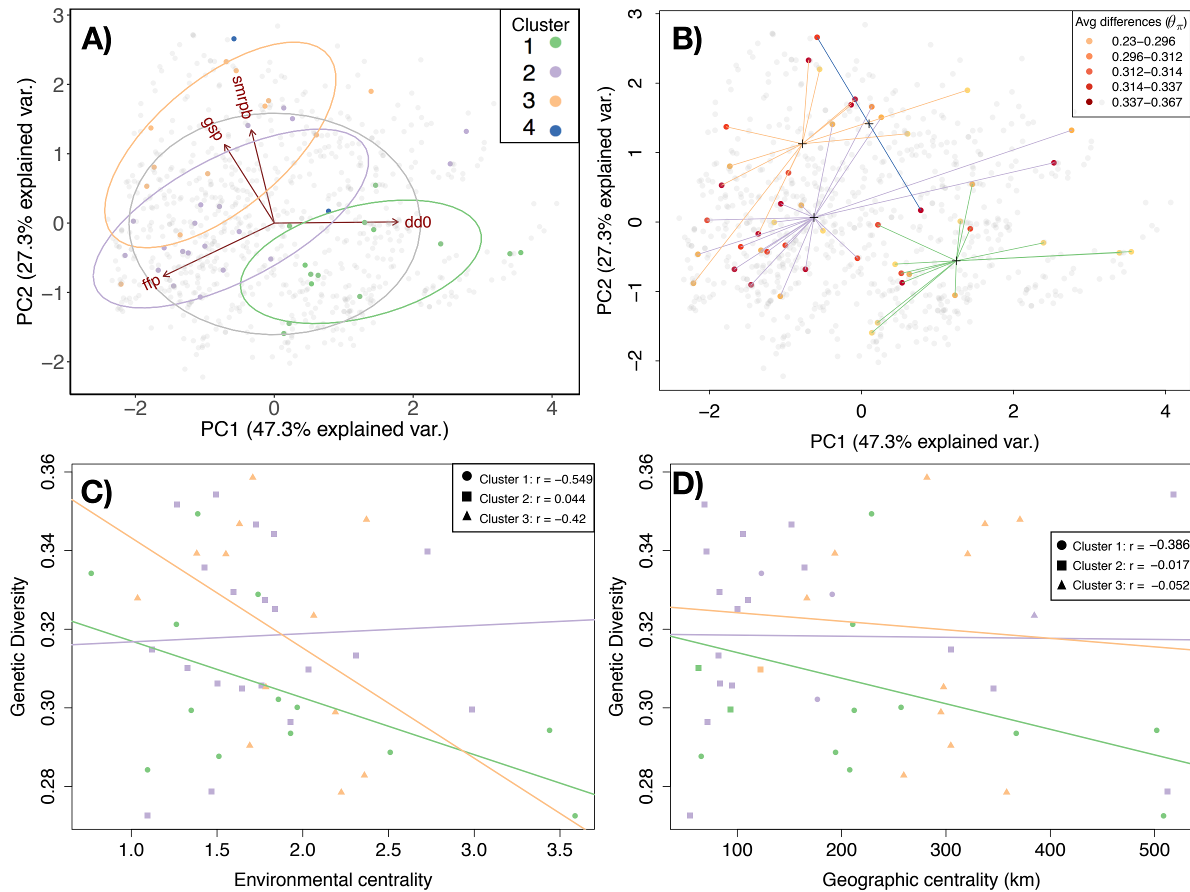

Figure 4: Principal component analysis of growing season precipitation (GSP), summer precipitation balance (smrpb), frost-free period (ffp) and degree days below zero Celsius (dd0). A) The genetic cluster assignment in environmental PCA space, B) shows genetic diversity for each population (point) and 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.