esa

ECOSPHERE

Provisional methods to guide species-specific seed transfer in ecological restoration

TAYLOR M. CROW,[1,]† SHANNON E. ALBEKE,[2] C. ALEX BUERKLE,[3] AND KRISTINA M. HUFFORD[1]

[1]*Ecosystem Science and Management, University of Wyoming, Laramie, Wyoming 82071 USA*
[2]*Wyoming Geographic Information Science Center, University of Wyoming, Laramie, Wyoming 82071 USA*
[3]*Department of Botany, University of Wyoming, Laramie, Wyoming 82071 USA*

Citation: 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(1):e02059. 10.1002/ecs2.2059

Abstract. Transferring plant material during ecological restoration has inherent risk. The use of seed transfer guidelines minimizes the possibility of introducing maladapted genotypes. We delineated biogeographic regions relevant to the distribution of *Cercocarpus montanus* for the purpose of creating provisional seed transfer zones for ecological restoration. We also modeled seed transfer guidelines using quantitative estimates of environmental tolerance and thresholds. Analyses identified broadscale environmental patterns relevant for seed transfer success. First, a species distribution model was used to identify the distribution of *C. montanus*. Next, we used non-metric multidimensional scaling to investigate the structure of environmental data, and hierarchical cluster analysis to delineate biogeographic regions (i.e., environmental discontinuities) using species distribution data. Finally, we calculated measures of environmental tolerance and thresholds for *C. montanus* to model the probability of seed transfer success with multiple logistic regression. Biogeographic regionalization of *C. montanus* resulted in four major clusters, which agreed with ordination methods. Logistic regression was implemented using estimates of environmental tolerance and threshold data to model seed transfer success. We compared our species-specific seed transfer zones and guidelines with other provisional seed transfer zone methods and found that our species-specific methods performed better at explaining phenotypic variation of *C. montanus* in four out of six cases. Seed transfer zones are useful for restoration planning; however, zonal models fail to reflect much of the environmental heterogeneity present across the range of *C. montanus*. Continuous models for seed transfer success using environmental tolerance and thresholds enhance the development and use of seed transfer guidelines because they reflect landscape heterogeneity at a fine scale, and the results are relative to restoration sites of interest. Herein, we describe a methodology to construct provisional seed transfer zones and continuous seed transfer guidelines using species-specific distribution models and multivariate analyses.

Key words: *Cercocarpus montanus;* cluster analysis; environmental niche model; logistic regression; seed transfer guidelines; seed zone; species distribution models.

Received 19 October 2017; accepted 23 October 2017; final version received 17 November 2017. Corresponding Editor: Debra P. C. Peters.

Copyright: © 2018 Crow et al. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

† **E-mail:** tcrow3@uwyo.edu

INTRODUCTION

Understanding the factors that define the spatial extent of local adaptation is a central question in the study of evolution (Chevin et al. 2010). Natural selection often results in adaptive genetic differentiation among populations of plant species to local environmental conditions (Endler 1973). Variation within species for adaptive traits may be continuous and the result of environmental gradients (Gavrilets and Losos 2009). Alternatively, plant ecotypes may form distinct genetic lineages adapted to heterogeneous environments (Turesson 1922, Liancourt

et al. 2013). Understanding intraspecific variation in fitness in relation to the environment is important for the translocation of plant material in ecological restoration and can assist efforts to reduce the potential for introduction of maladapted genotypes.

Rangeland restoration often requires large-scale introductions of native seed, and the genetic origin or provenance of plant materials affects revegetation success (Leimu and Fischer 2008, St. Clair et al. 2013). Seeds derived from long geographic distances from a restoration site and dissimilar environments may lead to low rates of germination, poor seedling establishment, and a decline in survival and fecundity of restored populations (Keller et al. 2000, Hufford and Mazer 2003, Hereford 2009). Seed transfer zones are geographic regions within which seeds may be moved with minimized risks of mean population fitness (Hufford and Mazer 2003), and represent resources to guide the selection of plant material for introduction to restoration sites (Ying and Yanchuk 2006). Seed zones are typically delineated using data from common garden experiments, where adaptive traits with evidence of population differentiation are correlated with variation in climate (Johnson et al. 2004, Saenz-Romero and Tapia-Olivares 2008). Climate variables that explain the largest proportion of heritable phenotypic variation among seed sources are then delineated to map seed transfer zones (Campbell and Sorensen 1978, Rehfeldt 1995). However, seed zones are not commonly characterized in plant species targeted for restoration due to costs and time constraints of common garden studies. In the absence of quantitative studies of plant fitness, provisional methods for creating seed zones include: seed zone delineation through analysis of genetic marker variation with geographic distance (Bussell et al. 2006, Krauss and He 2006), or geography and climate (Hamann et al. 2011), ecoregions (Miller et al. 2011), climate matching (Parker 1992), and generalized provisional zones that combine climate and ecoregion data (Bower et al. 2014). These studies expand the use of seed zones for species that do not have empirically delineated seed zones; however, they either require genetic marker data input (costly and time-consuming), or are broadscale summaries of vegetation or climate or both, which are not species specific.

Species occurrence and environmental data have become widely available, making analysis of species distribution accessible (Elith and Leathwick 2007). Such environmental and taxonomic patterns can be analyzed using ordination and cluster methodologies to delineate biogeographic regions (Kreft and Jetz 2010). Biogeographic regionalization methods provide a replicable framework for capturing discontinuities in environmental factors as they relate to species community data (Hattab et al. 2015). In this study, we apply biogeographic regionalization models to ecological restoration and may represent a useful method for delineation of species-specific provisional seed transfer zones.

Species distribution modeling is a useful tool for restoration and can help identify suitable habitat for threatened plant communities and endangered species (Pearce and Lindenmayer 1998, Riordan and Rundel 2009), as well as predict the effects of climate change for species persistence (Etterson and Shaw 2001, Hamann and Wang 2006, Parida et al. 2015). The environmental variables that are most correlated with species' distributions also help to elucidate range-limiting ecological factors (Pearson 2010, Aguirre-Gutiérrez et al. 2013). Range-limiting factors can influence the adaptive evolution of a species more than non-limiting factors, as populations in peripheral habitat often evolve quickly to the more extreme environmental conditions relative to centralized populations (Hill et al. 2011, Swaegers et al. 2015). The peripheries of suitable habitat can cause rapid trait evolution (Swaegers et al. 2015), and delineating seed transfer zones using range-limiting environmental factors will be important for conserving locally adapted populations.

We describe an approach using a combination of distribution modeling methods and multivariate statistics to map species-specific provisional seed transfer zones and guidelines for *Cercocarpus montanus*, a shrub species with a broad distribution in a region heavily impacted by human development. First, we used species distribution models (SDM) to map the likelihood of a species occurrence and identified environmental variables that best explain variation in the species distribution. Second, we delineated seed transfer zones using cluster analysis and explored the results further with non-metric multidimensional

scaling (NMDS) analysis. Cluster algorithms and NMDS use pairwise distance matrices to delineate data clusters by minimizing variation within groups and maximizing variation among them. Finally, we modeled continuous seed transfer guidelines using quantitative estimates of environmental tolerance and threshold limits. Our objective was to test the utility of species distribution modeling for the development of species-specific provisional seed transfer zones. These methodologies may be useful for plant species targeted for restoration when seed transfer recommendations are not available, or alternative methods for predicting successful transfer are desirable.

Methods

Study species and occurrence data

Alder leaf mountain mahogany, hereafter mountain mahogany, (*Cercocarpus montanus* Raf. Rosaceae), is a shrub species with a widespread distribution native to western North America. Mountain mahogany shrublands commonly occur in mountainous areas from the Black Hills, South Dakota to northern Mexico, and from the Texas panhandle to western Arizona. Plants are usually found between 1200 and 3000 m in elevation and often grow in rocky, shallow soils. Mountain mahogany hosts a nitrogen-fixing actinomycete in root nodules and is valuable as winter forage and habitat for wildlife and livestock (Greenwood and Brotherson 1978, Brotherson 1992). Foothill regions dominated by mountain mahogany are threatened by human development, and mountain mahogany shrublands are increasingly targeted for restoration. Approximately 2600 mountain mahogany occurrence points were obtained from the Southwest Environmental Information Network, (http//:swbiodiversity.org/seinet/index.php), and the Rocky Mountain Herbarium (http://www.rmh.uwyo.edu).

Environmental data

To fit distribution models to species occurrence data, we identified a set of potential environmental variables relevant for the study region and mountain mahogany. Thirty-year normal monthly and annual temperature and precipitation estimates for the reference period from 1961 to 1990 were obtained from thin-plate spline surfaces (http://forest.moscowfsl.wsu.edu/climate), and represent interpolated climate surfaces. A digital elevation model (DEM) was obtained from global 30 arc-second elevation data (GTOPO30; USGS, Washington, D.C.; available at: https://lta.cr.usgs.gov/GTOPO30), with a resolution of 1 km^2 to match climate rasters. To avoid collinearity and deflating variable importance, we removed redundant variables using qr matrix decomposition (threshold = 0.05; Becker et al. 1988). A complete list of variables is available in Table 1.

Species distribution model

We used the Random Forests (RF) algorithm (Breiman 2001) to model the distribution of *C. montanus*. We also tested Maxent (Phillips and Dudík 2008) for SDM of mountain mahogany and found that RF had better model performance for seven out of eight metrics (Appendix S1: Table S1). Random Forests is a type of classification and regression tree machine-learning algorithm, which incorporates complex nonlinear interactions among variables, and is not sensitive to autocorrelation found in environmental and species distributional data (De'ath and Fabricius 2000, Cutler et al. 2007). Environmental variable selection was based on optimal model fit with the least number of environmental variables following Murphy et al. (2010). Each variable was given a model improvement ratio (MIR) calculated as $I_n/I_{\max}$ where I_n is an importance score based on the number of times a variable decreases mean squared error (MSE) of the model, and $I_{\max}$ is the maximum model improvement score. The MIR thresholds were then evaluated to retain a minimum number of variables above a prescribed threshold, minimize MSE, and maximize the percentage of variation explained. We used the RFUTILITIES package (Evans and Murphy 2014) for model selection with the RANDOMFOREST package (Liaw and Wiener 2002) in R (R Core Team 2015) for distribution modeling. The RF model was implemented using 501 trees (bootstrap iterations), with 30% data withheld from each tree as a method of external model validation to calculate the out-of-bag error estimate (Liaw and Wiener 2002). Additionally, 30% of the occurrence and pseudo-absence points were withheld prior to RF

Table 1. Complete list of environmental variables used for model selection prior to species distribution modeling.

Raster layer	Description
D100	Julian date the sum of degree-days >5°C reaches 100
DD0	Degree-days <0°C (based on mean monthly temperature)
DD5	Degree-days >5°C (based on mean monthly temperature)
FDAY	Julian date of the first freezing date of autumn
FFP	Length of the frost-free period (days)
GSDD5	Degree-days >5°C accumulating within the frost-free period
GSP	Growing season precipitation, April–September
MAP	Mean annual precipitation
MATTENTHS	Mean annual temperature
MMAXTENTHS	Mean maximum temperature in the warmest month
MMINTENTHS	Mean minimum temperature in the coldest month
MMINDD0	Degree-days <0°C (based on mean minimum monthly temperature)
MTCMTENTHS	Mean temperature in the coldest month
MTWMTENTHS	Mean temperature in the warmest month
SDAY	Julian date of the last freezing date of spring
SMRP	Summer precipitation: (Jul + Aug)
SMRPB	Summer precipitation balance: (Jul + Aug + Sep)/(Apr + May + Jun)
SMRSPRPB	Summer/Spring precipitation balance: (Jul + Aug)/(Apr + May)
SPRP	Spring precipitation: (Apr + May)
WINP	Winter precipitation: (Nov + Dec + Jan + Feb)
ADI	Annual dryness index, DD5/MAP
SMI	Summer dryness index, GSDD5/GSP
PRATIO	Ratio of summer precipitation to total precipitation, GSP/MAP
AHMI	Annual heat:moisture index, ((MAT + 15)/MAP) × 10
DEM	Digital elevation model

Notes: Annual heat moisture index, mean annual precipitation, mean maximum temperature of the warmest month, and summer moisture index along with the digital elevation model were used in the final model. Climate rasters were downloaded from USDA Forest Service, Rocky Mountain Research Station, Moscow Forestry Sciences Laboratory.

modeling to calculate model performance metrics including: model accuracy, as well as the omission rate, sensitivity, specificity, proportion correct, and the Kappa statistic (Fielding and Bell 1997, Hernandez et al. 2006).

The SDM represented the probability of occurrence of *C. montanus* across the entire study area. However, presence and absence points were required for further analyses. We therefore transformed the probabilistic SDM map into a binary surface representing predicted presence and absence points for *C. montanus*. Model accuracy statistics were used to select the model threshold at which to transform the data (Escalante et al. 2013). We bootstrapped analyses with replacement 100 times to build an error distribution around model performance statistics.

Ordination

Ordination reduced the dimensionality of the environmental distance matrix, and provided a first step for visualization of biogeographic transition zones (Legendre and Legendre 2012). We used NMDS to perform an ordination of a distance matrix derived from the environmental datasets for the distribution of mountain mahogany. Due to the large number of possible pairwise distances (9,939,301,536) between all occurrence points, we randomly sampled 20,000 occurrence points to calculate a distance matrix and used the same matrix for ordination and cluster analyses. Distance matrices were derived independently for each environmental variable used in the SDM, normalized using a *z*-score, and combined by taking the mean pairwise distance of all predictors for each point pair. Non-metric multidimensional scaling makes no assumption about the linearity or normality of the data (Ludwig and Reynolds 1988), and plots similar samples closer together and less similar cases farther apart (Kreft and Jetz 2010). We performed a two-dimensional NMDS with 100 random starts using the metaMDS function in the VEGAN package (Oksanen et al. 2015). Stress values, the sum of square differences between the fitted and original distance values, were used to assess the fit between the NMDS results and original distance matrix (Legendre and Legendre 2012). The results from the NMDS were used to describe the relationship between points in ordination and geographic space. We also used the NMDS plot to visualize how groups from the cluster analysis were represented in ordination space following Hattab et al. (2015).

Species-specific provisional seed transfer zones

We used a multidimensional cluster analysis to partition mountain mahogany presence data

into groups, and identify discontinuities among the environmental data (herein species-specific provisional seed transfer zones). We used agglomerative hierarchical clustering because biogeographic regions can be considered to be hierarchically related (McLaughlin 1992). The performance of clustering linkage methods and validity of clustering results were evaluated using the cophenetic correlation coefficient, which measured the correlation between the original pairwise distance matrix and the cophenetic distances matrix of the dendrogram. Specifically, the coefficient measured how much of the pairwise distance information was retained in the dendrogram (Sokal and Rohlf 1962, Sneath and Sokal 1973), and the information was used to select the linkage method with the highest cophenetic correlation score. To select the most likely number of clusters, we calculated the Kelley-Gardner-Sutcliffe penalty function (KGS), which maximized the differences between the groups as well as the cohesiveness within groups to determine the optimal number of biogeographic clusters (Kelley et al. 1996). We used the MAPTREE package to run the KGS analysis (White and Gramacy 2012). A graphical representation of our workflow is available in Appendix S1: Fig. S1.

Species-specific seed transfer guidelines

Zonal statistics such as cluster analyses define broadscale biogeographic variation important for seed source evaluation and selection. However, significant levels of environmental variation can also occur within a single seed zone. Furthermore, categorical analyses generally have lower power to explain variation (here species occurrence) than do continuous predictor variables. To address this problem, we developed empirical estimates of environmental tolerance and thresholds to model species-specific seed transfer guidelines. To do this, we calculated the environmental and geographic distance between two occurrence points to measure environmental tolerance and between one occurrence point and one absence point to quantify environmental threshold. We randomly sampled this approximation of environmental tolerance and threshold data with replacement 50,000 times. The environmental distance was measured as the absolute difference between point pairs for each environmental variable, and centered on the mean of all pairwise distances of the same variable.

We used estimates of environmental tolerance and threshold to build a reasonable model of seed suitability for a given restoration site. The environmental variables that were included in the best fitting SDM as well as geographic distance served as potential predictors for multiple logistic regression. The binary response outcomes were tolerance (success) and threshold (failure). We randomly sampled 2000 points with replacement and ran logistic regression iteratively 100 times to create an error distribution for model parameters and performance statistics. We withheld 30% of the data to test model accuracy. The bestglm function from the BESTGLM package (McLeod and Xu 2014), and the Bayesian information criterion were used for model selection. To demonstrate the value of this method compared to fixed zones we chose a hypothetical restoration site within the range of mountain mahogany and modeled seed transfer success of every occurrence point relative to a focal (single) point, representing a potential restoration site.

Comparison among seed zone methods

We compared our species-specific provisional zones and guidelines with other provisional seed transfer zone methods including: generalized provisional seed transfer zones developed by Bower et al. (2014), and Omernik's Level III and Level IV ecoregions (Omernik 1987). We used phenotypic data from a common garden experiment described in Crow (2016). Briefly, common gardens were installed in 2015 in three locations across the range of mountain mahogany in a complete randomized block design with 10 plants from 18 seed sources planted in each of three blocks in every location. Phenotypic data were collected for two growing seasons. We used random-effects models to compare the phenotype variation explained by each seed zone method individually. We used lmer (Bates et al. 2014) in R to extract model variance components. Seed zones, common garden locations, and seed source accession nested within common garden locations were modeled as random effects. We calculated the amount of variation explained by each seed zone method as: $(\sigma^2_Z/[\sigma^2_Z + \sigma^2_{L(P)} + \sigma^2_L]) \times 100$. Where σ^2_Z is the amount of

variation explained by seed zone, σ^2_L represents common garden location, and $\sigma^2_{L(P)}$ is seed source accession nested within common garden location. Phenotypes included in the models were plant survival, final height, leaf area, leaf length, leaf width and leaf perimeter.

Results

Five environmental variables were retained in the best fitting RF model as predictors for species distribution modeling. These included annual heat moisture index (a measure of aridity), the DEM, mean maximum temperature of the warmest month, mean annual precipitation, and the summer dryness index (see Table 1 for details). The RF model n-fold cross-validation had an out-of-bag error rate of 0.246, percentage correctly classified (pcc) of 75.4%, absence class error of 0.170, and a presence class error of 0.391. The resulting SDM represented the probability of occurrence for the distribution of mountain mahogany (Appendix S1: Fig. S2). We used the holdout dataset to select the SDM probability of occurrence threshold and found that 0.6 yielded the highest model accuracy score (true positive + true negative)/(total predicted) (0.754 ± 0.002). The model threshold of 0.6 was therefore used as the cutoff for the presence/absence binary map in further analyses, where any cell in our study area with a probability of occurrence below 0.6 was considered an absence point, and any cell equal or >0.6 was considered a presence point.

The NMDS function had a stress value of (0.157), similar to previous studies (Kreft and Jetz 2010, Hattab et al. 2015, Manfrin et al. 2016), indicating an adequate projection of the dissimilarity matrix. The NMDS plot and map show the continuous transition zones between regions (Fig. 1), without distinct geographic clusters or disjunctions with unoccupied ordination space.

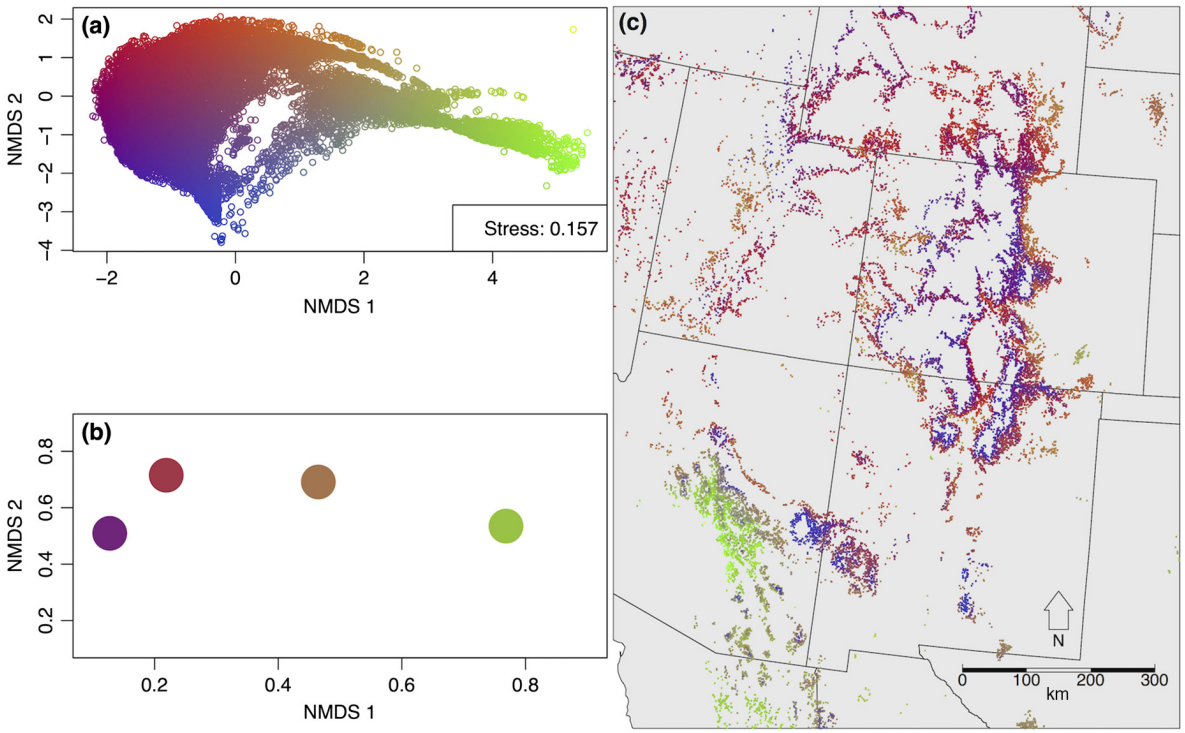


Fig. 1. Biogeographic patterns of environmental variables across the distribution of mountain mahogany using non-metric multidimensional scaling (NMDS). (a) Each point in the plot represents the ordination score of a single occurrence point of mountain mahogany. (b) We averaged the ordination scores for each of the four clusters and averaged their color scheme for visualization and (c) color coded each geographic point by their NMDS score.

We found differences in model performance among clustering algorithms. The unweighted pair-group method using arithmetic averages performed the best, while the single linkage method performed the worst. The two algorithms had cophenetic scores of 0.618 and 0.108, respectively, representing a similar range to previous studies (Table 2; Kreft and Jetz 2010, Hattab et al. 2015). The KGS penalty function identified four clusters as the optimal number of groups (Appendix S1: Fig. S3). Clusters were grouped without regard for spatial distance (Fig. 2), with each cluster covering nearly the entire distribution of the species. The most geographically restricted cluster is the dominant southwestern cluster, likely due to the unique environmental conditions relative to the rest of mountain mahogany's distribution (shown as black in Fig. 2).

Multiple logistic regression was used to model environmental tolerance and thresholds to create a probability function for seed transfer (Fig. 3). The model is summarized in Table 3. After model selection, all variables from previous analyses were used except the mean maximum temperature of the warmest months. The model had an external Area under the receiver operator curve (AUC) of 0.716 $\pm$ 0.00133, a similar internal AUC and a sensitivity score of 0.78 $\pm$ 0.00145. Model accuracy statistics were calculated for logistic regression models using a testing dataset where 30% of the data were withheld to test model performance, and compared with internal statistics to evaluate model overfitting. No significant difference was found between external and internal model validation, and we determined that model overfitting was not an issue. With all predictors held constant in our model, a one unit increase in annual heat moisture index away from a focal point decreases the odds of seed transfer success by a factor of 3.88 (odds ratio = 0.258, 95% CI = 0.206–0.321), while a one unit increase in summer moisture index increased the chances of seed transfer success by approximately 17% (odds ratio = 1.174, 95% CI = 1.127–1.224). The differences in odds ratio for our variables may be in part due to the discrepancy in the range of environmental variable units (e.g., meters vs. degree Celsius), as well as their overall effect on the dependent variable. For this reason, the standardized coefficients are included in Table 3.

Species-specific provisional seed transfer guidelines explained the highest phenotypic variation four out of six times (Table 4). Level IV ecoregions explained the most variation in leaf length and perimeter. Among the zonal methods, our species-specific provisional seed zones captured the most variation in leaf width, and the second most variation in survival. Level IV ecoregions explained the most variation in plant

Table 2. Hierarchical clustering linkage methods and associated cophenetic scores.

Clustering method (synonym)	Acronym	Description	Cophenetic correlation
Ward's method (minimum variance)	Ward	Uses an ANOVA approach and minimizes the sum of squares of any two clusters that can be formed at each step	0.215
Single linkage (nearest neighbor)	SL	Distance between two clusters is the minimum distance between all objects of each cluster	0.108
Complete linkage (furthest neighbor)	CL	Distance between two clusters is the maximum distance between all objects of each cluster	0.579
Unweighted pair-group method using arithmetic averages (average linkage)	UPGMA	Distance between two clusters is the average distance between all objects of each cluster	0.618
Weighted pair-group method using arithmetic averages (McQuitty's method)	WPGMA	Distance between two clusters is the arithmetic average distance between objects of each cluster weighted by the number of objects in each cluster	0.508
Weighted pair-group method using centroids (median linkage)	WPGMC	Distance between two clusters is the Euclidean distance between their weighted centroids	0.410
Unweighted pair-group method using centroids (centroid linkage)	UPGMC	Distance between the centroids ("mean point" or center of gravity) of each cluster	0.607
Divisive hierarchical method	DIANA	Starting with all observations forming one single large cluster, groups are repeatedly divided until each cluster contains only a single observation. At each stage, the cluster with the largest diameter in terms of the maximum observed dissimilarity is selected	0.390

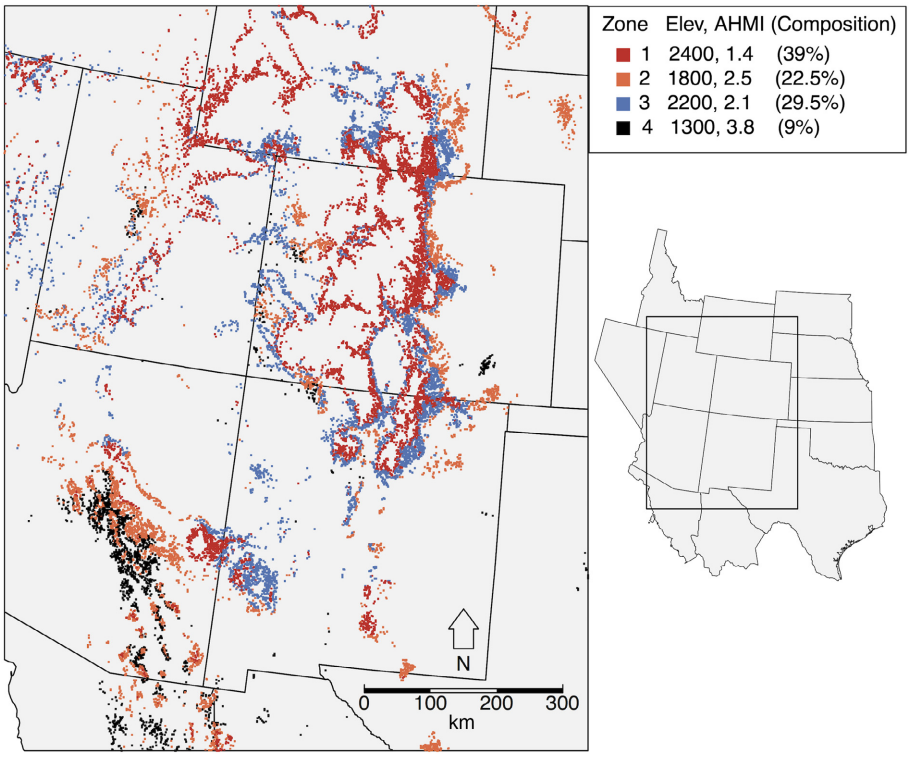


Fig. 2. Four seed transfer zones based on cluster analysis of environmental variables across range of *Cercocarpus montanus*. The average elevation and aridity (higher AHMI values equal higher aridity) of each cluster is included in the legend, as well as the relative percent composition. Points are enlarged for easier visualization.

height and three out of four leaf traits, and generalized provisional seed zones and captured the most variation in survival followed by species-specific provisional seed zones.

Discussion

Translocation of seeds in restoration has inherent risks of population viability (Vander Mijnsbrugge et al. 2010). Restored populations may have lower fitness relative to local populations, and introduction of plant propagules can result in the loss of co-adapted gene complexes, as a result of gene flow between introduced and local plants (Hufford and Mazer 2003, Kawecki and Ebert 2004, Lofflin and Kephart 2005). Despite the potential for revegetation failure, few plant species targeted for restoration have seed transfer guidelines. Our goal was to develop methods for quickly and inexpensively delineating species-specific provisional seed transfer zones and guidelines using a combination of SDM, cluster algorithms, and logistic regression analyses. Our approach is based on the hypothesis that delineating biogeographic regions within a species range can mitigate the risk of transferring seed materials that are maladapted to environmental conditions at the restoration site.

We used the environmental metrics that have the highest variable importance for the SDM in further analyses for two reasons: (1) Using model-selected environmental variables is a reproducible method for selection of predictors for subsequent use in modeling, and (2) variables identified by the SDM represent the best predictors of niche space and may detect the factors that limit a species distribution (Aguirre-Gutiérrez et al. 2013). Our method incorporates environmental variables that are correlated with a species range in order to include relevant species-specific data. Species evolve under contrasting environmental selection pressures. Therefore,

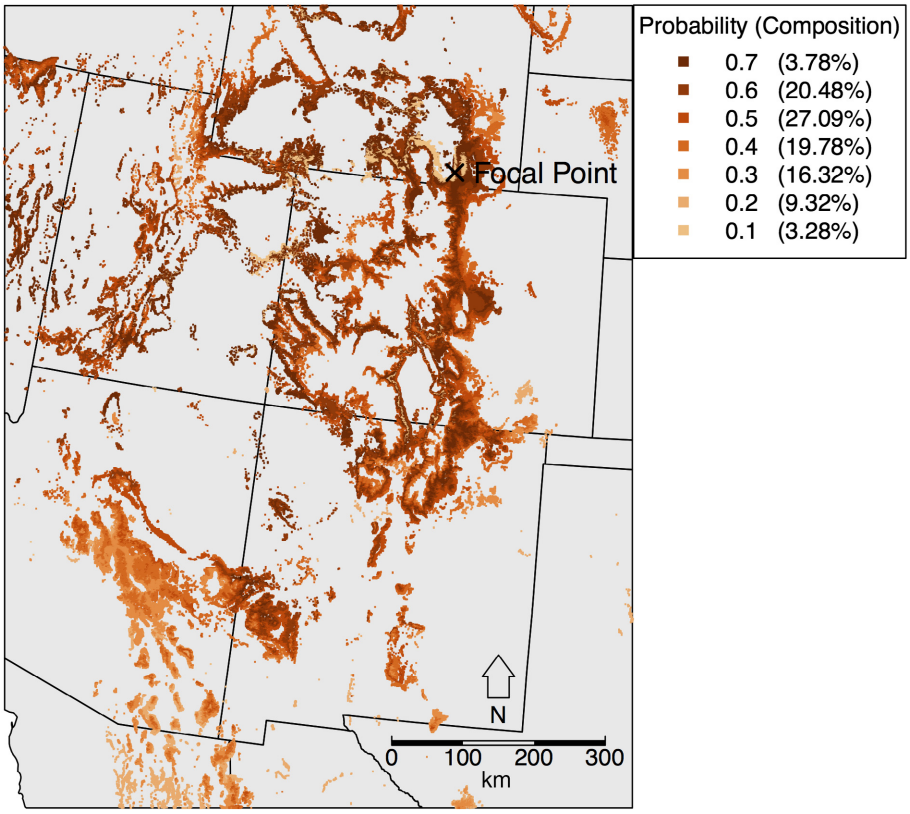


Fig. 3. Logistic regression model results projected onto the study area. The map represents the probability of seed transfer success of every predicted presence point of mountain mahogany relative to a focal point location marked by the X symbol. High probability corresponds to 0.7, decreasing to 0.1. Points are enlarged for easier visualization.

Table 3. Results of logistic regression where species tolerance and threshold values were modeled as a function of environmental and geographic distance (GEO-DIS).

Predictors	Coefficient	Standardized coefficients	*P*-value	Odds ratio
GEO-DIS	<−0.01	−2.06	0.039*	0.9999
AHMI	−1.35	−11.93	<0.01**	0.2584
MAP	<−0.01	−1.89	0.058	0.9987
SMI	0.16	7.65	<0.01**	1.174
DEM	<−0.01	−4.83	<0.01**	0.99

Note: AHMI, Annual heat moisture index; MAP, Mean annual precipitation; SMI, Summer moisture index; DEM, Digital elevation model.
$^{*}P < 0.05$, $^{**}P < 0.01$.

including the same environmental variables for all species (e.g., mean annual precipitation) will likely not represent patterns relevant to the biology of a single species that are critical for growth, survival, and reproduction.

Delineating seed transfer zones using cluster analysis is a novel application of biogeographic regionalization models described in Kreft and Jetz (2010). Cluster analysis, or methods that result in zones, can improve the process of seed source planning and decision-making by delineating broadscale environmental regions. We found that cluster results generally agreed with results from the ordination analysis, and transition zones were similar between the two methods. Hierarchical cluster analysis defined species-specific provisional seed transfer zones based on species distribution data. However, clustering continuous environmental data can be problematic. First, clustering over a large spatial area will group regions into the same cluster without spatial constraints. For example, similar

Table 4. Comparison of provisional seed transfer methods.

Trait	Percent variation explained				
	SS Prov STG	SS Prov SZ	Level III	Level IV	Prov SZ
Number of zones	NA	3	4	8	4
Survival	**13.72**	10.41	0.00	5.68	11.07*
Plant height	**47.70**	22.62	7.43	31.98*	4.75
Leaf area	**12.42**	0.28	1.34	7.74*	0.00
Leaf length	6.94	0.00	8.99	**19.04***	0.00
Leaf width	**15.97**	1.08*	0.00	0.00	0.00
Leaf perimeter	9.22	0.00	7.10	**12.19***	0.61

Notes: The percent variation explained by seed transfer guidelines and zones out of the total variation from random-effects models. Bold text represents the method that explains the most variation for each trait, and an asterisk denotes which of the zonal methods explained the most variation.

SS Prov STG = Species-specific provisional seed transfer guidelines; SS Prov SZ = species-specific provisional seed zones; levels III and IV = Omernik's ecoregions; Prov SZ = Bower et al. (2014) generalized provisional seed zones.

regions for transfer occurred at distances as wide as Wyoming and New Mexico. Geographic distance is an important driver of population genetic differentiation and local adaptation (Walser and Haag 2012). However, imposing geographic restrictions on cluster algorithms is arbitrary without a more thorough understanding of intraspecific gene flow. Second, clustering environmental data with no clear transition zones may not capture fine-scale differences in habitat.

We used logistic regression to create a parametric function that modeled the probability of seed sources crossing environmental thresholds between all potential donor locations and a single focal point representing a restoration site. Focal point seed zones have an advantage because geographic distance can be included more intuitively than in cluster methods. Our logistic regression model of seed transfer success or failure was based in the distribution of *Cercocarpus montanus* and constructed with quantitative estimates of environmental tolerance and thresholds. The model was geographically projected as a probability surface and is a function of the pairwise environmental and geographic distances between the focal point and all potential donor sites. Continuous probability surfaces avoid the arbitrary delineation of categorical seed zones and better reflect the continuous genetic structure found among populations of plant species across environmental gradients. This approach has the potential to assist practitioners in choosing collection areas or commercial seed sources most likely to establish at restoration sites, given local environmental conditions. The probabilities derived from this method reflect the environmental similarities between the restoration site and the plant materials' source location, but do not ensure revegetation success.

The use of environmental data alone to predict species adaptive divergence is an oversimplification of the factors that result in intraspecific variation. For example, Kozak and Wiens (2006) found that sister taxa of salamanders occupied similar niche space, suggesting that evolutionary divergence as a result of local adaptation may occur without corresponding divergence in the occupied niche (Peterson et al. 1999, Peterson and Vieglais 2001). In contrast, species with large distributions (such as mountain mahogany) often have clinal patterns of genetic or phenotypic structure along environmental gradients (Avise et al. 1987). For example, Kalkvik et al. (2012) found that genetic divergence in the *Peromyscus maniculatus* species group was correlated with niche divergence. Another study by Gapare et al. (2005) discovered that Sitka Spruce had a higher frequency of rare alleles at the periphery of its range, and Mimura and Aitken (2010) detected evidence of local adaptation at the range periphery of Sitka Spruce populations relative to core habitat. In both examples, the difference in occupied habitat was correlated with genetic divergence. Natural selection within species can overcome the homogenizing effects of gene flow (Gonzalo-Turpin and Hazard 2009), and is a central assumption driving the development of seed zones using environmental data. Cluster analysis may therefore be more applicable to species with broad geographic distributions along environmental gradients.

Focal point methods use a reference site to construct seed transfer models and have been shown to increase the potential seed collection area, while maintaining the relative risk of translocation (Ukrainetz et al. 2011). Parker (1992) developed focal point seed zones by modeling species- and site-specific adaptive traits from several populations. Seed zones were then

delineated using graphical analyses of principal component scores. Another method for seed transfer guidelines was described by Monserud et al. (1990) where an "expert system" modeled a similarity index between a given region and a specified site. This system incorporated results from Rehfeldt (1990) to map similar environments between donor and recipient populations of *Pinus ponderosa*. McKenney et al. (1999) used a Gower similarity metric to map areas with environmental characteristics of a "regeneration area" and developed the software "seedwhere" for public access. Bower et al. (2014) developed generalized provisional seed zones for plant species in the continental United States by intersecting climate data and overlapping Level III ecoregions (Omernik 1987). A recent and similar approach to ours has been developed by Doherty et al. (2017), where SDMs and cluster analysis were used to identify species-specific "climate centers" that optimize seed collection or plant materials development and use environmental and geographic distance to identify seed sources that are likely to compliment commercially available seed sources. The authors found that they could approximate the environmental distance between restoration site and seed donor sites with fewer geographic clusters than empirically based seed zones. Our approach differs in that we utilize variable selection before SDM and estimate environmental tolerance and thresholds to model continuous provisional seed transfer guidelines.

Zonal and focal point seed zones incorporate environmental data, plant adaptive differentiation data, or both. Provisional seed zones based solely on environmental data assume that natural selection is stronger than of the homogenizing effects of gene flow and that plants will perform better at their site of origin than non-local plants, which is not always the case (Galloway and Fenster 2000). However, species perform significantly better than non-local plants when grown at their site of origin, known as home site advantage, in 71% of studied taxa (Leimu and Fischer 2008). Rapid development of seed transfer guidelines based on genetic marker data uses various multivariate techniques to explain genetic variance with geography or climate or both (Krauss and Koch 2004, Bussell et al. 2006, Hamann et al. 2011). These studies are effective tools for guiding seed transfer and have become more valuable as plant sequence data become available.

Traditional seed transfer zones are delineated using data from common gardens and represent direct tests of adaptive differentiation (Wilson et al. 2008, Johnson et al. 2010, St. Clair et al. 2013). In these experiments, multiple plant populations are grown in a common environment and phenotypic traits are measured to quantify heritable variation among populations relevant for seed transfer guidelines (Johnson et al. 2004, Saenz-Romero and Tapia-Olivares 2008, St. Clair et al. 2013). While common garden experiments are preferable to indirect measures of environmental factors that may contribute to intraspecific divergence, they are limited by the number of populations included in the study and can take decades to ascertain meaningful results. Indirect methods for delineation of seed transfer zones, such as cluster and focal point analyses, are complementary to experimental common gardens. The hypothetical seed zones that result from these models can be tested in the field, and contribute to restoration guidelines for a wide variety of plant species. The use of SDM provides a promising method to improve provisional seed zones by including environmental data relevant for the presence or absence of a target species. Focal point analysis further expands the utility of provisional methods by providing the practitioner with a range of collection area and corresponding probabilities that seeds collected within those regions will result in establishment success. Further testing will be needed to determine the efficacy of provisional seed zone methods, including long-term monitoring of restored vegetation using seed sources designated by provisional models.

Acknowledgments

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 no. 228001. Research was also supported by the Nature Conservancy through the Nebraska Chapter's J.E. Weaver Competitive Grants Program. This paper was improved following comments by two anonymous reviewers.

Literature Cited

Aguirre-Gutiérrez, J., L. G. Carvalheiro, C. Polce, E. E. van Loon, N. Raes, M. Reemer, and J. C. Biesmeijer. 2013. Fit-for-purpose: Species distribution model performance depends on evaluation criteria–Dutch hoverflies as a case study. PLoS ONE 8:e63708.

Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb, and N. C. Saunders. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annual Review of Ecology and Systematics 18:489–522.

Bates, D., M. Maechler, B. Bolker, and S. Walker. 2014. lme4: linear mixed-effects models using eigen and s4. R package version 1:1–23.

Becker, R. A., J. M. Chambers, and A. R. Wilks. 1988. The new S language: a programming environment for data analysis and graphics. Wadsworth and Brooks/Cole Advanced Books & Software, Monterey, California, USA.

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

Breiman, L. 2001. Random forests. Machine Learning 45:5–32.

Brotherson, J. D. 1992. Mineral-nutrient concentrations in the mountain mahogany species *Cercocarpus montanus* and *Cercocarpus intricatus* and in their associated soils. Journal of Plant Nutrition 15: 49–67.

Bussell, J. D., P. Hood, E. A. Alacs, K. W. Dixon, R. J. Hobbs, and S. L. Krauss. 2006. Rapid genetic delineation of local provenance seed-collection zones for effective rehabilitation of an urban bushland remnant. Austral Ecology 31:164–175.

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.

Chevin, L.-M., R. Lande, and G. M. Mace. 2010. Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. PLoS Biology 8:e1000357.

Crow, T. M. 2016. Adaptive variation of *Cercocarpus montanus*. Dissertation. University of Wyoming, Laramie, Wyoming, USA.

Cutler, D. R., T. C. Edwards Jr., K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007. Random forests for classification in ecology. Ecology 88:2783–2792.

De'ath, G., and K. E. Fabricius. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81:3178–3192.

Doherty, K. D., B. J. Butterfield, and T. E. Wood. 2017. Matching seed to site by climate similarity: techniques to prioritize plant materials development and use in restoration. Ecological Applications 27:1010–1023.

Elith, J., and J. Leathwick. 2007. Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. Diversity and Distributions 13:265–275.

Endler, J. A. 1973. Gene flow and population differentiation studies of clines suggest that differentiation along environmental gradients may be independent of gene flow. Science 179:243–250.

Escalante, T., G. Rodríguez-Tapia, M. Linaje, P. Illoldi-Rangel, and R. González-López. 2013. Identification of areas of endemism from species distribution models: threshold selection and Nearctic mammals. TIP 16:5–17.

Etterson, J. R., and R. G. Shaw. 2001. Constraint to adaptive evolution in response to global warming. Science 294:151–154.

Evans, J. S., and M. A. Murphy. 2014. rfutilities. R package version 1.0-0. http://CRAN.R-project.org/package=rfUtilities

Fielding, A. H., and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24:38–49.

Galloway, L. F., and C. B. Fenster. 2000. Population differentiation in an annual legume: local adaptation. Evolution 54:1173–1181.

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.

Gavrilets, S., and J. B. Losos. 2009. Adaptive radiation: contrasting theory with data. Science 323:732–737.

Gonzalo-Turpin, H., and L. Hazard. 2009. Local adaptation occurs along altitudinal gradient despite the existence of gene flow in the alpine plant species *Festuca eskia*. Journal of Ecology 97:742–751.

Greenwood, L. R., and J. D. Brotherson. 1978. Ecological relationships between pinyon juniper and true mountain mahogany stands in the Uintah Basin, Utah. Journal of Range Management Archives 31:164–167.

Hamann, A., T. Gylander, and P. Chen. 2011. Developing seed zones and transfer guidelines with multivariate regression trees. Tree Genetics & Genomes 7:399–408.

Hamann, A., and T. Wang. 2006. Potential effects of climate change on ecosystem and tree species distribution in British Columbia. Ecology 87:2773–2786.

Hattab, T., C. Albouy, F. Ben Rais Lasram, L. Loc'h, F. Guilhaumon, and F. Leprieur. 2015. A biogeographical regionalization of coastal Mediterranean fishes. Journal of Biogeography 42:1336–1348.

Hereford, J. 2009. A quantitative survey of local adaptation and fitness trade-offs. American Naturalist 173:579–588.

Hernandez, P. A., C. H. Graham, L. L. Master, and D. L. Albert. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29:773–785.

Hill, J. K., H. M. Griffiths, and C. D. Thomas. 2011. Climate change and evolutionary adaptations at species' range margins. Annual Review of Entomology 56:143–159.

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.

Johnson, R., V. J. Erickson, N. L. Mandel, J. B. St Clair, and K. W. Vance-Borland. 2010. Mapping genetic variation and seed zones for *Bromus carinatus* in the Blue Mountains of eastern Oregon, USA. Botany-Botanique 88:725–736.

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.

Kalkvik, H. M., I. J. Stout, T. J. Doonan, and C. L. Parkinson. 2012. Investigating niche and lineage diversification in widely distributed taxa: phylogeography and ecological niche modeling of the *Peromyscus maniculatus* species group. Ecography 35:54–64.

Kawecki, T. J., and D. Ebert. 2004. Conceptual issues in local adaptation. Ecology Letters 7:1225–1241.

Keller, M., J. Kollmann, and P. Edwards. 2000. Genetic introgression from distant provenances reduces fitness in local weed populations. Journal of Applied Ecology 37:647–659.

Kelley, L. A., S. P. Gardner, and M. J. Sutcliffe. 1996. An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. Protein Engineering 9:1063.

Kozak, K. H., and J. J. Wiens. 2006. Does niche conservatism promote speciation? A case study in North American salamanders. Evolution 60:2604–2621.

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.

Krauss, S. L., and J. M. Koch. 2004. Methodological insights: rapid genetic delineation of provenance for plant community restoration. Journal of Applied Ecology 41:1162–1173.

Kreft, H., and W. Jetz. 2010. A framework for delineating biogeographical regions based on species distributions. Journal of Biogeography 37:2029–2053.

Legendre, P., and L. F. Legendre. 2012. Numerical ecology. Volume 24. Elsevier, Amsterdam, The Netherlands.

Leimu, R., and M. Fischer. 2008. A meta-analysis of local adaptation in plants. PLoS ONE 3:1–8.

Liancourt, P., L. A. Spence, D. S. Song, A. Lkhagva, A. Sharkhuu, B. Boldgiv, B. R. Helliker, P. S. Petraitis, and B. B. Casper. 2013. Plant response to climate change varies with topography, interactions with neighbors, and ecotype. Ecology 94:444–453.

Liaw, A., and M. Wiener. 2002. Classification and regression by randomForest. R News 2:18–22.

Lofflin, D. L., and S. R. Kephart. 2005. Outbreeding, seedling establishment, and maladaptation in natural and reintroduced populations of rare and common *Silene douglasii* (Caryophyllaceae). American Journal of Botany 92:1691–1700.

Ludwig, J. A., and J. F. Reynolds. 1988. Statistical ecology: a primer in methods and computing. Volume 1. John Wiley & Sons, New York, New York, USA.

Manfrin, A., L. Traversetti, F. Pilotto, S. Larsen, and M. Scalici. 2016. Effect of spatial scale on macroinvertebrate assemblages along a Mediterranean river. Hydrobiologia 765:185–196.

McKenney, D. W., B. G. Mackey, and D. Joyce. 1999. Seedwhere: a computer tool to support seed transfer and ecological restoration decisions. Environmental Modelling & Software 14:589–595.

McLaughlin, S. P. 1992. Are floristic areas hierarchically arranged? Journal of Biogeography 19:21–32.

McLeod, A., and C. Xu. 2014. bestglm: best subset glm. R package version 0.34. http://CRAN.R-project.org/package=bestglm

Miller, S. A., A. Bartow, M. Gisler, K. Ward, A. S. Young, and T. N. Kaye. 2011. Can an ecoregion serve as a seed transfer zone? Evidence from a common garden study with five native species. Restoration Ecology 19:268–276.

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

Monserud, R. 1990. An expert system for determining compatible seed-transfer locations. AI Applications in Natural Resource Management 4:44–50.

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.

Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner. 2015. vegan: community ecology package. R package version 2.3-0. http://CRAN.R-project.org/package=vegan

Omernik, J. M. 1987. Ecoregions of the conterminous United States. Annals of the Association of American geographers 77:118–125.

Parida, M., A. Hoffmann, and M. Hill. 2015. Climate change expected to drive habitat loss for two key herbivore species in an alpine environment. Journal of Biogeography 42:1210–1221.

Parker, W. H. 1992. Focal point seed zones: site-specific seed zone delineation using geographic information systems. Canadian Journal of Forest Research 22:267–271.

Pearce, J., and D. Lindenmayer. 1998. Bioclimatic analysis to enhance reintroduction biology of the endangered helmeted honeyeater (*Lichenostomus melanops cassidix*) in southeastern Australia. Restoration Ecology 6:238–243.

Pearson, R. G. 2010. Species distribution modeling for conservation educators and practitioners. Lessons in Conservation 3:54–89.

Peterson, A., J. Soberón, and V. Sánchez-Cordero. 1999. Conservatism of ecological niches in evolutionary time. Science 285:1265–1267.

Peterson, A. T., and D. A. Vieglais. 2001. Predicting species invasions using ecological niche modeling: New approaches from bioinformatics attack a pressing problem. BioScience 51:363–371.

Phillips, S. J., and M. Dudík. 2008. Modeling of species distributions with maxent: new extensions and a comprehensive evaluation. Ecography 31:161–175.

R Core Team. 2015. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rehfeldt, G. 1990. Genetic differentiation among populations of *Pinus ponderosa* from the upper Colorado River Basin. Botanical Gazette 151:125–137.

Rehfeldt, G. E. 1995. Genetic variation, climate models and the ecological genetics of *Larix occidentalis*. Forest Ecology and Management 78:21–37.

Riordan, E. C., and P. W. Rundel. 2009. Modelling the distribution of a threatened habitat: the California sage scrub. Journal of Biogeography 36:2176–2188.

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.

Sneath, P. H., and R. R. Sokal. 1973. Numerical taxonomy. The principles and practices of numerical classification. Volume 573. WF Freeman and Co., San Francisco, California, USA.

Sokal, R. R., and F. J. Rohlf. 1962. The comparison of dendrograms by objective methods. Taxon 11: 33–40.

St. Clair, J. B., F. F. Kilkenny, R. C. Johnson, N. L. Shaw, and G. Weaver. 2013. Genetic variation in adaptive traits and seed transfer zones for *Pseudoroegneria spicata* (bluebunch wheatgrass) in the northwestern United States. Evolutionary Applications 6:933.

Swaegers, J., J. Mergeay, A. Van Geystelen, L. Therry, M. Larmuseau, and R. Stoks. 2015. Neutral and adaptive genomic signatures of rapid poleward range expansion. Molecular Ecology 24:6163–6176.

Turesson, G. 1922. The species and the variety as ecological units. Hereditas 3:100–113.

Ukrainetz, N. K., G. A. O'Neill, and B. Jaquish. 2011. Comparison of fixed and focal point seed transfer systems for reforestation and assisted migration: a case study for interior spruce in British Columbia. Canadian Journal of Forest Research 41:1452–1464.

Vander Mijnsbrugge, K., A. Bischoff, and B. Smith. 2010. A question of origin: where and how to collect seed for ecological restoration. Basic and Applied Ecology 11:300–311.

Walser, B., and C. R. Haag. 2012. Strong intraspecific variation in genetic diversity and genetic differentiation in *Daphnia magna*: the effects of population turnover and population size. Molecular Ecology 21:851–861.

White, D., and R. B. Gramacy. 2012. maptree: mapping, pruning, and graphing tree models. R package version 1.4-7. http://CRAN.R-project.org/package=maptree

Wilson, B. L., D. C. Darris, R. Fiegener, R. Johnson, M. E. Horning, and K. Kuykendall. 2008. Seed transfer zones for a native grass (*Festuca roemeri*) genecological evidence. Native Plants Journal 9:287–303.

Ying, C. C., and A. D. Yanchuk. 2006. The development of British Columbia's tree seed transfer guidelines: purpose, concept, methodology, and implementation. Forest Ecology and Management 227:1–13.

Supporting Information

Additional Supporting Information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/ecs2.2059/full