*Ecological Applications*, 27(3), 2017, pp. 1010–1023

# Matching seed to site by climate similarity: Techniques to prioritize plant materials development and use in restoration

Kyle D. Doherty,[1,3] Bradley J. Butterfield,[2] and Troy E. Wood[2]

[1]*U.S. Geological Survey, Southwest Biological Science Center, 2255 North Gemini Drive, Flagstaff, Arizona 86001 USA*
[2]*Department of Biological Sciences and Merriam-Powell Center for Environmental Research,*
*Northern Arizona University, Flagstaff, Arizona 86011 USA*

*Abstract*. Land management agencies are increasing the use of native plant materials for vegetation treatments to restore ecosystem function and maintain natural ecological integrity. This shift toward the use of natives has highlighted a need to increase the diversity of materials available. A key problem is agreeing on how many, and which, new accessions should be developed. Here we describe new methods that address this problem. Our methods use climate data to calculate a climate similarity index between two points in a defined extent. This index can be used to predict relative performance of available accessions at a target site. In addition, the index can be used in combination with standard cluster analysis algorithms to quantify and maximize climate coverage (mean climate similarity), given a modeled range extent and a specified number of accessions. We demonstrate the utility of this latter feature by applying it to the extents of 11 western North American species with proven or potential use in restoration. First, a species-specific seed transfer map can be readily generated for a species by predicting performance for accessions currently available; this map can be readily updated to accommodate new accessions. Next, the increase in climate coverage achieved by adding successive accessions can be explored, yielding information that managers can use to balance ecological and economic considerations in determining how many accessions to develop. This approach identifies sampling sites, referred to as climate centers, which contribute unique, complementary, climate coverage to accessions on hand, thus providing explicit sampling guidance for both germplasm preservation and research. We examine how these and other features of our approach add to existing methods used to guide plant materials development and use. Finally, we discuss how these new methods provide a framework that could be used to coordinate native plant materials development, evaluation, and use across agencies, regions, and research groups.

*Key words: climate analogues; climate gradients; climate similarity; cluster analysis; conservation; local adaptation; native plants; restoration strategies; seed transfer zones.*

## INTRODUCTION

Increases in the severity and frequency of disturbance on public lands and changes in policy have bolstered demand for locally adapted native plant materials for use in restoration (Westerling et al. 2006, Seager et al. 2007, Mijnsbrugge et al. 2010, Wood et al. 2015). However, locally sourced native plant propagules are often unavailable, and commercial availability is the deciding factor for seed sourcing (Peppin et al. 2010). Consequently, a small number of seed sources, including many that have undergone intensive selection during development, are used across very broad and heterogeneous geographic areas, a reality that conflicts with research-derived guidance that recommends more conservative transfer of materials (Johnson et al. 2012, Bradley St Clair et al. 2013, Bower et al. 2014, Johnson et al. 2015). To address this, coordinated efforts are underway by agencies such as the Plant Conservation Alliance in the United States and elsewhere, to increase the availability of more diverse and/or locally adapted native plant materials for use in restoration and reclamation (National Seed Strategy; https://www.blm.gov/style/medialib/blm/wo/Planning_and_Renewable_Resources/fish__wildlife_and/plants/seedstrategy.Par.66250.File.dat/SeedStrategy081215.pdf). These efforts represent an important advance toward more effective collaboration across governmental agencies, their partners, and their regional programs and are aimed at producing a cost-effective, sustainable, and diverse supply of native plant materials that can be efficiently used to restore resilient native plant communities.

Current guidelines for the use of native plant materials recommend limiting movement within seed transfer zones, where source and site are predicted to be reasonably matched. Development of seed transfer guidelines has been motivated by decades of research that demonstrate that plant populations are often best adapted to their local habitat conditions (Langlet 1971, Linhart and Grant 1996, Leimu and Fischer 2008, Hereford 2009). Local adaptation may be driven by

Manuscript received 12 January 2016; revised 11 November 2016; accepted 20 December 2016. Corresponding Editor: Elizabeth A. Leger.

[3]Present address: School of Forestry, 200 East Pine Knoll Drive, P.O. Box 15018, Flagstaff, Arizona 86011 USA. e-mail: kyledaviddoherty@gmail.com

abiotic environmental conditions such as soil properties and climate, and/or biotic interactions such as coevolution with pollinators or fungal symbiotes (Joshi et al. 2001, Elzinga et al. 2007, Macel et al. 2007, Rutter and Fenster 2007, Johnson et al. 2010, Turner et al. 2010). The relative importance of these factors varies by species and functional group (Joshi et al. 2001, Macel et al. 2007). Currently, predictions of biotic and abiotic factors are limiting, and only global climate models are of sufficient extent and accuracy to study local adaptation at larger scales (Hijmans et al. 2005). Thus, species-specific, "empirical" seed transfer zones are derived from trait-by-climate relations, where traits are measured in a common garden (e.g., Johnson et al. 2012, Bradley St Clair et al. 2013). "Provisional" transfer zones, which can be applied to species lacking empirical analysis, are based on, for example, the partitioning of bivariate climate space (Bower et al. 2014). Because of the high cost of common garden studies, empirical guidelines are available for relatively few species. Broader assessments of potentially adaptive climate gradients, including seasonal variables, would likely improve outcomes from provisional seed transfer guidance (Butterfield 2015). Additionally, topographically complex regions, such as the Colorado Plateau of the western United States, exhibit wide-ranging environmental gradients, and it is unlikely that fine-scale environmental partitioning could be matched with corresponding numbers of plant materials for a sufficient number of species due to economic and logistic constraints (Hijmans et al. 2005, Bower et al. 2014). Because the diversity of available plant materials remains low, it would be advantageous to adopt a strategy that can assess the relative utility of available materials, and identify areas where plant material development would have the most utility. We present a method for generating provisional, species-specific seed transfer guidelines that can incorporate diverse drivers of plant adaptation that may predict the relative performance of accessions, absent empirical data, and may be used to balance ecologic and economic concerns to inform all phases of native plant materials development.

This method is an expansion of recently developed tools that assess the site-specific utility of agricultural plant materials (Ramirez-Villegas et al. 2011). These authors described an approach for identifying climate analogues by calculating Euclidean distance across multivariate climate space. This analytic tool was formulated and applied in an agricultural context where crop performance was predicted to be similar at climate analogues. The underlying assumption of this approach reflects the dependency of crop performance on climate (Lobell and Field 2007) and a large body of empirical studies of natural plant populations that demonstrate adaptive divergence along climatic axes (e.g., Manel et al. 2010). More specifically, relativized Euclidean climate distance has been used to study local adaptation to climate in forage plants (Joshi et al. 2001). Here, we adapt these frameworks to native plant materials development.

The key parameter of our analyses is a climate similarity index, which is used to predict performance of source-identified accessions at a restoration site. We demonstrate how this index can be calculated for an accession throughout a geographic region of interest and compared against indices for other accession, thus enabling mapping the closest climate match of accessions on hand. We demonstrate the use of this tool in guiding sampling efforts for plant performance evaluations by first generating a species distribution model (SDM) that defines the relevant geographic and climatic range. We then identify potential accessions that capture unique and representative regions of a species' climate space for a target number of accessions, and a means to quantify climate coverage of those accessions. Consequently, our method provides insight into the marginal benefits of developing additional accessions for a given species. We then evaluate our approach by comparing it with empirically derived seed transfer guidelines, assessing the ability of both approaches to select seed sources that minimize movement across biologically relevant climate axes (Johnson et al. 2012). Finally, we use climate similarity analyses to identify complementary seed sources to commercially available accessions that approximate maximum climate coverage. Importantly, our framework for informing plant materials collection, development, and use is directed at range-wide evaluation, and thus provides a platform for integrating efforts across land management agencies and eco-regional programs. Our work is timely, coming on the heels of the Plant Conservation Alliance's National Seed Strategy, which calls for significant expansion of efforts to develop genetically diverse native plant materials for use on public lands in the United States. The methods we describe here, which only require freely available data sets, could greatly improve the cost-effectiveness and efficiency of such efforts.

## METHODS

### *Analytical software and data*

All analyses were conducted in R version 3.2 (R Core Team 2015). Scripts, data, and documentation to reproduce our analyses are available at the Dryad Digital Repository: https://doi.org/10.5061/dryad.43bv0 (Doherty et al. 2017).

### *Selection of variables and calculation of climate similarity*

Although our method can incorporate any continuous environmental variable, for the current study, we chose seven from the Worldclim bioclimatic data set: mean annual temperature, mean diurnal range, temperature seasonality, mean temperature wettest quarter, mean annual precipitation, precipitation seasonality, and precipitation

---

[4] http://www.worldclim.org

of warmest quarter (Hijmans et al. 2005; data *available online*).[4] We used the Worldclim data set because its extent is global, allowing us to model the full range of a species. The above were chosen after conducting a principal components analysis with varimax rotation on the full set of bioclimatic variables, where seven principal components were selected based on an eigenvalue cutoff greater than one. A representative variable (having a correlation greater than 0.7) was selected from each orthogonal principal component axis, yielding a set of reasonably independent variables for the study extent of western North America (130°–90° W longitude and 15°–60° N latitude). A representative variable was used rather than the principal component itself for the sake of interpretability. Selecting uncorrelated variables guards against assigning unequal importance to any one variable, while capturing unique axes of multivariate climate space in the Worldclim bioclimatic data set. Ultimately, variable selection is best determined by careful examination of patterns of covariance within a target extent.

Our analytic method is based on calculation of a "climate similarity" index between two or more map cells (Fig. 1). To produce a climate similarity index, we first calculate Euclidean distance between two points in a defined climate space (bounded by a species distribution model, ecoregion, or other map unit) and then divide by the maximum Euclidean distance in that space. Dividing by maximum Euclidean distance has the effect of converting raw Euclidean distance values to a percentage of maximum climate distance for the extent of interest. We estimate maximum Euclidean climate distance from 10000 samples of 0.1% of cells of the extent for the set of climate variables. We convert this measure of dissimilarity to a measure of similarity by subtracting the percentage of maximum climate distance from one. This conversion relativizes climate distances, allowing for simplified comparison within the study extent. Thus, we may say that the similarity between two points in an extent is 1.0 for equivalent climates, 0.5 for points spanning half the distance of the climate space, and at or near 0.0 for the least similar points. While this index allows for comparison within a study extent, it precludes comparison of values across different extents. This is because different extents will vary in maximum Euclidean distance, and a given climate similarity value will represent a greater breadth in climate space in extents with greater maximum Euclidean distance.

Calculation of climate similarity allows for a comparison of predicted performance between two or more source-identified plant material accessions. If an accession has the greatest climate similarity with a map cell relative to all other alternatives, we predict that it will have the best performance at that cell. Using this climate similarity index, one can generate maps that indicate, in a visually accessible format, which accession of a target species to use for a given restoration site (Fig. 1). For each accession, we calculate the climate similarity index for all map grid cells in a study extent (Fig. 1C), then determine which accession has maximum climate similarity at each cell, and finally we assign each cell to its closest climate match based upon the index (Fig. 1E). We can then calculate the range-wide mean of these maxima to assess climate coverage of a set of accessions, and interpret a higher mean as greater coverage. The results from these analyses will report the closest match even if all accessions under considerations are a poor match, thus, we map the accession to which each cell is assigned as color and climate similarity of that accession to its assignment as opacity to provide a visual and quantitative assessment of assignment and variance from assignment within the defined climate space. By calculating climate similarity indices of multiple seed sources to a restoration site we can also produce a ranked list of climate similarity. If seed increase is an objective, the highest-ranking candidates could be evaluated, and the best agronomic performer could be entered into development. If accessions are seeded directly, a practitioner may choose a few of the highest-ranking candidates to increase the chances of matching other important sources of local adaptation, such as soils.

Achieving high mean climate similarity across an extent that has wide ranging climate gradients may prove more difficult than for those with narrow gradients, assuming a regular distribution in climate space for both extents. However, the distribution of climate space is rarely regular, which has important implications for the realized climate coverage of a set of accessions. In Fig. 1A, the scatterplot depicts a bivariate climate space defined by mean annual temperature and mean annual precipitation, where each of the points in the scatterplots corresponds to a 1-km grid cell in the state of Utah in the United States. We may map the climate similarity for an accession throughout this extent (Fig. 1D), and do so for three hypothetical accessions. In Fig. 1E, the contribution of Accession 1 (purple) to the overall climate coverage is high, because of the abundance of geographic areas possessing similar climate conditions in the state of Utah. In contrast, Accession 2 (green), occupies a region of climate space less abundant in this extent, and thus provides minimal gains in mean climate similarity, or coverage leant by the three accessions. Note that each accession may occupy comparable area in climate space (volume in higher dimensions), but it is the abundance of similar sites in that space that determines climate coverage. The patterns of multivariate distributions become increasingly complex in higher dimensions of climate space, and we present an algorithmic approach to partition ecologically relevant climate space to identify geographic areas within these partitions that maximize mean climate similarity in the following set of analyses.

### *Species distribution modeling to define ranges*

We first defined species-specific extents for analyses by generating species distribution models (SDMs) for six grass and five forb species (Table 1) that are characteristic of the cool deserts of Colorado Plateau and neighboring arid ecoregions of western North America. The species chosen for these analyses were identified as priorities for

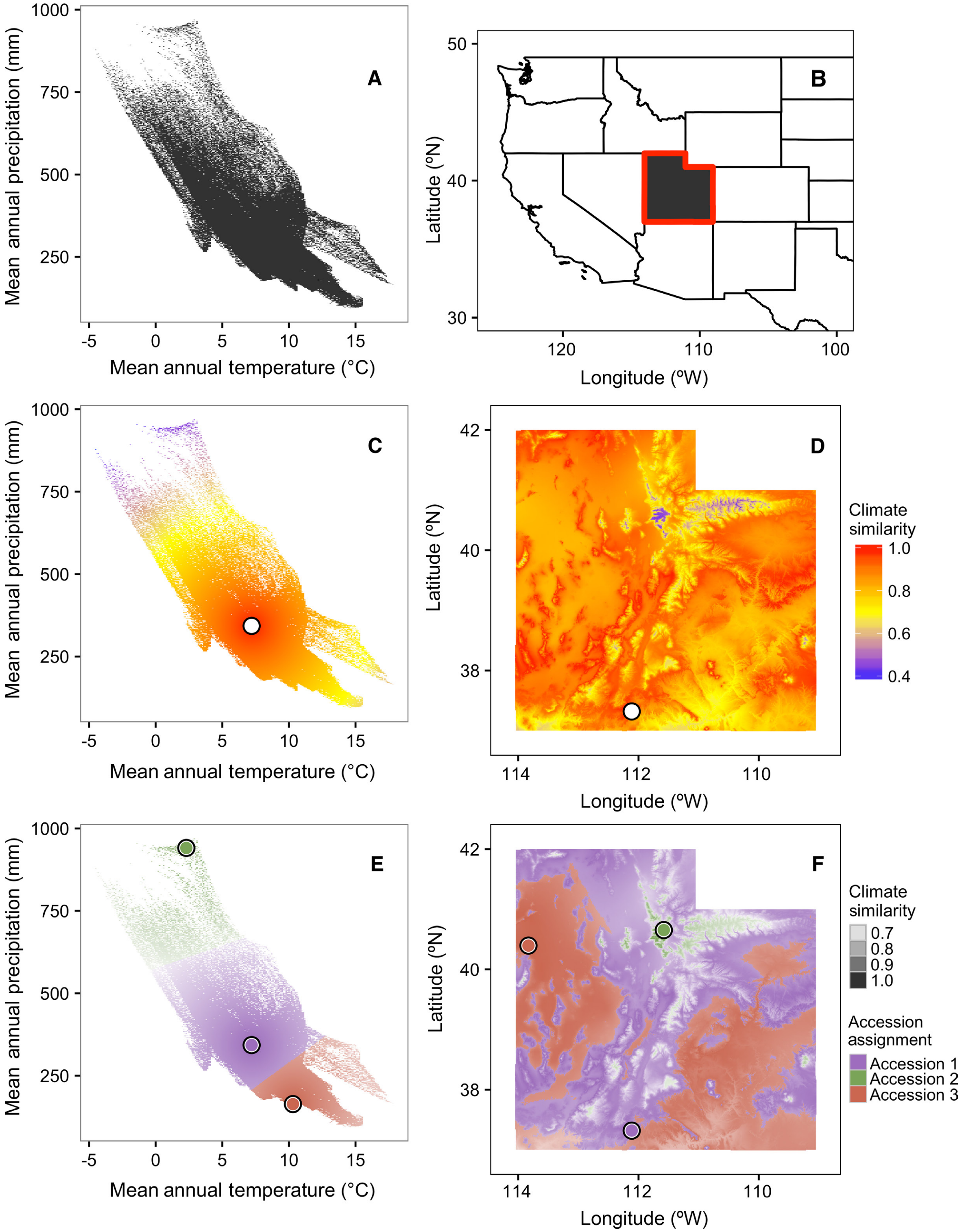

FIG. 1. As a simplified bivariate example, (A) we first extract 1-km resolution mean annual temperature and precipitation data for (B) the state of Utah in the United States. (C, D) From these data, we may calculate and map the climate similarity of an accession (white circle) to all other grid cells in this defined extent. (E, F) We may perform this exercise for multiple accessions (colored circles), and may then assign the accession with highest climate similarity to each grid cell, mapped as color, and visualize variance in (accession × grid cell) similarity as opacity within each color. Note that the breadth of climate space assigned to an accession (E) will not always be positively correlated to its geographic coverage in terms of area (F). [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1. Taxon, model fit, TSS (True Skill Statistic, a measure of model accuracy) cutoff value, and variable importance values for the 11 species distribution models.

| Species | AUC | TSS | Mean annual temp. | Diurnal range | Temp. seasonality | Temp. wettest quarter | Mean annual precip. | Precip. seasonality | Precip. warmest quarter |
|---|---|---|---|---|---|---|---|---|---|
| *Achnatherum hymenoides* | 0.919 | 342 | 0.175 | 0.219 | 0.190 | 0.144 | 0.051 | 0.055 | 0.372 |
| *Astragalus lonchocarpus* | 0.971 | 439 | 0.137 | 0.106 | 0.063 | 0.307 | 0.269 | 0.724 | 0.241 |
| *Bouteloua gracilis* | 0.934 | 186 | 0.258 | 0.288 | 0.471 | 0.160 | 0.075 | 0.081 | 0.139 |
| *Cleome lutea* | 0.96 | 416 | 0.274 | 0.051 | 0.096 | 0.134 | 0.161 | 0.207 | 0.377 |
| *Elymus elymoides* | 0.933 | 451 | 0.479 | 0.392 | 0.335 | 0.119 | 0.059 | 0.077 | 0.186 |
| *Heliomeris multiflora* | 0.964 | 374 | 0.180 | 0.210 | 0.342 | 0.146 | 0.124 | 0.259 | 0.293 |
| *Koeleria macrantha* | 0.938 | 493 | 0.414 | 0.329 | 0.360 | 0.107 | 0.073 | 0.075 | 0.164 |
| *Machaeranthera canescens* | 0.932 | 299 | 0.389 | 0.163 | 0.182 | 0.208 | 0.063 | 0.209 | 0.432 |
| *Pleuraphis jamesii* | 0.943 | 402 | 0.275 | 0.141 | 0.082 | 0.297 | 0.196 | 0.337 | 0.167 |
| *Sphaeralcea parvifolia* | 0.964 | 286 | 0.474 | 0.032 | 0.048 | 0.241 | 0.050 | 0.413 | 0.181 |
| *Sporobolus cryptandrus* | 0.923 | 432 | 0.492 | 0.133 | 0.116 | 0.121 | 0.092 | 0.087 | 0.099 |

*Note:* Temp., temperature; precip., precipitation.

development by the Colorado Plateau Native Plant Program (Wood et al. 2015). Many of these species are also under evaluation by other ecoregional programs and/or have been released as cultivars (cultivated native sources) for use in restoration and reclamation in western North America. Georeferenced occurrences for these eleven species were extracted from two databases of digitized herbaria records: SEINet (Southwest Environmental Science Network 2015) and GBIF (Global Biodiveristy Information Facility 2015; data *available online*).[5, 6]

Climate data corresponding to our seven bioclimatic variables chosen earlier were acquired for each study species' set of occurrences and used as the predictors for species distribution modeling (SDM). SDMs were developed using maximum entropy (Maxent; Phillips et al. 2004) in the biomod2 package (Thuiller et al. 2014), which provides support for numerous modeling techniques in an R environment. Background extent for pseudo-absence selection was defined for each species using a climatic bounding box approach. Model fit was very good in all cases (Table 1, all AUC > 0.9). Presence/absence thresholds delimiting a species climate niche were determined by calculating the True Skill Statistic cutoff, or the probability of occurrence value at which error of commission is equal to error of omission (TSS Cutoff, Table 1). Following thresholding (binary suitable/unsuitable conversion) of SDM geographic projections, we examined agreement with overlays of occurrence records, and excluded large, contiguous regions of model over-prediction, or areas predicted as suitable that had no occurrences present. The resultant raster layers were used to circumscribe species-specific extents for climate similarity analyses. For a detailed treatment of the MaxEnt modeling technique refer to Elith et al. (2011).

### *Sampling of plant materials to maximize range-wide climate similarity*

We extracted bioclimatic data sets corresponding to each of the studied species modeled ranges by overlaying the SDM layers onto the bioclimate layers and removing the climate data from outside of SDMs, using the mask function of the raster package in R v3.2 (Hijmans 2015). The resulting layers were then converted to a data frame using the rasterToPoints function, where each raster cell is stored as a row in a table with coordinates and associated bioclimatic values. The climate values were centered and scaled prior to Euclidean distance calculations to account for differences in disparate units.

We performed hierarchical cluster analysis (stats package) to partition the multivariate climate space of each species. This was an intermediate step to identify a climatically diverse set of geographic areas from which we may select potential accessions yielding high climate coverage. Hierarchical cluster analysis is an algorithmic method used to partition multivariate data sets into *k* number of clusters

---

[5] www.swbiodiversity.org/seinet/
[6] www.gbif.org

(Ward 1963). Cluster analysis has also been used to identify ecoregions quantitatively (Hargrove and Hoffman 2004). There are numerous algorithms that may be used for this exercise, but we found that different algorithms yielded similar results, including the *k*-means algorithm, and we used the Ward method (D2) because it tends to partition multivariate space into units of equivalent volume (Legendre and Legendre 2012). Hierarchical cluster analysis can yield an analytic solution for entire data sets, however, an analytic solution requires a matrix of all pair-wise Euclidean distances as input. Generation of complete matrices for all point contrasts within the SDM ranges of the broadly distributed species evaluated here was computationally intractable. Thus, we sampled 0.1% of cells without replacement, and performed clustering on this sample.

From each sample, we generated 1 through 25 ($k = 1–25$) cluster solutions. We identified the medoid of each cluster within a given solution set (e.g., 14 medoids for a 14 cluster solution), where the medoid is the point with the smallest total pairwise Euclidean distance from all other points in its cluster, and thus possesses the highest mean climate similarity. Hereafter we refer to medoids as "climate centers." We then calculated climate similarity of all map cells in a species' range to all climate centers, for each cluster solution set, then assigned each cell to its closest center, the step determining best climate match (Fig. 2). Mean climate similarity across assignments was compared for each set of climate centers identified in cluster analysis to evaluate gain in climate coverage as the number of climate centers increases. Sampling, clustering, and mean climate similarity calculations were replicated one hundred times for each cluster solution (100 × 1–25). As discussed earlier, climate space varies in density, where regions of high density may be interpreted as large geographic areas sharing similar climate (Fig. 1). Thus, density of assignments within and across solutions will vary, allocating climate centers in some solutions to distinct, but potentially geographically irrelevant, climate space. Because our ultimate goal is to develop seed sources that have high geographic utility, we determined the set of climate centers providing maximum mean climate

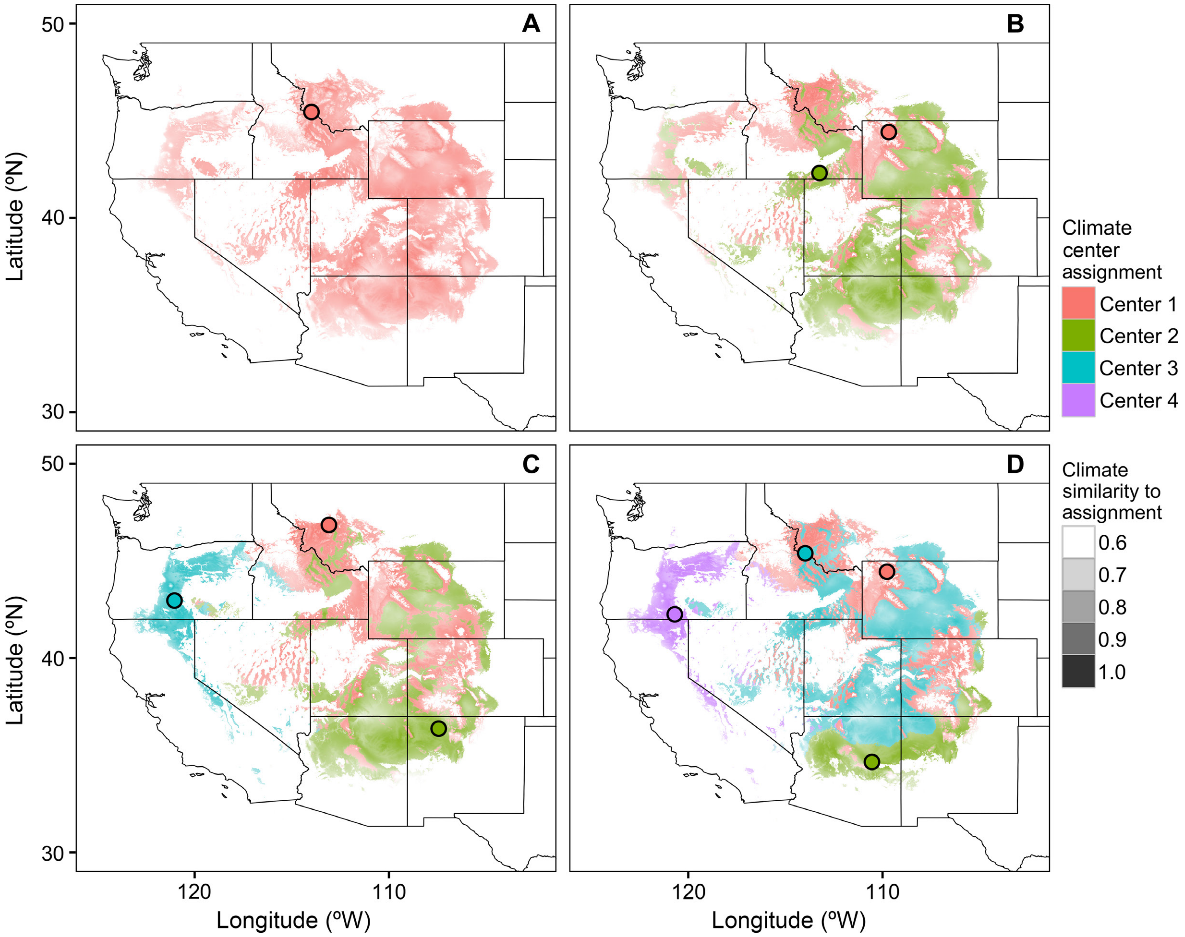

FIG. 2. (A–D) Increasing numbers (one through four) of climate centers are identified for the forb *Machaeranthera canescens* that maximize climate coverage throughout this species' modeled range. Note that climate centers differ across solutions; those in panel B are not equivalent (geographically or climatically) in panels C or D, for example. Thus, the optimal seed collection locations change based on the number of accessions that are desired and/or practical for a given species. [Colour figure can be viewed at wileyonlinelibrary.com]

similarity (maximum climate coverage) across replicates for each set of climate centers.

To contrast our climate similarity approach with a climate naïve approach, we randomly sampled 25 cells of each range-wide data set without replacement, and calculated mean climate similarity for 1 through 25 agglomerative samples, positing each sampled cell as a point of plant materials collection. Climate naïve sampling and mean climate similarity calculations were replicated one hundred times to generate 95% confidence intervals. The climate naïve sampling procedure provided a null model against which to compare the climate informed procedure, and provided an expectation of the difference between a "random search" for accession sampling to one that optimizes climate similarity.

### *Comparison of empirical seed transfers zones and climate similarity analyses*

Here we explicitly evaluate hypothetical accessions developed from climate similarity based sampling compared to those sampled from within an empirical seed transfer zone framework. For this comparison, we used a study of *Achnatherum hymenoides,* a broadly distributed perennial grass species, which presented a twelve-zone seed transfer map spanning eight Omernik level III ecoregions of the western United States (Omernik 1987, Johnson et al. 2012). These empirical seed zones were derived from the modeling of ordinated phenotypic measures in relation to annual and monthly temperature and precipitation data. For each of the 12 zones, we identified the climate center within its corresponding climate space, positing these centers as developed plant materials accessions. We then calculated the absolute differences between these hypothetical accessions and all map cells within their zone for each of the seven climate variables to quantify contrasts between home conditions and conditions at all possible restoration sites within a seed zone.

For our contrast, we extracted climate data for the species-wide extent as defined in Johnson et al. (2012) and identified climate centers for 5 and 12-cluster solutions within that extent. For each of the seven climate variables, we calculated absolute differences between these climate centers and their assignments. We chose a 5-cluster climate partition based on subjective assessment of the saturation point of mean climate similarity curves derived in the previous analysis, and we chose a 12-cluster partition to match the number of recommended seed transfer zones for *A. hymenoides*. For each of the bioclimate variables, we then compared the distribution of differences within empirical zones and climate similarity derived assignments to their respective centers, to assess the potential climate transfers that might occur when developing accessions by each method.

### *Identifying complements to developed seed sources*

Because a substantial number of restoration species already have one or more releases available on the market, we also used our approach to find the best complement to an accession of *Machaeranthera canescens*, recently released by the United States Natural Resources Conservation Service under the name Amethyst (USDA Amethyst Release Brochure 2014). We first calculated climate similarity of this developed accession across its species-wide extent. We then binned and clustered cells below a climate similarity cutoff value derived from the Amtheyst calculations, increasing the cutoff from 0.7 to 1.0 at 0.01 increments. We repeated this process for 100 iterations, calculating range-wide mean climate similarity at each cutoff point when including both the climate centers (hypothetical seed sources) and Amethyst, to identify the complement of accessions (climate centers) that maximizes climate coverage.

## RESULTS

### *Sampling of plant materials to maximize range-wide climate similarity*

Across all species, which represent both warm- and cool-season grasses and five families of forbs, we observed a rapid initial increase in maximum mean climate similarity values with the addition of climate centers identified by cluster analysis (Fig. 3). This initially high rate of return on climate coverage was followed by decreasing gains in mean similarity as the number of climate centers increased. Most species approached an asymptote in climate similarity gain prior to a 25-cluster solution. While not assessed formally, some species, e.g., *Pleurahpis jamesii* and *Astragalus lonchocarpus*, exhibited a distinct elbow, or pronounced decrease in the slope (climate similarity gain) of the curves (at three, two, and two climate centers, respectively). However, other species, e.g., *Elymus elymoides*, *Koeleria macrantha*, and *Cleome lutea*, had smoother gain responses to increasing climate center number. Each species displayed a unique curve profile, differing in initial value, rate of change in diminishing returns, and saturation point. The 95% confidence interval curves of random (climate-naïve) samples had both lower initial and asymptotic mean climate similarity values, and the 95% confidence intervals of the random samples and climate informed solutions did not overlap for any of the species (Fig. 3).

### *Comparison of empirical seed transfers zones and climate similarity analyses*

We found that using a climate similarity method as guidance for plant materials sampling for development could approximate the result of guidance from an empirical common garden study where derived seed zones were used to minimize transport of accessions along individual, trait-linked, climate axes. The distribution of absolute differences in climate variables within the 12 empirical zones was comparable to that within the 5 model-based cluster solution (Fig. 4). The model-based 12-cluster solution performed better than the empirical

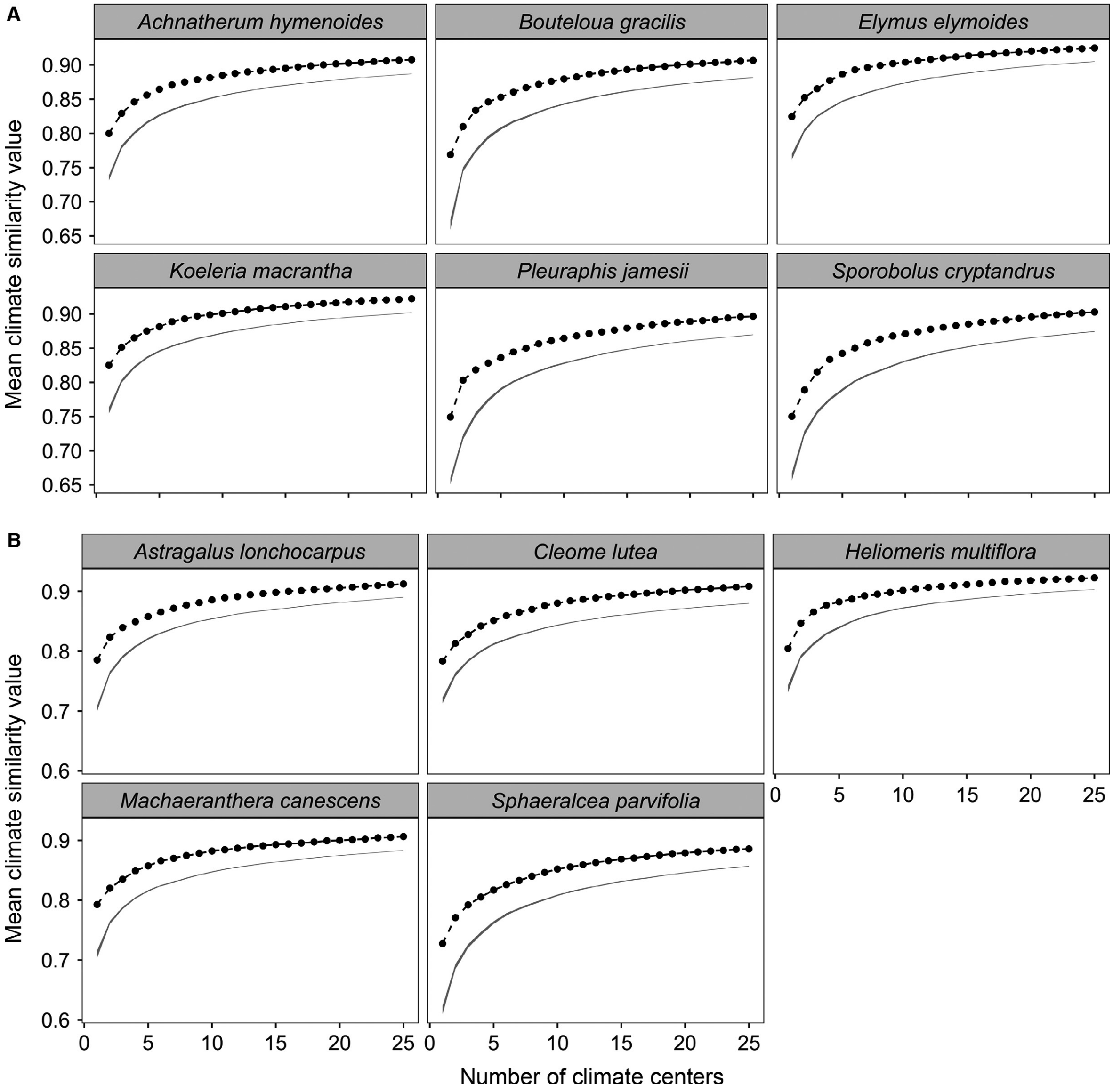

FIG. 3. Curves depicting the maximum mean climate similarity values derived from increasing numbers of climate centers (dashed-dotted lines) and the 95% confidence interval for increasing numbers of climate naïve random samples (gray ribbons, below) for (A) six grass species and (B) five forb species.

solution for all climate variables. This analysis also demonstrates the effect of increasing mean climate similarity on the potential transport along individual climate axes, as observed in the distribution differences between 5 climate centers and 12 climate centers. These results suggest that if one were to pick a single accession to develop from within each of the twelve empirically derived zones for *Achnatherum hymenoides*, these accessions would not be the optimal choice for minimizing climate transfer from source to all possible target sites, and comparable climate coverage could be achieved by developing five accessions drawn from sites at or near the five climate centers derived from climate similarity analyses.

### *Climate complementation of existing seed lines*

Our results from the complementation analysis illustrate how a suitable complement to an existing accession, or accessions, can be determined from climate similarity analyses (Fig. 5). We suggest four climate centers as potential accessions to develop as complements to the one existing release for *Machaeranthera canescens* (Fig. 6). Our proposed four accessions, together with the release, Amethyst, approach the maximum climate coverage provided by a five-climate-center solution (Fig. 5). The number of accessions within a complement set can be increased or decreased to reflect resources available and desired coverage.

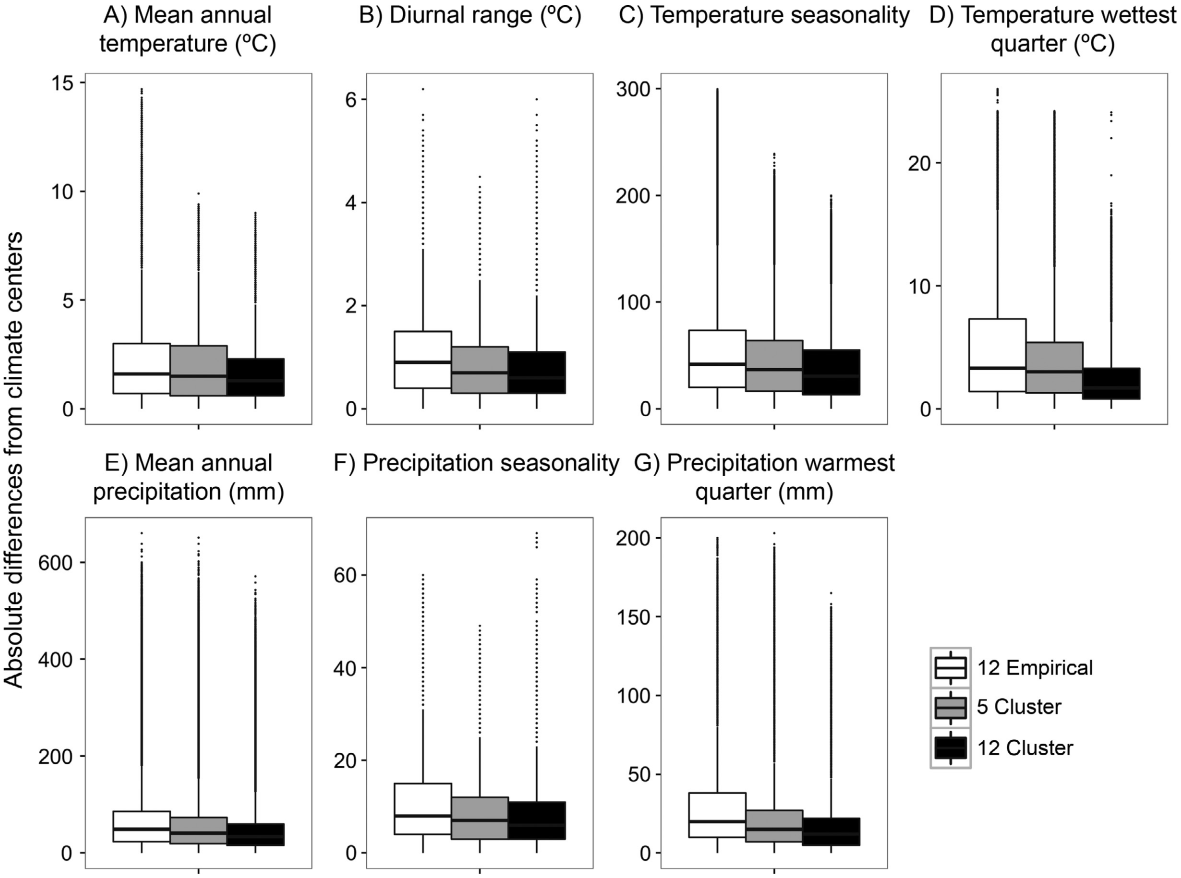

FIG. 4. We present hypothetical climate transfers that might occur when accessions are sourced from empirical seed transfer zones or from provisional climate similarity analyses for the perennial grass species, *Achnatherum hymenoides* (Johnson et al. 2012). We measure absolute differences between representative grid cells (multivariate medians) of each empirical seed transfer zone to all other grid cells in parent zones (white boxes), and the absolute differences between cluster analysis derived climate centers and their assignments for a 5- (gray) and 12- (black) cluster solution. Midlines indicate medians, box boundaries indicate the first (upper) and third (lower) quartiles, whiskers represent 1.5 times the inter-quartile range, and dots represent outliers. Temperature and precipitation seasonality are the standard deviation multiplied by 100 and the coefficient of variation, respectively (both unit-less conventions of the Worldclim bioclimate data set).

## DISCUSSION

It is widely recognized that plant materials may not be locally adapted to the restoration sites at which they are used, i.e., there are genotypes of a given species that are expected to perform better at a given site but are not available for use. This mismatch reflects an obvious constraint: Plants exhibit notoriously fine-scale adaptation in nature, but we cannot produce locally adapted plant materials at this same scale. The methods we present here allow for balancing the competing demands of economic constraints and a desire to restore natural adaptive structure. Our climate similarity index represents a prediction of accession performance that contrasts home climatic conditions with the conditions at all potential sites in a defined extent, such as an ecoregion, species range, or other geographic unit. This approach can rank the predicted relative performance of two or more accessions and identify optimal locations for additional collections if those available do not sufficiently represent the climatic conditions of the area to be restored. Thus, climate similarity analyses represent an explicit hypothesis and testing framework for local adaptation to climatic conditions. These tools may be used provisionally to select among plant materials, absent empirical data, with the assumptions that climate is an important driver of local adaptation, that the climate variables considered have equal importance to plant fitness, and that fitness responses to movement along climate gradients are linear. Later in the discussion we outline how field trial results can be used to test these assumptions and to calculate an empirical similarity index that reflects the relative importance of climate variables. An important feature of our approach is its ability to contingently rank available accessions. Such ranking addresses an important logistic reality of restoration: Materials from within a targeted seed transfer zone are often not available (Peppin et al. 2010). The calculations involved in ranking the climate match of available seed sources are computationally inexpensive, and an accessible mapping and

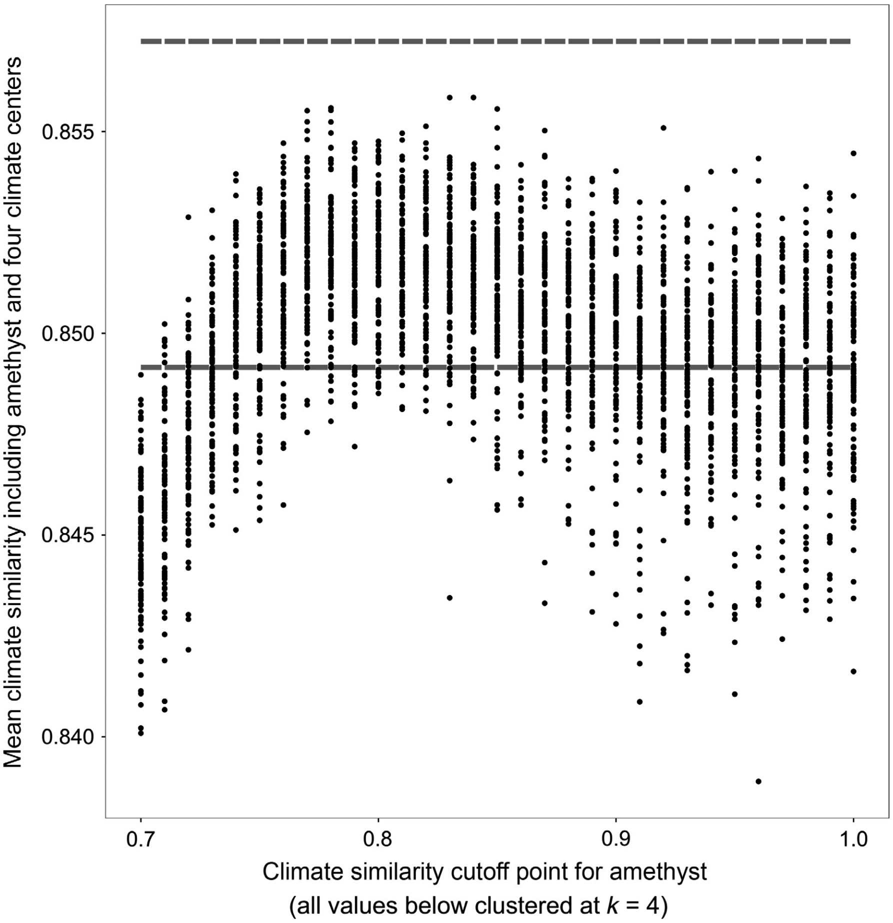

FIG. 5. Results of a simulation exercise with Amethyst, a United States Natural Resources Conservation Service "Selected Class Germplasm" release of the forb *Machaeranthera canescens*. Our goal was to identify four climate centers (hypothetical accessions) that best complement Amethyst to achieve high mean climate similarity. We excluded areas of *M. canescens'* modeled range that varied in climate similarity to Amethyst ($x$-axis, 0.7–1.0 at 0.01 intervals; cells equal to or greater than the cutoff point excluded) and conducted 100 simulations (solving for $k$ = 4 climate centers) in each retained portion at each interval. Individual points of the scatter plot represent the results of each simulation: the mean climate similarity value ($y$-axis) of four accessions plus Amethyst. Mean climate similarity of four (lower gray solid line) and five (upper gray dashed line) accession solutions that ignore Amethyst are presented for comparison. The points between the two lines demonstrate that the complementary approach, i.e., four hypothetical accessions plus Amethyst, can achieve better climate coverage than a set of four hypothetical accessions alone and similar climate coverage to that of five accessions alone.

ranking tool for this purpose has been developed and is under testing. To fully realize the utility of this tool, a database that provides current information on source-identified seed availability and that is accessible across agencies is, of course, essential.

We have defined a flexible framework for developing species-specific transfer guidelines that can inform all phases of plant materials development and guide wildland seed collection. In its most basic application, our approach of identification of climate centers can guide ongoing seed-banking efforts (e.g., BLM Seeds of Success targets, Kew Gardens Millennium Seed Bank Project, Australian Seed Bank Partnership) to ensure that diverse, representative germplasm of native species is available, both for basic conservation and to facilitate restoration research. For species targeted for restoration, we can identify the set of accessions, or climate centers, that maximize coverage across a set of climate axes for a target number of accessions. By running simulations where we sequentially increased target number of accessions, we were able to assess gains in climate coverage, measured in maximum mean climate similarity (Figs. 2 and 3). Our analyses identified rapid initial gains in mean climate similarity with addition of climate centers, followed closely by saturation of each focal species' climate space.

These species were chosen because of their relevance to past and ongoing native plant materials development efforts in the western United States and perception as "workhorse" species in this region (Wood et al. 2015), but our analyses are readily applied to other regions and species with proven or predicted restoration potential, or species with more specific conservation value. Across

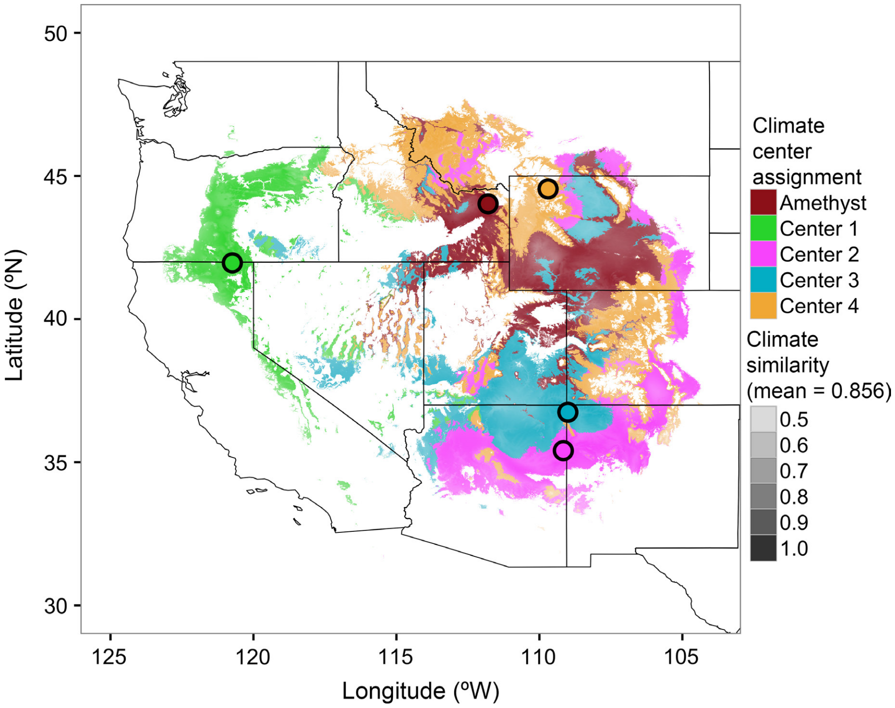

FIG. 6. A map of the four climate center complements (hypothetical accessions) to Amethyst and their assignments that maximize climate coverage throughout the species' modeled range. [Colour figure can be viewed at wileyonlinelibrary.com]

these broadly distributed species, we observed variability in the rates of increase in mean climate similarity with addition of climate centers, which is likely due to an interaction between the breadth of climate niche and the density distribution of climate space unique to each taxon (Fig. 1E). In the perennial grass species, *Pleuraphis jamesii*, we observed a marked elbow, or reduction in rate of increase in mean climate similarity when two climate centers were considered, which we did not observe in other grass species (Fig. 3). This result might indicate that this species occupies two disjunct regions in climate space that are separated by unsuitable niche space for *P. jamesii*. Future study may reveal other unique profiles for each species, or generalizable trends across groups.

We suggest that practitioners determine a practical number of accessions to evaluate on a species-by-species basis that balances climate coverage and investment in research and development by examining the relationship between climate similarity and accession number (Fig. 3). In cases of severe resource limitation, or for a species used only rarely in restoration, a single representative climate center may be identified. Importantly, our method can be modified to incorporate accessions that are already developed (cultivar releases, source-identified germplasm, etc.) or known to be under development (Figs. 5 and 6). Thus, our approach provides a framework for unifying native plant materials development efforts across spatial scales and management jurisdictions.

The first step within this framework is establishing a target number of accessions to develop *across a species' entire range*, where target number is informed by climate coverage (Fig. 3). This step is critical because having a specific goal enables more cost-effective research strategies and aligns efforts across regions. Next, accessions near a species' climate centers are collected in a manner that captures local genetic diversity. While accessions from the exact coordinates of climate centers may not be obtainable, numerous, alternatives can be found within climate similarity maps (seen as darker regions in Figs. 1C, 2, and 6), are distributed across large geographic areas, and thus provide a robust sampling strategy for field evaluations of climate transfer.

Such evaluations would be conducted to relate plant performance to differences between source climate of an accession and yearly conditions at a field site or sites, as in Rutter and Fenster (2007). Ideally, these evaluations would be conducted at an array of field sites that reflects the climatic diversity and breadth of accessions selected for study. It is expected that field evaluations would demonstrate that some climate variables are better predictors of plant performance than others (Anderson et al. 2015). In this case, the climate distance matrices for cluster analysis and climate similarity indices would be recalculated with climate variables weighted by their ability to predict performance. A second round of cluster analysis may reveal unsampled climate centers in weighted climate space and necessitate a secondary round of sampling. Having established biologically relevant climate axes, we would then enter into development accessions climatically similar to weighted climate centers. This framework would be less expensive than current methods of plant materials development, which use empirical seed zones derived from common garden study (estimated at US$150000 per species, according to the Bureau of Land Management Native Seed Strategy Business Plan 2015–2020), to guide field evaluations. However, it is important to note that empirical seed zones provide a more intuitive foundation for guiding field evaluations, by yielding more discrete hypotheses, e.g., accessions have reduced fitness outside of their seed zone of origin.

In the absence of field evaluation data, which are expensive to obtain, other sources of inference on the relation between climate and adaptive divergence can be used to refine species-specific climate similarity analysis. For example, biologically relevant climate axes that potentially correlate with adaptive divergence can be identified with population genetic study (e.g., Shryock et al. 2015), which is generally less expensive than analysis of phenotypic variation, and these results used to weight variables. In addition, when available, published reciprocal transplant, common garden studies, or other plant performance evaluations of focal species may suggest that some climate variables are more important than others in driving adaptive differentiation within a species. For example, data collected by agencies such as the Natural Resource Conservation Service (i.e., NRCS) to rank accessions and determine areas of adaptation of accessions entered into development would be very useful, although these data can be difficult to access and interpret (Hardegree et al. 2016).

As suggested above, the outcome of matching seed to a target site using the climate similarity method presented here is dependent on the validity of the assumptions about the impacts of different variables on fitness. For example, maximizing similarity for variables that do not have a strong correlation with performance could lead to the use of accessions that are maladapted to target sites. In addition, the method is, as deployed here, wholly dependent on interpolated climate data, which have clear limitations, e.g., microclimatic variability of a magnitude similar to that separating climate centers. One possible strategy to mitigate some of the error intrinsic to efforts to match source to site could be to pool top ranking accessions to increase the chance of planting at least some well-adapted seeds. However, the scale at which such pooling would be beneficial is relatively unexplored, although our approach suggests a method to quantify scale, in terms of climatic dispersion of pooled accessions.

Another potential problem is that climate similarity analyses can map regions that span large latitudinal intervals as climatically similar (see Fig. 6), a phenomenon shared with some empirical and provisional seed transfer zones (see zone M1L2H3 of Johnson et al. [2012], zones 1 and 3a from Bradley St Clair et al. [2013], and western zones of Bower et al. [2014]). For some species, transport across these intervals may result in phenological mismatch. In addition, because geographic and neutral genetic distance are generally correlated in plants, introduction of distant plant materials may result in deleterious outcomes such as outbreeding depression for remnant populations, which can be caused by the accumulation of (otherwise) genetically neutral incompatibilities to crossing (Hufford and Mazer 2003, Moyle and Nakazato 2010). Furthermore, biotic and abiotic variables not assessed here, like plant community associates and soil attributes, can drive adaptive divergence among populations (Bischoff et al. 2006, Macel et al. 2007). While resolution of data layers of soil characteristics that may impact plant fitness is generally coarser than that of interpolated climate data, recent developments allow for the ready incorporation of uncertainty based on spatial variance in soil map resolution (Hijmans 2015). Advances in modeling and remote sensing of edaphic variables could soon provide data of sufficient resolution, as well (Mulder et al. 2011, Hengl et al. 2014, 2015; NASA SMAP mission). While earlier efforts in soils modeling incorporated a great deal of error, Hengl et al. (2015) have developed accurate random forest models that predict soil chemical and physical properties at 250-m resolution at a regional extent, and we expect these models to be projected globally in the near future. When available, spatial variation in soil properties can be readily incorporated to yield a more comprehensive environmental similarity measure.

Finally, a primary assumption is that individual plant genotypes perform best when grown in climatic conditions that match their source. While this assumption is in general accord with studies of local adaptation in plants, an appreciable fraction of reciprocal transplant studies find no evidence for a home site advantage (Bischoff et al. 2006, Leimu and Fischer 2008). This observation highlights the importance of testing the relative performance of different sources across sites that vary in their climatic characteristics. A review of recent evidence suggests that modeled climate niche limits are often in fact range limits, however, and moving species outside of their climate niche is likely to result in poor performance (Lee-Yaw et al. 2016). We point out that commercially available plant materials may be selected for their exceptional performance in an agronomic context and are widely used; yet, their exact origins are often unreported. It would be informative to learn if these plant materials were sourced from climate centers identified by our method. We suggest that plant materials developers document provenance as it will further the discussion of local adaptation. As discussed above, our approach can guide tests by identifying source populations for testing, and also a tractable, representative set of evaluation sites.

It is also important to note that the climate centers we present (e.g., Fig. 6) are derived from historical climate data. The geographic utility of an accession is expected to shift with climate change (Ledig et al. 2010, Breed et al. 2013). The magnitude of these changes could be readily assessed through calculation and environmental partitioning of SDMs derived from multiple climate change scenarios, and such assessments would be highly useful to assisted migration efforts and longer-term restoration planning (Vitt et al. 2010). Comparisons of current and future utility of accessions could be used to estimate long-term return on investment and to schedule the appropriate development and introduction of new germplasm lines in step with climate change.

## IMPLICATIONS

Climate similarity analyses form the basis of a flexible, cost-effective toolset that provides robust and timely evaluation and development guidance for targeted restoration species. This guidance informs all phases of the development process, including collection, evaluation, and use. Furthermore, because climate similarity analyses can evaluate the range-wide climate match of source-identified seed, the approach facilitates the coordination of native plant materials efforts across research groups and ecoregions. Such integration could yield greater overall efficiency and cost savings in the development process and expedite progress toward multiple goals outlined in concerted conservation efforts, such as the Plant Conservation Alliance's National Seed Strategy (Oldfield and Olwell 2015). While production of species distribution models may be a barrier presently, SDMs are becoming increasingly accessible. For practitioners seeking a more generalized approach in the absence of an SDM, our methods may be analogously applied to any desired extent, including areas of likely disturbance. We hope that broad adoption of the approach outlined here will increase efficiency and restoration success and potentially help prioritize the collection of native plant materials to enable strategies for mitigating climate change.

## ACKNOWLEDGMENTS

We would like to thank Peter Chuckran for assembling plant occurrence records, the Bureau of Land Management's Colorado Plateau Native Plant Program for funding, Scott Jensen, Seth Munson, and the Kramer lab at the Chicago Botanic Garden/Northwestern University for providing suggestions that improved the manuscript. Any use of trade names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

## LITERATURE CITED

Anderson, J. T., N. Perera, B. Chowdhury, and T. Mitchell-Olds. 2015. Microgeographic patterns of genetic divergence and adaptation across environmental gradients in *Boechera stricta* (Brassicaceae)*. American Naturalist 186(S1):S60–S73.

Bischoff, A., et al. 2006. Detecting local adaptation in widespread grassland species – the importance of scale and local plant community. Journal of Ecology 94:1130–1142.

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

Bradley St Clair, J., 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 6:933–948.

Breed, M. F., M. G. Stead, K. M. Ottewell, M. G. Gardner, and A. J. Lowe. 2013. Which provenance and where? Seed sourcing strategies for revegetation in a changing environment. Conservation Genetics 14:1–10.

Butterfield, B. J. 2015. Environmental filtering increases in intensity at both ends of climatic gradients, though driven by different factors, across woody vegetation types of the southwest USA. Oikos 124:1374–1382.

Doherty, K., B. J. Butterfield, and T. E. Wood. 2017. Data from: Matching seed to site by climate similarity: Techniques to prioritize plant materials development and use in restoration. Dryad Digital Repository. https://doi.org/10.5061/dryad.43bv0

Elith, J., S. J. Phillips, T. Hastie, M. Dudík, Y. E. Chee, and C. J. Yates. 2011. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions 17:43–57.

Elzinga, J. A., A. Atlan, A. Biere, L. Gigord, A. E. Weis, and G. Bernasconi. 2007. Time after time: flowering phenology and biotic interactions. Trends in Ecology and Evolution 22: 432–439.

Hardegree, S. P., T. A. Jones, B. A. Roundy, N. L. Shaw, and T. A. Monaco. 2016. Assessment of range planting as a conservation practice. Rangeland Ecology and Management 69:237–247.

Hargrove, W. W., and F. M. Hoffman. 2004. Potential of multivariate quantitative methods for delineation and visualization of ecoregions. Environmental Management 34:S39–S60.

Hengl, T., et al. 2014. SoilGrids1 km—global soil information based on automated mapping. PLoS ONE 9:e105992.

Hengl, T., et al. 2015. Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. PLoS ONE 10:1–26.

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

Hijmans, R. J. 2015. raster: geographic data analysis and modeling. R package version 2.4-20. http://CRAN.R-project.org/package=raster

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25:1965–1978.

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

Johnson, N. C., G. W. T. Wilson, M. A. Bowker, J. A. Wilson, and R. M. Miller. 2010. Resource limitation is a driver of local adaptation in mycorrhizal symbioses. Proceedings of the National Academy of Sciences USA 107:2093–2098.

Johnson, R. C., M. J. Cashman, and K. Vance-Borland. 2012. Genecology and seed zones for Indian ricegrass collected in the southwestern United States. Rangeland Ecology and Management 65:523–532.

Johnson, R. C., M. E. Horning, E. K. Espeland, and K. Vance-Borland. 2015. Relating adaptive genetic traits to climate for Sandberg bluegrass from the intermountain western United States. Evolutionary Applications 8:172–184.

Joshi, J., et al. 2001. Local adaptation enhances performance of common plant species. Ecology Letters 4:536–544.

Langlet, O. 1971. Revising some terms of intra-specific differentiation. Hereditas 68:277–280.

Ledig, T. F., G. E. Rehfeldt, C. Sáenz-Romero, and C. Flores-López. 2010. Projections of suitable habitat for rare species under global warming scenarios. American Journal of Botany 97:970–987.

Lee-Yaw, J. A., H. M. Kharouba, M. Bontrager, C. Mahony, A. M. Csergő, A. M. E. Noreen, Q. Li, R. Schuster, and A. L. Angert. 2016. A synthesis of transplant experiments and ecological niche models suggests that range limits are often niche limits. Ecology Letters 19:710–722.

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

Leimu, R., and M. Fischer. 2008. A meta-analysis of local adaptation in plants. PLoS One 3:e4010.

Linhart, Y. B., and M. C. Grant. 1996. Evolutionary significance of local genetic difference in plants. Annual Review of Ecology and Systematics 27:237–277.

Lobell, D. B., and C. B. Field. 2007. Global scale climate–crop yield relationships and the impacts of recent warming. Environmental Research Letters 2:014002.

Macel, M., et al. 2007. Climate vs. soil factors in local adaptation of two common plant species. Ecology 88:424–433.

Manel, S., B. N. Poncet, P. Legendre, F. Gugerli, and R. Holderegger. 2010. Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. Molecular Ecology 19:3824–3835.

Mijnsbrugge, V. 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.

Moyle, L. C., and T. Nakazato. 2010. Hybrid incompatibility "snowballs" between *Solanum* species. Science 329:1521–1523.

Mulder, V. L., S. de Bruin, M. E. Schaepman, and T. R. Mayr. 2011. The use of remote sensing in soil and terrain mapping—a review. Geoderma 162:1–19.

Oldfield, S., and P. Olwell. 2015. The right seed in the right place at the right time. BioScience 65:955–956.

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

Peppin, D. L., P. Z. Fulé, J. C. Lynn, A. L. Mottek-Lucas, and C. Hull Sieg. 2010. Market perceptions and opportunities for native plant production on the Southern Colorado plateau. Restoration Ecology 18:113–124.

Phillips, S. J., M. Dudík, and R. E. Schapire. 2004. A Maximum Entropy Approach to Species Distribution Modeling. Twenty-first international conference on machine learning. ICML 2004:83.

R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Ramirez-Villegas, J., C. Lau, J. Hooker, A. Jarvis, K. Ann-Kristin, N. Arnell, S. Johannes, and O. Tom. 2011. Climate analogues: finding tomorrow's agriculture today. CGIAR research program on climate change, agriculture and food security. Working Paper 12. CCAFS, Copenhagen, Denmark.

Rutter, M. T., and C. B. Fenster. 2007. Testing for adaptation to climate in *Arabidopsis thaliana*: a calibrated common garden approach. Annals of Botany 99:529–536.

Seager, R., et al. 2007. Model projections of an imminent transition to a more arid climate in southwestern North America. Science 316:1181–1184.

Shryock, D. F., C. A. Havrilla, L. A. DeFalco, T. C. Esque, N. A. Custer, and T. E. Wood. 2015. Landscape genomics of *Sphaeralcea ambigua* in the Mojave Desert: a multivariate, spatially explicit approach to guide ecological restoration. Conservation Genetics 16:1303–1317.

Thuiller, W., D. Georges, and R. Engler. 2014. biomod2: Ensemble platform for species distribution modeling. R package version 3.1-64. https://cran.r-project.org/web/packages/biomod2/index.html

Turner, T. L., E. C. Bourne, E. J. Von Wettberg, T. T. Hu, and S. V. Nuzhdin. 2010. Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. Nature Genetics 42:260–263.

U.S. Department of Agriculture, Natural Resources Conservation Service. 2014. Release brochure for amethyst germplasm hoary tansyaster (*Machaeranthera canescens*). Aberdeen Plant Materials Center, Aberdeen, Idaho, USA.

U.S. Department of the Interior, Bureau of Land Management. 2015*a*. National Seed Strategy for Rehabilitation and Restoration (Report Number BLM/WO/GI-15/012 + 7400). http://www.blm.gov/ut/st/en/prog/more/CPNPP/0/seedstrategy.html

U.S. Department of the Interior, Bureau of Land Management. 2015*b*. National Seed Strategy Draft Business Plan 2015–2020. http://www.blm.gov/style/medialib/blm/wo/Planning_and_Renewable_Resources/fish__wildlife_and/plants/seedstrategy.Par.56981.File.dat/Seed_Strategy_Business_Plan_2015.pdf

Vitt, P., K. Havens, A. T. Kramer, D. Sollenberger, and E. Yates. 2010. Assisted migration of plants: changes in latitudes, changes in attitudes. Biological Conservation 143:18–27.

Ward, J. 1963. Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association 57:269–306.

Westerling, A. L., H. G. Hidalgo, D. R. Cayan, and T. W. Swetnam. 2006. Warming and earlier spring increase western U.S. forest wildfire activity. Science 313:940–943.

Wood, T. E., K. Doherty, and W. Padgett. 2015. Colorado plateau ecosystems development of native plant materials for restoration and rehabilitation of Colorado plateau ecosystems. Natural Areas Journal 35:134–150.

## DATA AVAILABILITY

Data associated with this paper have been deposited in a Dryad digital repository https://doi.org/10.5061/dryad.43bv0