

## RESEARCH ARTICLE

## UN Decade on Ecosystem Restoration

## Spatially explicit management of genetic diversity using ancestry probability surfaces

Rob Massatti<sup>1</sup>  | Daniel E. Winkler<sup>2</sup> <sup>1</sup>US Geological Survey, Southwest  
Biological Science Center, Flagstaff, AZ,  
USA<sup>2</sup>US Geological Survey, Southwest  
Biological Science Center, Tucson, AZ,  
USA

## Correspondence

Rob Massatti

Email: [rmassatti@usgs.gov](mailto:rmassatti@usgs.gov)

## Funding information

U.S. Bureau of Land Management, Grant/  
Award Number: L18PG00152

Handling Editor: Luisa Carvalheiro

## Abstract

1. Ecological restoration and conservation efforts are increasing worldwide and the management of intraspecific genetic variation in plants and animals, an important component of biodiversity, is increasingly valued. As a result, tailorable, spatially explicit approaches to map genetic variation are needed to support decision-making and management frameworks related to the recovery of threatened and endangered species and the maintenance of genetic resources in species utilized by humans, such as for restoration or agricultural purposes.
2. Here, we describe and demonstrate a workflow to spatially interpolate patterns of genetic differentiation using novel functions in the R package POPMAPS (Population Management using Ancestry Probability Surfaces). Our approach uses empirical genetic data to estimate ancestry coefficients across a user-defined landscape correlated with patterns of differentiation in the focal species. The resulting surface, which we term the ancestry probability surface, includes two components: hard population boundaries and estimations of uncertainty that represent confidence in population assignments (i.e. ancestry probabilities).
3. An ancestry probability surface developed for *Hilaria jamesii*, an important graminoid utilized in restoration across the western United States, demonstrates the functionality of POPMAPS. Genetic distances among empirical sites correlated better with least-cost distances across suitable habitat than with geographical distances, informing the surface over which the interpolation was conducted (i.e. a model indicating habitat suitability). A jackknifing procedure identified parameter values resulting in robust population assignments across the species' range, which were utilized in downstream analyses to estimate ancestry coefficients from empirical data. Ancestry coefficients were translated into ancestry probabilities, which tended to be low for cells that were intermediate in distance between empirical sampling locations representing different populations or when influenced by empirical sampling locations with mixed genetic ancestry.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Published 2022. This article is a U.S. Government work and is in the public domain in the USA. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

4. POPMAPS allows users to tailor parameter values and analytical approaches and thereby incorporate species-specific biological characteristics and desired levels of uncertainty into maps illustrating patterns of genetic differentiation. Ancestry probability surfaces may be used to guide management or investigate further ecological or evolutionary hypotheses. We discuss how maps produced by POPMAPS can inform multiple management challenges including species recovery planning and the utilization of commonly used species in restoration.

#### KEYWORDS

conservation unit, evolutionary significant unit, least-cost distance, management unit, native plant material, population genetics, population structure, seed transfer zone

## 1 | INTRODUCTION

Understanding the distribution of genetic variation within species is essential to the maintenance of biodiversity (e.g. CBD, 2010; Hoban et al., 2020) and ecosystem function (e.g. Sgrò et al., 2011), which are fundamental goals of worldwide efforts to conserve species and restore ecosystems (UN, 2019). For instance, spatially explicit genetic approaches can resolve patterns of functional connectivity across fragmented habitats (Auffret et al., 2017; Beninde et al., 2016), which may coincide with the identification of populations across an evolutionary continuum (Coates et al., 2018); identify genetically based ecological interactions among species (Bucharova et al., 2021); and uncover adaptive variation influencing the breadth of environmental space available to species under current and future climate scenarios (Fitzpatrick & Keller, 2015). Furthermore, understanding and maintaining genetic diversity has been at the forefront of conservation of wild relatives of important crops (Castañeda-Álvarez et al., 2016; Zhang et al., 2017) and calls to action have recently been made for applying these goals to *all* species to avoid substantial economic, social and cultural losses (Hoban et al., 2020). As genetic data become widely available, the development of tools to assist the translation of these data to management guidance becomes critical.

Genetic data have been integrated into decision-making and management frameworks in a variety of ways (e.g. Breed et al., 2019; Hanson et al., 2020). For species threatened by extirpation or extinction, genetic data have been used to investigate biological factors such as species delimitation (Stanton et al., 2019), gene flow (Rathmacher et al., 2010) and inbreeding depression (Moraes et al., 2018). This knowledge is critical for guiding the development of recovery and/or management plans such that distinct or independent populations may be protected and, if necessary, restored (e.g. Lesica et al., 2016). Commonly occurring species may also have complex management needs that can only be addressed by considering genetic data. For example, species important for human sustenance may require knowledge of population structure (Bradbury et al., 2018) or the genetic basis of trait variation (Varshney et al., 2019), while patterns of adaptive variation to regional climate

gradients (Carvalho et al., 2021) and knowledge of hybridization (Winkler & Massatti, 2020) are important for those utilized in ecological restoration. Given the abundance of applications of genetic data to management problems, translating spatially explicit genetic patterns into tangible products, such as maps, that may guide decision-making based on intraspecific genetic variation is critically important.

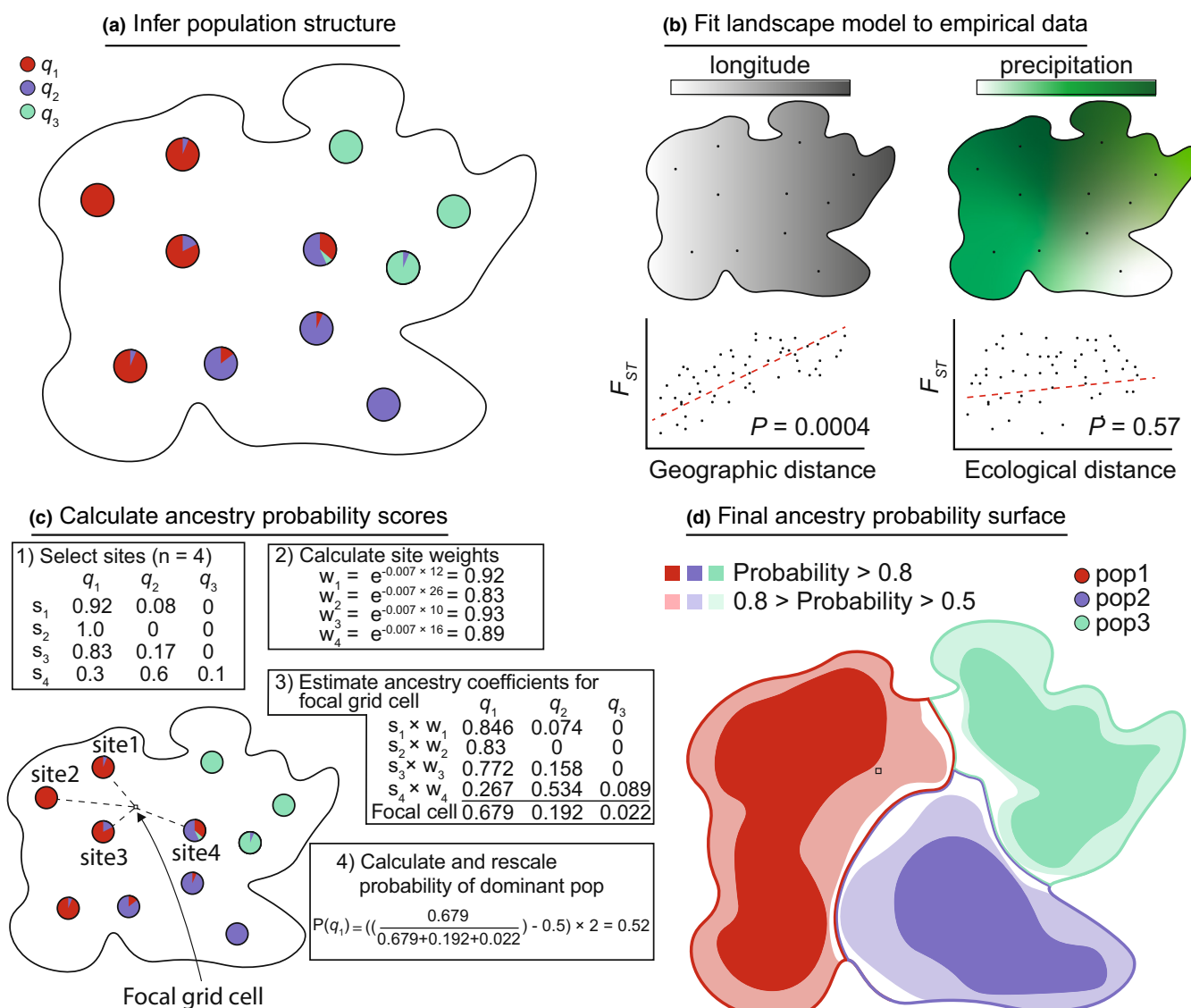
Methodologies to visualize spatially explicit genetic patterns have been actively developed over the past two decades to support ecological and evolutionary inferences and the translation of results into actionable products. For example, methods focusing on the identification of genetic discontinuities, such as Monmonier's algorithm (Manni et al., 2004), facilitate inference of biologically important regions influencing differentiation across species' ranges (Morris et al., 2010). Other methods produce spatially continuous interpolations of intraspecific variation using pairwise genetic differences (Rosauer et al., 2014; Tarroso et al., 2019; Vandergast et al., 2011) or assignment scores (Caye et al., 2016; Guillot et al., 2005; Murphy et al., 2008), which have been useful for developing conservation plans for threatened species (Hendricks et al., 2017), delineating biodiversity hotspots (Carnaval et al., 2014) and uncovering introductions of invasive species (Winkler et al., 2019), among others. Despite the variety of available methodologies, few incorporate the flexibility to modify the geospatial surfaces over which genetic patterns are interpolated (e.g. to make interpolations more biologically realistic; e.g. Storfer et al., 2010) or generate products that directly provide probability estimations of genetic identity across complicated patterns of differentiation, which are a critical component of management decisions. Furthermore, managers may desire the flexibility to create maps imparting higher or lower certainty in spatial interpolations, requiring adjustable parameters unavailable in similar methodologies to facilitate management goals.

To support the incorporation of spatially explicit patterns of genetic variation in species management, we demonstrate a workflow to visualize genetic patterns that includes novel functions available in the R package POPMAPS (Population Management using Ancestry Probability Surfaces). The workflow is highly customizable with

respect to geospatial surfaces and analyses required to develop prerequisite datasets and includes opportunities for parameter modification that allows system-specific biological characteristics and management goals to influence estimations of uncertainty. We validate and illustrate the methodology using empirical genetic data for *Hilaria jamesii* (Torr.) Benth. (synonym *Pleuraphis jamesii*, Poaceae; commonly called James' galleta grass), a priority restoration species used across degraded dryland habitats in the Intermountain West of the western United States (US). Through this framework and by virtue of the empirical example used, we highlight the effectiveness of the method across disciplines where the conservation of genetic diversity is valued as a primary management goal.

## 2 | MATERIALS AND METHODS

Herein, we demonstrate how empirical genetic data can be used to interpolate an ancestry probability surface, which is composed of population boundaries drawn over a surface of values imparting estimations of confidence. The first step in constructing this surface involves inferring two datasets: (a) ancestry coefficients from population structure analyses (Figure 1a) and (b) pairwise genetic distance data between sampling locations. Next, the genetic distance data are compared to spatial surfaces that reflect hypotheses regarding factors influencing species' patterns of differentiation (Figure 1b); this step informs how distances among sampling locations are



**FIGURE 1** Overview of a POPMAPS spatial interpolation workflow. Steps include (a) collecting empirical genetic data and inferring population structure; (b) investigating the factors influencing genetic differentiation to inform the surface over which distances among locations will be calculated; (c) estimating ancestry coefficients and probability scores at every grid cell across the user-defined area; and (d) visualizing the ancestry probabilities (i.e. coloured polygons) and hard boundaries (i.e. coloured lines surrounding polygons). Calculations in (c) accurately estimate ancestry coefficients and an ancestry probability at a focal grid cell using hypothetical data and assuming  $num\_tested = 4$  and  $popmod = -0.007$ . Note that steps a and b do not utilize POPMAPS functions and may be accomplished using a variety of user-defined methods.

calculated during the interpolation process. These steps occur outside of the POPMAPS framework and can be completed however investigators see fit. After prerequisite datasets are generated, novel POPMAPS functions are utilized in R to (a) test how empirical sampling locations influence the calculation of ancestry coefficients, allowing species-specific life-history characteristics to influence probability estimations and (b) calculate ancestry coefficients across the entire, user-defined geographical extent of a species (Figure 1c). Once complete, the ancestry probability surface can be visualized (Figure 1d) and utilized for downstream geospatial analyses. Below, we describe core POPMAPS functions and demonstrate the workflow using *H. jamesii*.

## 2.1 | The core POPMAPS methodology: Estimating ancestry coefficients and ancestry probabilities

The goal of POPMAPS is to estimate an ancestry probability surface, which requires estimating ancestry coefficients at every location (hereafter cell) across a spatial surface (hereafter grid) and translating these values into assignment probabilities (i.e. ancestry probabilities). POPMAPS requires the grid to be an object of the class *RasterLayer*. The ancestry probability of a cell represents the probability that, if individuals are found at that specific geographical location, they will belong to the same genetically defined cluster as individuals of the closest empirical sampling location; hereafter, we use genetic cluster and population interchangeably. Across a grid, cells for which no empirical genetic data are present will be situated in a variety of ways compared to the empirical sampling locations. For example, a cell may be close to one empirical sampling location with no others near, equidistant between two, or varying distances among multiple sampling locations. Furthermore, empirical sampling locations may have complex mixtures of ancestry between two or more of the inferred populations (e.g. Figure 1a). As a result, a primary goal of POPMAPS is to determine the influence an empirical sampling location should have on the calculation of ancestry coefficients at a location for which no empirical data exist.

Two POPMAPS parameters heavily impact the estimation of ancestry coefficients from empirical data. First, the number of empirical sampling locations used in the estimation of ancestry coefficients at cells (*num\_tested*) can be modified. For example, *num\_tested* = 2 will only allow the genetic identities of the two closest sampling locations to influence calculated ancestry coefficients, while *num\_tested* = (total number of sampling locations) will allow every sampling location to influence coefficients. Second, a parameter modifying the equation used to calculate the magnitude of contribution of an empirical sampling location to the estimation of the ancestry coefficients at a cell can be modified (*popmod*). While any equation can be used in relevant POPMAPS functions (Table 1), an exponential function is employed by default:

$$y = e^{(popmod * D)},$$

where increasingly negative values of *popmod* decrease the magnitude of contribution of the empirical location to the focal cell as distance increases (Figure 2a). As such, the default function is similar to other inverse distance weighting approaches (e.g. Vandergast et al., 2011). The distance between a focal cell and an empirical sampling location (*D*) can be represented by either geographical (Euclidean) distance [i.e. isolation-by-distance (IBD)] or least-cost distance (i.e. a distance accounting for the friction of the landscape); determining a landscape over which to calculate the distances used in the equation to calculate empirical sites' contributions is a step that occurs before employing the POPMAPS analytical framework (see empirical demonstration below). Tailoring these parameters to the dataset in hand results in a more accurately estimated ancestry probability surface that reflects species-specific biological characteristics. For example, highly dispersible species often display strong IBD patterns across landscapes at scales proportional to their dispersal kernels (e.g. wind-dispersed plants; Duminil et al., 2007); in such situations, geographical distance, in addition to larger values of *popmod* and higher *num\_tested*, may predict genetic ancestry most accurately. Alternatively, migration may be tightly constrained to suitable habitat distributed across a landscape (e.g. around rather than over mountain ranges), resulting in genetic patterns that are better predicted by least-cost paths (Graves et al., 2014), more negative values of *popmod*, and lower values of *num\_tested*.

Given the influence of *num\_tested* and *popmod* in estimating ancestry probabilities, we provide the function *jackknife()* to help users understand the impact of parameter combinations on predictive power and estimations of uncertainty. This function uses the same methodology as the function *popmaps()* (explained in detail below), but ancestry coefficients are only estimated for locations at which empirical data are present and not across the entire geographical extent of the grid. At every location with empirical genetic data, every combination of parameters (defined in two vectors: *num\_tested\_vec* and *popmod\_vec*; Table 1) is used to estimate ancestry coefficients without including the local empirical data. As such, each calculation follows a leave-one-out approach. This process results in a matrix composed of estimated ancestry coefficients for each genetic cluster and for every parameter combination and empirical sampling site. The function *jackknife\_viz()* takes this matrix and fits linear models to the estimated and empirical data, per genetic cluster and parameter combination, using *lm()* in R [i.e. *lm*(known ancestry coefficients across sampling locations) ~ (predicted ancestry coefficients across sampling locations for a parameter combination)]; ancestry coefficients are transformed with *logit()* prior to fitting linear models to ensure values are not bounded by 0 and 1. *R*<sup>2</sup> values are stored for each linear model and visualized per genetic cluster using heatmaps. Alternatively, the matrices of *R*<sup>2</sup> values for each genetic cluster can be averaged and visualized to guide parameter selection based on overall predictive power.

The core POPMAPS function *popmaps()* estimates ancestry coefficients and probabilities across the extent of a user-defined spatial surface (i.e. the grid). The calculation of the ancestry coefficient for each genetically defined population at a cell (i.e.  $q_k^{(c)}$ , or the average ancestry coefficient across individuals located in cell *c* that

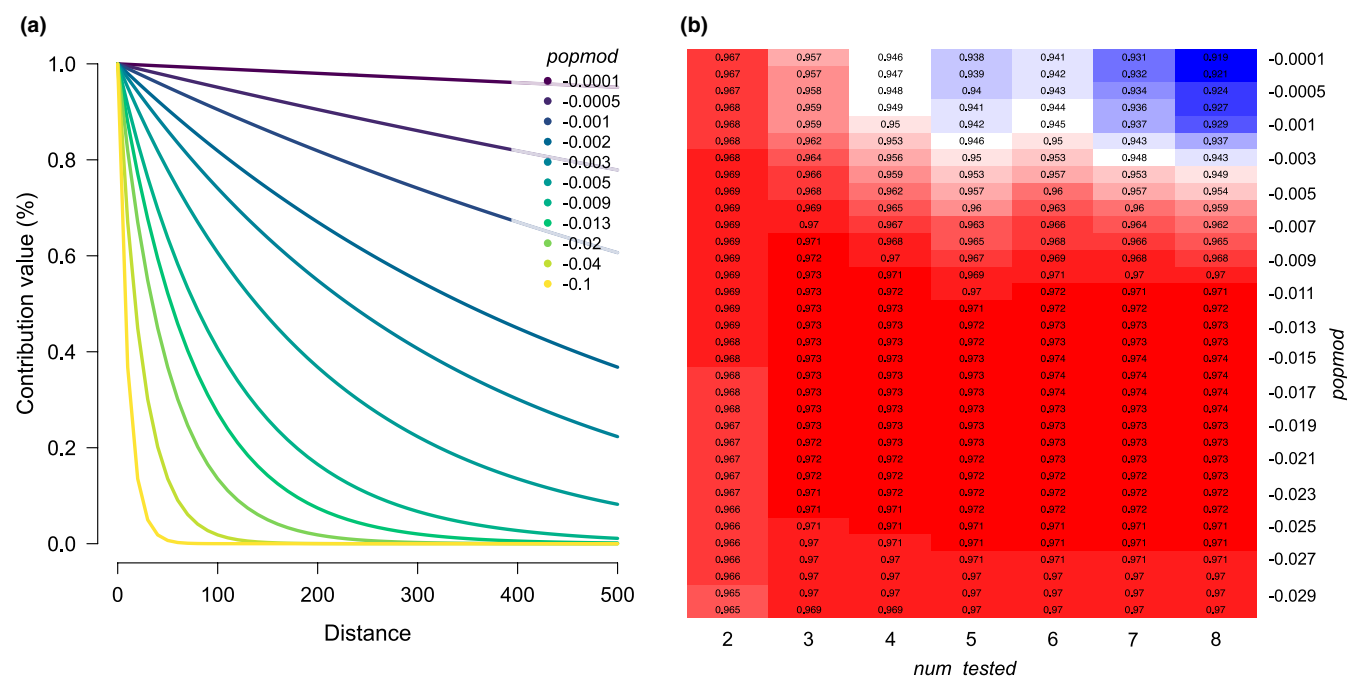
TABLE 1 Detailed description of the arguments available in POPMAPS functions, functions in which arguments are used and default values.

Argument	Function	Default	Description
<i>axis</i>	jackknife_viz()	0	An integer indicating which genetic cluster to visualize $R^2$ values as a heatmap for all parameter combinations. The default value averages $R^2$ values across all axes
<i>boundary_width</i>	popmap_viz()	-0.015	A float modifying the hard boundaries drawn on top of a probability surface. The scale of the analysis may require this variable to be modified so that adjacent hard boundaries connect (see Figure 3)
<i>crs</i>	popmap_viz()	"+init = epsg:5070"	A string defining a mapping projection. The default defines the Albers Equal Area Conic projection suitable for the contiguous United States
<i>dec_lat</i>	anc_extract()	—	A float specifying the decimal latitude from which to extract estimated ancestry coefficients
<i>dec_long</i>	anc_extract()	—	A float specifying the decimal longitude from which to extract estimated ancestry coefficients
<i>dist_prob_func</i>	jackknife(), popmaps()	function(popmod_ temp,distance) {exp(popmod_ temp*distance)}	A function defining the relationship between distance and the contribution of an empirical site's ancestry coefficients to the estimation of ancestry coefficients at an inference cell. The default equation defines the relationship in Figure 2
<i>empirical_pt_dist</i>	jackknife(), popmaps()	5	An integer representing the minimum distance that all sites in <i>num_tested</i> must be separated by. If <i>num_tested</i> = 3, the closest two empirical sites to the cell are determined first. If the two sites are closer than this value, the closest will be kept and the second discarded. The next closest site will then be selected and compared to the first; this process repeats until the two sites are farther than <i>empirical_pt_dist</i> . Subsequently, the third empirical site will be selected using the same process. This rarefaction procedure may reduce high spatial autocorrelation expected in genetic patterns among proximate empirical sites that may mislead estimations of ancestry coefficients. The parameter should be informed by the distribution of empirical sites and the biology of the focal species
<i>input_locs</i>	anc_extract(), jackknife(), jackknife_viz(), popmap_viz(), popmaps()	—	An R object (rows = total # empirical sites, columns = total # genetic axes +3)) with column 1: site name; column 2: decimal longitude; column 3: decimal latitude; column 4...column x: ancestry coefficients for genetic cluster 1... genetic cluster x. Functions depend on this precise format—see example data <i>hija_struc</i>
<i>input_raster</i>	anc_extract(), jackknife(), popmap_viz(), popmaps()	—	An R RasterLayer object defining the geographical extent for the spatial interpolation. Values in the cells will be used to calculate distance used in the <i>dist_prob_func</i> (see above) if <i>surface</i> = 'C' (see below). See example data <i>hija_raster</i>
<i>jackknife_data</i>	jackknife_viz()	—	An R object resulting from executing the function jackknife()
<i>maptype</i>	popmap_viz()	—	A string (either 'bound' or 'ancestry') that defines the type of map to be drawn. Specifying 'bound' will draw the hard boundaries only, while specifying 'ancestry' will draw the hard boundaries and ancestry probability surface
<i>ncore</i>	popmaps()	4	An integer allowing the user to specify how many cores on the local machine should be used to facilitate spatial interpolation
<i>num_sites</i>	jackknife(), popmaps()	10	An integer representing the pool of empirical sites considered when selecting <i>num_tested</i> sites to estimate ancestry coefficients. If empirical sites are highly clustered and rarefaction due to <i>empirical_pt_dist</i> causes many to be discarded, this variable will likely need to be increased at the cost of computing time
<i>num_tested</i>	popmaps()	3	An integer representing the number of empirical sampling sites that will contribute to the estimation of ancestry coefficients
<i>num_tested_vec</i>	jackknife()	c(2,3,4,5,6,7,8)	A vector of integers representing the values of <i>num_tested</i> that should be tested during jackknifing. All pairwise combinations with <i>popmod_vec</i> will be tested
<i>pie_radius</i>	popmap_viz()	0.15	A float modifying the size of the pie charts depicting empirical ancestry coefficients drawn on top of the probability surface (see Figure 3)
<i>popmod</i>	popmaps()	-0.001	A float used in <i>dist_prob_func</i> that modifies the relationship between the contribution of an empirical site to the estimation of ancestry coefficients at a cell and distance (see Figure 2). Using the default <i>dist_prob_func</i> , values ranging from -0.0001 to -0.9 have the most influence on the shape of the curve. Modifying <i>dist_prob_func</i> will require users to determine how to set <i>popmod</i> accordingly

(Continues)

TABLE 1 (Continued)

Argument	Function	Default	Description
<i>popmod_vec</i>	jackknife()	c(-0.001,-0.01,-0.05,-0.1,-0.15)	A vector of floats representing the values of <i>popmod</i> that should be tested during jackknifing. All pairwise combinations with <i>num_tested_vec</i> will be tested
<i>pop_raster_list</i>	anc_extract() popmap_viz()	—	An R object resulting from executing the function <i>popmaps()</i>
<i>surface</i>	jackknife(), popmaps()	'G'	A string (either 'G' or 'C') that determines how <i>input_raster</i> is used to calculate ancestry coefficients. If 'G', the spatial attributes of <i>input_raster</i> will be used to calculate geographical (i.e. Euclidean) distances between empirical sites and inference sites. If 'C', least-cost distances are calculated using the values contained in the raster cells
<i>threshold</i>	popmaps()	0	A float (0.0–1.0) determined during species distribution modelling inference. If >0, ancestry coefficients will not be estimated in raster cells with values < threshold, which saves computation time by avoiding estimating ancestry coefficients for cells in which the focal species would never be expected to occur



**FIGURE 2** (a) Influence of the parameter *popmod* on the default exponential equation used to determine how empirical sampling locations influence the estimation of ancestry coefficients at cells as distance increases. At lower values of *popmod*, the influence of empirical locations on estimations at focal grid cells decline quickly as distance increases. (b) Jackknife heatmap for *Hilaria jamesii*. Numbers within cells are averages of  $R^2$  values resulting from, per genetically defined population, linear models fit to estimated and empirical ancestry coefficients. Darker red colours represent higher  $R^2$  values and indicate parameter combinations that predict empirical ancestry coefficients well.

originated from population  $k$ ; modified from Pritchard et al., 2000) follows the equation:

$$q_k^{(c)} = \frac{\sum_{i=1}^{num\_tested} (q_k^{(i)} * e^{(popmod * D_{(i)})})}{num\_tested},$$

where  $q_k^{(i)}$  is the ancestry coefficient at empirical sampling location  $i$  for population  $k$  and  $D_{(i)}$  is the distance between the focal cell and empirical sampling location  $i$  (see example calculation in Figure 1c boxes 1–3). Estimated ancestry coefficients are then translated into two values

that together create the ancestry probability surface. First, every population inferred in the empirical dataset is arbitrarily assigned a unique integer value, and the integer value corresponding to the predominant population of the closest empirical sampling location is recorded for each cell. These values, when visualized across the entire grid, represent hard boundaries of the spatially interpolated populations under the assumption that individuals at any cell will simply belong to the same population as that of the closest empirical sampling location (see examples of hard boundaries in Figure 1d). Second, a rescaled probability, or ancestry probability score, is recorded for each cell. These values



represent the confidence assigned to the hard boundary hypothesis; when visualized over the entire grid, they compose the ancestry probability surface. The ancestry probability for a cell is calculated by determining the relative contribution of the dominant ancestry coefficient followed by rescaling relativized values of 0.5-1 to 0-1:

$$P(q_{\max}^{(c)}) = \left( \left( \frac{q_{\max}^{(c)}}{\sum_{k=1}^{num\_pops} (q_k^{(c)})} \right) - 0.5 \right) \times 2,$$

where *num\_pops* is the total number of populations in the empirical genetic dataset and  $q_{\max}^{(c)}$  is the ancestry coefficient with the highest value in the cell (see example calculation in Figure 1c box 4). This value is rescaled under the premise that confidence in the identity of the principal population at a location can only be attained when a population represents >50% of the total genetic identity. Any cell that has a negative relativized ancestry probability (i.e. with no population representing >50% of the total genetic identity) is reassigned a value of 0%. As such, in cells where high levels of admixture among populations are common (e.g. if ancestry coefficients among four populations are 20%, 25%, 15% and 40%, or even if one population is slightly dominant, such as 55%, 5%, 20% and 30%), confidence in estimating the principal population will always be low. Admixture zones among populations illustrate the importance of collecting consistent empirical data across the distribution of a species where an ancestry probability surface will be inferred, especially when a priori knowledge of genetic differentiation is unavailable.

After *popmaps()* completes calculations across the spatial grid, results can be visualized with *popmap\_viz()*. The spatial extent of the grid will influence the drawing of empirical sampling locations' pie charts representing ancestry coefficients and the hard boundaries of the genetically defined populations. As such, variables within *popmaps\_viz()* can be modified to meet user preferences (Table 1). Furthermore, estimated ancestry coefficients may be extracted from any point across the grid using *anc\_extract()*; these values represent quantitative hypotheses to support management applications. Finally, maps can be produced with or without probability estimates depending on end goals of the project, and any grid resulting from *popmaps()* (i.e. population-specific or combined) can be saved as a raster file to be used in downstream geospatial inferences.

## 2.2 | Inferring an ancestry probability surface for *H. jamesii*

Genetic data for *H. jamesii* (369 individuals representing 66 locations) were generated using a modified ddRAD library preparation protocol (Peterson et al., 2012) to meet the goals of adjacent research projects. Specifically, the evolutionary history and patterns of adaptive genetic variation of *H. jamesii* were investigated to inform restoration seed transfer guidelines (Massatti, 2020; Massatti & Knowles, 2020) and elucidate patterns of hybridization

with a closely related species, *H. mutica* (Buckley) Benth. (Winkler & Massatti, 2020). RADseq loci and single nucleotide polymorphisms (SNPs) were identified with the STACKS v2.53 (Catchen et al., 2013) pipeline. See Massatti and Knowles (2020) for data processing details, including the selection of processing parameters and quality filtering. To inform the selection of a landscape model (see below), pairwise  $F_{ST}$ -values between sampling locations were calculated using the POPULATIONS program in STACKS using a *p*-value correction with the default cut-off value ( $p < 0.05$ ).

Genetic structure was inferred using a Bayesian clustering approach implemented in STRUCTURE v2.3.4 (Falush et al., 2003; Pritchard et al., 2000). STRUCTURE was run across *K*-values ranging from 1 to 10 without assigning population membership a priori and using an admixture model with correlated allele frequencies. In all, 20 independent runs per *K*-value were conducted, each with 500,000 Markov chain Monte Carlo iterations following a burn-in of 150,000 iterations, and the most probable *K*-value was chosen using the  $\Delta K$  method (Evanno et al., 2005). Note that any assignment method may be used (to meet the biological assumptions of a given system or to incorporate spatially explicit information; e.g. Caye et al., 2016; Guillot et al., 2005), so long as a text file can be created documenting the ancestry coefficients aligning with each defined cluster for each sampling location. Furthermore, sampling locations can consist of one or multiple individuals. For STRUCTURE-defined populations, we calculated summary statistics using the POPULATIONS program of STACKS including:  $F_{ST}$ -values incorporating a *p*-value correction ( $p < 0.05$ ), nucleotide diversity ( $\pi$ ), Wright's inbreeding coefficient ( $F_{IS}$ ), private alleles, and observed and expected heterozygosity ( $H_{OBS}$  and  $H_{EXP}$ ).

With genetic data in hand, the next step in our methodology involved identifying a landscape model that best explained variation in patterns of differentiation. To incorporate and test alternative models (i.e. geographical vs. least-cost distance), we utilized a species distribution model (SDM) under the hypothesis that higher habitat suitability is positively correlated with gene flow across a landscape, as has been theoretically and empirically demonstrated during phylogeographical and landscape genetic investigations (Auffret et al., 2017; Brown & Knowles, 2012; Knowles & Massatti, 2017; Ortego & Knowles, 2020). However, users may tailor their landscape surface to meet project specific goals, for example as may be accomplished using a resistance surface produced by CIRCUITSCAPE (McRae, 2006). Here, we used MAXENT v3.4.1 (Phillips et al., 2021) to create an SDM raster file with logistic values by following the methods outlined in SDMTOLBOX v2.4 (<http://www.sdmtoolbox.org>; Brown et al., 2017). Prerequisite inputs for MAXENT included location data, which we identified by vetting herbarium records available on the SEINet data portal (<https://swbiodiversity.org/seinet/>), and environmental data represented by the 19 bioclimatic variables available from Worldclim v2 (Fick & Hijmans, 2017); because we do not interpret specific variables influencing the distribution of *H. jamesii*, all variables were incorporated into the final SDM to maximize the model's predictive power (Alvarado-Serrano & Knowles, 2014; Brown et al., 2017). According to the best practices recommended by SDMTOLBOX, we developed an SDM for *H. jamesii* that (a) utilized

spatially rarefied occurrence data accounting for climate heterogeneity; (b) applied a species-specific bias file created using a minimum convex polygon; (c) spatially jackknifed models to calibrate model parameters; and (d) independently evaluated feature class parameters and regularization multipliers. A variety of SDM approaches are available, and we advise users to utilize the abundant literature on this topic to inform their approaches (e.g. Hao et al., 2020).

Following the inference of an SDM, we utilized `GDISTANCE` v1.6 (van Etten, 2017) in R to select a best fit landscape model. Specifically, we used `pointDistance()` to generate geographical distance matrices and `cor()` to calculate the correlation between pairwise genetic distances and geographical distances. To calculate least-cost distances, we created transition layers using *H. jamesii*'s SDM using `transition()`, with `transitionFunction` = 'mean' and `directions` = 8. The transition layer was corrected using `geoCorrection()` with `type` = 'c' and `scale` = TRUE. Finally, least-cost distances were calculated using `costDistance()`, and correlations were calculated between least-cost and genetic distances using `cor()`. The best fit landscape model (i.e. the model with the highest correlation) was used in subsequent steps of this analytical framework to calculate ancestry probabilities across space. All steps up to this point must be performed prior to implementation of `POPMAPS`, and datasets matching those provided within the package as examples (`hija_raster` and `hija_struc`) must be loaded into R as objects that can be passed to `POPMAPS` functions. Other approaches to calculate distance values or test the fit between distance matrices may provide more reliable support for an optimum landscape model, such as those that account for potential autocorrelated structure of matrices (e.g. Clarke et al., 2002; Tarroso et al., 2019) or that utilize partial regressions (Ferrier et al., 2007; Worth et al., 2021); we encourage users to take advantage of this flexibility to produce high-quality `POPMAPS` products that are specific to users' intended goals.

To infer an ancestry probability surface for *H. jamesii*, the predictive power of all combinations of parameter values in `num_tested_vec` (ranging from 2 to 8 by intervals of 1) and `popmod_vec` (−0.001 to −0.030 by intervals of 0.001, in addition to −0.0001, −0.00025, −0.0005, −0.00075) was calculated using `jackknife()`. After selecting values of `num_tested` and `popmod` with the highest  $R^2$  value based on heatmaps generated by `jackknife_viz()`, `popmaps()` was initiated using the following parameter values: `num_tested` = 6; `popmod` = −0.015; `empirical_pt_dist` = 15; `threshold` = 0.1015; `surface` = 'C'. Additional iterations of `popmaps()` were executed to illustrate the influence of parameter combinations on the estimation of probability surfaces.

## 2.3 | Code development and dependencies

All `POPMAPS` functions and analyses were developed in R version 4.0.4 (R Core Team, 2021). Functions to estimate ancestry coefficients and probabilities (`popmaps()` and `jackknife()`) rely on the `RASTER` (Hijmans & van Etten, 2012), `GDISTANCE`, and `SP` v1.4-5 (Pebesma & Bivand, 2005) packages. Parallelization in `popmaps()` was achieved using the packages `PARALLEL` v4.0.4, `DOPARALLEL` v1.0.16 (Microsoft Corporation & Weston, 2020a) and `FOREACH` v1.5.1 (Microsoft Corporation &

Weston, 2020b). Visualization functions (`popmap_viz()` and `jackknife_viz()`) utilize functions in the packages `RASTER`, `GEOPLOTS` v3.1.1 (Warnes et al., 2020), `VIRIDIS` v0.5.1 (Garnier et al., 2018), `RGEOS` v0.5-5 (Bivand & Rundel, 2020), `PLOTRIX` v3.8-1 (Lemon, 2006) and `MAPS` v3.3.0 (Brownrigg et al., 2018).

## 3 | RESULTS

Identification of *Hilaria jamesii* population structure patterns and genetic diversity statistics were based on  $>6.4 \times 10^8$  reads across 369 individuals (average of  $1.35 \times 10^6$  reads per individual after quality filtering). `STRUCTURE` analyses supported  $K = 3$  genetic clusters (i.e. populations; Figure 3) using a dataset containing 9,534 unlinked SNPs. Pairwise sampling location  $F_{ST}$ -values ranged from 0.011 to 0.081 (Massatti & Knowles, 2020). Genetic diversity statistics are consistent across *H. jamesii*'s populations, which are putatively in mutation-drift equilibrium as supported by the empirical 95% quantiles for Tajima's  $D$  overlapping with 0 (Table 2).  $F_{ST}$ -values between population pairs are intermediate between the range of values identified between sampling locations (Figure 3).

The final SDM for *H. jamesii* had an AUC = 0.78 and a Fixed Cumulative 1 Threshold = 0.1015 (informative downstream if user wants to exclude unsuitable habitat when developing ancestry probability surfaces; Table 1 and Figure S1). Testing the correlation between patterns of genetic differentiation and landscape surfaces identified higher influence of least-cost paths ( $r = 0.40$ ) compared to geographical distances ( $r = 0.37$ ). While these values are similar, we note that similar analyses for wind-dispersed plant species overwhelmingly support geographical distances compared to least-cost paths (data not presented).

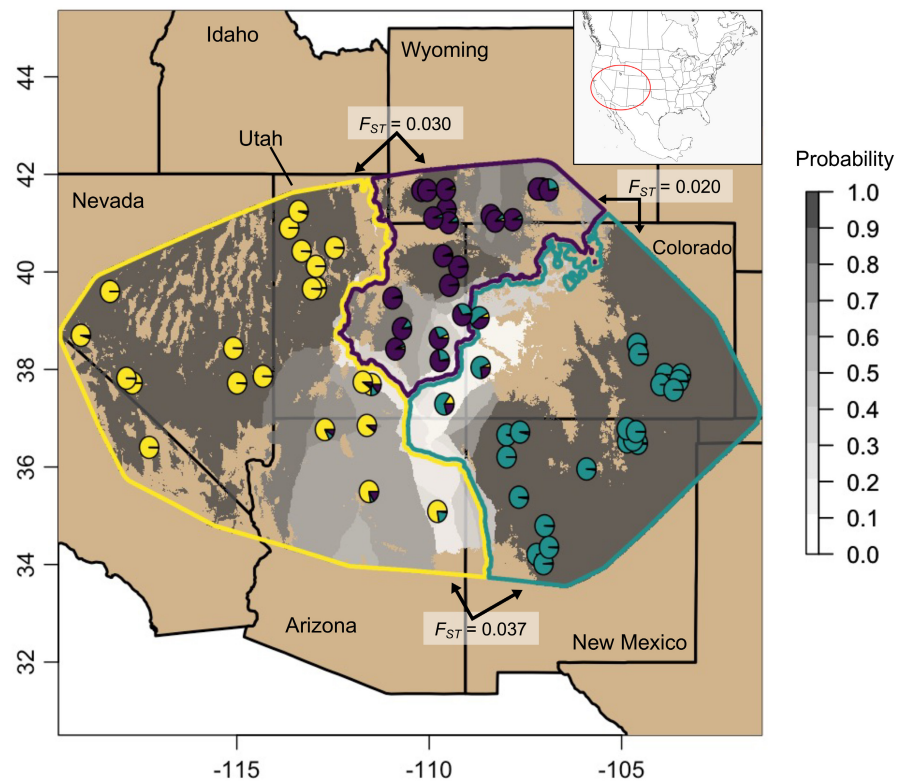
Given landscape correlation results, least-cost distances were used to determine contribution values of empirical sampling locations in the estimation of ancestry coefficients. Jackknife analyses identified an array of parameter combinations that estimated ancestry coefficients at empirical sampling locations well (highest average  $R^2 = 0.974$ ; Figure 2b). As such, we chose parameter values resulting in the highest  $R^2$  using the lowest `num_tested`, and highest `popmod` (`num_tested` = 6; `popmod` = −0.015). In addition, we illustrate ancestry probability surfaces across a smaller grid using all combinations of `num_tested` = 2, 4, 6, 8 and `popmod` = −0.0001, −0.001, −0.005, −0.015, −0.05 to help users understand how variation in different combinations influences resulting ancestry probability surfaces (Figure S2). As may be predicted, ancestry probabilities tend to be low for cells that are intermediate in distance between empirical sampling locations representing different populations or when influenced by empirical sampling locations with mixed genetic ancestry.

## 4 | DISCUSSION

As the intensity of restoration and conservation efforts increases worldwide (Chazdon et al., 2017; UN, 2019; Verdone & Seidl, 2017),



**FIGURE 3** Ancestry probability surface for *Hilaria jamesii*. Pie charts depict proportional ancestry coefficients (averaged across individuals) at empirical sampling locations, which were the data used to estimate assignment probabilities (grayscale colours) across the species' range. Probabilities represent the confidence that plants found at a specific location belong to the same population as the coloured boundary that they fall within. Grayscale colours are not present in locations where the individuals are not predicted to occur (e.g. higher elevations in mountain ranges denoted by the light brown background within the modelled region).



**TABLE 2** Summary statistics for genetically defined *Hilaria jamesii* populations. Results are presented for variant nucleotide positions and include mean values of observed heterozygosity ( $H_{OBS}$ ), expected heterozygosity ( $H_{EXP}$ ), nucleotide diversity ( $\pi$ ), Wright's inbreeding coefficient ( $F_{IS}$ ) and Tajima's  $D$ . Parenthetical population colours refer to pie charts in Figure 3.

Species	Population (colour)	Private alleles	$H_{OBS}$ (variance)	$H_{EXP}$ (variance)	$\pi$ (variance)	$F_{IS}$ (variance)	Tajima's $D$ (95% quantile)
<i>H. jamesii</i>	East (blue)	4,598	0.103 (0.027)	0.157 (0.033)	0.168 (0.038)	0.171 (0.034)	0.428 (−1.078 to 2.793)
<i>H. jamesii</i>	North (purple)	4,613	0.107 (0.030)	0.158 (0.034)	0.168 (0.039)	0.160 (0.036)	0.425 (−1.070 to 2.713)
<i>H. jamesii</i>	West (yellow)	4,312	0.105 (0.028)	0.154 (0.034)	0.166 (0.040)	0.153 (0.028)	0.467 (−1.159 to 2.703)

incorporating species-specific genetic knowledge into management frameworks becomes vitally important. Managing by genetically defined populations created through evolutionary processes shares similarities with strategies such as the primary restoration gene pool (Jones, 2003), evolutionarily significant units (Moritz, 1999) and management units (Palsbøll et al., 2007). Spatial interpolations generated by POPMAPS may be interpreted differently based on the subset of loci used for inference (e.g. adaptive versus neutral; Barbosa et al., 2018), highlighting a key benefit of our method's flexibility. When utilizing ancestry probability surfaces to support management activities, accounting for intraspecific variation may (a) protect evolutionarily significant diversity resulting from species' interactions with historical events (i.e. an important component of biodiversity; CBD, 2010; Coates et al., 2018); (b) avoid genetic processes that can influence management success in early generations following treatments (e.g. outbreeding depression; Frankham et al., 2011; Hufford et al., 2012; Trask et al., 2021); (c) protect unique adaptive variation, as adaptation to the same environmental conditions can occur using different genetic pathways across populations (Gould

& Stinchcombe, 2017); and (d) support interactions among species within communities (Vandegheuchte et al., 2011; Whitham et al., 2006). Our demonstration of the POPMAPS framework in a restoration context provides a concrete example that may help the science and management communities identify similar applications and promote further methodological development.

#### 4.1 | The popmaps methodology

The ancestry probability surface estimated for *Hilaria jamesii* depicts three populations having broad distributions across the western US; each population displays similar levels of diversity and differentiation and is putatively in mutation-drift equilibrium (Figure 3; Table 2). Large seeds that are likely gravity or animal dispersed support the inference that least-cost distances fit patterns of genetic differentiation better than geographical distance. In other words, seeds do not have adaptations to be distributed long distance across large expanses of unsuitable habitat (e.g. similar to wind-dispersed species; Duminil

et al., 2007), but rather are dependent on dispersing incrementally across adjacent, suitable habitat. Within each population, ancestry probabilities are uniformly high except where the three lineages come into contact near the Four Corners region in the Intermountain West (Figure 3). Additional sampling across low probability areas for *H. jamesii* could either reinforce admixture zones between the populations, knowledge of which may be useful to test evolutionary processes such as selection across clines (Gompert et al., 2012), or increase certainty where low probabilities were due only to insufficient empirical data. Patterns of differentiation likely result from migration out of a single, southeastern refugium following the Last Glacial Maximum, with coalescent-based demographic modelling supporting highly isolated contemporary populations that diverged at least thousands of generations in the past (Massatti & Knowles, 2020). This history supports management by populations, as the time since divergence combined with the species' extensive geographical distribution makes population-specific evolutionary (e.g. coadapted gene complexes) and ecological (e.g. community interactions) interactions likely (Frankham et al., 2011; McKay et al., 2005).

Demonstrating POPMAPS on *H. jamesii* highlights the utility of this mapping framework versus previously developed methodologies. The ancestry probability surface simultaneously provides information regarding the geographical distribution of dominant genetically defined populations (i.e. the hard boundaries in Figure 3) and a visualization of certainty that can be modified according to the intended use of the product (i.e. the shading in Figures 3 and S2). For example, managers may desire conservative guidance when considering moving organisms to meet management goals, which can be modelled using parameters and/or a function that results in lower contributions values of empirical sites as distances increase (i.e. Figure 2). In addition, the flexibility of the interpolation surface (i.e. any geospatial surface identified to influence genetic differentiation during prior analyses) allows biologically relevant environmental factors to directly influence ancestry probabilities, as may be hypothesized according to phyllogeographical (e.g. He et al., 2013) and landscape genetic studies (e.g. Eckert et al., 2010). For example, using an SDM as the interpolation surface can result in estimated ancestry coefficients that are influenced by unsuitable habitat, as happens with mountains ranges in topographically complex areas like the western United States. While other approaches facilitate spatial interpolations of genetic patterns by estimating and mapping ancestry coefficients (Caye et al., 2016; Guillot et al., 2005), identifying important regions associated with genetic differentiation (Manni et al., 2004), or even testing among alternative landscape surfaces (Tarroso et al., 2019), none incorporates the flexibility that is built into the POPMAPS analytical pipeline.

## 4.2 | Using popmaps to guide the restoration of degraded communities

Most current seed transfer guidance matches environmental adaptations of species to restoration sites (i.e. to consider local adaptation; Baughman et al., 2019; Leimu & Fischer, 2008) without

considering species' long-term histories (e.g. since the last glacial period, as is evident in many organisms' genomes; Hewitt, 2000). At a regional scale in North America, a geographical filter is imposed by overlaying physiographic regions (e.g. Omernik & Griffith, 2014) on estimations of adaptation (e.g. Bower et al., 2014; Shryock et al., 2017; St. Clair et al., 2013) as an attempt to avoid genetic processes due to species' histories that may influence restoration outcomes (e.g. McKay et al., 2005); similar approaches are used internationally (e.g. Prasse et al., 2010). However, geographical constraints directed by physiographic regions should not be expected to reflect species' patterns of genetic differentiation (Lesica et al., 2016; Massatti et al., 2020). An alternative restoration model illustrated by virtue of POPMAPS products replaces physiographic regions with spatially interpolated populations, which are then overlaid on inferences of adaptation (in the case of *H. jamesii*, as was inferred in Massatti & Knowles, 2020). In this context, the hard boundaries inferred by POPMAPS would geographically constrain seed transfer, and environmentally delimited polygons within these boundaries would further guide seed use to consider adaptation. Similarly, genetically defined populations can guide native plant material (i.e. germplasm intended for restoration purposes; NPM) development strategies by providing clear geographical boundaries delimiting the extent over which seeds may be pooled to develop new NPMs (e.g. Bucharova et al., 2019).

Ancestry probabilities as estimated by POPMAPS allow an additional, heretofore unaccounted for dimension to guide restoration. For example, after assessing hard boundaries within a management framework, ancestry probabilities beneath a certain threshold could be used to identify areas where seeds from multiple populations may be used for restoration. In *H. jamesii*, for example, seeds sourced from the West or East population may be suitable in low probability locations along the border of Arizona and New Mexico (Figure 3). Alternatively, guidance may dictate that low probability locations are to be avoided when collecting seeds intended for NPM development, as the resulting materials would only be suitable for those same areas, if one goal of management using populations is to protect the genetic integrity and patterns of biodiversity. Given the investment required to develop one NPM (McCormick et al., 2021) and the number of new NPMs needed to support restoration (e.g. Oldfield & Olwell, 2015), managing by species' innate patterns of diversity can decrease the complexity of restoration planning and simultaneously supports successful outcomes, as there are typically fewer genetically defined populations for a species compared to the number of physiographic regions it is distributed across (e.g. Massatti et al., 2018; Massatti & Knowles, 2020).

## 4.3 | Extending popmaps to other management applications

While demonstrated in a restoration framework, POPMAPS may be used for any species where protecting biodiversity and incorporating genetic

processes are recognized management goals (Hohenlohe et al., 2021). For example, ancestry probability surfaces may benefit (a) common, wild-harvested species that require spatially explicit guidance to manage populations (Bradbury et al., 2018; Love Stowell et al., 2020); (b) rare species in need of recovery planning that require management units (Funk et al., 2012) to guide protection of unique genetic diversity and/or mitigate the effects of small populations (Frandsen et al., 2020; Hoffmann et al., 2020; Weeks et al., 2011; Whiteley et al., 2015); and (c) organisms impacted by human-modified landscapes that require assistance, such as migration corridors through heavily used landscapes (e.g. Holderegger & Di Giulio, 2010; Skarin et al., 2015), to promote their continued existence. These types of management goals may best be supported by identification of populations that are demographically independent. While determining migration rates at which populations become demographically linked is not straightforward (Coates et al., 2018; Palsbøll et al., 2007), coalescent methods combined with information-rich datasets, such as those generated by next-generation sequencing, can provide estimates of species' demographic histories (Jones et al., 2021). Continued development of methods to estimate contemporary migration rates (e.g. Al-Asadi et al., 2019) and to understand the ramifications of demographic history on genetic incompatibilities that may influence intraspecific interactions (Lucek & Willi, 2021) remain critical needs. Ultimately, the application of POPMAPS, given its innate flexibility, will depend on researchers' knowledge of their study systems and ingenuity in applying different methodologies to explore spatial patterns, given that species have unique distributions, biologies, histories and adaptations.

## 5 | CONCLUSIONS

Overlooking spatial patterns of intraspecific genetic diversity may ultimately hamper species management efforts, reinforcing the importance of transitioning to decision-making frameworks with genetically based foundations. The POPMAPS analytical framework operationalizes the inclusion of genetic diversity when generating species-specific management plans, whether they be for the workhorses of restoration meant to rebuild resilient natural communities (NASEM, 2020) or species on the edge of extinction. Furthermore, the flexibility of POPMAPS to generate ancestry probability surfaces with datasets created under a range of ecological or evolutionary assumptions highlights how the method will remain pertinent as new analytical approaches are generated.

## AUTHORS' CONTRIBUTIONS

R.M. conceived the idea, designed the methodology, developed the functions, analysed the data and led the writing of the manuscript; R.M. and D.W. edited the manuscript and produced figures. Both authors gave final approval for publication.

## ACKNOWLEDGEMENTS

We are grateful for assistance and feedback provided by Matt Jones and colleagues at the Southwest Biological Science Center.

Support from the BLM Colorado Plateau Native Plant Program, BLM Plant Conservation Program and USGS Ecosystems Mission Area made this research possible. This manuscript has been greatly improved by comments from Dr Pedro Tarroso and two anonymous reviewers. Any use of trade, product or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/2041-210X.13902>.

## DATA AVAILABILITY STATEMENT

The popmaps R package, which includes example data, is available at GitLab (<https://code.usgs.gov/GWRC/popmaps/>) and in the U.S. Geological Survey ScienceBase Repository (<https://doi.org/10.5066/P96VLOA5>; Massatti, 2022).

## ORCID

Rob Massatti  <https://orcid.org/0000-0001-5854-5597>

Daniel E. Winkler  <https://orcid.org/0000-0003-4825-9073>

## REFERENCES

- Al-Asadi, H., Petkova, D., Stephens, M., & Novembre, J. (2019). Estimating recent migration and population-size surfaces. *PLoS Genetics*, 15, e1007908.
- Alvarado-Serrano, D. F., & Knowles, L. L. (2014). Ecological niche models in phylogeographic studies: Applications, advances and precautions. *Molecular Ecology Resources*, 14, 233–248.
- Auffret, A. G., Rico, Y., Bullock, J. M., Hooftman, D. A. P., Pakeman, R. J., Soons, M. B., Suárez-Esteban, A., Traveset, A., Wagner, H. H., & Cousins, S. A. O. (2017). Plant functional connectivity - integrating landscape structure and effective dispersal. *Journal of Ecology*, 105(6), 1648–1656. <https://doi.org/10.1111/1365-2745.12742>
- Barbosa, S., Mestre, F., White, T. A., Paupério, J., Alves, P. C., & Searle, J. B. (2018). Integrative approaches to guide conservation decisions: Using genomics to define conservation units and functional corridors. *Molecular Ecology*, 27, 3452–3465.
- Baughman, O. W., Agneray, A. C., Forister, M. L., Kilkenny, F. F., Espeland, E. K., Fiegner, R., Horning, M. E., Johnson, R. C., Kaye, T. N., Ott, J., St. Clair, J. B., & Leger, E. A. (2019). Strong patterns of intraspecific variation and local adaptation in Great Basin plants revealed through a review of 75 years of experiments. *Ecology and Evolution*, 9(11), 6259–6275. <https://doi.org/10.1002/ece3.5200>
- Beninde, J., Feldmeier, S., Werner, M., Peroverde, D., Schulte, U., Hochkirch, A., & Veith, M. (2016). Cityscape genetics: Structural vs. functional connectivity of an urban lizard population. *Molecular Ecology*, 25, 4984–5000.
- Bivand, R., & Rundel, C. (2020). *rgeos: Interface to Geometry Engine - Open Source ('GEOS')*. R package version 0.5-5. <https://CRAN.R-project.org/package=rgeos>
- Bower, A. D., St. Clair, J. B., & Erickson, V. (2014). Generalized provisional seed zones for native plants. *Ecological Applications*, 24, 913–919.

- Bradbury, I. R., Wringe, B. F., Watson, B., Paterson, I., Horne, J., Beiko, R., Lehnert, S. J., Clément, M., Anderson, E. C., Jeffery, N. W., Duffy, S., Sylvester, E., Robertson, M., & Bentzen, P. (2018). Genotyping-by-sequencing of genome-wide microsatellite loci reveals fine-scale harvest composition in a coastal Atlantic salmon fishery. *Evolutionary Applications*, 11(6), 918–930. <https://doi.org/10.1111/eva.12606>
- Breed, M. F., Harrison, P. A., Blyth, C., Byrne, M., Gaget, V., Gellie, N. J. C., Groom, S. V. C., Hodgson, R., Mills, J. G., Prowse, T. A. A., Steane, D. A., & Mohr, J. J. (2019). The potential of genomics for restoring ecosystems and biodiversity. *Nature Reviews Genetics*, 20(10), 615–628. <https://doi.org/10.1038/s41576-019-0152-0>
- Brown, J. L., Bennett, J. R., & French, C. M. (2017). SDMtoolbox 2.0: The next generation Python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *PeerJ*, 5, e4095.
- Brown, J. L., & Knowles, L. L. (2012). Spatially explicit models of dynamic histories: Examination of the genetic consequences of Pleistocene glaciation and recent climate change on the American Pika. *Molecular Ecology*, 21, 3757–3775.
- Brownrigg, R., Minka, T. P., & Deckmyn, A. (2018). *maps: Draw geographical maps*. R package version 3.3.0. Retrieved from <https://CRAN.R-project.org/package=maps>
- Bucharova, A., Bossdorf, O., Hölzel, N., Kollmann, J., Prasse, R., & Durka, W. (2019). Mix and match: Regional admixture provenancing strikes a balance among different seed-sourcing strategies for ecological restoration. *Conservation Genetics*, 20, 7–17.
- Bucharova, A., Lampei, C., Conrady, M., May, E., Matheja, J., Meyer, M., & Ott, D. (2021). Plant provenance affects pollinator network: Implications for ecological restoration. *Journal of Applied Ecology*, 0, 1–11.
- Carnaval, A. C., Waltari, E., Rodrigues, M. T., Rosauer, D., VanDerWal, J., Damasceno, R., Prates, I., Strangas, M., Spanos, Z., Rivera, D., Pie, M. R., Firkowski, C. R., Bornschein, M. R., Ribeiro, L. F., & Moritz, C. (2014). Prediction of phylogeographic endemism in an environmentally complex biome. *Proceedings of the Royal Society B: Biological Sciences*, 281(1792), 20141461. <https://doi.org/10.1098/rspb.2014.1461>
- Carvalho, C. S., Forester, B. R., Mitre, S. K., Alves, R., Imperatriz-Fonseca, V. L., Ramos, S. J., Resende-Moreira, L. C., Siqueira, J. O., Trevelin, L. C., Caldeira, C. F., Gastauer, M., & Jaffé, R. (2021). Combining genotype, phenotype, and environmental data to delineate site-adjusted provenance strategies for ecological restoration. *Molecular Ecology Resources*, 21(1), 44–58. <https://doi.org/10.1111/1755-0998.13191>
- Castañeda-Álvarez, N. P., Khoury, C. K., Achicanoy, H. A., Bernau, V., Dempewolf, H., Eastwood, R. J., Guarino, L., Harker, R. H., Jarvis, A., Maxted, N., Müller, J. V., Ramirez-Villegas, J., Sosa, C. C., Struik, P. C., Vincent, H., & Toll, J. (2016). Global conservation priorities for crop wild relatives. *Nature Plants*, 2(4). <https://doi.org/10.1038/nplants.2016.22>
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22, 3124–3140.
- Caye, K., Deist, T. M., Martins, H., Michel, O., & François, O. (2016). TESS3: Fast inference of spatial population structure and genome scans for selection. *Molecular Ecology Resources*, 16(2), 540–548.
- Chazdon, R. L., Brancalion, P. H., Lamb, D., Laestadius, L., Calmon, M., & Kumar, C. (2017). A policy-driven knowledge agenda for global forest and landscape restoration. *Conservation Letters*, 10, 125–132.
- Clarke, R. T., Rothery, P., & Raybould, A. F. (2002). Confidence limits for regression relationships between distance matrices: Estimating gene flow with distance. *Journal of Agricultural, Biological, and Environmental Statistics*, 7(3), 361–372.
- Coates, D. J., Byrne, M., & Moritz, C. (2018). Genetic diversity and conservation units: Dealing with the species-population continuum in the age of genomics. *Frontiers in Ecology and Evolution*, 6, 1–13.
- Convention on Biological Diversity (CBD). (2010). Strategic plan for biodiversity 2011–2020 and the Aichi Targets. *Secretariat of the convention on biological diversity*. Retrieved from <https://www.cbd.int/sp/>
- Duminil, J., Fineschi, S., Hampe, A., Jordano, P., Salvini, D., Vendramin, G. G., & Petit, R. J. (2007). Can population genetic structure be predicted from life-history traits? *The American Naturalist*, 169(5), 662–672.
- Eckert, A. J., Van Heerwaarden, J., Wegrzyn, J. L., Nelson, C. D., Ross-Ibarra, J., González-Martínez, S. C., & Neale, D. B. (2010). Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics*, 185(3), 969–982.
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software Structure: A simulation study. *Molecular Ecology*, 14, 2611–2620.
- Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, 164, 1567–1587.
- Ferrier, S., Manion, G., Elith, J., & Richardson, K. (2007). Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions*, 13(3), 252–264.
- Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37, 4302–4315.
- Fitzpatrick, M. C., & Keller, S. R. (2015). Ecological genomics meets community-level modelling of biodiversity: Mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters*, 18, 1–16.
- Frandsen, P., Fontseré, C., Nielsen, S. V., Hanghøj, K., Castejon-Fernandez, N., Lizano, E., Hughes, D., Hernandez-Rodriguez, J., Korneliussen, T. S., Carlsen, F., Siegmund, H. R., Mailund, T., Marques-Bonet, T., & Hvilsom, C. (2020). Targeted conservation genetics of the endangered chimpanzee. *Heredity*, 125(1–2), 15–27. <https://doi.org/10.1038/s41437-020-0313-0>
- Frankham, R., Ballou, J. D., Eldridge, M. D., Lacy, R. C., Ralls, K., Dudash, M. R., & Fenster, C. B. (2011). Predicting the probability of outbreeding depression. *Conservation Biology*, 25(3), 465–475.
- Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, 27, 489–496.
- Garnier, S., Ross, N., Rudis, B., Sciaini, M., & Scherer, C. (2018). *viridis: Default Color Maps from 'matplotlib'*. R package version 0.5.1, 2018.
- Gompert, Z., Lucas, L. K., Nice, C. C., Fordyce, J. A., Forister, M. L., & Buerkle, C. A. (2012). Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution: International Journal of Organic Evolution*, 66, 2167–2181.
- Gould, B. A., & Stinchcombe, J. R. (2017). Population genomic scans suggest novel genes underlie convergent flowering time evolution in the introduced range of *Arabidopsis thaliana*. *Molecular Ecology*, 26, 92–106.
- Graves, T., Chandler, R. B., Royle, J. A., Beier, P., & Kendall, K. C. (2014). Estimating landscape resistance to dispersal. *Landscape Ecology*, 29, 1201–1211.
- Guillot, G., Mortier, F., & Estoup, A. (2005). GENELAND: A computer package for landscape genetics. *Molecular Ecology Notes*, 5(3), 712–715.
- Hanson, J. O., Marques, A., Veríssimo, A., Camacho-Sanchez, M., Velo-Antón, G., Martínez-Solano, Í., & Carvalho, S. B. (2020). Conservation planning for adaptive and neutral evolutionary processes. *Journal of Applied Ecology*, 57, 2159–2169.
- Hao, T., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2020). Testing whether ensemble modelling is advantageous for maximising



- predictive performance of species distribution models. *Ecography*, 43(4), 549–558.
- He, Q., Edwards, D. L., & Knowles, L. L. (2013). Integrative testing of how environments from the past to the present shape genetic structure across landscapes. *Evolution*, 67(12), 3386–3402.
- Hendricks, S., Epstein, B., Schönfeld, B., Wiench, C., Hamede, R., Jones, M., Storfer, A., & Hohenlohe, P. (2017). Conservation implications of limited genetic diversity and population structure in Tasmanian devils (*Sarcophilus harrisii*). *Conservation Genetics*, 18(4), 977–982. <https://doi.org/10.1007/s10592-017-0939-5>
- Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, 405(6789), 907–913.
- Hijmans, R. J., & van Etten, J. (2012). *raster: Geographic analysis and modeling with raster data*. R package version 2.0-12.
- Hoban, S., Bruford, M., Jackson, J. D. U., Lopes-Fernandes, M., Heuertz, M., Hohenlohe, P. A., ... Laikre, L. (2020). Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. *Biological Conservation*, 248, 108654.
- Hoffmann, A. A., Miller, A. D., & Weeks, A. R. (2020). Genetic mixing for population management: From genetic rescue to provenancing. *Evolutionary Applications*, 14, 634–652.
- Hohenlohe, P. A., Funk, W. C., & Rajora, O. P. (2021). Population genomics for wildlife conservation and management. *Molecular Ecology*, 30, 62–82.
- Holderegger, R., & Di Giulio, M. (2010). The genetic effects of roads: A review of empirical evidence. *Basic and Applied Ecology*, 11, 522–531.
- Hufford, K. M., Krauss, S. L., & Veneklaas, E. J. (2012). Inbreeding and outbreeding depression in *Stylidium hispidum*: Implications for mixing seed sources for ecological restoration. *Ecology and Evolution*, 2, 2262–2273.
- Jones, M. R., Winkler, D. E., & Massatti, R. (2021). Demographic modeling informs functional connectivity and management interventions in Graham's beardtongue. *Conservation Genetics*, 22, 993–1003.
- Jones, T. A. (2003). The restoration gene pool concept: Beyond the native versus non-native debate. *Restoration Ecology*, 11, 281–290.
- Knowles, L. L., & Massatti, R. (2017). Distributional shifts—not geographic isolation—as a probable driver of montane species divergence. *Ecography*, 40(12), 1475–1485.
- Leimu, R., & Fischer, M. (2008). A meta-analysis of local adaptation in plants. *PLoS ONE*, 3, e4010.
- Lemon, J. (2006). Plotrix: A package in the red light district of R. *R-news*, 6, 8–12.
- Lesica, P., Adams, B., & Smith, C. T. (2016). Can physiographic regions substitute for genetically-determined conservation units? A case study with the threatened plant, *Silene spaldingii*. *Conservation Genetics*, 17, 1041–1054.
- Love Stowell, S. M., Gagne, R. B., McWhirter, D., Edwards, W., & Ernest, H. B. (2020). Bighorn sheep genetic structure in Wyoming reflects geography and management. *The Journal of Wildlife Management*, 84, 1072–1090.
- Lucek, K., & Willi, Y. (2021). Drivers of linkage disequilibrium across a species' geographic range. *PLoS Genetics*, 17, e1009477.
- Manni, F., Guérard, E., & Heyer, E. (2004). Geographic patterns of (genetic, morphologic, linguistic) variation: How barriers can be detected by using Monmonier's algorithm. *Human Biology*, 76, 173–190.
- Massatti, R. (2020). *Hilaria jamesii* data for the Colorado Plateau of the southwestern United States – U.S. Geological Survey data release. <https://doi.org/10.5066/P9CNFWOX>
- Massatti, R. (2022). Data from: POPMAPS: An R package to estimate ancestry probability surfaces: U.S. Geological Survey Software Release, <https://doi.org/10.5066/P96VLOA5>
- Massatti, R., & Knowles, L. L. (2020). The historical context of contemporary climatic adaptation: A case study in the climatically dynamic and environmentally complex southwestern United States. *Ecography*, 43, 735–746.
- Massatti, R., Prendeville, H. R., Larson, S., Richardson, B. A., Waldron, B., & Kilkenny, F. F. (2018). Population history provides foundational knowledge for utilizing and developing native plant restoration materials. *Evolutionary Applications*, 11, 2025–2039.
- Massatti, R., Shriver, R. K., Winkler, D. E., Richardson, B. A., & Bradford, J. B. (2020). Assessment of population genetics and climatic variability can refine climate-informed seed transfer guidelines. *Restoration Ecology*, 28, 485–493.
- McCormick, M. L., Carr, A. N., Massatti, R., Winkler, D. E., De Angelis, P., & Olwell, P. (2021). How to increase the supply of native seed to improve restoration success: The US native seed development process. *Restoration Ecology*, 29(8). <https://doi.org/10.1111/rec.13499>
- McKay, J. K., Christian, C. E., Harrison, S., & Rice, K. J. (2005). 'How local is local?' – A review of practical and conceptual issues in the genetics of restoration. *Restoration Ecology*, 13, 432–440.
- McRae, B. H. (2006). Isolation by resistance. *Evolution*, 60, 1551–1561.
- Microsoft Corporation, & Weston, S. (2020a). *doParallel: Foreach parallel adaptor for the 'parallel' package*. R package version 1.0.16. Retrieved from <https://CRAN.R-project.org/package=doParallel>
- Microsoft Corporation, & Weston, S. (2020b). *foreach: Provides foreach looping construct*. R package version 1.5.1. Retrieved from <https://CRAN.R-project.org/package=foreach>
- Moraes, M. A., Kubota, T. Y. K., Rossini, B. C., Marino, C. L., Freitas, M. L. M., Moraes, M. L. T., Silva, A. M., Cambuim, J., & Sebbenn, A. M. (2018). Long-distance pollen and seed dispersal and inbreeding depression in *Hymenaea stigonocarpa* (Fabaceae: Caesalpinioideae) in the Brazilian savannah. *Ecology and Evolution*, 8(16), 7800–7816. <https://doi.org/10.1002/ece3.4253>
- Moritz, C. (1999). Conservation units and translocations: Strategies for conserving evolutionary processes. *Hereditas*, 130, 217–228.
- Morris, A. B., Graham, C. H., Soltis, D. E., & Soltis, P. S. (2010). Reassessment of phylogeographical structure in an eastern North American tree using Monmonier's algorithm and ecological niche modelling. *Journal of Biogeography*, 37(9), 1657–1667.
- Murphy, M., Evans, J., Cushman, S., & Storfer, A. (2008). Evaluation of a novel approach for representing 'populations' as continuous surfaces in landscape genetics. *Ecography*, 31, 685–697.
- National Academies of Sciences, Engineering, and Medicine (NASEM). (2020). *An assessment of the need for native seeds and the capacity for their supply: Interim report*. The National Academies Press.
- Oldfield, S., & Olwell, P. (2015). The right seed in the right place at the right time. *Bioscience*, 65, 955–956.
- Omerik, J. M., & Griffith, G. E. (2014). Ecoregions of the conterminous United States: Evolution of a hierarchical spatial framework. *Environmental Management*, 54, 1249–1266.
- Ortego, J., & Knowles, L. L. (2020). Incorporating interspecific interactions into phylogeographic models: A case study with Californian oaks. *Molecular Ecology*, 29(23), 4510–4524.
- Palsbøll, P. J., Berube, M., & Allendorf, F. W. (2007). Identification of management units using population genetic data. *Trends in Ecology & Evolution*, 22, 11–16.
- Pebesma, E., & Bivand, R. S. (2005). S classes and methods for spatial data: The sp package. *R News*, 5, 9–13.
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, 7, e37135.
- Phillips, S. J., Dudík, M., & Schapire, R. E. (2021). *Maxent software for modeling species niches and distributions (Version 3.4.1)*. Retrieved from [http://biodiversityinformatics.amnh.org/open\\_source/maxent/](http://biodiversityinformatics.amnh.org/open_source/maxent/)
- Prasse, R., Kunzmann, D., & Schröder, R. (2010). Development and practical implementation of minimal requirements for the verification of origin of native seeds of herbaceous plants (in German). In Cooperation with Verband Deutscher Wildsamen- und Wildpflanzenproduzenten. DBU, reference no. 23931.



- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.
- R Core Team. (2021). *R: A language and environment for statistical computing* v.4.0.4. R Foundation for Statistical Computing.
- Rathmacher, G., Niggemann, M., Köhnen, M., Ziegenhagen, B., & Bialozyt, R. (2010). Short-distance gene flow in *Populus nigra* L. accounts for small-scale spatial genetic structures: Implications for in situ conservation measures. *Conservation Genetics*, 11, 1327–1338.
- Rosauer, D. F., Ferrier, S., Williams, K. J., Manion, G., Keogh, J. S., & Laffan, S. W. (2014). Phylogenetic generalised dissimilarity modelling: A new approach to analysing and predicting spatial turnover in the phylogenetic composition of communities. *Ecography*, 37(1), 21–32.
- Sgrò, C. M., Lowe, A. J., & Hoffmann, A. A. (2011). Building evolutionary resilience for conserving biodiversity under climate change. *Evolutionary Applications*, 4, 326–337.
- Shryock, D. F., Havrilla, C. A., DeFalco, L. A., Esque, T. C., Custer, N. A., & Wood, T. E. (2017). Landscape genetic approaches to guide native plant restoration in the Mojave Desert. *Ecological Applications*, 27, 429–445.
- Skarin, A., Nellemann, C., Rönnegård, L., Sandström, P., & Lundqvist, H. (2015). Wind farm construction impacts reindeer migration and movement corridors. *Landscape Ecology*, 30, 1527–1540.
- St. Clair, B. J., Kilkenny, F. F., Johnson, R. C., Shaw, N. L., & Weaver, G. (2013). Genetic variation in adaptive traits and seed transfer zones for *Pseudoroegneria spicata* (bluebunch wheatgrass) in the north-western United States. *Evolutionary Applications*, 6, 933–948.
- Stanton, D. W. G., Frandsen, P., Waples, R. K., Heller, R., Russo, I. M., Orozco-terWengel, P. A., Pedersen, C. T., Siegmund, H. R., & Bruford, M. W. (2019). More grist for the mill? Species delimitation in the genomic era and its implications for conservation. *Conservation Genetics*, 20(1), 101–113. <https://doi.org/10.1007/s10592-019-01149-5>
- Storfer, A., Murphy, M. A., Spear, S. F., Holderegger, R., & Waits, L. P. (2010). Landscape genetics: Where are we now? *Molecular Ecology*, 19(17), 3496–3514.
- Tarrosso, P., Carvalho, S. B., & Velo-Antón, G. (2019). Phylin 2.0: Extending the phylogeographical interpolation method to include uncertainty and user-defined distance metrics. *Molecular Ecology Resources*, 19, 1081–1094.
- Trask, A. E., Ferrie, G. M., Wang, J., Newland, S., Canessa, S., Moehrensclager, A., Laut, M., Duenas, L. B., & Ewen, J. G. (2021). Multiple life-stage inbreeding depression impacts demography and extinction risk in an extinct-in-the-wild species. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-020-79979-4>
- United Nations General Assembly (UN). (2019). Resolution adopted by the General Assembly on 1 March 2019: 73/284. United Nations Decade on Ecosystem Restoration (2021–2030). Retrieved from [undocs.org/A/RES/73/284](https://undocs.org/A/RES/73/284)
- van Etten, J. (2017). R package gdistance: Distances and routes on geographical grids. *Journal of Statistical Software*, 76, 1–21.
- Vandeghechuchte, M. L., De La Peña, E., & Bonte, D. (2011). Contrasting covariation of above- and belowground invertebrate species across plant genotypes. *Journal of Animal Ecology*, 80, 148–158.
- Vandergast, A. G., Perry, W. M., Lugo, R. V., & Hathaway, S. A. (2011). Genetic landscapes GIS Toolbox: Tools to map patterns of genetic divergence and diversity. *Molecular Ecology Resources*, 11, 158–161.
- Varshney, R. K., Thudi, M., Roorkiwal, M., He, W., Upadhyaya, H. D., Yang, W., ... Liu, X. (2019). Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. *Nature Genetics*, 51, 857–864.
- Verdone, M., & Seidl, A. (2017). Time, space, place, and the Bonn Challenge global forest restoration target. *Restoration Ecology*, 25, 903–911.
- Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., & Venables, B. (2020). *gplots: Various R programming tools for plotting data*. R package version 3.1.1. Retrieved from <https://CRAN.R-project.org/package=gplots>
- Weeks, A. R., Sgro, C. M., Young, A. G., Frankham, R., Mitchell, N. J., Miller, K. A., Byrne, M., Coates, D. J., Eldridge, M. D. B., Sunnucks, P., Breed, M. F., James, E. A., & Hoffmann, A. A. (2011). Assessing the benefits and risks of translocations in changing environments: A genetic perspective. *Evolutionary Applications*, 4(6), 709–725. <https://doi.org/10.1111/j.1752-4571.2011.00192.x>
- Whiteley, A. R., Fitzpatrick, S. W., Funk, W. C., & Tallmon, D. A. (2015). Genetic rescue to the rescue. *Trends in Ecology & Evolution*, 30, 42–49.
- Whitham, T. G., Bailey, J. K., Schweitzer, J. A., Shuster, S. M., Bangert, R. K., LeRoy, C. J., Lonsdorf, E. V., Allan, G. J., DiFazio, S. P., Potts, B. M., Fischer, D. G., Gehring, C. A., Lindroth, R. L., Marks, J. C., Hart, S. C., Wimp, G. M., & Wooley, S. C. (2006). A framework for community and ecosystem genetics: From genes to ecosystems. *Nature Reviews Genetics*, 7(7), 510–523. <https://doi.org/10.1038/nrg1877>
- Winkler, D. E., Chapin, K. J., François, O., Garmon, J. D., Gaut, B. S., & Huxman, T. E. (2019). Multiple introductions and population structure during the rapid expansion of the invasive Sahara mustard (*Brassica tournefortii*). *Ecology and Evolution*, 9(14), 7928–7941.
- Winkler, D. E., & Massatti, R. (2020). Unexpected hybridization reveals the utility of genetics in native plant restoration. *Restoration Ecology*, 28, 1047–1052.
- Worth, J. R. P., Tamaki, I., Tsuyama, I., Harrison, P. A., Sugai, K., Sakio, H., Aizawa, M., & Kikuchi, S. (2021). Genetic distinctiveness but low diversity characterizes rear-edge *Thuja standishii* (Gordon) Carr. (Cupressaceae) populations in southwest Japan. *Diversity*, 13(5), 185. <https://doi.org/10.3390/d13050185>
- Zhang, H., Mittal, N., Leamy, L. J., Barazani, O., & Song, B. H. (2017). Back into the wild—Apply untapped genetic diversity of wild relatives for crop improvement. *Evolutionary Applications*, 10, 5–24.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Massatti, R., & Winkler, D. E. (2022). Spatially explicit management of genetic diversity using ancestry probability surfaces. *Methods in Ecology and Evolution*, 00, 1–14. <https://doi.org/10.1111/2041-210X.13902>