# Individual-based landscape genomics for conservation: an analysis pipeline

**Running header:** Individual-based landscape genomics

E. Anne Chambers[1,2*†] | Anusha P. Bishop[1,2†] | Ian J. Wang[1,2]

[1]*Department of Environmental Science, Policy, and Management, University of California Berkeley, Berkeley, CA 94720, USA*

[2]*Museum of Vertebrate Zoology, University of California Berkeley, Berkeley, CA 94720, USA*

**Correspondence to be sent to: Department of Environmental Science, Policy, and Management, University of California Berkeley, Berkeley, CA 94720, USA; E-mail: eachambers@berkeley.edu*

*† E. Anne Chambers and Anusha P. Bishop should be considered joint first authors.*

## Abstract

Landscape genomics can harness environmental and genetic data to inform conservation decisions by providing essential insights into how landscapes shape biodiversity. The massive increase in genetic data afforded by the genomic era provides exceptional resolution for answering critical conservation genetics questions. The accessibility of genomic data for non-model systems has also enabled a shift away from population-based sampling to individual-based sampling, which now provides accurate and robust estimates of genetic variation that can be used to examine the spatial structure of genomic diversity, population connectivity, and the nature of environmental adaptation. Nevertheless, the adoption of individual-based sampling in conservation genetics has been slowed due, in large part, to concerns over how to apply methods developed for population-based sampling to individual-based sampling schemes. Here, we discuss the benefits of individual-based sampling for conservation and describe how landscape genomic methods, paired with individual-based sampling, can answer fundamental conservation questions. We have curated key landscape genomic methods into a user-friendly, open-source workflow, which we provide as a new R package, A Landscape Genomics Analysis Toolkit in R (ALGATR). The ALGATR package includes added functionality for all of the included methods and extensive vignettes designed with the primary goal of making landscape genomic approaches more accessible and explicitly applicable to conservation biology.



**KEYWORDS:** Conservation Genetics, Conservation Biology, Landscape Genomics, Spatial Analysis, Population Structure, Genetic Diversity

# 1 | INTRODUCTION

The Anthropocene is an era of unprecedented global change, making the question of how best to conserve biodiversity more important than ever. Although broad conservation policies are typically enacted at the species level, conservation actions and management decisions regularly take place at the landscape level (Fiedler et al., 2022). Protecting species diversity and the processes that maintain it ultimately requires conserving the geographic regions that support the ecological and evolutionary processes intrinsic to population viability (Malcom & Carter, 2021). Particularly with accelerating environmental change, land-use conversion, and habitat loss and degradation, biological conservation has increasingly become an inherently spatial problem (Shaffer et al. 2022). Uniting landscape ecology and population genetics in a single framework, landscape genetics provides a suite of spatially-explicit approaches for addressing this challenge (Balkenhol et al., 2015; Keller et al., 2015; Manel et al., 2003, 2010; Segelbacher et al., 2010; Shaffer et al., 2022; Storfer et al., 2007; van Strien et al., 2014). Spatial approaches in landscape genetics involve the integration of geographic, environmental, and genetic data to understand how spatial patterns of genetic variation are influenced by landscape features, environmental factors, and dispersal barriers. Broadly, these include methods to delimit population boundaries, quantify the drivers of genetic differentiation, estimate landscape resistance to gene flow, characterize patterns of genetic diversity, and identify genes involved in adaptation to local environmental variation (Balkenhol et al., 2015; Storfer et al., 2018; Wagner & Fortin, 2013).

Genomic data provide valuable information for identifying population boundaries, demographics, and connectivity, all of which are important for conservation (Funk et al., 2012; Hohenlohe et al., 2021; Keller et al., 2015; Manel & Holderegger, 2013; Segelbacher et al., 2010). Genetic diversity itself is increasingly a target of conservation action especially because

of the role it can play in mitigating the impacts of ongoing climate change (Hoban et al., 2020; Schmidt et al., 2023). Prior to the availability and accessibility of genomic data for non-model systems, landscape genetics studies relied on population-based sampling, in which many individuals from each location were genotyped for a handful of genetic markers. Now that we can sequence many thousands of loci from across the genome – allowing us to capture more of the complexity that exists in natural systems (Forester & Lama, 2022; Holliday et al., 2017; Shafer et al., 2015; Supple & Shapiro, 2018) – landscape genetics has increasingly shifted towards individual-based sampling. With genomic data, just a single individual per location can provide robust inferences about intraspecific genetic variation (Hohenlohe et al., 2021; Selmoni et al., 2020). This is advantageous for conservation efforts as well, as individual-based sampling allows for broader geographic and environmental coverage, provides greater spatial resolution, and minimizes the overall impact of sampling on each population (Shaffer et al., 2022; Wang & Bradburd, 2014).

Despite the benefits of individual-based sampling, a broader shift to these sampling schemes has been slowed, in large part, by methodological concerns. Landscape genetics is replete with analytical approaches, but the very breadth of those choices, and the many decisions to be made in implementing each method present a certain challenge, particularly for those seeking entry into landscape genetics but even for experienced practitioners as well. Moreover, because many landscape and population genetic methods were originally designed for population-based sampling, their validity for individual-based sampling has sometimes come into question, further complicating the question of how best to implement a landscape genomic framework for conservation. Here, we present a pipeline for performing cutting-edge landscape genomic analyses with individual-based sampling, discuss the key conservation-oriented

questions it can answer, and detail the considerations for each method it includes (Table 1). Our pipeline is implemented as an R package, A Landscape Genomics Analysis Toolkit in R (ALGATR), that is publicly available on GitHub (https://github.com/TheWangLab/algatr). ALGATR includes detailed walkthroughs and documentation to make it easily accessible to anyone eager to use landscape genomics to achieve actionable conservation impacts.

## 2 | INDIVIDUAL-BASED LANDSCAPE GENOMICS FOR CONSERVATION

Having many independent loci is advantageous because each serves as an independent instantiation of the coalescent process (Leitwein et al., 2020), so sequencing large numbers of loci can provide robust estimates of genetic differentiation between localities, even when the sample size at each locus is small. For this reason, the number of loci that can be captured by next-generation sequencing makes individual-based sampling tenable for landscape genetics. Large genomic datasets provide strong statistical power for inferring spatial patterns of genetic variation, including the detection of genotype-environment associations (GEA), geographic barriers and corridors, and populations or genetic clusters without *a priori* assignments (Manel et al., 2003; Patterson et al., 2006). Individual-based, genomic-scale datasets also contribute distinct advantages for practical conservation biology – in particular, distributing sampling across more sites provides greater geographic and environmental coverage, better spatial resolution, and lower impact on natural populations.

### 2.1 | Greater geographic and environmental coverage

For the same total sample size, individual-based sampling schemes are able to include more localities than population-based schemes, providing greater geographic coverage and

environmental breadth. Broadening the spatial extent of a project captures greater landscape heterogeneity and more of the genetic variation across a species' range. It also allows projects to cover more areas that may be of interest for conservation and land management efforts, which is valuable for evaluating currently designated protected areas, for preserving genetic diversity and population connectivity, and for assessing the potential contributions of new management areas.

Expanding coverage over environmental space provides increased power to estimate response curves that capture the relationships between a species and its environment (Shaffer et al., 2022; Wang & Bradburd, 2014), including the detection of ecologically important patterns of genetic variation and the environmental drivers of that variation (Manel et al., 2012; Selmoni et al., 2020). This increases the likelihood of capturing alleles involved in local adaptation, particularly if sampling includes environments near the edge of species' tolerance limits (Lotterhos & Whitlock, 2014; Rellstab et al., 2015; Storfer et al., 2018; Stucki et al., 2017). These alleles may play important roles in resilience to future climate change, making them important targets for conservation action. Even analyses that identify environmental drivers of neutral genetic variation can contribute important information on which environmental axes are important for maintaining adaptive potential and how environmental variation influences genetic structure and population connectivity.

## 2.2 | Greater spatial resolution

Individual-based sampling schemes also allow for greater sampling density, relative to population-based schemes, providing finer spatial resolution for inferences of population structure, gene flow, and the distribution of genetic diversity (Balkenhol & Fortin, 2015; Manel et al., 2003; Shaffer et al., 2022). By reducing the gaps between sampling localities, individual-based sampling designs allow for more accurate identification of where population genetic

elements, like genetic breaks or corridors of connectivity, occur on the landscape. By doing so, we can also understand how population genetic elements intersect with landscape features, like habitat transitions and potential physical barriers to gene flow. Homing in on where population boundaries, corridors of connectivity, and regions supporting greater genetic diversity are found can assist conservation practitioners and land managers in determining which areas are most valuable for designating critical habitat and which contribute most to enabling population connectivity and persistence. This is especially important because many land management decisions take place over relatively small areas and prioritizing conservation efforts requires understanding the precise relationships between landscape elements and biodiversity.

## 2.3 | Lower impact on populations

Many species are of conservation concern because they have small or heavily fragmented populations, which puts them at greater risk from population fluctuations or disturbances. In trying to understand which species or populations to prioritize for conservation management, researchers must strike a balance between collecting adequate sample sizes to obtain crucial information while limiting the overall impacts of sampling on population health. In the most extreme cases, population-based sampling may not even be feasible for rare or highly imperiled species because finding sufficient numbers of individuals may prove prohibitively difficult (Supple & Shapiro, 2018). By sampling only one, or a few, individuals from each locality, individual-based sampling minimizes the impact of sampling efforts on natural populations, conveying clear benefits for population health and sustainability and even enabling studies of some highly imperiled species that would otherwise be impossible. Sampling single individuals can also reduce the effect of repeated sampling when species management plans call for repeated genetic monitoring over time.

## 2.4 | Potential downsides of individual-based sampling

Because many population genetic analysis methods were developed for a traditional population-based framework, their application to individual-based sampling can present certain challenges. At a minimum, knowing which population genetic methods can be applied to individual-based sampling requires understanding the assumptions inherent to each method, such as whether they assume normal distributions of allele frequencies per locality or whether they use population designation as an informative prior. Some methods can also be sensitive to the metrics used as inputs, and several common metrics of genetic distance and diversity cannot be calculated on an individual basis (e.g., *F*-statistics). However, there are reliable metrics that can be calculated from individual-based sampling that should not bias results (e.g., Shirk et al., 2017), several of which can be calculated in our pipeline (ALGATR).

Sampling fewer individuals per location also increases the binomial sampling noise around estimates of local allele frequencies (Wang & Bradburd, 2014). This is consequential for some methods that require precise allele frequency estimates. For example, genotype-environment association (GEA) methods were developed for population-based sampling and may not perform as well with allele frequency estimates based on only a single individual per a locality (Rellstab et al., 2015). Although not exclusive to individual-based sampling, researchers should also be cognizant of how the spatial distribution of their sampling may affect spatial autocorrelation in their sampling design. Reflecting a common concern in spatial ecological studies, when data points have a high degree of non-independence, analyses that do not include an autoregressive component may produce inaccurate results (Dale & Fortin, 2014; Hurlbert, 1984).

Finally, individual-based sampling inherently limits the information acquired about the variation present within a population. This represents a reasonable tradeoff for landscape genetics, which is primarily concerned with between-population variation, and even for conservation efforts that are concerned with landscape-level processes, but it does present a limitation for some objectives, such as those concerned with demographic processes. For example, individual-based sampling can only provide coarse data on population sizes and would likely not be an optimal choice for genetic monitoring of within-population changes through time. Despite these limitations, the set of methods that have assumptions compatible with individual-based sampling leaves researchers with many strong analytical options.

## 3 | A LANDSCAPE GENOMICS PIPELINE FOR CONSERVATION

Researchers must choose from a wealth of methodological options when designing a landscape genomic study. The best path will depend on a complex set of factors, ranging from the goals and questions of the study to the assumptions and requirements for each potential method to be employed to computational costs and limitations that balance available resources and methods with dataset size. Important methodological considerations in landscape genomics begin with the broad consideration of which type of analysis to include, followed by deciding which method to perform, and then making decisions on parameter settings and options for that specific method. In addition to methodological considerations, input data must often be processed (e.g., calculating genetic or environmental distances). Understanding and prioritizing tradeoffs can be challenging even for experienced landscape geneticists and especially for conservation practitioners or other biologists who may be approaching landscape genetics for the

first time. We aim to address these issues by providing a user-friendly, accessible workflow which is implemented in the ALGATR pipeline.

The ALGATR pipeline includes a curated set of methods that can be applied to individual-based sampling, providing robust results under most realistic scenarios, and covering the core areas of landscape genomics that are of particular interest for conservation. To make landscape genomics methodology more approachable, we have included transparent documentation in ALGATR for each method, including vignettes that provide step-by-step guidance for processing input data, running analyses, producing relevant summary statistics, interpreting results, and generating figures. Although ALGATR makes use of (for the most part) existing packages, we have added new functionality to each of these methods and adapted, when necessary, methods to accommodate individual-based sampling (Table S1).

Two of the main barriers to performing any analysis are ensuring input files are in the correct format and determining the proper parameter settings for respective landscape genomic methods; ALGATR provides functions for formatting input files and testing parameters for each method included in the pipeline. Performing any of the methods in ALGATR requires only a file with variant calls (i.e., a VCF file) and sampling coordinates. ALGATR also includes a set of utility functions that provide users with options for customization if they so desire, including functions for downloading bioclimatic data, pruning SNPs based on linkage disequilibrium, imputing missing genotypes, and calculating a variety of genetic and geographic distance metrics (Table S1). We provide guidance on the application of these utility functions so that users can visualize and evaluate the effects of different ways of processing their data.

Below, we outline four key questions for conservation that landscape genomic approaches can answer and describe methods for doing so that are included in the ALGATR

pipeline (Table 1). For each method, we describe its key components, practical considerations, and relevance to conservation questions.

## 3.1 | How do we delineate population units for management?

Species management plans frequently call for the designation of evolutionarily significant units, populations, or genetic groups that contain some particular value to the species as a whole (Allendorf et al., 2022; Turbek et al., 2023). For example, some conservation plans target the protection of populations that harbor unique genetic variants, subspecies that exhibit geographic variation (e.g., Teixeira & Huber, 2021), or groups that maximize adaptive potential (Funk et al., 2019). A critical first step in delimiting conservation units is inferring genetic clusters based on population structure or genetic similarity.

Identifying genetic clusters is straightforward when individuals are structured into discrete, isolated populations, but many species instead exhibit continuous distributions of individuals (Bradburd & Ralph, 2019; Hohenlohe et al., 2021; Manel et al., 2003). Methods to infer genetic clusters, including approaches like ADMIXTURE (Alexander et al., 2009), STRUCTURE (Pritchard et al., 2000) sNMF (Frichot et al., 2014), TESS (Caye et al., 2016), and conStruct (Bradburd et al., 2018), use clustering techniques (e.g., *K*-means clustering) to assign individuals to genetic clusters based on how neutral genetic variation is distributed. The approaches that are most apt for landscape genetics are those that are spatially explicit, often using sample location as prior information. The ALGATR pipeline implements one such method, TESS (Caye et al., 2016), which infers the optimal numbers of clusters and then calculates an assignment probability or ancestry coefficient of each individual to each cluster. We have added functions to ALGATR to generate interpolated maps of ancestry coefficients using the *autoKrige* function from the AUTOMAP package (Hiemstra et al., 2009), which differs from the interpolation

done by default in TESS in that it produces raster maps that can then be used in downstream analyses (Figure 1).

## 3.2 | How is genetic variation distributed?

Genetic diversity provides the raw material on which selection can act, allowing populations or species to respond to new threats or perturbations, ranging from novel pests and pathogens to ongoing climate change (Hoffmann et al., 2017). Loss of genetic diversity can lead to inbreeding depression, diminished adaptive potential, and increased extinction risk, making it a key target for conservation action (Kardos et al., 2021; O'Grady et al., 2004). Understanding how genetic diversity is distributed across the landscape is, therefore, critical for examining the drivers of genetic variation, identifying areas that harbor greater genetic diversity, and quantifying regional vulnerability to environmental change (Segelbacher et al., 2010; Sommer et al., 2013).

One challenge for individual-based landscape genomics has been how to calculate metrics of genetic diversity, which are traditionally calculated at the population level, when samples are distributed across the landscape without discretely-bounded clusters. ALGATR uses a recently developed method for mapping local estimates of genetic diversity based on continuous individual samples implemented in the R package WINGEN (Bishop et al., 2023). WINGEN generates continuous maps of heterozygosity, allelic richness, and nucleotide diversity using moving windows and spatial interpolation (Figure 1). WINGEN also includes options to account for any unevenness in sampling design using rarefaction and allows users to customize the size of the moving window and the resolution of the output maps, which are provided as rasters that can be imported for further analysis in any commonly-used GIS software. These rasters thus

provide detailed views of genomic diversity across the landscape that can inform land management strategies aimed at protecting genetic diversity.

## 3.3 | What are the drivers of population connectivity?

One of the primary goals of systematic conservation planning is to ensure that protected areas promote population viability and persistence, and key to this is identifying how and where critical population processes occur on the landscape (Nielsen et al., 2022; Segelbacher et al., 2010). Understanding the drivers of gene flow and genetic differentiation can help conservation practitioners understand the impact of habitat fragmentation, land-use conversion, and environmental change on population connectivity (Hall & Beissinger, 2014; van Strien et al., 2014). The population connectivity between points on a landscape can diminish when gene flow is reduced over greater geographic distances, resulting in a pattern of isolation-by-distance (IBD; Wright, 1943), or when gene flow is restricted between regions with increasingly different environmental conditions (due to selective or non-selective mechanisms), leading to a pattern of isolation-by-environment (IBE; Wang & Bradburd, 2014). The extent to which geographic and environmental isolation shape genetic divergence provides important information for maintaining functional connectivity, which is a key component of population viability, across the landscape (Segelbacher et al., 2010).

Generally, landscape genomics methods that investigate IBD and IBE rely on matrix regression, in which matrices of geographic and environmental distances serve as explanatory variables, and the response variable is a genetic distance matrix. For fitted models, the ratios between standardized regression coefficients (beta coefficients) for each explanatory variable can approximate the relative contributions of each environmental variable (or aggregated environmental distance) and geographic distance to explain variation in genetic distances.

Variable selection can be performed using backwards elimination in which explanatory variables are incrementally removed and the model refitted until only variables with statistically significant effects remain (Ferrier et al., 2002; Wang, 2013).

ALGATR implements two methods for estimating IBD and IBE: generalized dissimilarity modeling (GDM; Ferrier, 2002; Ferrier et al., 2002, 2007) and multiple matrix regression with randomization (MMRR; Wang, 2013) (Figure 1). GDM models turnover in the compositional dissimilarity between pairs of sites by transforming explanatory variables using a set of I-spline basis functions (Ferrier, 2002; Ferrier et al., 2002, 2007; Fitzpatrick & Keller, 2015). GDM's main advantage is that it can account for nonlinear relationships between variables (Mokany et al., 2022). The shape of the resulting relationships can then be used to identify threshold values that may represent important areas of genetic turnover because the shape of the I-spline functions indicate how genetic dissimilarity changes across an environmental gradient for each explanatory variable (Storfer et al., 2018). The environmental data layers used in the analysis can also be transformed based on these fitted relationships and combined into a map of genetic compositional dissimilarity across the study region (see Figure 1).

Similarly, MMRR performs linear matrix regression on genetic and environmental distances and allows for multiple independent variables (environmental and geographic distances) to be examined simultaneously. This approach provides less flexibility than GDM (Table S2) but may be less prone to overfitting, a tradeoff some researchers may prefer. MMRR performs significance testing through random permutations of the rows and columns of the dependent matrix, which is necessary because of the non-independence of values in pairwise distance matrices. MMRR provides individual regression coefficients and *p*-values for each explanatory variable and for the fitted model.

One remaining set of analyses in this category are those that seek to understand the extent to which landscape features act to restrict gene flow, a phenomenon that results in a pattern known as isolation-by-resistance (IBR; McRae, 2006). The workflow for investigating IBR includes first generating a resistance surface based on prior information or a hypothesis about how landscape elements contribute to differential resistance to movement for the study organism. Resistance distances can then be calculated using least-cost path analysis (Wang et al., 2009; Wang & Shaffer, 2017) or circuit theory analysis (Dickson et al., 2019; McRae & Beier, 2007). Because the parameterization of the resistance surface is highly complex, requiring decisions that often have a significant impact on the estimate of resistance distances (Koen et al., 2012; Spear et al., 2010; Spear et al., 2015), and because the optimal approach to resistance surface parameterization depends strongly on the study design and objectives (Peterman et al., 2019; Spear et al., 2010; Zeller et al., 2012), ALGATR does not include an automated approach for generating resistance surfaces. However, if a user has generated a resistance surface with their preferred method, which could include parameterization based on habitat suitability models (Wang et al., 2008), genetic algorithms (Peterman, 2018), or gradient forest analysis of allele frequencies (Vanhove & Launey, 2023), then resistance distances can be calculated using ALGATR's *geo_dist()* function, specifying "resistance" with the type argument (type = "resistance"). This function uses the *commuteDistance()* function in the GDISTANCE package (van Etten & Hijmans, 2010). The resulting resistance distances can then be used in downstream analyses, including using them in MMRR or GDM to quantify IBR.

## 3.4 | How can we identify and protect adaptive genetic variation?

The accessibility of genomic data for natural systems brings not only greater resolution but also opportunities to examine spatial patterns of adaptive genetic variation (Manel et al.,

2010; Parisod & Holderegger, 2012; Schoville et al., 2012). Identifying ecologically-important, functional genetic variation – genes that convey fitness benefits under different conditions – and the environmental forces underlying that variation provides a new dimension of valuable information for conservation efforts. Characterizing genes involved in adaptation to local environmental conditions can help guide local reintroduction efforts, assess the feasibility of genomic rescue for genetically depauperate populations, evaluate strategies for assisted gene flow, and quantify climate resilience and vulnerability (Browne et al., 2019; Frankham et al., 2014; Seaborn et al., 2021; Thurman et al., 2020).

A critical step for these objectives is to identify genes associated with specific environmental variables. Genotype-environment association (GEA) analyses quantify statistical associations between allele frequencies and environmental variables to test the hypothesis that allelic variation at a locus reflects adaptation to the local environment (Capblancq et al., 2018; Lotterhos, 2023). Because spurious genotype-environment associations could result from a variety of factors other than local adaptation, including population structure and demography, the signal of environmental selection must be parsed from the background level of neutral divergence resulting from population structure (Ahrens et al., 2018; De Mita et al., 2013; Lotterhos & Whitlock, 2014, 2015; Rellstab et al., 2015; Storfer et al., 2018). Approaches for doing so include using latent factors to represent population structure (Caye et al., 2019; Rellstab et al., 2015), conditioning on neutral variation using principal components (Duforet-Frebourg et al., 2016), using Moran eigenvector maps to decompose spatial relationships (Forester et al., 2018), and identifying sets of loci that are putatively neutral and incorporating them as covariates in the model (Dauphin et al., 2022; Meirmans, 2015; Storfer et al., 2018).

Two GEA methods are included in ALGATR, latent factor mixed modeling (LFMM; Caye et al., 2019) and redundancy analysis (RDA; Figure 1). LFMM is a univariate method that uses latent factors to account for unobserved variables, including population structure (Caye et al., 2019). This approach is advantageous because it provides controls on factors, other than selection, that may incidentally covary with allele frequencies. The challenge is determining the appropriate number of latent factors for any given dataset (called *K* values). ALGATR provides four options for doing so: a Tracy-Widom test (Frichot et al., 2013), a "quick elbow" test (an approach similar to examining a scree plot), the TESS clustering algorithm (in which latent factors correspond to some measure of population structure), or K-means clustering (Jombart et al., 2010).

RDA is a constrained ordination method that models linear, multivariate relationships (Rellstab et al., 2015). By simultaneously testing multiple loci against multiple environmental variables, RDA is able to detect multilocus selection (Forester et al., 2018; Rellstab et al., 2015) while also accounting for covariation in allele frequencies between loci. Optionally, RDA can perform variable selection using forward selection until a specified threshold is met (typically using a permutation test and adjusted $R^2$ values). The RDA method can also utilize covariates, such as population structure or geographic distance, alongside environmental predictors, an approach known as a partial RDA (pRDA). Finally, variance partitioning can be performed using RDA to quantify the independent contributions of each explanatory variable as well as variation that is explained by a combination of explanatory variables (i.e., confounded variance; Capblancq & Forester, 2021). ALGATR includes RDA and pRDA with and without variance partitioning, and it implements two approaches for determining which loci should be considered

significant outliers, one that uses Z-scores (Forester et al., 2018) and another that transforms RDA loadings into *p*-values (Capblancq, Morin, et al., 2020; Capblancq & Forester, 2021).

## 3.5 | Caveats

Like any analysis, landscape genomic methods have various limitations and assumptions, and even preparing data for these methods carries some potential pitfalls. Below, we briefly discuss some potentially important concerns that may arise during landscape genomic analysis and describe how options and solutions for addressing these issues can be implemented (see also Table S2). The ALGATR documentation aims to provide all of the information necessary to run each of the analyses and guidance on the decisions that must be made for the options offered by each method. It also provides references to several extensive and recent reviews that explore the methodology, assumptions, and theory underlying many of these methods (e.g., Capblancq, Fitzpatrick, et al., 2020; Capblancq & Forester, 2021; Fenderson et al., 2020; Forester et al., 2018; Lotterhos, 2023; Rellstab et al., 2015). The package itself also contains tools that help to evaluate different options for each method implemented in the pipeline.

GDM and MMRR take in genetic, environmental, and geographic distance matrices as input – various distance metrics are available – and the choice of metric may influence downstream results. Different genetic distance metrics, in particular, may result in different outcomes. ALGATR provides options to calculate several metrics of genetic, geographic, and environmental distances, and users should consider and test the potential impacts of different metrics on their results (Shirk et al., 2017; Wang, 2020).

Several landscape genomic methods, including RDA, do not allow for any missing genotypes, meaning that studies must either remove loci containing missing data or perform imputation to fill in missing genotypes. Because removing sites containing missing data often

results in greatly reduced dataset size, imputation is commonly performed to maximize data retention. Different types of imputation exist, including population structure-based imputation (Caye et al., 2016), maximum likelihood-based imputation (D'Angelo et al., 2010), or imputation based on the mean, median, or most common genotype at each site (e.g., Capblancq & Forester, 2021), each of which introduces different assumptions (e.g., Money et al., 2015; Shi et al., 2018; Yi & Latch, 2022). To deal with missing values, ALGATR performs two types of imputation. The first is a per site median-based imputation (similar to that of Capblancq & Forester, 2021), although we do not recommend this simplified approach for general use because artificially inflated *p*-values can result if missing values are non-random (i.e., if there is allelic bias in missing data). We have also implemented a population structure-based imputation method that uses non-negative matrix factorization (sNMF; Frichot et al., 2014)) to assign missing values. This utilizes functions in the LEA package (Frichot & François, 2015) and provides a more sophisticated (albeit computationally slower) imputation method compared to the median-based approach. As with many similar population structure methods, it also requires a user-selected value for *K* clusters.

Researchers should keep in mind that determining the number of *K* values that best describe their data should be done with care. This is relevant to selecting the number of latent factors (for LFMM) and *K* clusters (for TESS). For TESS, ALGATR can perform manual and automatic *K*-selection. Automatic *K*-selection is provided mainly for simulation studies or meta-analyses where *K*-values have to be selected for a large quantity of datasets such that manual selection is not feasible. For LFMM, ALGATR implements four methods for automatically selecting the number of latent factors (also represented using the term *K*): by performing a

Tracy-Widom test, a "quick elbow" test, using cross-validation scores from TESS, or K-means clustering using adegenet's *find.clusters* function.

Linkage disequilibrium (LD) results in collinearity among SNPs, a pattern that can misinform landscape genomic analyses. For example, estimates of population structure can be overinflated as more SNPs appear to independently support the same pattern (Rellstab et al., 2015). This is, fortunately, fairly straightforward to address by pruning sites that are putatively in LD. A common approach involves calculating correlations between SNPs in sliding windows of custom sizes across the genome (Ahrens et al., 2018) which can be done before or after performing the GEA analysis (Capblancq & Forester, 2021). ALGATR performs LD-pruning using the SNPRELATE package (Zheng et al., 2012), which also avoids the need for phased input data, and provides options to specify window size, window overlap, and the LD threshold.

## 4 | CONCLUSIONS

The implementation of genetically-informed conservation actions is critical to understand how best to conserve biodiversity. Landscape genomic approaches can provide important insight into how genetic variation is spatially distributed, allowing conservation practitioners to prioritize areas for conservation efforts. Genomic-scale data have enabled a shift towards individual- rather than population-based sampling because they provide more genetic resolution than was previously feasible. This shift towards individual-based sampling reduces the impact on populations that may already be in rapid decline, while also allowing researchers to capture greater geographic and environmental coverage for their datasets and to achieve higher resolution results.

Individual-based sampling provides a number of distinct benefits for biodiversity conservation, and landscape genomic approaches are poised to seize upon these advantages to

provide actionable information for conservation and management efforts. Our R package, ALGATR, provides an easily accessible and user-friendly pipeline that uses individual-based genomic datasets to provide fine-scale genomic and spatial resolution to answer fundamental conservation questions. Landscape genomics provides an ever-increasing resolution into the processes and patterns that shape the genetic variation of Earth's biodiversity, thereby increasing our understanding of how best to protect it.

## ACKNOWLEDGEMENTS

We thank Matt Fitzpatrick for providing input on and updates to the GDM package and Rachael Bay, Erik Enbody, Ryan Harrigan, Jay Sexton, Brad Shaffer, Victoria Sork, and Erin Toffelmier for productive conversations about methodology to include in ALGATR. We also thank the editor and two anonymous reviewers for their feedback which improved the quality of the manuscript.

## FUNDING INFORMATION

This work was supported by the California Conservation Genomics Project, with funding provided to the University of California by the State of California, State Budget Act of 2019 (UC Award ID RSI-19-690224) and by a National Science Foundation grant (DEB-1845682) awarded to IJW.

## AUTHOR CONTRIBUTIONS

All authors conceived of the study; APB and EAC wrote the ALGATR package, with contributions from IJW; EAC wrote and APB co-wrote the manuscript; all authors reviewed and edited the manuscript.

## CONFLICT OF INTEREST

The authors declare that they have no competing interests.

## DATA AVAILABILITY STATEMENT

The ALGATR package can be installed from GitHub: https://github.com/TheWangLab/algatr. A containerized version of the package can also be run using Docker; instructions for its installation can be found in the package's README on GitHub.

# REFERENCES


Ahrens, C. W., Rymer, P. D., Stow, A., Bragg, J., Dillon, S., Umbers, K. D. L., & Dudaniec, R. Y. (2018). The search for loci under selection: trends, biases and progress. *Molecular Ecology*, *27*(6), 1342–1356.

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664.

Allendorf, F. W., Funk, C. W., Aitken, S. N., Byrne, M., & Luikart, G. (Eds.). (2022). *Conservation and the Genomics of Populations*. Oxford University Press.

Balkenhol, N., Cushman, S., Storfer, A., & Waits, L. (Eds.). (2015). *Landscape genetics: concepts, methods, applications*. Wiley-Blackwell.

Balkenhol, N., & Fortin, M.-J. (2015). Basics of study design: sampling landscape heterogeneity and genetic variation for landscape genetic studies. In N. Balkenhol, S. A. Cushman, A. T. Storfer, & L. P. Waits (Eds.), *Landscape Genetics* (pp. 58–76). Wiley.

Bay, R. A., Harrigan, R. J., Underwood, V. L., Gibbs, H. L., Smith, T. B., & Ruegg, K. (2018). Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science*, *359*(6371), 83–86.

Bishop, A. P., Chambers, E. A., & Wang, I. J. (2023). Generating continuous maps of genetic diversity using moving windows. *Methods in Ecology and Evolution*, *14*(5), 1175–1181.

Bradburd, G. S., Coop, G. M., & Ralph, P. L. (2018). Inferring continuous and discrete population genetic structure across space. *Genetics*, *210*(1), 33–52.

Bradburd, G. S., & Ralph, P. L. (2019). Spatial population genetics: It's about time. *Annual Review of Ecology, Evolution, and Systematics*, *50*(1), 427–449.

Browne, L., Wright, J. W., Fitz-Gibbon, S., Gugger, P. F., & Sork, V. L. (2019). Adaptational

lag to temperature in valley oak (*Quercus lobata*) can be mitigated by genome-informed assisted gene flow. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(50), 25179–25185.

Capblancq, T., Fitzpatrick, M. C., Bay, R. A., Exposito-Alonso, M., & Keller, S. R. (2020). Genomic prediction of (mal)adaptation across current and future climatic landscapes. *Annual Review of Ecology, Evolution, and Systematics*, *51*(1), 245–269.

Capblancq, T., & Forester, B. R. (2021). Redundancy analysis: A Swiss Army Knife for landscape genomics. *Methods in Ecology and Evolution*, *12*(12), 2298–2309.

Capblancq, T., Luu, K., Blum, M. G. B., & Bazin, E. (2018). Evaluation of redundancy analysis to identify signatures of local adaptation. *Molecular Ecology Resources*, *18*(6), 1223–1233.

Capblancq, T., Morin, X., Gueguen, M., Renaud, J., Lobreaux, S., & Bazin, E. (2020). Climate-associated genetic variation in *Fagus sylvatica* and potential responses to climate change in the French Alps. *Journal of Evolutionary Biology*, *33*(6), 783–796.

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.

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.

Caye, K., Jay, F., Michel, O., & François, O. (2018). Fast inference of individual admixture coefficients using geographic data. *The Annals of Applied Statistics*, *12*(1), 586–608.

Caye, K., Jumentier, B., Lepeule, J., & François, O. (2019). LFMM 2: Fast and accurate inference of gene-environment associations in genome-wide studies. *Molecular Biology and Evolution*, *36*(4), 852–860.

Cortellari, M., Barbato, M., Talenti, A., Bionda, A., Carta, A., Ciampolini, R., Ciani, E., Crisà, A., Frattini, S., Lasagna, E., Marletta, D., Mastrangelo, S., Negro, A., Randi, E., Sarti, F. M., Sartore, S., Soglia, D., Liotta, L., Stella, A., … Crepaldi, P. (2021). The climatic and genetic heritage of Italian goat breeds with genomic SNP data. *Scientific Reports*, *11*(1), 10986.

D'Angelo, G. M., Kamboh, M. I., & Feingold, E. (2010). A likelihood-based approach for missing genotype data. *Human Heredity*, *69*(3), 171–183.

Dale, M. R. T., Fortin, M.-J. (2014). *Spatial analysis: a guide for ecologists.* Cambridge University Press.

Dauphin, B., Rellstab, C., Wüest, R. O., Karger, D. N., Holderegger, R., Gugerli, F., & Manel, S. (2022). Re-thinking the environment in landscape genomics. *Trends in Ecology & Evolution*, *38*(3), 261–274.

De Mita, S., Thuillet, A.-C., Gay, L., Ahmadi, N., Manel, S., Ronfort, J., & Vigouroux, Y. (2013). Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology*, *22*(5), 1383–1399.

Dickson, B. G., Albano, C. M., Anantharaman, R., Beier, P., Fargione, J., Graves, T. A., Gray, M. E., Hall, K. R., Lawler, J. J., Leonard, P. B., Littlefield, C. E., McClure, M. L., Novembre, J., Schloss, C. A., Schumaker, N. H., Shah, V. B., & Theobald, D. M. (2019). Circuit-theory applications to connectivity science and conservation. *Conservation Biology*, *33*(2), 239–249.

Duforet-Frebourg, N., Luu, K., Laval, G., Bazin, E., & Blum, M. G. B. (2016). Detecting genomic signatures of natural selection with principal component analysis: application to the 1000 Genomes data. *Molecular Biology and Evolution*, *33*(4), 1082–1093.

Fenderson, L. E., Kovach, A. I., & Llamas, B. (2020). Spatiotemporal landscape genetics: Investigating ecology and evolution through space and time. *Molecular Ecology*, *29*(2), 218–246.

Ferrier, S. (2002). Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? *Systematic Biology*, *51*(2), 331–363.

Ferrier, S., Drielsma, M., & Manion, G. (2002). Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. II. Community-level modelling. *Biodiversity and Conservation*, *11*, 2309–2338.

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 & Distributions*, *13*(3), 252–264.

Fiedler, P. L., Erickson, B., Esgro, M., Gold, M., Hull, J. M., Norris, J. M., Shapiro, B., Westphal, M., Toffelmier, E., & Shaffer, H. B. (2022). Seizing the moment: The opportunity and relevance of the California Conservation Genomics Project to state and federal conservation policy. *The Journal of Heredity*, *113*(6), 589–596.

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), 1–16.

Forester, B. R., & Lama, T. M. (2022). The role of genomics in the future of ESA decision-making. In B. L. Organ J (Ed.), *The Codex of the Endangered Species Act: The Next Fifty*

*Years – Volume II* (pp. 1–33).

Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Molecular Ecology*, *27*(9), 2215–2233.

Forester, B. R., Murphy, M., Mellison, C., Petersen, J., Pilliod, D. S., Van Horne, R., Harvey, J., & Funk, W. C. (2022). Genomics-informed delineation of conservation units in a desert amphibian. *Molecular Ecology*, *31*(20), 5249–5269.

Frankham, R., Bradshaw, C. J. A., & Brook, B. W. (2014). Genetics in conservation management: Revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. *Biological Conservation*, *170*, 56–63.

Freedman, A. H., Thomassen, H. A., Buermann, W., & Smith, T. B. (2010). Genomic signals of diversification along ecological gradients in a tropical lizard. *Molecular Ecology*, *19*(17), 3773–3788.

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

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, *196*(4), 973–983.

Frichot, E., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, *30*(7), 1687–1699.

Funk, W. C., Forester, B. R., Converse, S. J., Darst, C., & Morey, S. (2019). Improving conservation policy with genomics: a guide to integrating adaptive potential into U.S. Endangered Species Act decisions for conservation practitioners and geneticists.

*Conservation Genetics*, *20*(1), 115–134.

Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, *27*(9), 489–496.

Hall, L. A., & Beissinger, S. R. (2014). A practical toolbox for design and analysis of landscape genetics studies. *Landscape Ecology*, *29*(9), 1487–1504.

Hiemstra, P. H., Pebesma, E. J., Twenhöfel, C. J. W., & Heuvelink, G. B. M. (2009). Real-time automatic interpolation of ambient gamma dose rates from the Dutch radioactivity monitoring network. *Computers & Geosciences*, *35*(8), 1711–1721.

Hoban, S., Bruford, M., D'Urban Jackson, J., Lopes-Fernandes, M., Heuertz, M., Hohenlohe, P. A., Paz-Vinas, I., Sjögren-Gulve, P., Segelbacher, G., Vernesi, C., Aitken, S., Bertola, L. D., Bloomer, P., Breed, M., Rodríguez-Correa, H., Funk, W. C., Grueber, C. E., Hunter, M. E., Jaffe, R., … 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., Sgrò, C. M., & Kristensen, T. N. (2017). Revisiting adaptive potential, population size, and conservation. *Trends in Ecology & Evolution*, *32*(7), 506–517.

Hohenlohe, P. A., Funk, W. C., & Rajora, O. P. (2021). Population genomics for wildlife conservation and management. *Molecular Ecology*, *30*(1), 62–82.

Holliday, J. A., Aitken, S. N., Cooke, J. E. K., Fady, B., González-Martínez, S. C., Heuertz, M., Jaramillo-Correa, J.-P., Lexer, C., Staton, M., Whetten, R. W., & Plomion, C. (2017). Advances in ecological genomics in forest trees and applications to genetic resources conservation and breeding. *Molecular Ecology*, *26*(3), 706–717.

Hurlbert, S. H. (1984). Pseudoreplication and the design of ecological field experiments.

610 *Ecological Monographs*, *54*(2), 187–211.
611 Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components:
612 a new method for the analysis of genetically structured populations. *BMC Genetics*, *11*, 94.
613 Kardos, M., Armstrong, E. E., Fitzpatrick, S. W., Hauser, S., Hedrick, P. W., Miller, J. M.,
614 Tallmon, D. A., & Funk, W. C. (2021). The crucial role of genome-wide genetic variation in
615 conservation. *Proceedings of the National Academy of Sciences of the United States of*
616 *America*, *118*(48), e2104642118.
617 Keller, D., Holderegger, R., van Strien, M. J., & Bolliger, J. (2015). How to make landscape
618 genetics beneficial for conservation management? *Conservation Genetics* , *16*(3), 503–512.
619 Koen, E. L., Bowman, J., & Walpole, A. A. (2012). The effect of cost surface parameterization
620 on landscape resistance estimates. *Molecular Ecology Resources, 12*(4), 686–696.
621 Leitwein, M., Duranton, M., Rougemont, Q., Gagnaire, P.-A., & Bernatchez, L. (2020). Using
622 haplotype information for conservation genomics. *Trends in Ecology & Evolution*, *35*(3),
623 245–258.
624 Lotterhos, K. E. (2023). The paradox of adaptive trait clines with nonclinal patterns in the
625 underlying genes. *Proceedings of the National Academy of Sciences of the United States of*
626 *America*, *120*(12), e2220313120.
627 Lotterhos, K. E., & Whitlock, M. C. (2014). Evaluation of demographic history and neutral
628 parameterization on the performance of FST outlier tests. *Molecular Ecology*, *23*(9), 2178–
629 2192.
630 Lotterhos, K. E., & Whitlock, M. C. (2015). The relative power of genome scans to detect local
631 adaptation depends on sampling design and statistical method. *Molecular Ecology*, *24*(5),
632 1031–1046.

633 Malcom, J., & Carter, A. (2021). Better representation is needed in U.S. Endangered Species Act
634 implementation. *Frontiers in Conservation Science*, *2*, 650543.
635 Manel, S., Albert, C. H., & Yoccoz, N. G. (2012). Sampling in landscape genetics. In F.
636 Pompanon & A. Bonin (Eds.), *Data Production and Analysis in Population Genomics:*
637 *Methods and Protocols* (pp. 3–12). Humana Press.
638 Manel, S., & Holderegger, R. (2013). Ten years of landscape genetics. *Trends in Ecology &*
639 *Evolution*, *28*(10), 614–621.
640 Manel, S., Joost, S., Epperson, B. K., Holderegger, R., Storfer, A., Rosenberg, M. S., Scribner,
641 K. T., Bonin, A., & Fortin, M.-J. (2010). Perspectives on the use of landscape genetics to
642 detect genetic adaptive variation in the field. *Molecular Ecology*, *19*(17), 3760–3772.
643 Manel, S., Schwartz, M. K., Luikart, G., & Taberlet, P. (2003). Landscape genetics: combining
644 landscape ecology and population genetics. *Trends in Ecology & Evolution*, *18*(4), 189–197.
645 McRae, B. H., & Beier, P. (2007). Circuit theory predicts gene flow in plant and animal
646 populations. *Proceedings of the National Academy of Sciences of the United States of*
647 *America*, *104*(50), 19885–19890.
648 Medina, R., Wogan, G. O. U., Bi, K., Termignoni-García, F., Bernal, M. H., Jaramillo-Correa, J.
649 P., Wang, I. J., & Vázquez-Domínguez, E. (2021). Phenotypic and genomic diversification
650 with isolation by environment along elevational gradients in a neotropical treefrog.
651 *Molecular Ecology*, *30*(16), 4062–4076.
652 Meirmans, P. G. (2015). Seven common mistakes in population genetics and how to avoid them.
653 *Molecular Ecology*, *24*(13), 3223–3231.
654 Mokany, K., Ware, C., Woolley, S. N. C., Ferrier, S., & Fitzpatrick, M. C. (2022). A working
655 guide to harnessing generalized dissimilarity modelling for biodiversity analysis and

conservation assessment. *Global Ecology and Biogeography: A Journal of Macroecology*, *31*(4), 802–821.

Money, D., Gardner, K., Migicovsky, Z., Schwaninger, H., Zhong, G.-Y., & Myles, S. (2015). LinkImpute: Fast and accurate genotype imputation for nonmodel organisms. *G3*, *5*(11), 2383–2390.

Nielsen, E. S., Hanson, J. O., Carvalho, S. B., Beger, M., Henriques, R., Kershaw, F., & von der Heyden, S. (2022). Molecular ecology meets systematic conservation planning. *Trends in Ecology & Evolution*, *38*(2), 143–155.

O'Grady, J. J., Reed, D. H., Brook, B. W., & Frankham, R. (2004). What are the best correlates of predicted extinction risk? *Biological Conservation*, *118*(4), 513–520.

Ogbonna, A. C., Braatz de Andrade, L. R., Mueller, L. A., de Oliveira, E. J., & Bauchet, G. J. (2021). Comprehensive genotyping of a Brazilian cassava (*Manihot esculenta* Crantz) germplasm bank: insights into diversification and domestication. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, *134*(5), 1343–1362.

Ortego, J., Gugger, P. F., & Sork, V. L. (2015). Climatically stable landscapes predict patterns of genetic structure and admixture in the Californian canyon live oak. *Journal of Biogeography*, *42*(2), 328–338.

Parisod, C., & Holderegger, R. (2012). Adaptive landscape genetics: pitfalls and benefits. *Molecular Ecology*, *21*(15), 3644–3646.

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, *2*(12), e190.

Peterman, W. E. 2018. ResistanceGA: An R package for the optimization of resistance surfaces using genetic algorithms. *Methods in Ecology and Evolution*, *9*(6), 1638–1647.

679 Peterman, W. E., Winiarski, K. J., Moore, C. E., Carvalho, C. da S., Gilbert, A. L., & Spear, S. F.
680 (2019). A comparison of popular approaches to optimize landscape resistance surfaces.
681 *Landscape Ecology*, *34*(9), 2197–2208.
682 Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using
683 multilocus genotype data. *Genetics*, *155*(2), 945–959.
684 Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical
685 guide to environmental association analysis in landscape genomics. *Molecular Ecology*,
686 *24*(17), 4348–4370.
687 Schmidt, C., Hoban, S., Hunter, M., Paz-Vinas, I., & Garroway, C. J. (2023). Genetic diversity
688 and IUCN Red List status. *Conservation Biology*, *37*(4), e14064.
689 Schoville, S. D., Bonin, A., François, O., Lobreaux, S., Melodelima, C., & Manel, S. (2012).
690 Adaptive genetic variation on the landscape: methods and cases. *Annual Review of Ecology,*
691 *Evolution, and Systematics*, *43*(1), 23–43.
692 Seaborn, T., Griffith, D., Kliskey, A., & Caudill, C. C. (2021). Building a bridge between
693 adaptive capacity and adaptive potential to understand responses to environmental change.
694 *Global Change Biology*, *27*(12), 2656–2668.
695 Segelbacher, G., Cushman, S. A., Epperson, B. K., Fortin, M.-J., Francois, O., Hardy, O. J.,
696 Holderegger, R., Taberlet, P., Waits, L. P., & Manel, S. (2010). Applications of landscape
697 genetics in conservation biology: concepts and challenges. *Conservation Genetics* , *11*(2),
698 375–385.
699 Selmoni, O., Vajana, E., Guillaume, A., Rochat, E., & Joost, S. (2020). Sampling strategy
700 optimization to increase statistical power in landscape genomics: A simulation-based
701 approach. *Molecular Ecology Resources*, *20*(1), 154–169.

Sexton, J. P., Hufford, M. B., Bateman, A. C., Lowry, D. B., Meimberg, H., Strauss, S. Y., & Rice, K. J. (2016). Climate structures genetic variation across a species' elevation range: a test of range limits hypotheses. *Molecular Ecology*, *25*(4), 911–928.

Shafer, A. B. A., Wolf, J. B. W., Alves, P. C., Bergström, L., Bruford, M. W., Brännström, I., Colling, G., Dalén, L., De Meester, L., Ekblom, R., Fawcett, K. D., Fior, S., Hajibabaei, M., Hill, J. A., Hoezel, A. R., Höglund, J., Jensen, E. L., Krause, J., Kristensen, T. N., … Zieliński, P. (2015). Genomics and the challenging translation into conservation practice. *Trends in Ecology & Evolution*, *30*(2), 78–87.

Shaffer, H. B., Toffelmier, E., Corbett-Detig, R. B., Escalona, M., Erickson, B., Fiedler, P., Gold, M., Harrigan, R. J., Hodges, S., Luckau, T. K., Miller, C., Oliveira, D. R., Shaffer, K. E., Shapiro, B., Sork, V. L., & Wang, I. J. (2022). Landscape genomics to enable conservation actions: the California Conservation Genomics Project. *The Journal of Heredity*, *113*(6), 577–588.

Shirk, A. J., Landguth, E. L., & Cushman, S. A. (2017). A comparison of individual-based genetic distance metrics for landscape genetics. *Molecular Ecology Resources*, *17*(6), 1308–1317.

Shi, S., Yuan, N., Yang, M., Du, Z., Wang, J., Sheng, X., Wu, J., & Xiao, J. (2018). Comprehensive assessment of genotype imputation performance. *Human Heredity*, *83*(3), 107–116.

Shryock, D. F., Havrilla, C. A., DeFalco, L. A., Esque, T. C., Custer, N. A., & Wood, T. E. (2015). Landscape genomics of *Sphaeralcea ambigua* in the Mojave Desert: a multivariate, spatially-explicit approach to guide ecological restoration. *Conservation Genetics*, *16*(6), 1303–1317.

Sommer, S., McDevitt, A. D., & Balkenhol, N. (2013). Landscape genetic approaches in conservation biology and management. *Conservation Genetics*, *14*(2), 249–251.

Spear, S. F., Balkenhol, N., Fortin, M.-J., McRae, B. H., & Scribner, K. (2010). Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. *Molecular Ecology*, *19*(17), 3576–3591.

Storfer, A., Murphy, M. A., Evans, J. S., Goldberg, C. S., Robinson, S., Spear, S. F., Dezzani, R., Delmelle, E., Vierling, L., & Waits, L. P. (2007). Putting the "landscape" in landscape genetics. *Heredity*, *98*(3), 128–142.

Storfer, A., Patton, A., & Fraik, A. K. (2018). Navigating the interface between landscape genetics and landscape genomics. *Frontiers in Genetics*, *9*, 68.

Stucki, S., Orozco-terWengel, P., Forester, B. R., Duruz, S., Colli, L., Masembe, C., Negrini, R., Landguth, E., Jones, M. R., NEXTGEN Consortium, Bruford, M. W., Taberlet, P., & Joost, S. (2017). High performance computation of landscape genomic models including local indicators of spatial association. *Molecular Ecology Resources*, *17*(5), 1072–1089.

Supple, M. A., & Shapiro, B. (2018). Conservation of biodiversity in the genomics era. *Genome Biology*, *19*(1), 131.

Teixeira, J. C., & Huber, C. D. (2021). The inflated significance of neutral genetic diversity in conservation genetics. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(10), e2015096118.

Thurman, L. L., Stein, B. A., Beever, E. A., Foden, W., Geange, S. R., Green, N., Gross, J. E., Lawrence, D. J., LeDee, O., Olden, J. D., Thompson, L. M., & Young, B. E. (2020). Persist in place or shift in space? Evaluating the adaptive capacity of species to climate change. *Frontiers in Ecology and the Environment*, *18*(9), 520–528.

748 Turbek, S. P., Funk, W. C., & Ruegg, K. C. (2023). Where to draw the line? Expanding the
749 delineation of conservation units to highly mobile taxa. *The Journal of Heredity*, *114*(4),
750 300–311.
751 van Etten, J., & Hijmans, R. J. (2010). A geospatial modelling approach integrating
752 archaeobotany and genetics to trace the origin and dispersal of domesticated plants. *PloS*
753 *One*, *5*(8), e12060.
754 van Strien, M. J., Keller, D., Holderegger, R., Ghazoul, J., Kienast, F., & Bolliger, J. (2014).
755 Landscape genetics as a tool for conservation planning: predicting the effects of landscape
756 change on gene flow. *Ecological Applications*, *24*(2), 327–339.
757 Vanhove, M., & Launey, S. (2023). Estimating resistance surfaces using gradient forest and
758 allelic frequencies. *Molecular Ecology Resources.* [Online ahead of print]
759 Wagner, H. H., & Fortin, M.-J. (2013). A conceptual framework for the spatial analysis of
760 landscape genetic data. *Conservation Genetics*, *14*(2), 253–261.
761 Wang, I. J. (2013). Examining the full effects of landscape heterogeneity on spatial genetic
762 variation: a multiple matrix regression approach for quantifying geographic and ecological
763 isolation. *Evolution*, *67*(12), 3403–3411.
764 Wang, I. J. (2020). Topographic path analysis for modelling dispersal and functional
765 connectivity: Calculating topographic distances using the topoDistance R package. *Methods*
766 *in Ecology and Evolution*, *11*(2), 265–272.
767 Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, *23*(23),
768 5649–5662.
769 Wang, I. J., Savage, W. K., & Shaffer, H. B. (2009). Landscape genetics and least-cost path
770 analysis reveal unexpected dispersal routes in the California tiger salamander (*Ambystoma*

*californiense*). *Molecular Ecology*, *18*(7), 1365–1374.

Wang, I. J., & Shaffer, H. B. (2017). Population genetic and field-ecological analyses return similar estimates of dispersal over space and time in an endangered amphibian. *Evolutionary Applications*, *10*(6), 630–639.

Wang, Y.-H., Yang, K.-C., Bridgman, C., & Lin, L.-K. (2008). Habitat suitability modelling to correlate gene flow with landscape connectivity. *Landscape Ecology, 23*(8), 989–1000.

Wogan, G. O. U., Yuan, M. L., Mahler, D. L., & Wang, I. J. (2020). Genome-wide epigenetic isolation by environment in a widespread *Anolis* lizard. *Molecular Ecology*, *29*(1), 40–55.

Wright, S. (1943). Isolation by Distance. *Genetics*, *28*(2), 114–138.

Yi, X., & Latch, E. K. (2022). Nonrandom missing data can bias Principal Component Analysis inference of population genetic structure. *Molecular Ecology Resources*, *22*(2), 602–611.

Zeller, K. A., McGarigal, K., & Whiteley, A. R. (2012). Estimating landscape resistance to movement: a review. *Landscape Ecology*, *27*(6), 777–797.

Zhang, Y.-H., Wang, I. J., Comes, H. P., Peng, H., & Qiu, Y.-X. (2016). Contributions of historical and contemporary geographic and environmental factors to phylogeographic structure in a Tertiary relict species, *Emmenopterys henryi* (Rubiaceae). *Scientific Reports*, *6*, 24041.

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* , *28*(24), 3326–3328.

# TABLES

TABLE 1 Summary of key conservation-oriented questions that our landscape genomic analysis pipeline (ALGATR) can answer, including input requirements for the ALGATR package and recent empirical examples of each method.

| Question | Analysis category | Method | Input requirements | Empirical examples |
|---|---|---|---|---|
| How do we delineate population units for management? | Population structure | TESS (Caye et al., 2016) | Genotype dosage matrix (preferably LD-pruned), sampling coordinates, *optional: raster layer for mapping* | Ogbonna et al., 2021 |
| How is genetic variation distributed? | Genetic diversity | wingen (Bishop et al., 2023) | VCF, sampling coordinates, *optional: raster layer for mapping* | |
| What are the drivers of population connectivity? | Isolation by distance and isolation by environment | MMRR (Wang, 2013)<br><br>GDM (Ferrier et al., 2007; Fitzpatrick & Keller, 2015; Freedman et al., 2010) | Genetic distance matrix, environmental distance matrices or environmental layers plus sampling coordinates | Ortego et al., 2015; Sexton et al., 2016; Zhang et al., 2016<br><br>Bay et al., 2018; Medina et al., 2021; Shryock et al., 2015; Wogan et al., 2020 |
| How can we identify and protect adaptive genetic variation? | Genotype-environment association methods | RDA (Capblancq & Forester, 2021)<br><br>LFMM (Caye et al., 2019) | Genotype dosage matrix, environmental layers, sampling coordinates | Forester et al., 2022<br><br>Carvalho et al., 2021;<br><br>Cortellari et al., 2021 |

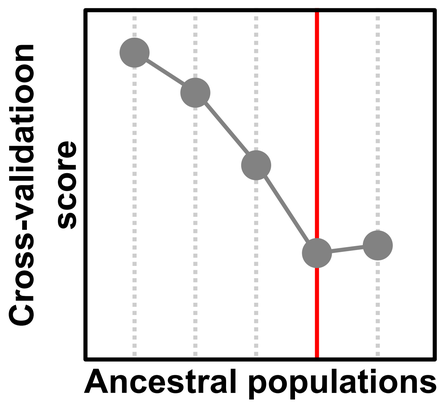
# FIGURE LEGENDS

FIGURE 1 Summary of outputs produced by the ALGATR pipeline with suggested methods to answer conservation questions. Methods included in ALGATR's workflow are TESS (Caye et al., 2016), wingen (Bishop et al., 2023), multiple matrix regression with randomization (MMRR; Wang, 2013), generalized dissimilarity modeling (GDM; Ferrier et al., 2007; Fitzpatrick & Keller, 2015; Freedman et al., 2010), redundancy analysis (RDA; Capblancq & Forester, 2021), and latent factor mixed modeling (LFMM; Caye et al., 2019).

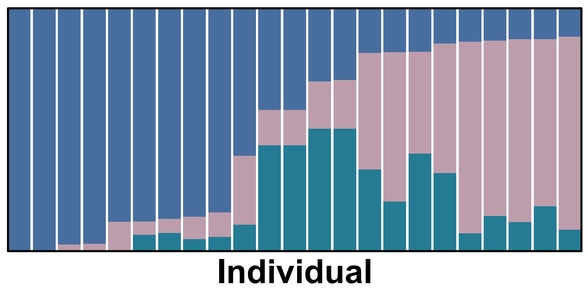
## How do we delineate population units for management?

TESS

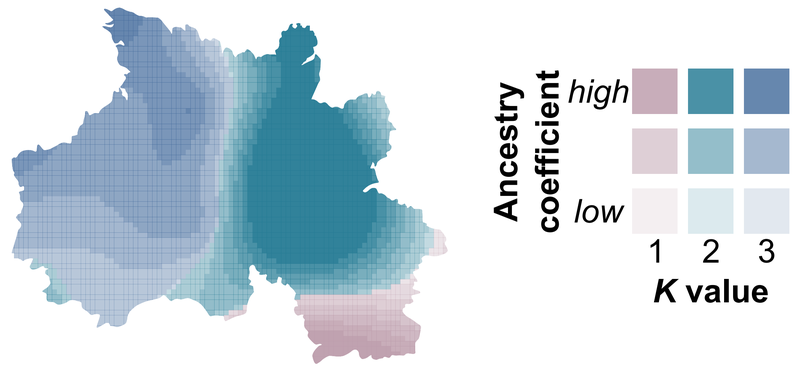
*K selection*




*Ancestry coefficients*




*Spatially interpolated ancestry coefficients*



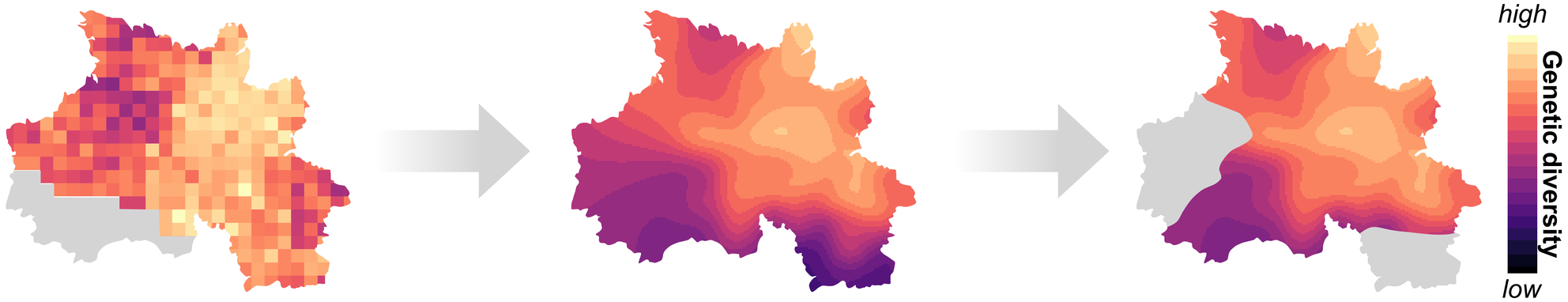

## How is genetic variation distributed?

wingen

*Genetic diversity across moving windows*

*Spatially interpolated genetic diversity*

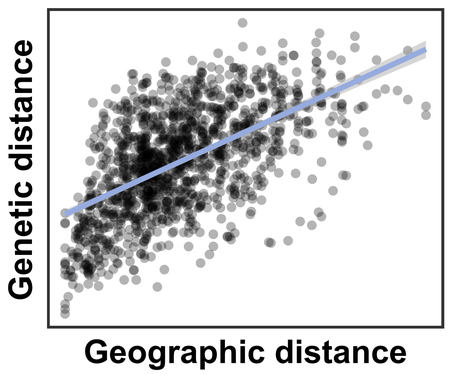
*Optional masking of undersampled areas*



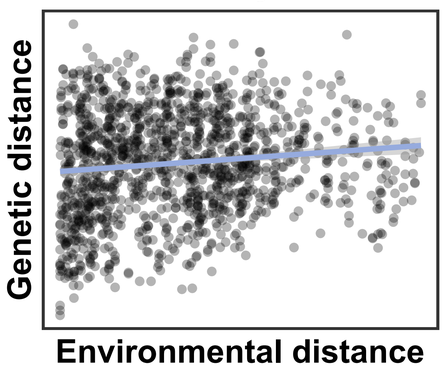

## What are the drivers of population connectivity?

MMRR

*Isolation by distance*

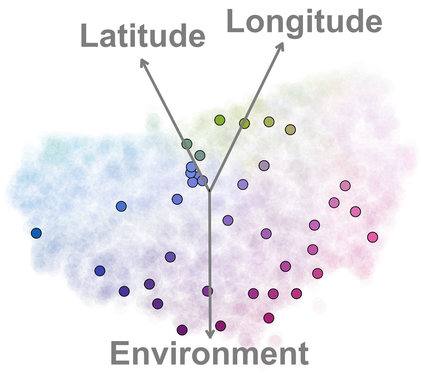



*Isolation by environment*

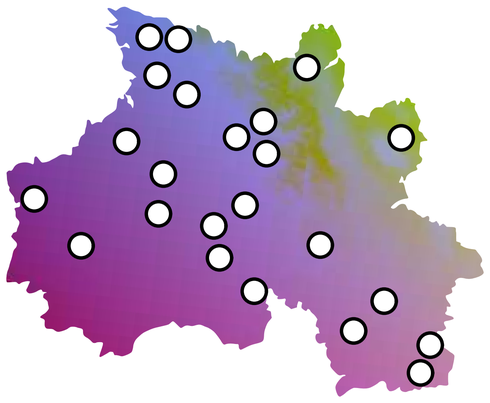



GDM

*Variable loadings*



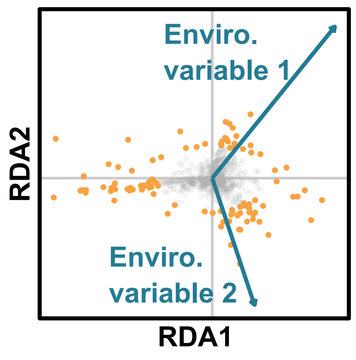

*Compositional dissimilarity*



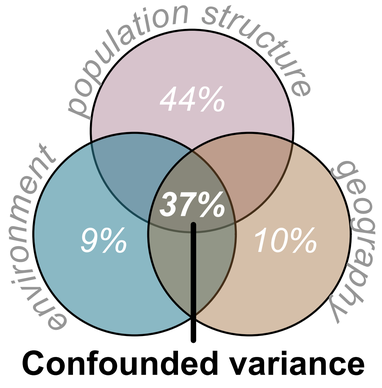
## How can we identify and protect adaptive genetic variation?

RDA

*RDA biplot of outlier SNPs*




*Variance partitioning*




LFMM

*Manhattan plot of outlier SNPs for each environmental variable*