Optimizing sampling design for landscape genomics

Anusha P. Bishop[1,2], Drew E. Terasaki Hart[1,3,4], Ian J. Wang[1,2]

Author affiliation(s):

[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

[3]The Nature Conservancy, Arlington, VA 22203, USA

[4]CSIRO Environment, Brisbane Queensland 4102, Australia

Corresponding author: Anusha Bishop (anusha.bishop@berkeley.edu), Department of Environmental Science, Policy, & Management, University of California, Berkeley, 130 Mulford Hall #3114, Berkeley, CA 94720

Running header: Landscape genomic sampling design

Keywords: ecological genetics, population genetics, simulation, genotype-environment association, isolation by environment, adaptation

Abstract

Landscape genomic approaches for detecting genotype-environment associations (GEA), isolation by distance (IBD), and isolation by environment (IBE) have seen a dramatic increase in use, but there have been few thorough analyses of the influence of sampling strategy on their performance under realistic genomic and environmental conditions. We simulated 24,000 datasets across a range of scenarios with complex population dynamics and realistic landscape structure to evaluate the effects of the spatial distribution and number of samples on common landscape genomics methods. Our results show that common analyses are relatively robust to sampling scheme as long as sampling covers enough environmental and geographic space. We found that for detecting adaptive loci and estimating IBE, sampling schemes that were explicitly designed to increase coverage of available environmental space matched or outperformed sampling schemes that only considered geographic space. When sampling does not cover adequate geographic and environmental space, such as with transect-based sampling, we detected fewer adaptive loci and had higher error when estimating IBD and IBE. We found that IBD could be detected with as few as nine sampling sites, while large sample sizes (e.g., greater than 100 individuals) were crucial for detecting adaptive loci and IBE. We also demonstrate that, even with optimal sampling strategies, landscape genomic analyses are highly sensitive to landscape structure and migration — when spatial autocorrelation and migration are weak, common GEA methods fail to detect adaptive loci.

1 | Introduction

Landscape genomics aims to characterize and interpret complex patterns of local adaptation and genetic structure based on a limited number of genetic samples (Manel et al. 2003; Storfer et al. 2007). The number and spatial distribution of these samples can have substantial effects on the power and accuracy of common analyses, and the strength of this effect is likely dependent on the environmental structure of the landscape and the population dynamics of the study system (Dauphin et al., 2023; De Mita et al., 2013; Forester et al., 2018; Lotterhos & Whitlock, 2015; Manel et al. 2012; Oyler-McCance et al., 2013; Selmoni et al., 2020). This dependence remains poorly understood, however, because the sensitivity of common methods to sampling strategy has not been evaluated across a realistic range of environmental, demographic, and genetic conditions.

Two major classes of landscape genomic analysis include (1) quantifying the drivers of genetic structure, including isolation by distance (IBD; Wright, 1943) and isolation by environment (IBE; Wang, 2013), and (2) identifying genes linked to environmental variation using genotype-environment association (GEA; Capblancq & Forester, 2021; Caye et al., 2019; Frichot et al., 2013; Forester et al., 2018). Previous work evaluating the

performance of GEA under different sampling schemes and demographic scenarios have found mixed support for the importance of sample distribution and have generally concluded that maximizing sample size is most important (De Mita et al., 2013; Forester et al., 2018; Lotterhos & Whitlock, 2015; Selmoni et al., 2020). These studies also showed that demographic conditions play a role in the performance of different methods and sampling strategies, but all of them were limited to testing a small number of demographic scenarios (two to four), likely due to computational constraints, and none evaluated the effects of landscape structure (though see Forester et al. 2016) nor tested methods for detecting IBD and IBE (De Mita et al., 2013; Forester et al., 2018; Lotterhos & Whitlock, 2015; Selmoni et al., 2020).

Here, we build off of existing studies to evaluate the effects of common sampling strategies on landscape genomic analyses under complex and realistic conditions simulated using Geonomics (Terasaki Hart et al., 2021). Our aim was to test a range of sampling strategies while varying key demographic and landscape parameters expected to affect landscape genomic results, including population size, migration rate, selection strength, autocorrelation within landscape layers, and correlation between landscape layers. We predicted that weak migration and small population size would reduce discovery rates for GEA by generating confounding population genetic structure (Hoban et al., 2016; Price et al., 2020; Li et al., 2017; Rellstab et al., 2015), but that these same conditions would positively affect the performance of methods to quantify IBD and IBE since decreased migration and small population size should increase signatures of isolation (Wang & Bradburd, 2014). We expected that greater selection strength would have a positive effect on both GEA and detection of IBE by increasing the strength of local adaptation. We hypothesized that the structure of the landscape, as defined by the correlations between environmental variables and the degrees of spatial autocorrelation within them, would also have a strong effect since landscapes drive spatially variable selection and, thereby, patterns of local adaptation (Forester et al., 2016; Lotterhos 2022; Rellstab et al., 2015). Specifically, we predicted that higher autocorrelation would have a positive effect on both GEA and the detection of IBE, while higher correlation between landscape layers would make it harder to disentangle which environmental variable is associated with an adaptive locus. Finally, we hypothesized that sampling design would interact with all of these factors to determine the power and accuracy of landscape genomic analyses.

We created 32 unique simulation scenarios by varying population genetic and landscape parameters and ran forward-time, spatially continuous, and individual-based Geonomics simulations (Terasaki Hart et al., 2021). We allowed populations to evolve freely on the simulated landscapes and sampled them using a total of seven individual- and site-based sampling schemes. We tested a suite of common landscape genomic methods for GEA

and for detecting IBE and IBD. By exploring the parameter space defined by the full crossing of simulation conditions and sampling schemes, we created a robust framework through which to understand the effects of sampling design, demography, and landscape structure on landscape genomic analysis.

2 | Materials and Methods

2.1 | Simulated landscapes

To evaluate the effects of landscape structure, we simulated environmental layers with different levels of spatial autocorrelation (i.e., the tendency of nearby points to exhibit more similar values) and correlation between layers. We used the R package "NLMR" version 1.1.1 (Sciaini et al., 2018) to create 100 by 100 cell layers using a midpoint displacement neutral landscape model, which generates landscapes with realistic fractal patterns and user-defined levels of spatial autocorrelation (Peitgen & Saupe, 1988). To simulate adaptation to two environmental variables, we used a pair of simulated landscapes in each simulation. We set spatial autocorrelation using the roughness parameter (H) to either 0.05 (weak autocorrelation; Moran's I of approximately 0.66) or 0.5 (strong autocorrelation; Moran's I of approximately 0.99). We set the correlation between layers to r = 0.3 (weak correlation) or r = 0.6 (strong correlation). We chose these levels of spatial autocorrelation and environmental correlation based on distributions we generated from global temperature, precipitation, and tree canopy cover rasters by randomly sampling windows of 100 by 100 cells and calculating Moran's I of the variables and the correlation between temperature and precipitation within each window (File S2; Fick & Hijmans, 2017). To ensure that our results were not dependent on a particular landscape configuration, we generated three independent replicates of our landscapes (see Figure S1 for example landscapes).

2.2 | Geonomics simulations

To test the effects of population size, migration rate, selection strength, spatial autocorrelation, and environmental correlation on the performance of different sampling strategies and methods, we simulated each of these parameters at a "low" and "high" level (see File S3 for the complete parameters file; Figure S2). We ran 10 replications of each simulation to capture variation in results due to stochasticity. Together with the three sets of simulated landscapes, this produced a total of 960 simulations (30 repetitions of 32 unique parametrizations).

We ran our simulations in Python version 3.9.7 (Van Rossum & Drake, 2009) using Geonomics version 1.3.9 (Terasaki Hart et al., 2021).

The simulated landscapes were provided to Geonomics as the environmental layers driving local adaptation. A single environmental layer acts as the selective force for a single continuous trait. To simulate multivariate adaptation (i.e., adaptation to multiple environmental gradients), we used two environmental layers to drive selection on two separate traits. Each of the two traits had phenotypic values determined by the additive effects of four independent, co-dominant loci, for a total of eight adaptive loci. The phenotype for a given trait was calculated as the "null" phenotype plus the sum of the allele dosages at all loci underlying that trait multiplied by their effect sizes. In our simulations, the "null" phenotype value was 0.50 and the effect sizes were 0.25 for two of the loci and -0.25 for the other two loci for each trait. Because the simulated species was diploid, there were eight total alleles underlying each trait, resulting in phenotype values ranging from 0 to 1 by increments of 0.125. We also simulated 10,000 neutral loci to simulate neutral population genetic structure. All adaptive and neutral loci were allowed to freely recombine and, therefore, are unlinked.

We varied population size, migration rate, and selection strength in our Geonomics simulations to create a range of population genetic scenarios (Table S1). We varied population size by changing the carrying capacity (K) of each cell in the landscape, with all cells having uniform values. This results in population sizes of approximately 3,000 - 4,000 when K = 1 and 5,000 - 7,000 when K = 2. We varied migration rate by changing the variance of the lognormal distributions used to model migration and dispersal distances (Nield et al., 2020; Russo et al., 2006; Uriarte et al., 2011). The mean of these distributions was set to 0 such that the expected movement and dispersal values when variance = 0.25 was approximately 1 cell width and when variance = 1.00 was approximately 2 cell widths. We varied selection strength by changing the phenotypic selection coefficient (ϕ), which dictates the strength of selection acting on all loci underlying a trait. We tested ϕ values of 0.5 and 1.0 for both traits to represent our lower and higher levels of selection, respectively. An individual's fitness is determined by the product of ϕ and the absolute difference between the individual's phenotype value and the environmental value at its location.

Each Geonomics simulation begins with a "burn-in" period where individuals move and reproduce, without selection, until an equilibrium in the population dynamics is achieved, as described in Terasaki Hart et al. (2021). Following this period, we ran our simulations for 6,000 time steps. We determined that 6,000 steps was sufficient time for the simulations to achieve equilibrium by running a subset of simulations for 10,000 timesteps and testing for demographic and population genetic equilibrium over time (File S4). We also

confirmed *post hoc* that the simulations had reached equilibrium in mean fitness and that local adaptation had occurred based on phenotype-environment correlations (File S5). We estimated the strength of genotype-environment associations by calculating the magnitude of the correlation between dosages for each adaptive locus and their corresponding environmental variable. We tested the effect of landscape and population genetic parameters on local adaptation by running linear models with the phenotype-environment and genotype-environment correlations as the response variables and the simulated parameters as the predictors. To estimate the confounding effects of neutral structure under the different simulated conditions, we calculated the correlation between each neutral locus and each environmental variable and compared the magnitudes across the different parameter levels.

2.3 | Effects of sampling

To quantify the effects of sampling strategy on the performance of landscape genomic methods, we varied the sampling unit (i.e., either individual- or site-based), sampling scheme (i.e., spatial distribution), and sample size. For individual-based sampling, we extracted allele dosages, environmental values, and coordinates for each individual. For site-based sampling, we extracted allele frequencies, average environmental values across individuals, and coordinates for each site.

Under individual-based sampling, we selected samples across the landscape using four different schemes: random, grid-based, transect-based, and a form of environmentally stratified sampling that we refer to hereafter as 'E-space sampling.' The random sampling scheme mirrors unstructured or opportunistic sampling, and the grid-based scheme mimics sampling efforts that aim for even geographic coverage. We used E-space sampling to maximize the range of environmental values covered while minimizing spatial autocorrelation for both environmental variables, thereby increasing the power to detect environmentally-driven patterns of genetic variation while decreasing the confounding effects of spatial autocorrelation, which can lead to false discoveries. This is similar to an approach described by Lotterhos & Whitlock (2015) that involves selecting sampling pairs that are geographically close but environmentally distant. However, instead of selecting pairs, we performed E-space sampling by taking sets of random samples and selecting the set that balanced maximizing the range of environmental values covered against minimizing spatial autocorrelation measured with Moran's I. This was done by calculating a score for each of the random sampling schemes by summing the scaled variance of each environmental variable with scaled *1 - Moran's I* and selecting the sample set with the highest score.

For each sampling scheme, we selected 36, 81, 144, or 225 samples for a total of 16 unique sampling strategies. These values were chosen to allow for even grid-based sampling since they are all square numbers (e.g., a 6 by 6 grid-based can be used to obtain 36 samples). Applied across all 960 simulations, this resulted in 15,360 individual-based landscape genomic datasets. We performed random sampling by picking random individual samples from across the landscape. We performed grid-based sampling by breaking up the landscape layer into a grid based on the number of points to be sampled and sampling each cell randomly. For example, to sample 36 points, we divided the landscape into a 6 by 6 grid and sampled one random point from each cell. We performed transect-based sampling by taking random samples along three parallel and equidistant transects across the landscape. We performed E-space sampling as described above, selecting from 1000 sets of random samples of a given sample size. Examples of each sampling scheme are shown in Figure 1.

Under site-based sampling, we chose samples around "sites" selected based on the three possible sampling schemes: random, grid, and E-space sampling. For each scheme, we used 9, 16, or 25 sites for a total of nine unique sampling strategies, and we selected the ten closest samples to each site. These values were chosen to allow for even grid-based sampling since they are all square numbers. Applied across all 960 simulations, this resulted in 8,640 site-based landscape genomic datasets. We performed random and E-space sampling the same way as for individual-based sampling. We performed grid-based sampling by choosing equidistant sites across the landscape based on the number of sites to be sampled. This scheme is analogous to the individual grid-based sampling scheme and approximates a sampling scheme with even coverage of geographic space. Examples of each sampling scheme are shown in Figure 1.

2.4 | Landscape genomic analyses

Using the 24,000 simulated landscape genomic datasets, we evaluated two GEA methods and two methods for estimating IBD and IBE. For GEA analyses, we implemented Latent Factor Mixed Modeling (LFMM2; Caye et al., 2019) and Redundancy Analysis (RDA; Capblancq & Forester, 2021; Forester et al., 2018). For estimating IBD and IBE, we chose Multiple Matrix Regression with Randomization (MMRR; Wang 2013) and Generalized Dissimilarity Modeling (GDM; Ferrier et al. 2007; Fitpzatrick & Keller, 2015; Mokany et al., 2022). We selected these methods based on their widespread use and applicability to individual-based sampling. We carried out our analyses using R version 4.3.0 (R Core Team, 2023).

2.4.1 | Genotype-environment association analyses

LFMM is a univariate GEA method that tests for significant associations between loci and environmental variables while correcting for unobserved confounders using latent factors (Caye et al. 2019; Frichot et al., 2013). We performed LFMM (Frichot et al., 2013) using the LFMM2 method, implemented in the R package 'lfmm' version 1.1 (Caye et al., 2019).

There are two methods within LFMM, “lasso” and “ridge”, which differ in the penalty used for regularized least-squares minimization (see Caye et al., 2019 and the “lfmm” package documentation). We tested both methods and found that the LFMM “ridge” method generally performed best in terms of TPR and FDR, so we selected the “ridge” method for our analysis (File S6). LFMM also requires the selection of *K* latent factors for use in the model. *K* can be selected based on estimates of population genetic structure from clustering algorithms (Frichot et al., 2013). We used TESS3 to perform *K* selection with the R package “tess3r” version 1.1.0 (Caye et al., 2016). TESS3 provides estimates of population genetic structure while accounting for geography (Caye et al., 2016). For each dataset, we evaluated *K* values from one to nine and automatically determined *K* based on where the TESS3 cross validation score plateaued (Hyseni, 2019; File S6).

For RDA, we followed the procedure described by Capblancq & Forester (2021) using the R package “vegan” version 2.6.4. RDA is a multivariate GEA method that uses constrained ordination to identify covarying allele frequencies associated with multiple environmental variables (Capblanq & Forester, 2021; Forester et al., 2018). In addition to standard RDA, we performed partial RDA (pRDA) conditioning on two genetic PCs to correct for population genetic structure as in Lotterhos (2022).

For both LFMM and RDA, we used a minor allele frequency filter of 0.05, following common practice used with real genomic datasets and because minor allele frequency filters can affect GEA results (Ahrens et al. 2021). We corrected the p-values produced by LFMM and RDA using a false discovery rate correction and determined significance based on an alpha level of 0.05. Since RDA computes p-values based on RDA axes, there is one set of p-values overall, rather than one for each environmental variable. Therefore, to distinguish which loci were associated with which environmental variable, we assigned each locus identified by RDA to the environmental variable it had the strongest correlation with, similarly to Capblancq & Forester (2021). The true positive rate (TPR; the number of correctly identified loci, divided by the total number of adaptive loci) and the false discovery rate (FDR; the number of loci identified incorrectly, divided by the total number of loci identified) were used as evaluation metrics. The total nuber of adaptive loci for the TPR was calculated based on the number of adaptive loci that passed the minor allele frequency filter (we did not count adaptive loci that were

essentially fixed). We calculated the TPR and FDR in two ways: (1) counting any adaptive loci identified as a true positive or (2) counting only adaptive loci identified with the correct environmental variable as a true positive. We will refer to these statistics as the relaxed and strict TPR and FDR, respectively.

2.4.2 | IBE and IBD analyses

We performed MMRR as described in Wang (2013). We performed GDM using the R package "gdm" version 1.5.0.9.1 (Fitzpatrick et al., 2022). Both of these methods estimate IBD and IBE based on the relationships between genetic, geographic, and environmental distances. The key difference between these methods is that GDM fits nonlinear functions, known as I-splines, while MMRR uses linear regression with randomization (Ferrier et al., 2007; Fitzpatrick & Keller, 2015; Mokany et al., 2022; Wang 2013). The coefficients output by these models were used to estimate the contributions of IBD and IBE, and the p-values from their significance procedures were used to determine detectability of IBD and IBE. We used Euclidean genetic, geographic, and environmental distances as inputs for both the MMRR and GDM models. For GDM, genetic distance values must be less than or equal to one, so we rescaled the euclidean genetic distances from zero to one. For variable significance testing we used a total of 50 permutations for both methods.

To evaluate the accuracy of these models, "full" models were built using 1000 randomly selected individuals to approximate "true" values for IBD and IBE using a consistent number of samples across all of the simulations, which varied in total population size. We *post hoc* confirmed that coefficient values plateaued by 1000 samples (File S7). The error of the coefficients of IBD and IBE were calculated by taking the difference between the full model values (i.e., the "true" values) and the sub-sampled model values (i.e., the "observed" values). Because the maximum coefficient value for MMRR is one while for GDM there is no upper limit, the coefficients, and therefore the errors, cannot be directly compared. To make the errors more comparable, we scaled the coefficients by dividing by the maximum full model coefficient value across all of the simulations for each method so the maximum possible coefficient value was one. We calculated the mean absolute error by taking the mean absolute value of the difference between the full and sub-sampled model coefficients. Because there were two environmental variables, we averaged the error for IBE across both environmental variables. We calculated the correlation between the full model coefficients obtained from MMRR and GDM to get an approximation of the agreement between the two methods. We also confirmed *post hoc* that IBD and IBE had time to develop by calculating the proportion of times that IBD and IBE were detected as significant using the full models (File S7).

We calculated the TPR and FDR for detection of IBD and IBE based on the bootstrapped p-values from each method with an alpha cut-off of 0.05. MMRR calculates bootstrapped p-values based on a null distribution of t-values from permuting the rows and columns of the genetic distance matrix. GDM calculates bootstrapped p-values based on a null distribution of deviance-explained values from permuting each predictor variable individually. The TPR was the proportion of times there was a positive detection in both the sub-sampled model and the full model. The FDR was the proportion of times there was a detection in the sub-sampled model that was not shared with the full model. The statistics for both environmental variables were averaged to get the overall IBE TPR and FDR. Sometimes GDM was unable to calculate p-values because (1) the variable coefficient was zero, (2) the variable permutation procedure could not be conducted because more than two variable coefficients in the model were zero, or (3) the variable permutation procedure failed because one of the models used in the calculation could not be fit. In the case of (1), the corresponding variable was treated as non-significant because the coefficient was zero. In the case of (2) and (3), the p-value based statistics TPR and FDR were not calculated since the variable permutation procedure could not be carried out.

2.4 | Summary analyses

We used linear mixed effect models to summarize the results of the landscape genomic analyses. The response variable for these models was the evaluation statistic of interest. The fixed effects were the parameter levels, the sampling scheme, and the number of samples. Separate models were run for each sampling unit (i.e., individual and site-based sampling). A random effect for the random seeds used to generate the sets of simulated landscapes was used to account for similarities in results between landscapes generated using the same random seed. We used a Type III Analysis of Variance (ANOVA) to test for significant differences across the predictor variables using Satterthwaite's degrees of freedom method with the R packages "lme4" version 1.1.34 (Bates et al., 2015) and "lmerTest" version 3.1.3 (Kuznetsova, Brockhoff, & Christensen, 2017). We used the "emmeans" version 1.8.7 (Lenth, 2023) R package to compare the estimated marginal means for the different sampling strategies with a Tukey adjustment for multiplicity.

3 | Results

3.1 | Simulation results

We confirmed that local adaptation occurred across almost all of our simulations based on phenotype-environment correlations. We found a significant association between phenotype and environment for at least

one trait in 93% of the simulations ($p < .05$; File S5). Of the 7% of simulations that did not have significant associations, 100% of them occurred when spatial autocorrelation was low and correlation between layers was high (File S5). We found significant correlations between the adaptive loci and their corresponding environmental variable 63% of the time ($p < .05$; File S5). Spatial autocorrelation had the greatest effect on the strength of local adaptation; phenotype-environment and genotype-environment correlations were much stronger when autocorrelation was high (mean $r_{phenotype} = 0.86 \pm 0.05$, mean $r_{genotype} = 0.32 \pm 0.04$) than low (mean $r_{phenotype} = 0.36 \pm 0.18$, mean $r_{genotype} = 0.09 \pm 0.05$; File S5). Phenotype-environment and genotype-environment correlations were also strengthened by larger population sizes, stronger selection, less migration, and lower environmental correlation ($p < .05$ for all effects), but the magnitude of the fixed effects for these variables were all ≤ 0.1 (File S5). We found that high spatial autocorrelation and low migration led to stronger environmentally-associated neutral structure; mean genotype-environment correlations at non-adaptive loci were stronger under high ($r = 0.12 \pm 0.04$) compared to low ($r = 0.06 \pm 0.03$) autocorrelation and stronger under low ($r = 0.13 \pm 0.04$) compared to high ($r = 0.06 \pm 0.02$) migration (File S5). Based on these results, it is evident that (1) low spatial autocorrelation weakens phenotype- and genotype-environment correlations and (2) low migration strengthens these correlations but generates stronger confounding correlations between neutral loci and the environment.

3.2 | GEA analyses

We evaluated how key landscape and population genetic parameters affected the performance of GEA methods under different sampling strategies. We found that the best conditions for detecting adaptive loci were when migration and spatial autocorrelation were high (Figure S4; Figure S3; File S6). Outside of these conditions, the TPR for LFMM and RDA were frequently zero across all schemes (Figure S4; Figure S5; File S6). The low TPR in scenarios with low spatial autocorrelation and low migration were likely because (1) the strength of local adaptation was weaker under low autocorrelation and (2) the confounding effects of neutral structure were stronger under low migration.

In cases where the TPR was not zero, we found that correction for population structure using partial RDA resulted in reduced TPR compared to standard RDA; the magnitude of the reduction depended on simulated conditions and sampling strategy, but on average the relaxed TPR was reduced by 0.08 using partial RDA (File S6). Given these results, we focus further discussion of RDA on regular and not partial RDA. We also found that, as expected, the relaxed TPR (i.e., based on any adaptive loci identified) were generally higher than the strict

TPR (i.e., based on only adaptive loci identified with the correct environmental variable); however, the difference between the two was relatively small (mean difference of 0.004 for LFMM and 0.01 for RDA; File S6). Our finding of little to no difference between strict and relaxed TPR was consistent even under high environmental correlation (mean difference of less than 0.02 for both methods; File S6), indicating that both methods were able to distinguish between the driving environmental variables, even when the variables were correlated. The effects of simulation conditions and sampling strategy were consistent between the strict and relaxed TPR and FDR, so henceforth we refer to the strict statistics, unless specified otherwise (File S6).

Our results showed that transect-based sampling performed poorly and had a significantly lower TPR than the other individual-based sampling schemes for both LFMM and RDA (Figure 2, Figure S5, Figure S6; File S6). The estimated difference in TPR between transect-based sampling and the other schemes was approximately 0.02-0.05 across both methods and all simulated conditions ($p < .001$; File S6). The differences were most prominent under the best conditions for GEA analysis (i.e., high autocorrelation, high migration, strong selection, large population size, and low correlation between layers); under these conditions, transect-based sampling led to an average reduction in the TPR of 0.08 for LFMM and 0.09 for RDA compared to other schemes (File S6). Under site-based sampling, the E-Space scheme performed better than random schemes for both methods (estimated difference of 0.03 for LFMM and 0.02 for RDA, $p < 0.001$; Figure 2; File S6). E-space sampling also performed better than grid-based sampling for LFMM (estimated difference of 0.02, $p < .001$; File S6) and had comparable TPR to grid-based sampling for RDA (estimated difference of 0.01, $p = 0.21$; File S6). Under both individual- and site-based sampling, differences in FDR between sampling schemes were insignificant or small (e.g., estimated differences of ≤ 0.05; Figure 2; Figure S5; Figure S6; File S6).

We found that large sample sizes (>100 samples) were necessary for strong GEA performance. However, under unfavorable conditions for GEA (i.e., low spatial autocorrelation and low migration), almost no adaptive loci were detected, even with the largest sample sizes (Figure S4; Figure S5; File S6). Outside of these conditions, the TPR generally increased with increasing sample size and plateaued around 200 samples with TPR of around 0.90 for LFMM and 0.70 for RDA under the best conditions for GEA (Figure 2; File S6). For site-based sampling, with increasing sample size the TPR stayed relatively constant for LFMM at around 0.40 and increased for RDA up to about 0.60 under the best conditions for GEA (Figure 2, Figure S5; File S6). For LFMM, the FDR decreased with increasing sample size for both individual- and site-based sampling but was generally above 0.50 under site-based sampling (Figure 2, Figure S5; File S6). For RDA, the FDR was

universally close to zero for individual-based sampling and less than 0.30 for site-based sampling regardless of sample size (Figure 2, Figure S5; File S6).

3.3 | IBD and IBE analyses

We found that both GDM and MMRR were almost universally able to detect IBD regardless of sampling strategy and simulation conditions (Figure 3, Figure S7; File S7). The best and worst performing scheme in terms of IBD error varied between different simulation conditions; for example, with MMRR, under the best conditions (characterized by lower migration), transect-based sampling produced consistently higher error, but, under the worst conditions (characterized by higher migration), performed no differently than other schemes (Figure 3, Figure S7). For GDM, under the best conditions, grid-based site-based sampling performed similarly to other schemes, but, under the worst conditions, had much lower error than other schemes (Figure 3, Figure S7). Otherwise, we did not observe any notable differences between schemes (Figure 3, Figure S6, Figure S7; File S7). IBD error decreased with increasing sample size and appeared to start plateauing around 225 individual-based samples and 25 site-based samples (Figure 3, Figure S7). Our test for concordance between MMRR and GDM coefficients of IBD found that they were strongly, positively correlated overall ($r = 0.91$, $p < .001$; File S7). The coefficients were most strongly correlated when migration was high and population size was large ($r = 0.60$, $p < .001$; File S7) and most weakly correlated when migration was low and population size was small ($r = 0.21$, $p = 0.001$; File S7).

Our results indicated that detection of IBE was very sensitive to simulation conditions, sample size, and method. The TPR for IBE was highest when migration was low and spatial autocorrelation was high (Figure S7; File S7). When migration was high and spatial autocorrelation was low, IBE did not establish in most cases, based on the full models (e.g., MMRR had a full model IBE detection rate of 13% under these conditions; File S7), as might be expected under these conditions, and therefore detection rates were low overall. GDM had much lower detection rates for IBE than did MMRR (Figure 4, Figure S8; File S7). MMRR was able to detect IBE with intermediate success (e.g., TPR of ~50 - 80% for individual-based sampling; Figure S8) when either migration rates were high and spatial autocorrelation was high or when migration rates were low and spatial autocorrelation was low, but GDM was infrequently able to detect IBE under these conditions (TPR < ~25%; Figure S8). We believe this difference is due to how these methods assess significant relationships (see Discussion). Another difference between results for the two methods is that MMRR sometimes produced negative coefficients for IBE. This is of note since the expectation for IBE is that increasing environmental

distance results in increasing genetic distance, so coefficients of IBE are expected to be positive. Only a small proportion (< 5%) of the negative coefficients for MMRR were significant (File S7). Negative coefficients of IBE were more common when spatial autocorrelation was weak and environmental correlation was high (File S7). Weaker selection strength and lower migration rates also resulted in a greater proportion of negative coefficients, but had a less notable effect (File S7). We found that the full model coefficients of IBE were strongly correlated between MMRR and GDM ($r = 0.67$, $p < .001$; File S7) and the strength of the relationship was consistent across different levels of migration and autocorrelation, suggesting concordance between the approaches (File S7).

We found that the best sampling scheme for detecting IBE varied across method and simulation conditions (Figure 4, Figure S6; Figure S9; File S7). Transect-based sampling had the highest error with MMRR (estimated difference of 0.007-0.008 compared to other schemes, $p < .001$; File S7); however, similarly to IBD error, this difference was mainly found under the best conditions for detecting IBE (Figure 4; File S7) and otherwise transect-based sampling performed similarly to other schemes (Figure S9). Individual-based E-space sampling had higher TPR than random and grid-based sampling (estimated difference of 0.08-0.09, $p < .001$; File S7) and comparable rates to transect-based sampling ($p = 0.09$; File S7); However, this pattern was most evident under the low spatial autocorrelation and low migration scenario shown in Figure S8 and was not consistent across other simulated conditions. Under site-based sampling, E-space sampling had the highest TPR for MMRR (estimated difference of 0.14 compared to grid-based sampling and 0.05 compared to random-based sampling, $p < .001$; File S7), but, again, this difference was inconsistent across simulated conditions (Figure S8; File S7).

For GDM, the TPR for IBE was often undefined because IBE was not detected in the full GDM models (i.e., the denominator for the TPR calculation was zero; Figure S9; File S7). The TPR was also occasionally undefined in cases where the variable significance procedure could not be carried out because more than two variable coefficients in the model were zero or because one of the models used in the calculation could not be fit (File S7). Overall, the TPR was only consistently defined when spatial autocorrelation was high and migration was low (Figure 4; File S7).

We found that detection of IBE required large sample sizes (>100 samples). When individual-based sample sizes were approximately 100 or greater, IBE was detected almost 100% of the time by MMRR and about 50% of the time at best by GDM (Figure 4; File S7), when migration was low and spatial autocorrelation was high. Under the same conditions, for the largest site-based sample sizes, IBE was detected less than 50% of

the time by MMRR and less than 10% of the time by GDM (Figure 4; File S7). We found that more migration and lower spatial autocorrelation resulted in lower detection rates using both MMRR (for individual-based sampling: $\beta_{migration} = -0.43$, $p < .001$; $\beta_{spatial\ autocorrelation} = 0.27$, $p = <.001$; Figure S3; File S7) and GDM (for individual-based sampling: $\beta_{migration} = -0.25$, $p < .001$; $\beta_{spatial\ autocorrelation} = 0.12$, $p < .001$; Figure S3; File S7). When migration was high and spatial autocorrelation was low, IBE did not consistently establish in our simulations, as expected under conditions of high gene flow and low environmental structure (Figure S8). For both MMRR and GDM, IBE error decreased with increasing sample size and plateaued around 225 individuals and 25 sites (Figure 4, Figure S9; File S7).

4 | Discussion

We found that landscape genomic methods were largely robust to sampling schemes. However, sampling to maximize environmental space and minimize sample spatial autocorrelation (i.e., E-space sampling) performed better than or comparably to other sampling schemes, while transect-based sampling led to consistently worse results. Sufficient sample size (>100 samples) was critical for detecting adaptive loci and IBE, but not for detecting IBD. Our results show that landscape structure and migration have significant effects on the performance of landscape genomic analyses. Even though local adaptation occurred across most of our simulated conditions, high environmental spatial autocorrelation and strong migration were the only conditions under which we were able to consistently identify adaptive loci and high spatial autocorrelation and weak migration were the only conditions under which we could consistently detect IBE.

4.1 | The effect of sampling strategy on GEA analyses

Transect-based sampling resulted in lower detection ability for GEA analyses (Figure 2). We believe this is due to the irregular distribution of geographic distances and potential gaps in coverage of the environmental gradient caused by sampling along transect lines. Otherwise, we found that the differences between schemes under individual-based sampling were minor in most cases (e.g., absolute differences in TPR of < 0.02; Figure 2; Figure S5), likely because they all covered large parts of geographic and environmental space. However, there are other schemes which we did not evaluate, because they have known inadequacies, such as highly clustered sampling or sampling with large gaps, which we expect would exhibit substantially worse performance than the schemes we tested here. Altogether, we suggest that landscape gnomic researchers may not need to be too

concerned about picking between sampling schemes (e.g., random versus E-space) so long as they gather enough samples to sufficiently cover environmental and geographic space.

Differences between sampling schemes were more substantial under site-based sampling, likely because there were gaps in coverage of geographic and environmental space caused by sampling fewer locations overall. For site-based sampling, we found that E-space sampling resulted in significantly higher TPR compared to random sampling (estimated difference of 0.03 for LFMM and 0.02 for RDA, $p < 0.001$ for both; Figure 2; File S6) and had comparable or higher TPR than grid-based sampling (estimated difference of 0.02 for LFMM and 0.01 for RDA, $p < .001$ and $p = 0.21$, respectively; Figure 2; File S6). These differences were especially pronounced under the best conditions for detecting genotype-environment associations (high autocorrelation and high migration); for example, with RDA using 25 sites, the average TPR was 0.62 for E-space sampling and 0.55 for random sampling (File S6). This is in line with findings of Lotterhos & Whitlock (2015) who tested an analogous approach of sampling pairs that were geographically close but environmentally distant and found that sampling to maximize adaptive differences and minimize neutral distances resulted in increased power for GEA. When researchers are aiming to capture patterns across more than a couple of environmental gradients, sampling to maximize environmental coverage may be even more important.

We found that RDA had lower FDR than LFMM, while LFMM had higher TPR than RDA (Figure 2; Figure S5). When the number of samples was small (i.e., less than 114 samples or 25 sites), LFMM suffered from extremely high FDR (Figure 2; Figure S5). In contrast, RDA had universally low FDR at large sample sizes (more than 114 samples or 25 sites) and moderate FDR at small sample sizes (Figure 2; Figure S5). Ahrens et al. (2021) similarly found that LFMM had high false positive rates. Ahrens et al. (2021) also found that RDA was unable to detect any loci, correctly or incorrectly, in their analyses, which is similar to our findings that RDA had generally lower TPR and FDR than LFMM. In terms of selecting between regular and partial RDA, we found that correcting for structure with PCs using partial RDA resulted in a loss of power, similarly to Forester et al. (2018) and Lotterhos (2022). TPR for both methods increased with increasing sample size and plateaued around 200 samples (Figure 2), which is in line with findings by Forester et al. (2018), Oyler-McCance et al. (2013), and Selmoni et al. (2020). Our comparison between the relaxed versus strict TPR revealed only minor differences between the two (average difference of < 0.01; File S6), demonstrating that these methods were able to distinguish between the environmental variables driving selection on each trait, even when the variables were correlated ($r = 0.6$). This is a promising finding as the ability to distinguish between environmental drivers of selection in scenarios where the variables of interest are correlated is of key interest in landscape genomics;

however we caution that at levels of environmental correlation higher than simulated here ($r > 0.6$), which can be the case for variables such as temperature and precipitation (File S2), it is still likely that GEA methods would not be able to distinguish between the environmental drivers.

We were unable to detect almost any adaptive loci when migration was weak or when the environmental variables had low levels of spatial autocorrelation (Figure S4; Figure S5). We found that under low spatial autocorrelation, the signal of local adaptation was much weaker based on phenotype-environment and genotype-environment correlations, while under low migration, the signal of confounding population genetic structure was stronger based on correlations between non-adaptive loci and the environment (File S5). Even with large sample sizes, we were unable to identify any loci correctly under these conditions (Figure S4; Figure S5). In these scenarios, it is more conservative to use RDA due to the lower likelihood of false discoveries. However, if the goal is maximizing the detection of adaptive loci and the tradeoff of false discoveries is acceptable, we advise using LFMM, since RDA is unlikely to detect any loci at all (Figure S4; Figure S5). Forester et al. (2016) similarly found that higher levels of spatial autocorrelation corresponded to stronger local adaptation and better GEA analysis performance. However, while Forester et al. (2016) found that high dispersal resulted in worse GEA analysis performance, we found the opposite effect (File S6). It is challenging to pinpoint the source of this difference because our simulations have several key differences; Forester et al. (2016) simulated a single adaptive locus governing a single trait on a binary environmental landscape, while we simulated multiple adaptive loci governing multiple traits on continuous landscapes. A likely explanation is that we did not simulate high enough migration rates to reach the level of swamping gene flow described by Forester et al. (2016) and that if we continued to increase the migration rate we would see similar results. In the same vein, it ipossible that Forester et al. (2016) did not uncover the strong confounding effects of population structure at lower levels of dispersal that we did because a single adaptive locus of strong effect is easier to detect than several loci with smaller effects (Lotterhos, 2023).

A limitation of our results is that we did not evaluate the effects of different genetic architectures, although the prevailing genetic architecture underlying traits that drive environmental adaptation remains largely unknown. In general, polygenic architectures comprising many alleles of small effect may be common in environmental adaptation (Savolainen et al., 2013), but prolonged divergent selection may give rise to 'clustered' architectures in which loci of small effect cluster into tightly linked haplotypes of larger effect (Yeaman, 2022). Our simulations approximate these clustered architectures as an oligogenic system. While we do not vary the numbers and effect sizes of the loci involved, and thus do not explore the influence of genetic architecture on our

results, we expect doing so would generate similar results to Lotterhos (2023): increasing polygenicity would likely decrease the appearance of clear clines and make GEA largely unsuccessful. Landscape genomic researchers should bear this in mind, as the genomic architecture underlying ecological traits of interest is often unknown but may frequently be polygenic (Savolainen et al., 2013), presenting a significant hurdle for GEA.

Another caveat of our study is that we do not know how well our simulated selection strengths align with real environmental selection, because the strength of environmental selection in nature remains difficult to quantify. We controlled selection strength in our simulations using the phenotypic selection coefficient (ϕ). In Geonomics, an individual's fitness is determined by the product of ϕ and the degree of phenotype-environment mismatch. We set the levels of ϕ at 0.5 and 1.0 to observe the effect of halving the strength of selection. We found that weaker selection corresponded to lower detection rates of adaptive loci, as expected, and selection strength does not appear to have any notable interacting effects with sampling scheme.

4.2 | The effect of sampling strategy on IBD and IBE methods

Across all of the simulated conditions and sampling regimes, we were able to detect IBD consistently using both MMRR and GDM, even with as few as nine sites (Figure 3; Figure S7). For MMRR, transect-based sampling resulted in higher error when IBD was stronger (e.g, under low migration and high spatial autocorrelation; Figure 3; Figure S7). For site-based sampling, there were no notable differences between schemes for either method, except grid-based sampling led to lower error for GDM when IBD was weaker (e.g., under high migration and low spatial autocorrelation; Figure 3; Figure S7). Error in estimating the coefficient of IBD decreased with increasing sample size and began plateauing around 225 individual-based samples and 25 site-based samples (Figure 3; Figure S7).

In contrast to IBD, detecting and estimating IBE proved to be more challenging. We found that detection of IBE was most successful when migration was low and spatial autocorrelation was high, which makes sense given that these conditions create greater population genetic structure and stronger local adaptation and therefore stronger IBE (Figure 4; File S7). When one of these conditions (i.e., low migration or high spatial autocorrelation) was met, but not the other, IBE was detected at intermediate frequency using MMRR but almost never using GDM (File S7). When neither condition was met, IBE was rarely detected by either method (Figure S9).

GDM generally had much lower levels of detection for IBE compared to MMRR (Figure 4; Figure S9). We believe this difference may be due to how these methods determine significance. MMRR calculates p-values

based on permuting the rows and columns of the genetic distance matrix while GDM calculates p-values based on permuting each predictor variable individually (Fitzpatrick et al., 2022; Wang, 2013). The GDM permutation test is likely more conservative than that of MMRR. Despite these differences in detection, the coefficients of IBE were strongly correlated between GDM and MMRR ($r > 0.6$; File S6), suggesting that estimates of IBE from the two methods are comparable.

To optimize the sampling approach for the detection and estimation of IBE, we recommend using E-space sampling. E-space sampling resulted in better or similar detection of IBE compared to other individual- and site-based sampling schemes (Figure 4; Figure S8; Figure S9). We suggest avoiding transect-based sampling to detect and estimate IBE, as this scheme had higher error in many cases, especially for MMRR (Figure 4). We recommend that 100 or more individual samples or 50 or more sites be used for detecting IBE with MMRR (Figure 4). For GDM, larger sample sizes are needed, but even with large sample sizes IBE detection rates may be low (Figure 4; Figure S9).

4.3 | Conclusions

The performance of landscape genomic analyses is shaped by the interactions between sampling strategy, population dynamics, and landscape structure. Through individual-based, forward-time simulations on realistic landscapes we show that, as long as sampling covers sufficient environmental and geographic space, differences between sampling schemes are likely to be minimal. To optimize performance, we recommend using E-space sampling, as it performs better than or comparably to other sampling schemes. We also recommend avoiding transect-based sampling, which consistently produces the worst results. We find that having over 100 samples, collected individually or by site, is essential for GEA analysis and for detecting and estimating IBE, but is not necessary for detecting and estimating IBD. When spatial autocorrelation and migration are weak, RDA and LFMM fail to detect adaptive loci, regardless of sample scheme and size, and when spatial autocorrelation is weak and migration is strong, GDM and MMRR fail to detect IBE, but still manage to detect IBD. Otherwise, when spatial autocorrelation and migration are high and sampling adequately covers environmental and geographic space, we find that GEA methods are able to detect around 80% of the loci underlying multivariate adaptation. Our results demonstrate the importance of simulating complex population genetic and landscape scenarios when evaluating sampling strategies, as the relative performance of different strategies varied dramatically, and often idiosyncratically, under different conditions. Altogether, we show that landscape genomic

practitioners should focus on maximizing overall sample size using any even sampling scheme, especially in scenarios where landscape structure is weak.

Acknowledgments

We thank K. Brock, E. A. Chambers, J. McLaughlin, B. Karin, E. Westeen, D. Oliveira, P. Kalhori, R. Nielsen, and E. Rosenblum for providing their feedback and insight on this manuscript. We thank M. Windu for their endurance throughout the long development of this study and the many, many compute hours. 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. APB was supported by the Philomathia Graduate Student Fellowship in Environmental Sciences (UC Berkeley), the William and Lisa Liu Fellowship for Environmental Studies (UC Berkeley), and the ARCS Scholar Award. DETH was partially supported by a Bezos Earth Fund grant awarded to The Nature Conservancy.

References

Ahrens, C. W., Jordan, R., Bragg, J., Harrison, P. A., Hopley, T., Bothwell, H., Murray, K., Steane, D. A., Whale, J. W., Byrne, M., Andrew, R., & Rymer, P. D. (2021). Regarding the F-word: The effects of data filtering on inferred genotype-environment associations. Molecular Ecology Resources, *21*(5), 1460–1474. https://doi.org/10.1111/1755-0998.13351

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

[dataset] Bishop, A. P., Terasaki Hart, D. E., & Wang, I. J.; 2; D023; Data from: Optimizing sampling design for landscape genomics; Dryad; doi:10.5061/dryad.63xsj3v8s

Capblancq, T., & Forester, B. R. (2021). Redundancy Analysis (RDA): A Swiss Army knife for landscape genomics. *Methods in Ecology and Evolution*, 2041-210X.13722. https://doi.org/10.1111/2041-210X.13722

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. https://doi.org/10.1111/1755-0998.12471

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. https://doi.org/10.1093/molbev/msz008

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*, S0169534722002774. https://doi.org/10.1016/j.tree.2022.10.010

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*. https://doi.org/10.1111/mec.12182

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. https://doi.org/10.1111/j.1472-4642.2007.00341.x

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*(12), 4302–4315. https://doi.org/10.1002/joc.5086

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. https://doi.org/10.1111/ele.12376
Fitzpatrick, M., Mokany, K., Manion, G., Nieto-Lugilde, D., & Ferrier, S. (2022). gdm: Generalized Dissimilarity Modeling [R package version 1.5.0-9.1]. Retrieved from https://CRAN.R-project.org/package=gdm
Forester, B. R., Jones, M. R., Joost, S., Landguth, E. L., & Lasky, J. R. (2016). Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Molecular Ecology*, *25*(1), 104–120. https://doi.org/10.1111/mec.13476
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. https://doi.org/10.1111/mec.14584
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. https://doi.org/10.1093/molbev/mst063
Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., Poss, M. L., Reed, L. K., Storfer, A., & Whitlock, M. C. (2016). Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *The American Naturalist*, *188*(4), 379–397. https://doi.org/10.1086/688018
Hyseni, C. (2019). Determining the Best K for Your Clustering Algorithm. Retrieved from https://chazhyseni.github.io/NALgen/post/determining_bestk/
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13
Lenth, R. (2023). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.8.7, https://CRAN.R-project.org/package=emmeans
Li, Y., Zhang, X.-X., Mao, R.-L., Yang, J., Miao, C.-Y., Li, Z., & Qiu, Y.-X. (2017). Ten Years of Landscape Genomics: Challenges and Opportunities. *Frontiers in Plant Science*, *8*, 2136. https://doi.org/10.3389/fpls.2017.02136
Lotterhos, K. E. (2023). The paradox of adaptive trait clines with nonclinal patterns in the underlying genes. *Proceedings of the National Academy of Sciences*, *120*(12), e2220313120. https://doi.org/10.1073/pnas.2220313120
Lotterhos, K. E., & Whitlock, M. C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology*, *24*(5), 1031–1046. https://doi.org/10.1111/mec.13100
Manel, S., Albert, C. H., & Yoccoz, N. G. (2012). Sampling in Landscape Genomics. In F. Pompanon & A. Bonin (Eds.), *Data Production and Analysis in Population Genomics* (Vol. 888, pp. 3–12). Humana Press. https://doi.org/10.1007/978-1-61779-870-2_1
Mokany, K., Ware, C., Woolley, S. N. C., Ferrier, S., & Fitzpatrick, M. C. (2022). A working guide to harnessing generalized dissimilarity modelling for biodiversity analysis and conservation assessment. *Global Ecology and Biogeography*, *31*(4), 802–821. https://doi.org/10.1111/geb.13459
Nield, A. P., Nathan, R., Enright, N. J., Ladd, P. G., & Perry, G. L. W. (2020). The spatial complexity of seed movement: Animal-generated seed dispersal patterns in fragmented landscapes revealed by animal movement models. *Journal of Ecology*, *108*(2), 687–701. https://doi.org/10.1111/1365-2745.13287
Oyler-McCance, S. J., Fedy, B. C., & Landguth, E. L. (2013). Sample design effects in landscape genetics. Conservation Genetics, *14*(2), 275–285. https://doi.org/10.1007/s10592-012-0415-1
Price, N., Lopez, L., Platts, A. E., & Lasky, J. R. (2020). In the presence of population structure: From genomics to candidate genes underlying local adaptation. *Ecology and Evolution*, *10*(4), 1889–1904. https://doi.org/10.1002/ece3.6002
R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/
Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, *24*(17), 4348–4370. https://doi.org/10.1111/mec.13322
Russo, S. E., Portnoy, S., & Augspurger, C. K. (2006). Incorporating Animal Behavior into Seed Dispersal Models: Implications for Seed Shadows. *Ecology*, *87*(12), 3160–3174. http://www.jstor.org/stable/20069345

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, *14*(11), 807–820. https://doi.org/10.1038/nrg3522
Sciaini, M., Fritsch, M., Scherer, C., & Simpkins, C. E. (2018). NLMR and landscapetools: An integrated environment for simulating and modifying neutral landscape models in R. *Methods in Ecology and Evolution*, *9*(11), 2240–2248. https://doi.org/10.1111/2041-210X.13076
Selmoni, O., Vajana, E., Guillaume, A., Rochat, E., & Joost, S. (2020). Sampling strategy optimization to increase statistical power in landscape genomics: A simulation-based approach. *Molecular Ecology Resources*. https://doi.org/10.1111/1755-0998.13095
Terasaki Hart, D. E., Bishop, A. P., & Wang, I. J. (2021). Geonomics: Forward-Time, Spatially Explicit, and Arbitrarily Complex Landscape Genomic Simulations. *Molecular Biology and Evolution*, msab175. https://doi.org/10.1093/molbev/msab175
Uriarte, M., Anciães, M., Da Silva, M. T. B., Rubim, P., Johnson, E., & Bruna, E. M. (2011). Disentangling the drivers of reduced long-distance seed dispersal by birds in an experimentally fragmented landscape. *Ecology*, *92*(4), 924–937. https://doi.org/10.1890/10-0709.1
Van Rossum, G., & Drake, F. L. (2009). Python 3 Reference Manual. Scotts Valley, CA: CreateSpace.
Wang, I. J. (2013). Examining the full effects of landscape heterogeneity on spatial genetic variation: A multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution*. https://doi.org/10.1111/evo.12134
Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*. https://doi.org/10.1111/mec.12938
Wright, S. (1943). Isolation by Distance. *Genetics*, *28*(2), 114–138.
Yeaman, S. (2022). Evolution of polygenic traits under global *vs* local adaptation. *Genetics*, *220*(1), iyab134. https://doi.org/10.1093/genetics/iyab134

Data Accessibility Statement

All code used in this study can be found on GitHub (https://github.com/AnushaPB/LandGenSamp) and is archived on Zenodo (doi:10.5281/zenodo.14009717). The final copies of the simulation files can be found on Dryad (doi:10.5061/dryad.63xsj3v8s; Bishop et al., 2024).

Benefit-Sharing Statement

Benefits from this research accrue from the public sharing of our data and results, as described above.

Author Contributions

APB and IJW conceived of the study. APB conducted the simulations and analyses, with input from DETH and IJW. APB wrote and DETH and IJW co-wrote the manuscript; all authors reviewed and edited the manuscript.

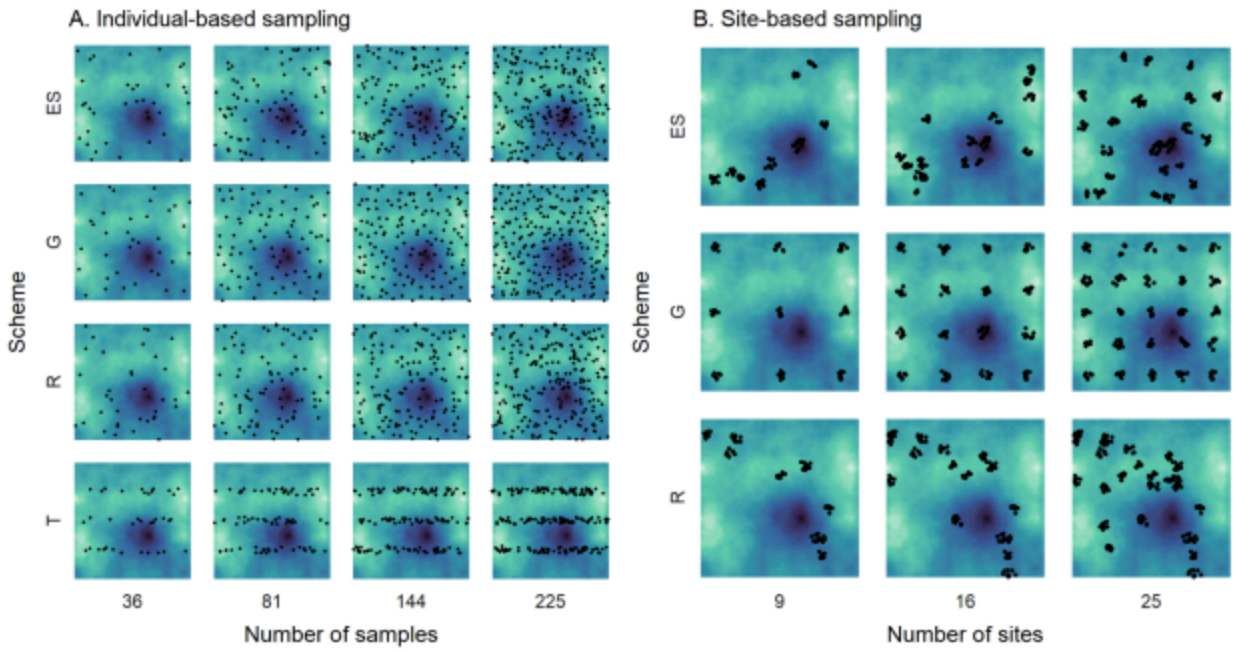

Figure 1.

Example of different (A) individual-based and (B) site-based sampling strategies on an example landscape with high spatial autocorrelation. Abbreviations: transect-based (T), random (R), grid-based (G), E-space (ES).

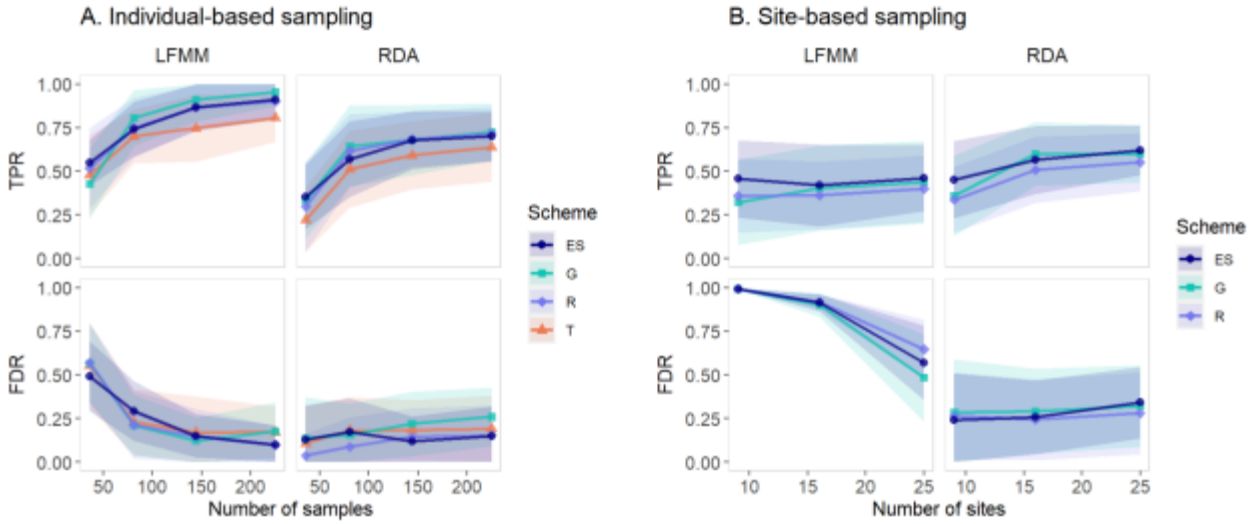

Figure 2. Selected simulation results showing the effect of sampling strategy on TPR and FDR for GEA analyses under the best conditions. Sampling strategy includes sampling units (i.e., individual and site-based; panels A and B, respectively), sampling scheme (e.g., random), and sample size. The best conditions are high spatial autocorrelation, low environmental correlation, high migration rate, large population size, and strong selection. To see results under the worst conditions, see Figure S5. Results for all other simulated conditions and statistics can be found in File S6. Each point represents the average of all of the simulation iterations and landscape seeds. The bands represent one standard deviation from the mean. Abbreviations: Redundancy Analysis (RDA), Latent Factor Mixed Models (LFMM), transect-based (T), random (R), grid-based (G), E-space (EG), strict true positive rate (TPR), and strit false discovery rate (FDR).

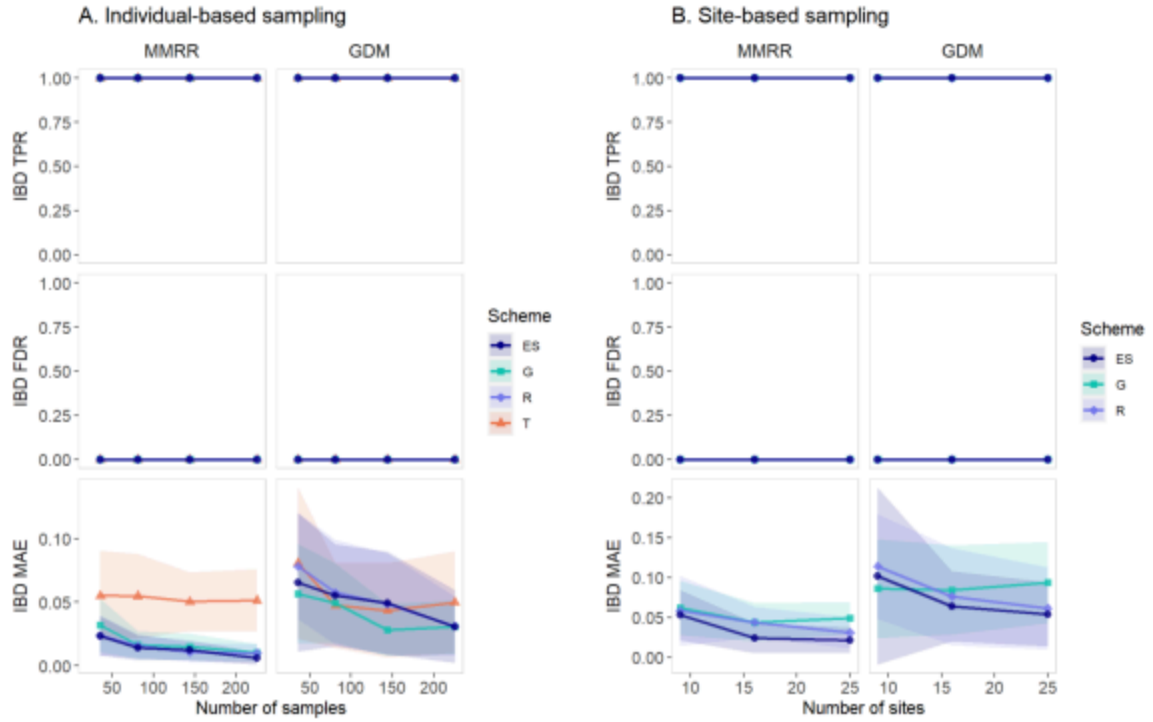

Figure 3. Selected simulation results showing the effect of sampling strategy on detection and estimation of IBD under the best conditions. Sampling strategy includes sampling units (i.e., individual and site-based; panels A and B, respectively), sampling scheme (e.g., random), and sample size. The best conditions are low correlation between layers, low migration rate, small population size, and strong selection. For MMRR, high spatial autocorrelation is also best (i.e., significantly reduces error), but the effect is small (< 0.01; File S7), and therefore for consistency with the IBE plots, high spatial autocorrelation is used in this figure. To see results under the worst conditions, see Figure S7. Results for all other simulated conditions and statistics can be found in File S7. Each point represents the average of all of the simulation iterations and landscape seeds. Abbreviations: Multiple Matrix Regression with Randomization (MMRR), Generalized Dissimilarity Model (GDM), transect-based (T), random (R), grid-based (G), E-space (ES), true positive rate (TPR), false discovery rate (FDR), and mean absolute error (MAE).

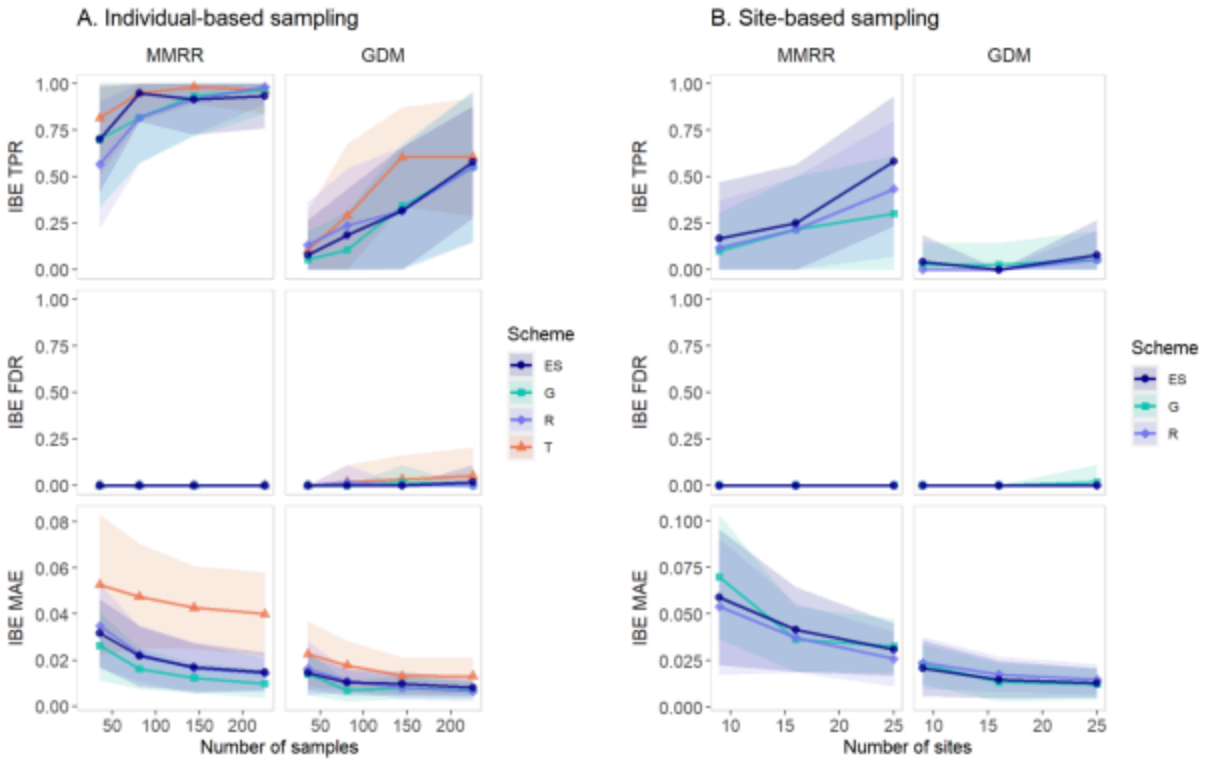

Figure 4. Selected simulation results showing the effect of sampling strategy on detection and estimation of IBE under the best conditions. Sampling strategy includes sampling units (i.e., individual and site-based; panels A and B, respectively), sampling scheme (e.g., random), and sample size. The best conditions are high spatial autocorrelation, low environmental correlation, low migration rate, small population size, and strong selection. To see results under the worst conditions, see Figure S9. Results for all other simulated conditions and statistics can be found in File S7. Each point represents the average of all of the simulation iterations and landscape seeds. Abbreviations: Multiple Matrix Regression with Randomization (MMRR), Generalized Dissimilarity Model (GDM), transect-based (T), random (R), grid-based (G), E-space (ES), true positive rate (TPR), false discovery rate (FDR), and mean absolute error (MAE).