Received: 3 April 2017 | Accepted: 19 December 2017

DOI: 10.1111/1755-0998.12747

RESOURCE ARTICLE

WILEY MOLECULAR ECOLOGY RESOURCES

Evaluating methods to visualize patterns of genetic differentiation on a landscape

Geoffrey L. House | Matthew W. Hahn

Indiana University Bloomington, Bloomington, IN, USA

Correspondence
Geoffrey L. House, Los Alamos National Laboratory, Los Alamos, NM, USA.
Emails: ghouse@lanl.gov; geo.house.7@gmail.com

Funding information
Strategic Environmental Research and Development Program (SERDP), Grant/Award Number: RC-2330

Abstract

With advances in sequencing technology, research in the field of landscape genetics can now be conducted at unprecedented spatial and genomic scales. This has been especially evident when using sequence data to visualize patterns of genetic differentiation across a landscape due to demographic history, including changes in migration. Two recent model-based visualization methods that can highlight unusual patterns of genetic differentiation across a landscape, *SpaceMix* and *EEMS*, are increasingly used. While *SpaceMix*'s model can infer long-distance migration, *EEMS*' model is more sensitive to short-distance changes in genetic differentiation, and it is unclear how these differences may affect their results in various situations. Here, we compare *SpaceMix* and *EEMS* side by side using landscape genetics simulations representing different migration scenarios. While both methods excel when patterns of simulated migration closely match their underlying models, they can produce either un-intuitive or misleading results when the simulated migration patterns match their models less well, and this may be difficult to assess in empirical data sets. We also introduce unbundled principal components (*un-PC*), a fast, model-free method to visualize patterns of genetic differentiation by combining principal components analysis (PCA), which is already used in many landscape genetics studies, with the locations of sampled individuals. *Un-PC* has characteristics of both *SpaceMix* and *EEMS* and works well with simulated and empirical data. Finally, we introduce *msLandscape*, a collection of tools that streamline the creation of customizable landscape-scale simulations using the popular coalescent simulator *ms* and conversion of the simulated data for use with *un-PC*, *SpaceMix* and *EEMS*.

KEYWORDS
conservation genetics, ecological genetics, landscape genetics, phylogeography

1 | INTRODUCTION

The combination of decreasing cost and increasing throughput of DNA sequencing has transformed many fields. This is especially true of landscape genetics, where improved access to sequencing has enabled increasingly ambitious studies that sample larger landscapes as well as more individuals across the landscape than was possible just a few years ago (e.g., Miller et al., 2016; Pagani et al., 2016; Roffler et al., 2016). For instance, for a growing number of species, it is now possible to observe landscape-scale patterns of genetic differentiation, which represent long-term records of evolutionary history. Analysing these patterns of genetic differentiation is important to help understand the demographic history of populations as they may relate to geographical features on the landscape (Richardson, Brady, Wang, & Spear, 2016), and this spatial information can be leveraged to improve the genetic clustering of individuals into populations (Bradburd, Coop, & Ralph, 2017).

The vanguard attempt to visualize patterns of genetic variation across a landscape was made 40 years ago. Menozzi, Piazza, and Cavalli-Sforza (1978) used principal components analysis (PCA) to summarize the patterns in multiple allozyme markers sampled from humans across Europe and then interpolated each individual's score from a single principal component (eigenvector) across the landscape. They used patterns found in these interpolated genetic maps to make inferences about past migration routes and timing. However, patterns arising from the interpolation of single principal components were later shown to be artefacts that consistently occur during PCA calculation, especially due to the orthogonality constraints imposed by PCA (DeGiorgio & Rosenberg, 2013; Novembre & Stephens, 2008). Instead, if two principal components, typically PC 1 and PC 2, are plotted against each other, which we refer to here as a "PCA biplot," these artefacts are lessened (Novembre et al., 2008). When the dispersal of organisms is spatially limited, the amount of genetic differentiation between populations increases with geographic distance, and PC 1 and PC 2 from the genetic data can become surrogates for latitude and longitude (or vice versa). Under these conditions, PCA biplots may provide a striking match with the geographic locations of sampled individuals (Figure 1; Novembre et al., 2008; Wang, Zöllner, & Rosenberg, 2012). Due to its combination of utility and speed of calculation, PCA is now commonly used as a diagnostic tool to examine patterns of genetic variation before model-based (and more computationally intensive) analyses are undertaken.

The advantage of model-based analyses is that they can locate areas on the landscape that have anomalous patterns of genetic differentiation; although there are model-free methods such as *MAPI* (Piry et al., 2016) that also aim to highlight such areas, PCA biplots by themselves cannot serve this purpose. Two recently introduced model-based methods that can use both thousands of genetic markers (e.g., single nucleotide polymorphisms; SNPs) and thousands of sampled individuals are *SpaceMix* (Bradburd, Ralph, & Coop, 2016) and *EEMS* (Petkova, Novembre, & Stephens, 2016). Other model-based methods not tested here include *LocalDiff* (Duforet-Frebourg & Blum, 2014) and *SPA* (Yang, Novembre, Eskin, & Halperin, 2012). Both *SpaceMix* and *EEMS* fit an underlying model to the observed data in order to identify populations that either have anomalously weak or strong genetic differentiation that may indicate differences in demographic history, such as changes in effective population size or migration. Although these two methods differ in their details, they both fit their models using Markov chain Monte Carlo (MCMC) sampling to explore the large parameter spaces required. *SpaceMix* is unique in inferring long-distance migration events that occurred between pairs of geographically separated populations on the landscape. This includes inferring the direction of migration and the proportion of admixture in the recipient population that was caused by long-distance migration from the source population. *EEMS*, on the other hand, is built to visualize patterns of genetic differentiation over shorter distances by fitting a continuous surface of estimated gene flow across the landscape. Both *EEMS* and *SpaceMix* are being used increasingly frequently in empirical studies (Montinaro et al., 2016; Richmond et al., 2017; Tsuda et al., 2016; Uren et al., 2016; Yoder et al., 2016), but there has been no comparison of the two methods to determine how they perform in situations where the underlying patterns of genetic differentiation poorly match their models, or how these situations appear in their visualizations. Here, we sought to determine how the two methods perform using simulated genetic data that represent a range of both short- and long-distance migration scenarios with the goal of better understanding the strengths and limitations of each method.

Because both *EEMS* and *SpaceMix* are model-based and use MCMC methods to explore their large parameter spaces, they are computationally intensive and slow. Therefore, before these methods are run, it would be useful to have a fast diagnostic method that can visualize major patterns of genetic differentiation across a landscape. To fill this gap, here we introduce unbundled principal components

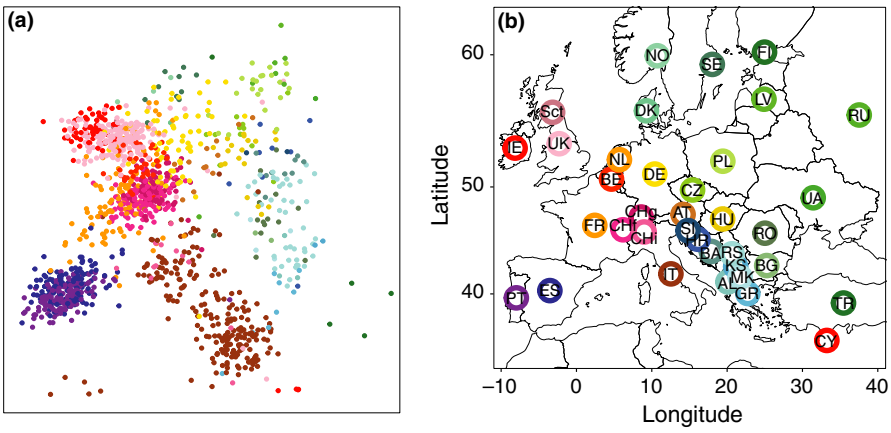

FIGURE 1 PCA biplot and map of countries or populations used in the analysis of the POPRES genetic data set, coloured as in Novembre et al. (2008). (a) Biplot of PC 1 and PC 2 with each point representing a different individual; points are coloured according to the individual's birth country or the birth country of their grandparents. (b) Map of Europe with countries coloured the same as in Panel (a), and with country abbreviations used in Novembre et al. (2008), including splitting Swiss individuals into French-speaking (CHf), Italian-speaking (CHi) and German-speaking (CHg) groups [Colour figure can be viewed at wileyonlinelibrary.com]

(*un-PC*), a model-free method implemented in R that extends PCA results by incorporating the geographic locations of the sampled individuals. *Un-PC* is quick to run and its visualizations can provide insight about genetic differentiation both across short distances, similar to *EEMS*, as well as across longer distances, similar to *SpaceMix*. *Un-PC* is not a substitute for model-based methods like *SpaceMix* or *EEMS*, but it provides a value-added diagnostic tool that leverages the PCA results that are typically already calculated in landscape genetics studies.

Differences in allele frequencies among populations sampled across a landscape, which form the basis for the visualizations of genetic differentiation produced by *EEMS*, *SpaceMix* and *un-PC*, are records of the sometimes complex evolutionary histories of populations. These differences in allele frequencies may be caused by migration, but may also be due to mutation, selection and drift. Although the relative effect of selection may be strong on specific loci, the effect of mutation is thought to be minimal overall (Whitlock & McCauley, 1999), and both factors are generally ignored in demographic inference (Wang & Whitlock, 2003). However, it is difficult if not impossible to separate the effects of drift from migration when studying patterns of genetic variation on a landscape. Because of this, the effective amount of gene flow between populations is often represented as the compound parameter $N_e\,m$, where N_e is the effective population size and m is the migration rate. Any indication of genetic differentiation between populations that is identified by the methods discussed here may therefore be caused by changes in the amount of drift between the populations (due to differences in N_e) rather than any differences in migration, and this would not be identifiable (Slatkin, 1987; Whitlock & McCauley, 1999).

It must also be mentioned that none of the methods described here directly estimate $N_e\,m$. The closest such estimate comes from *EEMS*, which estimates the relative amount of gene flow across the landscape, but not the absolute value of $N_e\,m$ between any populations. *SpaceMix*'s "geogenetic" visualizations should be proportional to $N_e\,m$, but they do not provide any numerical estimates. Results from *un-PC*, because they are not model-based, cannot be interpreted as values of $N_e\,m$.

Here, we use simulated landscape genetics data to compare the three analysis methods, but these simulations also make a number of simplifying assumptions with regard to migration and drift. First, the simulated populations are at migration–drift equilibrium. In nature, this equilibrium may be slow to develop if individuals are moving into new environments (Whitlock & McCauley, 1999). In such cases, visualizations of differences in allele frequencies across a landscape, which represent long-term demographic histories averaged across many generations, may not match the most recent patterns of migration among populations (Slatkin, 1987). Second, the simulated data keep N_e constant, so we are confident that the extent of genetic differentiation between populations identified by each method is due primarily to differences in migration and not due to differences in drift. Because of this constraint on the simulations used, however, we are unable to assess the robustness of the three tested methods to differences in N_e across populations, although *EEMS* appears relatively unaffected by changes in N_e (Petkova et al., 2016).

Using landscape-scale genetic simulations not only allows controlled comparisons of different analysis methods but can also give a better understanding of the factors that may affect patterns of genetic variation in empirical studies. Although simulation programs such as *SPLATCHE2* (Ray, Currat, Foll, & Excoffier, 2010) facilitate landscape-scale simulations, most empirical landscape genetics studies do not include such simulations, possibly because they can be tedious to produce. Both the original *SpaceMix* and *EEMS* papers use the popular coalescent simulator *ms* (Hudson, 2002) to test their methods. The program *ms* is ideal for large spatial simulations because it has no limit to the number of populations simulated, the simulation scenarios are highly customizable and its output format is easy to parse for downstream applications. These spatially explicit *ms* simulations mimic geographic landscapes by allowing different migration rates across them. Similar functionality is provided by *SPLATCHE2* (Ray et al., 2010), although it can also simulate the effects of variation in migration rates through time, which we do not evaluate here. Using *ms*, these landscape-scale simulations are produced by specifying all possible pairwise migration connections between each population to be simulated. Enumerating these connections by hand is laborious and error-prone, and is difficult to either graph or troubleshoot before simulation. To address these problems, here we also introduce *msLandscape*, a toolset to streamline the creation of landscape-scale genetic simulations with *ms*. *msLandscape* enables the easy simulation of essentially any landscape size, shape and sampling pattern using *ms*; simple graphing of the landscape configuration before simulation; and conversion of the *ms* output for subsequent use in PCA/*un-PC*, *EEMS* and *SpaceMix* analyses.

Our goal here is to better characterize the performance of existing methods to visualize patterns of genetic variation across the landscape using genetic data and to facilitate making these spatially explicit visualizations faster and easier. This paper addresses three major aims: (i) to compare the visualized patterns of genetic variation generated by *EEMS* and *SpaceMix* using a range of simulated migration scenarios, (ii) to introduce and evaluate *un-PC*, a model-free method meant as a diagnostic tool that uses PCA results together with sampling locations to quickly visualize overall patterns of genetic variation across the landscape and (iii) to facilitate the greater use of simulations in landscape genetics studies through *msLandscape*, a toolbox that streamlines the generation of landscape-scale simulations using the coalescent simulator *ms*.

2 | MATERIALS AND METHODS

2.1 | *ms* simulations

We generated landscape simulations using the coalescent simulator *ms* (Hudson, 2002), which was also used to test *SpaceMix* and *EEMS* when they were first introduced (Bradburd et al., 2016; Petkova et al., 2016). In empirical landscape genetics studies, the sampled

populations are usually interspersed among intervening, unsampled populations. To incorporate this in the simulations, we arranged 100 sampled populations, each with 10 sampled individuals, in a 10-by-10 grid with each sampled population surrounded by six unsampled populations, and two bordering rings of unsampled populations that enclosed the landscape (Figure 2a). For the simulations, populations were defined as all of the individuals simulated at each spatial location; individuals were not assigned to populations based on genetic clustering.

For the baseline scenario of constant migration, all populations (sampled or unsampled) exchange migrants with each of their adjacent populations (Figure 2a) using the migration parameter $4\ N_e\ m = 3.0$. We also simulated a wide belt of reduced migration, where $4\ N_e\ m$ through the middle of the landscape was reduced to 0.3, or 10% of its value on both edges of the landscape, with narrow transition zones having $4\ N_e\ m = 1.7$ (Figure 2b). Although migration across the middle of the landscape was still possible, for brevity we refer to this scenario as a migration barrier. This scenario could be analogous to a mountain range or an area of anthropogenic disturbance that interrupts suitable habitat patches. Finally, we simulated multiple independent long-distance migration scenarios where either 10%, 30%, 50%, 70% or 90% of the individuals from the sampled recipient population, located in the bottom right corner of the landscape, had arrived directly from the unsampled source population, located in the upper left corner of the landscape (Figure 2c), in the recent past ($0.01 \times 4\ N_e$ generations ago). The migration rates among all other populations were unchanged compared to the constant migration scenario.

For all scenarios, we sampled 1000 individuals (10 individuals from each of the 100 sampled populations) and simulated 10,000 independent SNPs in each individual. For the constant migration and migration barrier scenarios, we also simulated both random uneven sampling of individuals from each sampled population (min = 1, max = 18, mean = 10.2, *SD* = 4.9) and strongly patterned uneven sampling, where individuals were heavily sampled from the right side of the landscape but no individuals were sampled from the left side (min = 1, max = 48, mean = 21.2, *SD* = 14.9). In both cases, the total number of individuals sampled remained 1000 although the number of populations with sampled individuals was 98 for the random uneven sampling and 47 for the strongly patterned uneven sampling (Figure S1). See Table S1 for summary information about the different simulation scenarios.

We completed 100 independent simulation iterations for each scenario to incorporate the stochastic differences inherent in coalescent simulations; the simulated data for each iteration were then analysed independently using each analysis method. Here, we present analysis results and visualizations that are means of the 100 independent simulation iterations for each scenario. Variation between simulation iterations was generally moderate; see Figure S2 for a representative iteration of each scenario and the *msLandscape* website for visualizations from all iterations for each scenario.

2.2 | *msLandscape*

To facilitate the creation and analysis of large custom landscape genetics simulations, we wrote a collection of user-friendly programs in PYTHON and R that together form the *msLandscape* toolbox and workflow (see Figure S3 and the *msLandscape* documentation for additional description and examples). The workflow starts by automatically generating an *ms* parameter file that directs *ms* to simulate a rectangular landscape comprised of hexagonal tiles with fully customizable landscape dimensions. Each of these tiles represents seven populations, with migration simulated between each population and its nearest neighbours. The parameter files that are necessary for *ms*

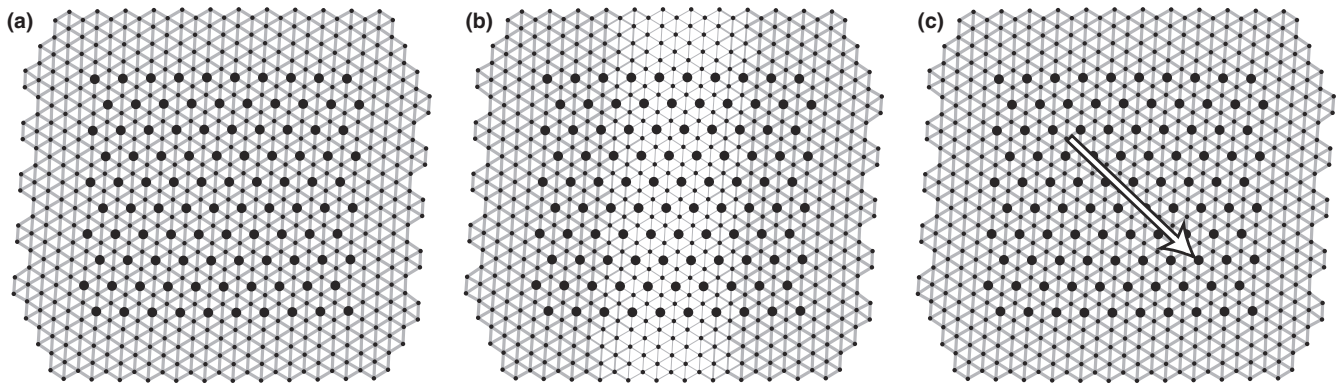

FIGURE 2 Simulated scenarios used to evaluate the different methods of visualizing genetic differentiation across the landscape. Equal numbers of individuals were selected from each sampled population on the landscape (indicated by large black dots), while surrounding unsampled populations are indicated by smaller black dots. Migration connections are indicated using grey lines connecting population pairs, where the width of the line is proportional to the migration rate. The scenarios were as follows: (a) constant migration across the landscape, (b) a vertical migration barrier with ten times lower migration through the middle of the landscape and (c) long-distance migration from an unsampled population in the upper left corner of the grid (source population) to a sampled population in the lower right corner of the grid (recipient population), as indicated by the location and direction of the arrow. All panels are direct output from the *msLandscape_networkPlotter* function in the *msLandscape* R package, except that the arrow denoting the long-distance migration in Panel (c) was manually added

to generate landscape-scale simulations are large and difficult to interpret directly (e.g., Appendices S1–S3 for the parameter files that generate the scenarios presented in Figure 2). To make these files easier to interpret and troubleshoot, *msLandscape* uses the IGRAPH package in R to visualize the landscape configuration of an *ms* parameter file as a network graph with populations as vertices and migration connections as edges. The size of each vertex is proportional to the number of individuals sampled from that population, and the width of each edge is proportional to the specified migration rate between the pair of populations that it connects. The populations in this graph can be optionally labelled using the numbers assigned in the *ms* parameter file, and these numbers can be used to identify populations that should be removed from the rectangular landscape in order to "sculpt" it into virtually any desired shape. This "sculpting" is achieved by automatically removing populations specified by the user from the *ms* parameter file; at the same time, the user can also specify changes to the number of individuals sampled in any population. The network graph can also verify changes to migration rates after manual editing of the *ms* parameter file. Finally, it highlights any population pairs with only a single migration connection between them, denoting unidirectional migration. Once the user is satisfied with the landscape characteristics represented in the *ms* parameter file, it is then ready for input to *ms*. Additional options are available in *msLandscape* to have *ms* produce multiple independent simulations for the same *ms* parameter file and to convert the *ms* output into the appropriate format for multiple downstream analysis tools (Figure S3).

2.3 | *Un-PC*

Un-PC is an analysis method that uses the principal components from any PCA method applied to genetic data (*smartpca* from *EIGENSOFT* (Patterson, Price, & Reich, 2006) is used here) in combination with geographic coordinates of the samples to create visualizations of genetic differentiation across the landscape. Here, we calculated individual-level PCAs, but then averaged the resulting coordinates in PC 1 and PC 2 space for all individuals in each sampled population to create population-level PCA coordinates that were used for *un-PC*. *Un-PC* first calculates the Euclidean distance between the PCA coordinates for each pair of populations to generate a PCA-based genetic distance. The pairwise geographic distance between populations is also calculated. For simulated data sets, the Euclidean distance between population coordinates was used as the geographic distance (though see below for an important modification to these coordinates that we used to help correct for distortion caused by PCA); for empirical data sets, the geographic distance was calculated using the Haversine formula to account for the Earth's curvature. The ratio of the genetic distance to the geographic distance for each pair of populations is the *un-PC* value for that pair. Larger *un-PC* values (larger genetic distance relative to geographic distance) indicate increased genetic differentiation between the populations, while smaller *un-PC* values indicate less genetic differentiation.

Once the *un-PC* values are calculated for each population pair, they collectively form a distribution where population pairs with *un-PC* values in the two tails of the distribution are the ones that represent the most extreme genetic differentiation or similarity and therefore are likely to be the most useful in visualizing patterns across the landscape (Figure S4). *Un-PC* produces two visualizations in geographic space, one highlighting each tail of this distribution. It does this by first connecting each population pair with an ellipse that is coloured based on the *un-PC* value for that population pair. Ellipses with *un-PC* values near the mean of the distribution are coloured white, while ellipses with higher *un-PC* values (greater differentiation) are coloured progressively more pink, and ellipses with lower *un-PC* values (less differentiation) are coloured progressively more green (Figure S4). We used a consistent colour scale for all *un-PC* analyses that used simulated data so that the visualizations from different migration scenarios can be directly compared. One of the *un-PC* visualizations is then produced by stacking the ellipses in order from least to greatest *un-PC* values so that the most genetically differentiated pairwise comparisons (the pinkest ellipses) are on top, and the other visualization is produced by reversing the stacking so that the ellipses with the smallest *un-PC* values, representing the least genetic differentiation (the greenest ellipses), are on top (Figure S4).

PCA biplots of PC 1 and PC 2 produce a known distortion, especially towards their edges, where points are crowded closer together than points in the centre (Figure 3a). This distortion can be well modelled using a two-dimensional discrete cosine transform (DCT; Novembre & Stephens, 2008). By determining the expected location of each population in PCA space using the DCT and then using those transformed coordinates instead of the raw coordinates when calculating the pairwise geographic distance for *un-PC*, the signal from this distortion is strongly reduced (Figure S5).

2.4 | *SpaceMix*

To run *SpaceMix* (Bradburd et al., 2016), we converted the *ms* simulation results for each individual into genotype count and sampling count files that were aggregated by population. We ran *SpaceMix* to estimate the location of each population in "geogenetic" space while also inferring the amount and direction of any long-distance migration between populations. The visualizations presented here use the "ellipses" option in *SpaceMix* to show the 95% credible interval for the location of each population in geogenetic space. They also use the "source" option to automatically plot arrows between populations with inferred long-distance migration, which we defined here to be populations where the inferred proportion of admixture due to long-distance migration was more than three standard deviations above the mean for all populations, and which had an absolute value >0.025. These modifications removed weak, erroneous signals of long-distance migration in the simulation results; no modifications were made for the empirical data set. We provided *SpaceMix* with the geographic coordinates of the populations to act as geographic priors we ran all *SpaceMix* analyses using 10 initial fast runs, each with 100,000 iterations followed by a long MCMC run of 2.2 million

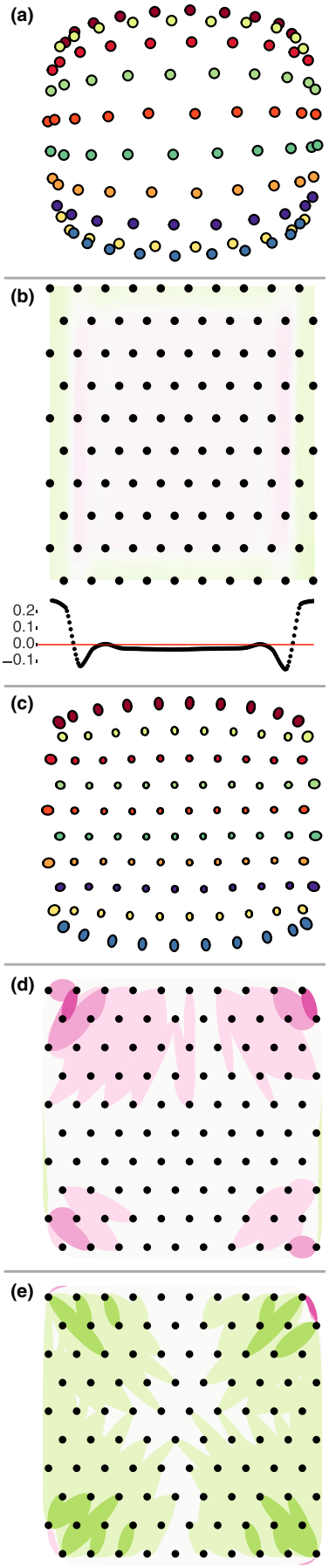

FIGURE 3 Visualizations of genetic differentiation for the constant migration scenario with even-sampling (Figure 2a). (a) Biplot of PC 1 and PC 2 with each row of 10 sampled populations marked using an arbitrarily chosen colour for clarity. (b) *EEMS* with areas of estimated decreased genetic differentiation in green, areas of estimated increased genetic differentiation in pink and areas with average estimated genetic differentiation in white; the darker the colour, the stronger the deviation from the estimated average. Black dots denote population locations. The graph below the landscape is the mean estimated relative gene flow on the landscape, averaged from top to bottom, with increases in estimated relative gene flow corresponding to decreases in estimated genetic differentiation and vice versa. (c) *SpaceMix* with each ellipse representing 95% credible intervals around the estimated location of each population in geogenetic space. Sampled populations are coloured using the same colours as the PCA biplot in Panel (a). (d) *Un-PC* with the most differentiated ellipses stacked on top; like *EEMS*, the darker pink colours represent population pairs with increased genetic differentiation (higher *un-PC* scores). (e) *Un-PC* with the least differentiated ellipses stacked on top; like *EEMS*, the darker green colours represent population pairs with decreased genetic differentiation (lower *un-PC* scores). Visualizations from *EEMS* and *un-PC* are expected to be uniformly white in this scenario, indicating no differences in genetic differentiation across the landscape. However, they both show artefacts at the edges of the landscapes, and therefore the visualizations from this scenario form a baseline to compare against the results from other simulated migration or sampling scenarios. All visualizations are the mean from 100 simulation iterations [Colour figure can be viewed at wileyonlinelibrary.com]

iterations that was sampled every 4,000 iterations. The first 200 samples were discarded as burn-in, leaving 350 samples per analysis run.

2.5 | *EEMS*

In order to run *EEMS* (Petkova et al., 2016), we converted the simulated genotype data into a matrix of individual-level pairwise genetic differences following the method used in Petkova et al. (2016) and implemented in R as part of the *msLandscape* toolbox. We specified the population membership of each individual, the population coordinates as geographic priors, similar to *SpaceMix*, and a triangular grid connecting all sampled populations. To generate more detailed visualizations, we directly used the raster output of *EEMS* that estimates the relative amount of genetic differentiation across the landscape (on a log scale) instead of the colour-based contour plots used in the original paper (Petkova et al., 2016). Raster cells having estimated values close to the mean across the landscape are coloured white; cells with higher estimated differentiation (lower relative gene flow) are coloured progressively darker pink, while cells with lower estimated differentiation (higher relative gene flow) are coloured progressively darker green. Below each landscape plot is a dot plot of the estimated relative gene flow averaged across each "column" of raster cells. *EEMS* can also generate visualizations that highlight areas of the landscape having significantly higher or lower

genetic differentiation as evaluated by posterior probabilities (Petkova et al., 2016). We used pilot runs with the constant migration, even-sampling simulations to tune the *EEMS* parameters such that the acceptance proportions were from 20% to 30%, as recommended in Petkova et al. (2016). We then used those parameters for all subsequent analyses. Each *EEMS* run consisted of 3 million MCMC iterations, with the first 1.25 million discarded as burn-in, and sampled every 2000 iterations, leaving 875 samples per analysis run. For both *EEMS* and *SpaceMix*, we visually confirmed that the MCMC chain showed evidence of mixing well (a jagged trace of posterior probabilities) and appeared to converge in the majority of analysis runs, where one run was performed for each simulation iteration.

2.6 | Testing all three methods using an empirical example

To further evaluate the analysis methods, we used individuals with European ancestry from the POPRES data set of human genetic variation (Nelson et al., 2008). This data set has been shown to display striking similarity between the PC 1/PC 2 biplot of genetic differentiation and geographic locations (Figure 1; Novembre et al., 2008). The sampled individuals self-identified the country of birth for either themselves or their grandparents and were genotyped using the Affymetrix 500K genotyping chip. We used the same SNPs (197,146) and individuals as in Novembre et al. (2008) with three exceptions: individuals 31645 and 32480, both with French ancestry, were not listed in the POPRES version 2, QC2 data set that we obtained, and individual 13011 reported Slovakian ancestry but was genetically a large outlier in the PCA biplot, as was noted in Novembre et al. (2008). This left 1384 individuals to analyse. We aggregated individuals by country (or region for individuals with Scottish, Swiss French, Swiss German, and Swiss Italian ancestry) for *SpaceMix* and *EEMS*, following the aggregation groups in Novembre et al. (2008). For *un-PC*, the PCA was run using individual-level data (Figure 1a) and PC scores were averaged for individuals from each country or region. For each analysis method, we used the same geographic location for each population as Novembre et al. (2008), which was generally the spatial centroid of each country except Norway, Sweden and Russia where the location of the capitals was used instead; we also combined Serbia and Montenegro, as in Novembre et al. (2008).

3 | RESULTS

3.1 | Constant migration

As a baseline condition, we simulated a scenario with constant migration between neighbouring populations across the landscape and even-sampling of individuals from each population (Figure 2a). The PCA biplot for the simulated data correctly located the central populations and showed the characteristic distortion of PCA for populations nearer the edge of the landscape (Figure 3a). To help correct for this distortion, the *un-PC* analysis used population positions derived from a two-dimensional DCT (Figure S5; see Methods). The DCT correction was generally effective at reducing distortion-based artefacts in the *un-PC* visualizations, especially near the central populations for the most differentiated ellipses and the corner populations for the least differentiated ellipses, as evidenced by most ellipses being close to the mean value (having light colours; Figure S5). However, after the DCT correction, there were a few more strongly differentiated ellipses, which occur because the correction does not exactly match the actual PCA distortion (Figure S5). *SpaceMix*'s assignment of population locations in geogenetic space (Figure 3c) was extremely similar to the raw PCA results (Figure 3a), except with less of the edge distortion present in the PCA. In this scenario, we expected the *EEMS* results to be white, indicating no systematic deviations from the average estimated genetic differentiation. Instead, we consistently observed a pattern of decreased genetic differentiation (indicated by darker green colours) in the edge populations and increased differentiation (indicated by darker pink colours) among the central populations (Figure 3b), and this pattern was associated with areas of the landscape that had significant posterior probabilities (Figure S6). This artefact resulted from including unsampled populations in the simulations; these strongly defined areas of apparent genetic differentiation disappeared, and the patterns of significant posterior probabilities were substantially weakened under unrealistic conditions with no intervening unsampled populations among the sampled populations, and no unsampled populations surrounding the sampled landscape (Figure S7). The three visualization methods took distinctly different amounts of time to run regardless of the simulation scenario: PCA and *un-PC* were consistently more than two orders of magnitude faster than *EEMS* and more than three orders of magnitude faster than *SpaceMix* (Table 1).

TABLE 1 Mean run times in hours for *un-PC* (with PCA), *EEMS* and *SpaceMix* over the 100 iterations of each of the three simulation scenarios represented in Figures 3–5. *Un-PC* was run to generate mean visualizations of genetic differentiation across all 100 PCA iterations; *un-PC* did this in <15 s for all simulated scenarios

Simulation scenario	Analysis method	Mean time (fractional hours)	Standard deviation (fractional hours)
Constant migration	PCA + *un-PC*	0.003	0.0002
	EEMS	0.86	0.005
	SpaceMix	5.05	0.019
Migration barrier	PCA + *un-PC*	0.004	0.0003
	EEMS	0.89	0.004
	SpaceMix	5.07	0.055
Long-distance migration	PCA + *un-PC*	0.003	0.0003
	EEMS	0.86	0.006
	SpaceMix	5.05	0.019

3.2 | Migration barrier

The PCA biplot for this scenario showed populations within the migration barrier having increased genetic distances between them compared to populations on either side of the barrier, resulting in a central "bubble" of distortion (Figure 4a). Scenarios where there are short-distance changes in migration rates across the landscape are well suited for analysis with the model that underlies *EEMS*. Indeed, *EEMS* clearly visualized the correct position of the migration barrier on the landscape (Figure 4b) and associated it with significantly increased genetic differentiation based on posterior probabilities (Figure S6). However, for *EEMS*, the artefact of decreased genetic differentiation around the landscape edges in the constant migration scenario (Figure 3b) was still present and was of the same magnitude as the estimated increase in genetic differentiation that was caused by the migration barrier (Figure 4b, bottom). *SpaceMix* depicted the populations inside the migration barrier as more differentiated (and therefore further apart in geogenetic space) than populations outside of it, making the barrier appear as a region of distortion (Figure 4c), similar to the pattern in the PCA biplot (Figure 4a). *SpaceMix* also had increased uncertainty in estimating the position of populations within the migration barrier compared to populations outside it, as indicated by larger 95% credible interval ellipses. Additionally, *SpaceMix* displayed a consistent artefact of placing the populations on either side of the migration barrier in a compact arch (Figure 4c). Both of the *un-PC* visualizations (stacking either the most or the least differentiated ellipses on top) showed complimentary information (Figure 4d,e): a band of increased genetic differentiation coinciding with the migration barrier and relatively less genetic differentiation among populations on either side of the barrier.

3.3 | Long-distance migration

Here as an example, we present results for the long-distance migration scenario where 70% of individuals in the recipient population had recently arrived directly from the source population; see Figure S8 for results from other amounts of long-distance migration. Long-distance migration was apparent in the PCA biplots, with the recipient population (the purple dot near the centre of the grid) being positioned closer to the source population than it actually is in geographic space (Figure 5a). *SpaceMix* is built to infer these kinds of long-distance migration events, and it consistently performed the best of the three methods in estimating the location of the source population, and uniquely in inferring long-distance migration and

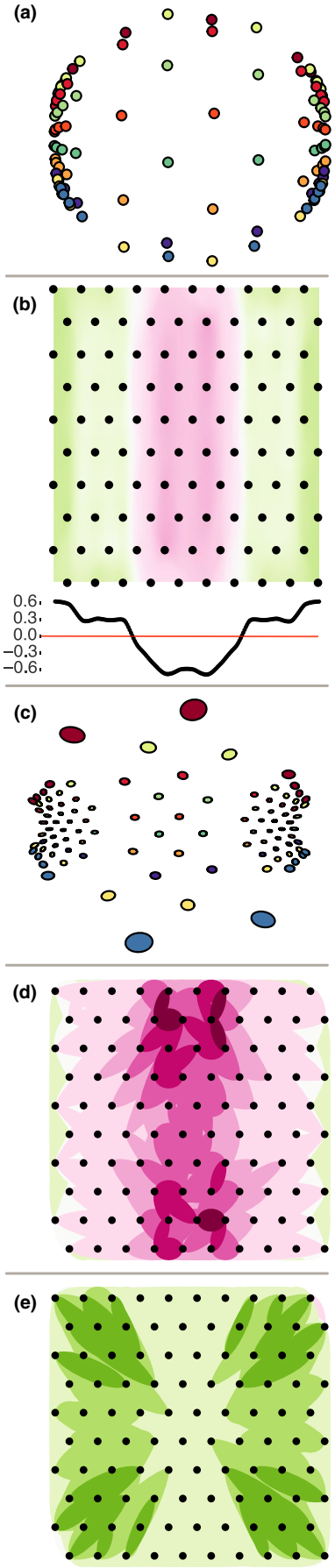

FIGURE 4 Visualizations of genetic differentiation for the migration barrier scenario with even-sampling (Figure 2b); all visualizations are coloured the same as in Figure 3. (a) Biplot of PC 1 and PC 2. (b) *EEMS*. (c) *SpaceMix*. (d) *Un-PC* with the most differentiated ellipses on top. (e) *Un-PC* with the least differentiated ellipses on top. All visualizations are the mean from 100 simulation iterations [Colour figure can be viewed at wileyonlinelibrary.com]

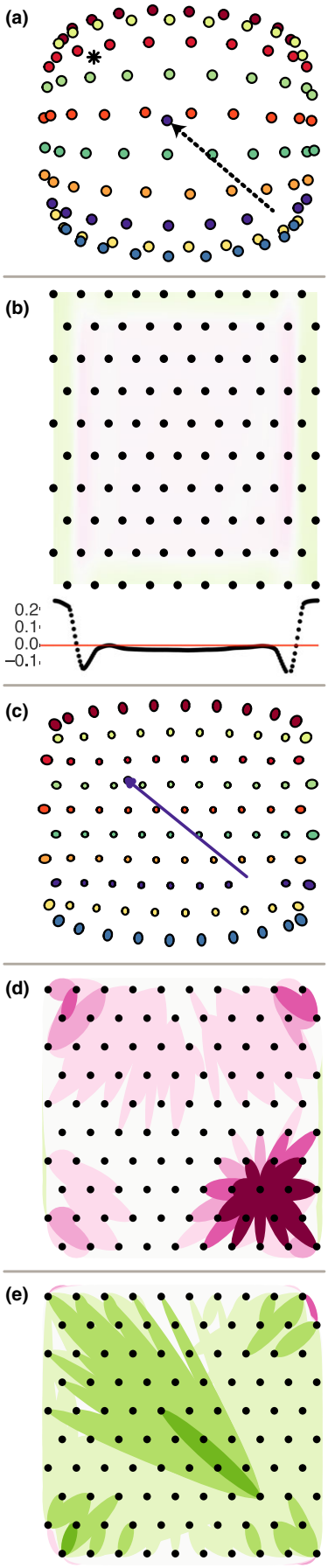

FIGURE 5 Visualizations of genetic differentiation for the long-distance migration scenario with even-sampling, where 70% of the individuals in the recipient population were long-distance migrants from the source population in the recent past (Figure 2c). All visualizations are coloured the same as in Figure 3. (a) Biplot of PC 1 and PC 2 with the dashed arrow added to indicate the geographic location of the recipient population on the landscape (the arrow tail) compared to its position in the biplot (the purple dot by the arrow head); the location of the unsampled source population is indicated by a black asterisk in the upper left corner. (b) *EEMS* fails to identify this long-distance migration event. (c) *SpaceMix* correctly identified the recipient population, plotting it as a purple ellipse near the location of the source population in geogenetic space, and automatically drawing an arrow that originates from the correctly inferred geographic location of the recipient population (lower right) and ends at its position in geogenetic space (upper left). (d) *Un-PC* with the most differentiated ellipses on top indicates an area of genetic differentiation around the recipient population because its individuals are genetically distinct from those in the neighbouring populations. (e) *Un-PC* with the least differentiated ellipses on top gives an indication of the direction of long-distance migration because the individuals in the recipient population are more genetically similar to populations in the area of the source population than they are to individuals elsewhere on the landscape. All visualizations are the mean from 100 simulation iterations [Colour figure can be viewed at wileyonlinelibrary.com]

determining its direction (Figures 5c and S8). *Un-PC* highlighted the recipient population as having increased genetic differentiation compared to its neighbouring populations, reflecting the fact that many of its individuals were not genetically similar to those of nearby populations (Figure 5d). *Un-PC* also gave an indication of the direction of the long-distance migration event due to the recipient population having less genetic differentiation compared to several populations in the vicinity of the source population (Figure 5e). Similar to *un-PC*, *EEMS* depicted the recipient population of the long-distance migration event as having increased genetic differentiation compared to adjacent populations, though *EEMS* was markedly less sensitive than *un-PC* in showing this (Figure 5b), with it only appearing in the visualizations when 90% of the individuals in the recipient population were long-distance migrants from the source population (Figure S8). The long-distance migration simulated in these scenarios all occurred in the recent past (Table S1); if the long-distance migration instead took place in the more distant past, then its resulting signal of admixture is expected to be reduced compared to what we demonstrate here.

3.4 | Effects of uneven sampling

Random uneven sampling of individuals from each population across the landscape did not qualitatively affect the patterns of genetic differentiation from any of the three analysis methods, in either the constant migration or the migration barrier scenarios (Figure S9). Uneven sampling only resulted in artefacts when it was strongly patterned across the landscape, forming a gradient in the number of sampled individuals (Figure S10). Generally, *SpaceMix* was most

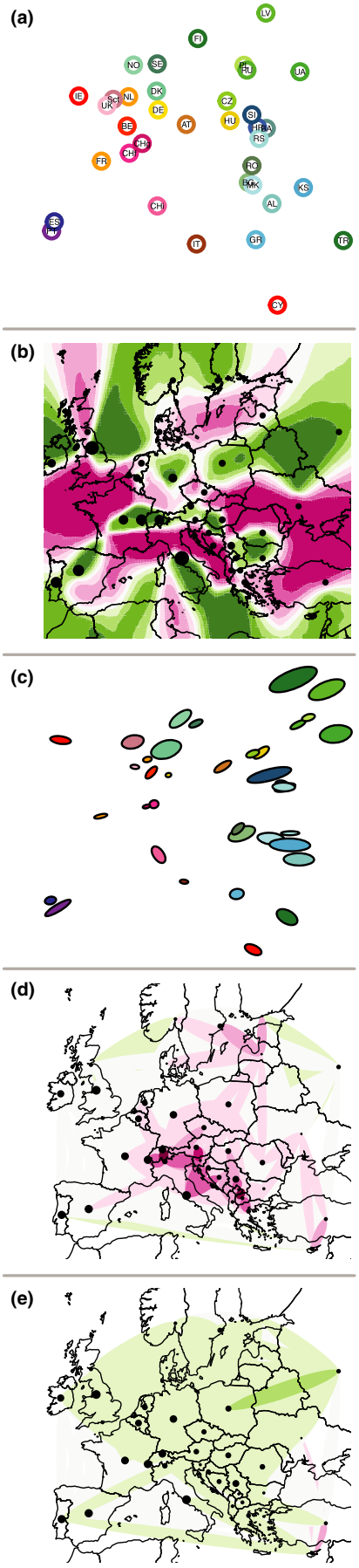

FIGURE 6 Analysis results using the empirical POPRES data set of human genetic differentiation in Europe following Novembre et al. (2008). (a) Biplot of PC 1 and PC 2 calculated for each individual (Figure 1a) after aggregation by country (or by language for Switzerland) using the same colours and labels as in Figure 1b. (b) *EEMS* visualization with darker pink representing areas of estimated increased genetic differentiation and darker green representing areas of estimated decreased genetic differentiation. (c) *SpaceMix* visualization locating individuals from each country or population in geogenetic space with colours following those in Figure 1 and Panel (a). (d) *Un-PC* with the most differentiated ellipses on top. (e) *Un-PC* with the least differentiated ellipses on top. For *EEMS* and *Un-PC* the size of each black dot is proportional to the number of individuals sampled from each country or region [Colour figure can be viewed at wileyonlinelibrary.com]

robust to artefacts due to strongly uneven sampling, *EEMS* had moderate artefacts and *un-PC* had the strongest artefacts when the uneven sampling was most extreme (Figure S10).

3.5 | Empirical example

We used a data set of landscape-scale genetic differentiation from humans in Europe to illustrate a comparison among the three methods. This data set was originally used by Novembre et al. (2008) to highlight the ability of individual-level PCA biplots to closely match geographic sampling locations (Figures 1 and 6a). *SpaceMix*'s placement of the populations in geogenetic space closely matched their geographic positions (Figure 6c), although the distance between the Iberian Peninsula populations (Spain and Portugal) and other sampled populations was large. The continuous surface produced by *EEMS* assigned areas of decreased genetic differentiation between most populations in Central Europe and areas of increased differentiation across the Alps and Dinaric Alps mountain ranges, as well as across portions of Eastern Europe (Figure 6b). The visualizations from *un-PC* also indicated increased genetic differentiation across the Alps and in the area of the Dinaric Alps (Figure 6d) and indicated decreased genetic differentiation primarily between Poland and Russia (Figure 6e).

4 | DISCUSSION

4.1 | Comparing visualizations from *EEMS* and *SpaceMix*

EEMS and *SpaceMix* each gave the most accurate visualizations of genetic differentiation when the migration scenarios were well matched to their underlying models but produced either less sensitive or less intuitive visualizations when the simulated migration conditions did not fit their model well. *SpaceMix*'s visualization is similar to a PCA biplot and is not always intuitive, especially with the distortion caused by the migration barrier scenario (Figure 4c). The distortion of populations within the migration barrier occurs because these populations exchange fewer migrants compared to populations

on either side of the barrier, making them more genetically differentiated and therefore placed further apart in geogenetic space. Without recognizing this distinctive pattern using simulations, however, it could be difficult to interpret a similar pattern if it appeared during the analysis of an empirical data set. We note that this distortion does not appear in the migration barrier simulation in the original *SpaceMix* paper (Bradburd et al., 2016), likely because that barrier was narrower and did not have any sampled populations within it. The barrier simulated here is wider and spans sampled populations, similar to the migration barrier simulation in the original *EEMS* paper (Petkova et al., 2016).

In contrast, *EEMS* performed especially well with the migration barrier scenario, as expected, but failed to detect long-distance migration in all but the most extreme cases where the number of individuals received from the source population was at least 90% of the recipient population (Figures 5b and S8). Because the model underlying *EEMS* is built to only visualize short-distance changes in gene flow, when there was an extreme amount of long-distance migration it was visualized as an area of increased genetic differentiation between the recipient population and its neighbouring populations (Figure S8); this characteristic was described in the original *EEMS* paper (Petkova et al., 2016). This area of apparently increased genetic differentiation was due to the fact that most of the individuals in the recipient population had been transplanted from the source population, and therefore had reduced genetic similarity with the neighbouring populations. However, as with *SpaceMix*, using known migration scenarios to understand these potentially nonintuitive patterns in the *EEMS* visualizations is important in helping to interpret patterns of genetic variation in empirical data sets. Furthermore, the presence of unsampled populations on the landscape can cause a characteristic artefact in *EEMS* visualizations (Figure S7), and this artefact may be analogous to anomalies in apparent migration caused by unsampled populations in pairwise migration matrices (Slatkin, 2005). While this artefact does not appear to be problematic when there is variation in simulated migration rates across the landscape (Figure 4), it is nevertheless important to document as it may affect *EEMS* visualizations from empirical studies.

4.2 | *Un-PC* has characteristics of both *EEMS* and *SpaceMix* but is faster to run

Un-PC performed well as a diagnostic method to visualize patterns of genetic differentiation that retained most of the important features of *EEMS* and *SpaceMix*, but in a fraction of the time (Table 1). With *un-PC* being model-free, its visualizations are not as detailed as those from the two model-based methods. However, *un-PC* clearly and intuitively identified the region of increased genetic differentiation that resulted from the simulated migration barrier, similar to *EEMS* (Figure 4), and perhaps better than *SpaceMix*. Conversely, because PCA provides a sensitive indication of long-distance migration (Ma & Amos, 2012), *un-PC* was able to visualize the general direction of long-distance migration similar to *SpaceMix* (Figure 5), and in a much clearer way than *EEMS*. Another model-free method to visualize patterns of genetic differentiation across the landscape, *MAPI*, has recently been introduced (Piry et al., 2016), and it appears to perform well in identifying areas of increased genetic differentiation that can result from migration barriers, similar to *EEMS* and *un-PC*. While *MAPI* can also detect signals of long-distance migration (S. Piry, personal communication), it is optimized for the visualization of short-distance changes in gene flow similar to *EEMS*, and its visualizations can be used directly in GIS applications (Piry et al., 2016). However, *MAPI* uses postgreSQL, which can require a relatively complicated database installation and configuration before it can be used. PCA is already typically run as a diagnostic step in landscape genetics studies to visualize genetic differences, and *un-PC* provides a lightweight extension on top of existing PCA results in order to quickly visualize general patterns of genetic differentiation on the landscape. These results should then be able to help guide the implementation of more computationally intensive model-based methods like *EEMS* or *SpaceMix*.

4.3 | Performance of all methods with an empirical data set

The PCA biplot of genetic differences from the POPRES data set provides a striking match to geography (Novembre et al., 2008), but there are relatively larger gaps between both Italy (brown) and the Balkans (light blues) compared to the rest of Central Europe (Figure 6a), and this increased genetic differentiation across the Alps and the Dinaric Alps is generally well represented by all three analysis methods (Figure 6). Conversely, all three methods identify areas of decreased genetic differentiation between Poland and Russia (Figure 6).

4.4 | Uneven sampling

A frequently cited criticism of PCA (that affects *un-PC* due to its reliance on PCA) is its sensitivity to uneven sampling of individuals from populations across the landscape (Bradburd et al., 2016; Petkova et al., 2016), and this sensitivity is mathematically unavoidable (McVean, 2009). However, the POPRES data set has strongly uneven sampling of individuals across populations (min = 1; max = 219; mean = 38), yet the PCA biplots representing the genetic differentiation are still strikingly similar to the geographic sampling locations (Figure 1a). Similar close matches between PCA biplots and geographic locations are also obtained in human populations sampled across Africa and Asia, albeit with a smaller range of uneven sampling (Wang et al., 2012). Like the example with the POPRES data set, when we randomly varied the number of individuals sampled from each population in our landscape simulations, the results from *un-PC* were relatively unaffected (Figure S9). However, we also note that this range of uneven sampling of individuals from each population is moderate (Figure S9). When there is a greater range of sample sizes, or when the pattern of uneven sampling coincides with the pattern of migration across the landscape, severe distortions of PCA

biplots can occur (Figure S10), and these are expected (McVean, 2009). Unfortunately, the amount of variance in sample sizes that will unduly distort the PCA results is not easily generalizable to other scenarios, and these distortions, when present, necessarily affect the accuracy of the *un-PC* results. If working with an existing data set where this distortion appears to be a problem, it can be mitigated by subsampling individuals from each population to roughly equal levels and rerunning the PCA. If genetic data has yet to be collected, then it is useful to be aware that fairly equal sample sizes are necessary for PCA and *un-PC* to work reliably.

4.5 | General recommendations

Given our results using simulated data sets that assume migration–drift equilibrium, *SpaceMix* appears to be an effective method to accurately infer long-distance migration in landscape genetics data sets (Figure 5), even when the proportion of long-distance migrants is small relative to the size of the recipient population (Figure S8). *SpaceMix* also performed well in the other simulated scenarios, but because its visualizations locate populations in geogenetic space rather than geographic space, its visualizations do not always have intuitive interpretations as demonstrated by the migration barrier scenario (Figure 4). In contrast, *EEMS* gives a more detailed view of relative genetic differentiation that is interpolated across the landscape. *EEMS* was particularly good at highlighting differences in genetic differentiation that occurred across large areas of the landscape, such as the migration barrier (Figure 4), but had little sensitivity in identifying long-distance migration events (Figure 5). For these reasons, *SpaceMix* and *EEMS* have different strengths in visualizing patterns of genetic differentiation across the landscape. The visualizations from *un-PC* cannot replicate the resolution provided by the model-based methods of *SpaceMix* and *EEMS*. However, because *un-PC* has characteristics of both the other methods: high sensitivity in detecting the genetic patterns produced by long-distance migration events similar to *SpaceMix*, and intuitive interpretation of its visualization for the migration barrier scenario similar to *EEMS*, it can help guide the effective use of the more computationally intensive analysis methods.

ACKNOWLEDGEMENTS

We thank the members of the Hahn laboratory for discussions and feedback that improved this work, and for their expertise in resolving technical problems. G. Bradburd, S. Piry, four anonymous reviewers and L. Waits all made suggestions that helped to improve the manuscript. G.L.H. was partially supported by Strategic Environmental Research and Development Program (SERDP) (RC-2330) to James D. Bever.

AUTHOR CONTRIBUTION

G.L.H. and M.W.H. designed the research, G.L.H. performed the research, G.L.H. analysed data in consultation with M.W.H., G.L.H. and M.W.H. wrote the manuscript.

DATA ACCESSIBILITY

The *msLandscape* toolbox, user manual, and example input files are available on GitHub (https://github.com/hahnlab/msLandscape). The *un-PC* R package, user manual, and example data sets are available on GitHub (https://github.com/hahnlab/un-PC). The *ms* parameter files for the three major migration scenarios evaluated here are provided as Appendices S1–S3; the *ms* parameter files for all other tested scenarios are available on GitHub (https://github.com/hahnlab/msLandscape). The collections and methods for the POPRES data are described by Nelson et al. (2008). The POPRES data used in the empirical analysis was a subset of data downloaded from the dbGaP website under study ID phs000145.v4.p2.

ORCID

Geoffrey L. House http://orcid.org/0000-0003-3330-0301

REFERENCES

Bradburd, G. S., Coop, G. M., & Ralph, P. L. (2017). Inferring continuous and discrete population genetic structure across space. bioRxiv. https://doi.org/10.1101/189688

Bradburd, G. S., Ralph, P. L., & Coop, G. M. (2016). A spatial framework for understanding population structure and admixture. *PLoS Genetics*, *12*(1), e1005703. https://doi.org/10.1371/journal.pgen.1005703

DeGiorgio, M., & Rosenberg, N. A. (2013). Geographic sampling scheme as a determinant of the major axis of genetic variation in principal components analysis. *Molecular Biology and Evolution*, *30*(2), 480–488. https://doi.org/10.1093/molbev/mss233

Duforet-Frebourg, N., & Blum, M. G. B. (2014). Nonstationary patterns of isolation-by-distance: Inferring measures of local genetic differentiation with Bayesian kriging. *Evolution*, *68*(4), 1110–1123. https://doi.org/10.1111/evo.12342

Hudson, R. R. (2002). Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics*, *18*(2), 337–338. https://doi.org/10.1093/bioinformatics/18.2.337

Ma, J., & Amos, C. I. (2012). Principal components analysis of population admixture. *PLoS ONE*, *7*(7), e40115. https://doi.org/10.1371/journal.pone.0040115

McVean, G. (2009). A genealogical interpretation of principal components analysis. *PLoS Genetics*, *5*(10), e1000686. https://doi.org/10.1371/journal.pgen.1000686

Menozzi, P., Piazza, A., & Cavalli-Sforza, L. (1978). Synthetic maps of human gene frequencies in Europeans. *Science*, *201*(4358), 786–792. https://doi.org/10.1126/science.356262

Miller, A. D., van Rooyen, A., Rasic, G., Ierodiaconou, D. A., Gorfine, H. K., Day, R., … Weeks, A. R. (2016). Contrasting patterns of population connectivity between regions in a commercially important mollusc *Haliotis rubra*: Integrating population genetics, genomics and marine LiDAR data. *Molecular Ecology*, *25*(16), 3845–3864. https://doi.org/10.1111/mec.13734

Montinaro, F., Busby, G. B. J., Gonzalez-Santos, M., Oosthuitzen, O., Oosthutizen, E., Anagnostou, P., … Capelli, C. (2016). Complex ancient genetic structure and cultural transitions in southern African populations. *Genetics*, https://doi.org/10.1101/043562

Nelson, M. R., Bryc, K., King, K. S., Indap, A., Boyko, A. R., Novembre, J., … Lai, E. H. (2008). The Population Reference Sample, POPRES: A resource for population, disease, and pharmacological genetics

research. *The American Journal of Human Genetics*, *83*(3), 347–358. https://doi.org/10.1016/j.ajhg.2008.08.005

Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A. R., Auton, A., … Bustamante, C. D. (2008). Genes mirror geography within Europe. *Nature*, *456*(7218), 98–101. https://doi.org/10.1038/nature07331

Novembre, J., & Stephens, M. (2008). Interpreting principal component analyses of spatial population genetic variation. *Nature Genetics*, *40*(5), 646–649. https://doi.org/10.1038/ng.139

Pagani, L., Lawson, D. J., Jagoda, E., Mörseburg, A., Eriksson, A., Mitt, M., … Metspalu, M. (2016). Genomic analyses inform on migration events during the peopling of Eurasia. *Nature*, *538*(7624), 238–242. https://doi.org/10.1038/nature19792

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, *2*(12), e190. https://doi.org/10.1371/journal.pgen.0020190

Petkova, D., Novembre, J., & Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, *48*(1), 94–100. https://doi.org/10.1038/ng.3464

Piry, S., Chapuis, M.-P., Gauffre, B., Papaïx, J., Cruaud, A., & Berthier, K. (2016). Mapping Averaged Pairwise Information (MAPI): A new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, *7*(12), 1463–1475. https://doi.org/10.1111/2041-210X.12616

Ray, N., Currat, M., Foll, M., & Excoffier, L. (2010). SPLATCHE2: A spatially explicit simulation framework for complex demography, genetic admixture and recombination. *Bioinformatics*, *26*(23), 2993–2994. https://doi.org/10.1093/bioinformatics/btq579

Richardson, J. L., Brady, S. P., Wang, I. J., & Spear, S. F. (2016). Navigating the pitfalls and promise of landscape genetics. *Molecular Ecology*, *25*(4), 849–863. https://doi.org/10.1111/mec.13527

Richmond, J. Q., Wood, D. A., Westphal, M. F., Vangergast, A. G., Leache, A. D., Saslaw, L. R., … Fisher, R. N. (2017). Persistence of historical population structure in an endangered species despite near-complete biome conversion in California's San Joaquin Desert. *Molecular Ecology*, *26*(14), 3618–3635. https://doi.org/10.1111/mec.14125

Roffler, G. H., Amish, S. J., Smith, S., Cosart, T., Kardos, M., Schwartz, M. K., & Luikart, G. (2016). SNP discovery in candidate adaptive genes using exon capture in a free-ranging alpine ungulate. *Molecular Ecology Resources*, *16*(5), 1147–1164. https://doi.org/10.1111/1755-0998.12560

Slatkin, M. (1987). Gene flow and the geographic structure of natural populations. *Science*, *236*(4803), 787–792. https://doi.org/10.1126/science.3576198

Slatkin, M. (2005). Seeing ghosts: The effect of unsampled populations on migration rates estimated for sampled populations. *Molecular Ecology*, *14*(1), 67–73. https://doi.org/10.1111/j.1365-294X.2004.02393.x

Tsuda, Y., Chen, J., Stocks, M., Källman, T., Sønstebø, J. H., Parducci, L., … Lascoux, M. (2016). The extent and meaning of hybridization and introgression between Siberian spruce (*Picea obovata*) and Norway spruce (*Picea abies*): Cryptic refugia as stepping stones to the west? *Molecular Ecology*, *25*(12), 2773–2789. https://doi.org/10.1111/mec.13654

Uren, C., Kim, M., Martin, A. R., Bobo, D., Gignoux, C. R., van Helden, P. D., … Henn, B. M. (2016). Fine-scale human population structure in southern Africa reflects ecogeographic boundaries. *Genetics*, *204*(1), 303–314. https://doi.org/10.1534/genetics.116.187369

Wang, J., & Whitlock, M. C. (2003). Estimating effective population size and migration rates from genetic samples over space and time. *Genetics*, *163*(1), 429–446.

Wang, C., Zöllner, S., & Rosenberg, N. A. (2012). A quantitative comparison of the similarity between genes and geography in worldwide human populations. *PLoS Genetics*, *8*(8), e1002886. https://doi.org/10.1371/journal.pgen.1002886

Whitlock, M. C., & McCauley, D. E. (1999). Indirect measures of gene flow and migration: F_{ST}=1/(4*Nm*+1). *Heredity*, *82*(2), 117–125. https://doi.org/10.1038/sj.hdy.6884960

Yang, W.-Y., Novembre, J., Eskin, E., & Halperin, E. (2012). A model-based approach for analysis of spatial structure in genetic data. *Nature Genetics*, *44*(6), 725–731. https://doi.org/10.1038/ng.2285

Yoder, A. D., Campbell, C. R., Blanco, M. B., dos Reis, M., Ganzhorn, J. U., Goodman, S. M., … Weisrock, D. W. (2016). Geogenetic patterns in mouse lemurs (genus *Microcebus*) reveal the ghosts of Madagascar's forests past. *Proceedings of the National Academy of Sciences*, *113*(29), 8049–8056. https://doi.org/10.1073/pnas.1601081113

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: House GL, Hahn MW. Evaluating methods to visualize patterns of genetic differentiation on a landscape. *Mol Ecol Resour*. 2018;18:448–460. https://doi.org/10.1111/1755-0998.12747