ORIGINAL ARTICLE

NONSTATIONARY PATTERNS OF ISOLATION-BY-DISTANCE: INFERRING MEASURES OF LOCAL GENETIC DIFFERENTIATION WITH BAYESIAN KRIGING

Nicolas Duforet-Frebourg[1] and Michael G.B. Blum[1,2]

[1]Laboratoire TIMC-IMAG, Centre National de la Recherche Scientifique, Université Joseph Fourier, Grenoble, France

[2]E-mail: michael.blum@imag.fr

Patterns of isolation-by-distance (IBD) arise when population differentiation increases with increasing geographic distances. Patterns of IBD are usually caused by local spatial dispersal, which explains why differences of allele frequencies between populations accumulate with distance. However, spatial variations of demographic parameters such as migration rate or population density can generate nonstationary patterns of IBD where the rate at which genetic differentiation accumulates varies across space. To characterize nonstationary patterns of IBD, we infer local genetic differentiation based on Bayesian kriging. Local genetic differentiation for a sampled population is defined as the average genetic differentiation between the sampled population and fictive neighboring populations. To avoid defining populations in advance, the method can also be applied at the scale of individuals making it relevant for landscape genetics. Inference of local genetic differentiation relies on a matrix of pairwise similarity or dissimilarity between populations or individuals such as matrices of F_{ST} between pairs of populations. Simulation studies show that maps of local genetic differentiation can reveal barriers to gene flow but also other patterns such as continuous variations of gene flow across habitat. The potential of the method is illustrated with two datasets: single nucleotide polymorphisms from human Swedish populations and dominant markers for alpine plant species.

KEY WORDS: Gene flow, genetic barrier, isolation-by-distance, landscape genetics, nonstationary kriging.

Characterizing patterns of genetic differentiation within a species is a recurring task in population genetics. Wright (1943) introduced the model of isolation-by-distance (IBD), which assumes that differences of allele frequencies between populations accumulate under the assumption of local spatial dispersal. Because of local dispersal, IBD models predict the *pattern of IBD* where population differentiation increases with increasing geographic distances (Slatkin 1993; Rousset 1997). This pattern is observed in many model and nonmodel organisms as well as in humans suggesting that local dispersal is a leading evolutionary force (Sharbel et al. 2000; Ramachandran et al. 2005; Hardy et al. 2006; Hellberg 2009).

However, the pattern of IBD can mask complex variations of demographic parameters resulting in differential increases of genetic differentiation in different regions of the habitat. Variations of demographic parameters can arise when population densities or migration rates vary across space (Slatkin 1985). With the advent of landscape genetics (Manel et al. 2003; Manel and Holderegger 2013), the spatial variation of demographic parameters is an important topic because spatial heterogeneity (or landscape characteristics) is now recognized to be a key factor to explain population differentiation and gene flow (McRae and Beier 2007). Examples of spatial heterogeneity influencing population differentiation include varying local subpopulation size (Serrouya et al. 2012) as well as fragmented landscapes in urban and agricultural area where there are "corridors" for gene flow (Arnaud 2003; Munshi-South 2012). Barriers to gene flow, which can be caused by anthropogenic or geographic factors, are also emblematic examples of spatial heterogeneity influencing population structure (e.g., Castella et al. 2000; Epps et al. 2005; Riley et al. 2006;
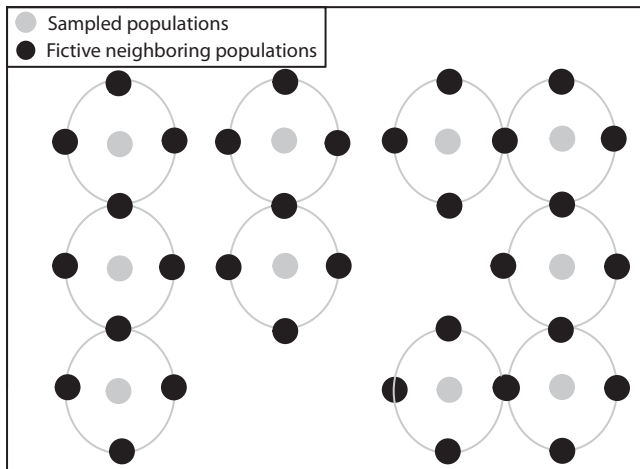
Figure 1. A two-dimensional range with putative sampled locations (in gray) and neighboring locations (in black). Local differentiation at each sampled location corresponds to the average (over neighbors) pairwise measure of differentiation between the population or individual at the sampled location and its neighbors.

Gauffre et al. 2008; Zalewski et al. 2009). Because the identification of barriers to gene flow has attracted considerable attention (Storfer et al. 2010), there is a large variety of statistical methods to detect them (Barbujani et al. 1989; Bocquet-Appel and Bacro 1994; Dupanloup et al. 2002; Manni et al. 2004; Cercueil et al. 2007; Crida and Manel 2007; Manel et al. 2007; Safner et al. 2011). Here, we propose a more general method that characterizes *non-stationary* patterns of IBD. A nonstationary pattern of IBD occurs when the rate at which differentiation between individuals or populations accumulates with distance depends on space. Nonstationary patterns of IBD arise for instance when there is a barrier to gene flow because genetic differentiation accumulates more rapidly with distance around the barrier but they can also occur on different situations such as continuous variations of gene flow across the species range.

To characterize nonstationary patterns of IBD, our approach provides a measure of local differentiation at each location where genetic data are available. The principle of the method is to estimate for each sampled location z_i, $i = 1, \ldots, n$, a local pairwise measure of population differentiation or of dissimilarity between the population located at the sampled location and fictive neighboring populations located at a fixed distance d of z_i (see Fig. 1). Considering for instance F_{ST} as a pairwise measure of genetic differentiation, the method provides estimates of F_{ST} for pairs of populations separated by a distance d and located in the vicinity of the sampling sites. The distance d has to be set in advance and should be small compared to the dimension of the region under study. Fictive neighboring populations are introduced as a mean to provide measures of local genetic differentiation—F_{ST} between popula-

tions separated by a distance d here—that are comparable between sampling sites. Compared to common tests for IBD (Hardy and Vekemans 1999), the method is more informative because it quantifies how local genetic differentiation varies across space; the rate at which genetic differentiation increases with distance may vary across space and the proposed approach provides a quantitative assessment of this variation. To determine if variation of local differentiation is sufficiently large to reject stationary IBD, we additionally provide an hypothesis-testing procedure based on simulations. The method is not restricted to pairwise population measurements and can also accommodate individual pairwise measures. Working at the scale of individuals is a desirable feature because using individuals as the operational unit avoids potential bias in identifying populations in advance and offers the opportunity to conduct studies at a finer scale (Manel et al. 2003, 2007). Using a detailed simulation study, we demonstrate that the method can correctly infer local variation of genetic differentiation and we present applications to human single nucleotide polymorphism (SNP) data (Humphreys et al. 2011) and amplified fragment length polymorphism (AFLP) from alpine plants (Gugerli et al. 2008).

Methods

For the sake of the presentation, we assume that the data consist of allele frequencies in each population and that the method relies on the empirical correlation matrix between populations. In the Results section, we show that the proposed approach is also appropriate with other pairwise matrices such as F_{ST} matrices between populations or correlation matrices between individuals.

To assess local genetic differentiation around a given sampled site, we estimate the correlation of allele frequencies between the sampled population and fictive populations located in the neighborhood of the sampling sites. Neighboring populations are located at a fixed and short distance from the sampled populations, and we measure the expected local correlation (averaged over neighbors) of allele frequencies between the sampled population and the neighboring fictive populations (Fig. 1). Because we aim at providing local genetic differentiation values that should be larger in regions of abrupt genetic changes, we consider one minus the local correlation as a measure of local differentiation.

We estimate local correlation using a Gaussian process approach (Bishop 2006), which is known as *kriging* in geostatistics (Cressie 1993). Kriging refers to a set of interpolation methods where a variable of interest is estimated at unsampled locations based on values measured at the sampling sites. Interpolation relies on a weighted average of the values measured at the sampling sites and the weights depend on a parametric function C, which describes how the correlation or the covariance decreases with distance (Cressie 1993). A direct application of kriging would

consist of interpolating the allele frequencies at the neighboring sites based on the allele frequencies estimated at the sampled sites. However, the proposed approach is nonstandard and requires methodological developments because we rather aim at estimating the correlation matrix between sampled and unsampled neighboring sites based on the correlation matrix between sampled sites. There is a vast literature of kriging procedures with nonstationary covariance when the function C describing the decay of correlation with distance varies in space (Nott and Dunsmuir 2002; Schmidt and O'Hagan 2003; Paciorek and Schervish 2006). The covariance between sampled and unsampled sites is usually estimated using a parametric model (Paciorek and Schervish 2006) or at least using a given functional model for the covariance function (Schmidt and O'Hagan 2003). However, compared to geostatics where only one or a few variables are observed at the sampling sites, we are in a favorable situation in population genetics to estimate how the covariance or the correlation varies across space. Because each locus is a statistical replicate, there is enough information to estimate the correlation between the sampled sites using the empirical correlation matrix for instance. Estimating local correlation amounts at interpolating the correlation between sampled and neighboring sites from the correlation matrix between sampled sites. We explain below how we perform the interpolation step.

THE KRIGING/GAUSSIAN PROCESS APPROACH

In the following, we denote by $\mathbf{X}$ and $\mathbf{Y}$ the vectors of allele frequencies at sampled and unsampled sites. We assume independence between loci and the vectors $\mathbf{X}$ and $\mathbf{Y}$ contain allele frequencies for an arbitrary locus. The objective of the kriging approach is to interpolate the covariance (or correlation) matrix between $\mathbf{X}$ and $\mathbf{Y}$ based on the empirical covariance matrix between sampled sites. The covariance matrix between $\mathbf{X}$ and $\mathbf{Y}$ is denoted $\mathbf{E}[(\mathbf{Y} - \mathbf{m})(\mathbf{X} - \mathbf{m})^T]$, where $\mathbf{m}$ is a constant mean over the range. The main principle is to use weighted means of covariance values between sampled sites to estimated covariance between sampled and unsampled sites. As usual in kriging, the weights depend on a parametric function C that gives the decay of correlation with distance. We explain below how we compute these weights.

The Gaussian process viewpoint of kriging is to model the joint values of the variable at sampled and unsampled sites as a multivariate Gaussian variable (Bishop 2006)

$$(\mathbf{X}, \mathbf{Y}) \rightsquigarrow \mathcal{N}(\mathbf{m}, \Psi), \qquad (1)$$

where

$$\Psi = \begin{pmatrix} \Psi_{\mathbf{xx}} & \Psi_{\mathbf{xy}} \\ \Psi_{\mathbf{xy}}{}^T & \Psi_{\mathbf{yy}} \end{pmatrix},$$

where $\Psi_{\mathbf{xx}}$ (respectively, $\Psi_{\mathbf{yy}}$) denote the covariance matrix between the sampled sites (respectively, unsampled sites) and $\Psi_{\mathbf{xy}}$ contain the covariances between the sampled and unsampled sites. The interpolation of the variable of unknown allele frequencies Y is obtained using the conditional distribution of $\mathbf{Y}$ given $\mathbf{X}$, which can be written in the following regression form

$$\mathbf{Y} - \mathbf{m} = \boldsymbol{\tau}_{\Psi}(\mathbf{X} - \mathbf{m}) + \boldsymbol{\epsilon}, \qquad (2)$$

where $\boldsymbol{\tau}_{\Psi} = \Psi_{\mathbf{xy}}{}^T \Psi_{\mathbf{xx}}{}^{-1}$ and $\boldsymbol{\epsilon}$ is a residual independent of $\mathbf{X}$ (Brown et al. 1994). A naive computation of local covariance would consist of simulating with equation (2) the vector $\mathbf{Y}$ containing the allele frequencies at the neighboring sites and then to evaluate numerically the empirical covariance between allele frequencies at sampled and at neighboring sites. Although it is a valid approach, we can actually derive what is the expected covariance between sampled and neighboring sites using equation (2) and we obtain

$$\mathbf{E}[(\mathbf{Y} - \mathbf{m})(\mathbf{X} - \mathbf{m})^T] = \boldsymbol{\tau}_{\Psi}\mathbf{E}[(\mathbf{X} - \mathbf{m})(\mathbf{X} - \mathbf{m})^T]. \qquad (3)$$

In the computations, we replace $\mathbf{m}$ by the empirical mean so that we estimate the covariance matrix with $\boldsymbol{\tau}_{\Psi}\mathbf{Var}(\mathbf{X})$, where $\mathbf{Var}(\mathbf{X})$ denotes the empirical covariance matrix of $\mathbf{X}$. The matrix $\boldsymbol{\tau}_{\Psi}$ provides the weights of the weighted means, which are used to interpolate the covariance values between sampled and unsampled sites based on the covariance values between sampled sites.

More generally, we can estimate local similarities by multiplying the weight matrix $\boldsymbol{\tau}_{\Psi}$ with a similarity matrix between sampled sites. In the Results section, we consider similarity matrices that are not correlation or covariance matrices, and we use the pairwise matrix of $(1 - F_{ST})$ values for instance. When using individuals as operational units, numerical problems can arise if they are multiple individuals by site because the matrix Ψ_{xx} can be difficult to invert. Potential solutions are to consider a population—with one or more individuals—at each sampling site or to add a small perturbation to the geographical coordinates of the individuals.

Providing the correlation instead of the covariance between sampled and unsampled sites requires the standardization of the covariance equation (3) and the renormalization formula is provided in Appendix A. The final estimate for the covariance matrix is finally obtained by averaging equation (3) over posterior replicates of $\boldsymbol{\tau}_{\Psi}$. The parametric model for Ψ, which is needed to generate the posterior distribution of $\boldsymbol{\tau}_{\Psi}$, is given below.

A MODEL FOR THE CORRELOGRAM

To compute the weight matrix $\boldsymbol{\tau}_{\Psi}$, we consider the standard model of *stationary* kriging that assumes that the correlation between two points only depends on the distance between these two points. Using these assumptions, we should model how the correlation

decreases with increasing distance. We assume that this function C, called the correlogram, decays exponentially

$$C(d) = ((1 - \alpha) + \alpha e^{-d/r} + \lambda 1_{d=0})/(1 + \lambda), \qquad (4)$$

where d is the distance between two points, 1 denotes the indicator function, α determines the sill, which measures the limiting value of the correlation function, r is the *range* parameter, and λ is the *regularization* parameter. The parameter λ is introduced for numerical reasons because it ensures that the matrix Ψ_{xx} is invertible, which is required for the computation of the weight matrix τ_Ψ (Bishop 2006). The range parameter r is inversely related to the rate at which correlation decays with distance. Denoting by d_{ij} the geographical distance between the ith and jth sites, then the entry of Ψ at the ith row and jth column is given by $C(d_{ij})$. We sample the triplet (α, λ, r) from the posterior distribution using an MCMC algorithm that contains both Gibbs and Metropolis–Hastings updating steps, and the details of the algorithm are provided in Appendix B (Handcock and Stein 1993).

HYPOTHESIS-TESTING PROCEDURE

We introduce two test statistics to test if the variation of local genetic differentiation is significant. The first test statistic is the coefficient of variation of local genetic differentiation values, that is the ratio between standard deviation and mean of local differentiation measures. The second statistic is the distance correlation statistic and it measures the dependence between local genetic differentiation and geographical coordinates. The distance correlation statistic extends Pearson correlation coefficient because it can measure nonlinear dependence (Székely et al. 2007). Because we use two test statistics, we consider the conservative Bonferroni correction and reject stationarity when one the two observed values of the test statistics is larger than the 97.5% quantile obtained for the null distribution of stationarity.

We consider two options for generating distribution of the test statistic under the null hypothesis. In the first option, we consider the parametric model of equation (4). We compute M pairwise covariance matrices $\mathbf{Var}(\mathbf{X})^i$, $i = 1, \ldots, M$, using the stationary correlogram of equation (4). The parameters $(\alpha_i, \lambda_i, r_i)$ of equation (4), which are used to compute the covariance matrices, are sampled according to the posterior distribution. The $2 \times M$ values of the tests statistics are then obtained after running the MCMC algorithm (Appendix B) M times for each of the simulated covariance matrix $\mathbf{Var}(\mathbf{X})^i$, $i = 1, \ldots, M$. When the sample size is too large, we have to limit the computational burden of the procedure, and we do not perform M MCMC runs. Instead, for the ith covariance matrix $\mathbf{Var}(\mathbf{X})^i$, we use the ith triplet $(\alpha_i, \lambda_i, r_i)$ to compute the weight matrix τ_Ψ^i and to obtain values of local genetic differentiation. However, equation (4) is only an approximation of the correlation pattern found for IBD models. It is exact, for

instance, in the one-dimensional stepping-stone model with infinite range (Kimura and Weiss 1964). To avoid the approximation of equation (4), we also consider explicit simulations of a stationary stepping-stone model using *ms* (Hudson 2002). We consider uniformly sampled migration rates such as $1 \leq 4N_0 m \leq 20$ and we choose a sampling scheme that mimics the sampling of the data.

Results

SIMULATION STUDY

In the simulation study, we consider two different models for generating nonstationary patterns of IBD. First, we consider non-homogeneous stepping stone models in one and two dimensions. We simulate with *ms* 2000 independent SNPs using spatially dependent effective migration rate $4N_0 m$, where N_0 is the population size of each deme and m is the migration rate per generation between two neighboring demes. Because we assume independence between SNPs, each SNP is simulated with a coalescent simulation that is conditioned on having one segregating site. The second model is analytic and has been developed for performing nonstationary kriging when the correlogram function (equation (4)) is assumed to vary across space (Paciorek and Schervish 2006). The range parameter r of equation (4), which measures the rate at which correlation decays with distance, is assumed to be a function of space. For the second model, zones of abrupt changes such as genetic barriers correspond to regions with a smaller range parameter because correlation decays more rapidly with distance in these regions.

Barrier in a one-dimensional model

We investigate an example of a one-dimensional model with a genetic barrier. We simulate a stepping stone model with 100 populations of effective sizes $N_0 = 1000$ diploid individuals. Depending on the simulations, we sample either 20 equidistant populations or 20 uniformly sampled populations. We consider 20 chromosomes in each of the population. Migrations are constant between neighboring populations and we consider $4N_0 m = 4$ and $4N_0 m = 20$. The barrier is located between populations 50 and 51 and arose 8 units of time ago ($4N_0 m = 0$) where time is counted in units of $4N_0$ generations. As similarity matrix, we consider the pairwise correlation of allele frequencies for 20 sampled populations. For each sampled deme, local genetic differentiation corresponds to one minus the expected correlation between the sampled deme and its two neighbors.

With equidistant sampling, we find that the parameters of the correlogram function (equation (4)) affect the estimated values of local genetic differentiation (Fig. S1). However, for all values of the correlogram parameters (α, λ, r) we consider, local differentiation is larger in the middle of the range, which is consistent with
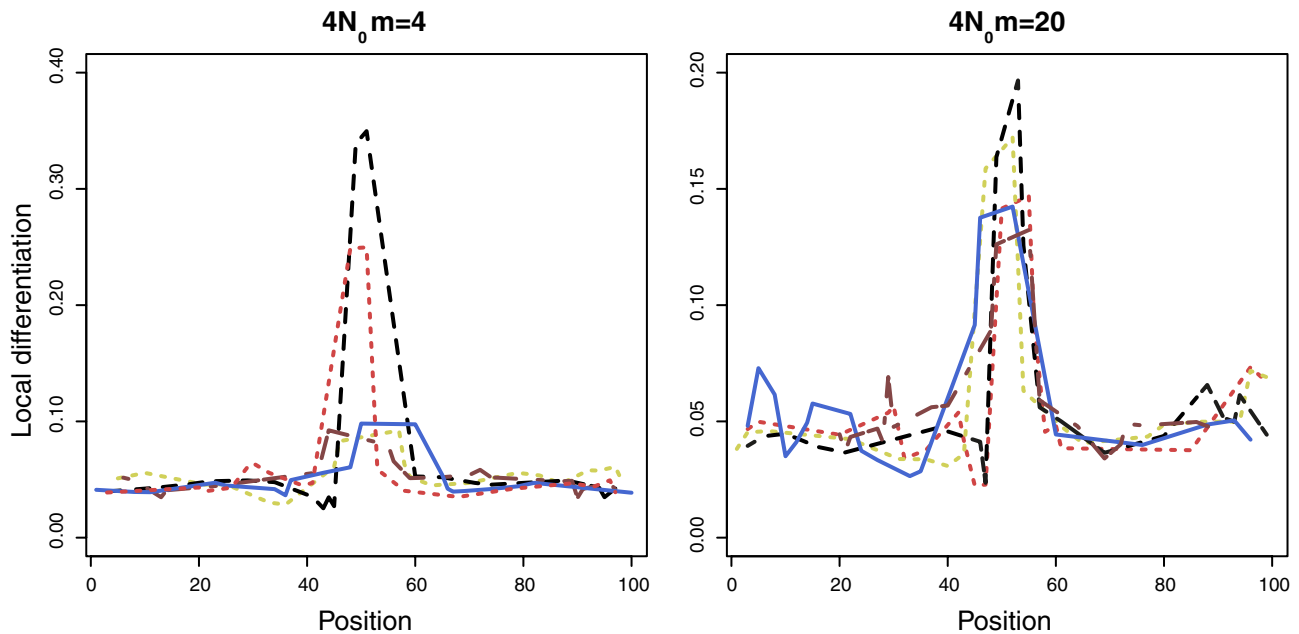
Figure 2. Estimation of local genetic differentiation for five population sets where each set contains 20 randomly picked populations among 100 populations. Simulations are performed in a one-dimensional stepping-stone model with a barrier to gene flow in the middle of the range ($4N_0m = 20$ or $4N_0m = 4$ except between population 50 and population 51 where $4N_0m = 0$). Local genetic differentiation corresponds to one minus the expected correlation of allele frequencies between sampled demes and their two closest neighboring demes.

the presence of a barrier to gene flow. Nonetheless the detailed trajectory of local differentiation as a function of space depends on the correlogram parameters and edge effects can be large for some parameter values (Fig. S1). To account for the uncertainty associated with the parameters of the correlogram function, we integrate the values of local genetic differentiation over the posterior distribution of (α, λ, r) (Fig. S1). To investigate if a barrier to gene flow is still detectable with irregular sampling, we also sampled randomly 20 populations among the 100 populations. For both intensities of barrier ($4N_0m = 20$ or $4N_0m = 4$ except at the barrier where $4N_0m = 0$) and for each replicate of population sampling, local differentiation is larger around the barrier to gene flow (Fig. 2). However, for the less stringent and more difficult to detect barrier, local differentiation increases less markedly around the barrier when sampling in the vicinity of the barrier is sparse (Fig. 2).

To provide a comparison, we apply multidimensional scaling (MDS), which is a commonly used method to represent differentiation between populations. Based on the pairwise matrix of correlation between allele frequencies computed for each of the 20 equally spaced populations, we apply MDS. Figure 3 displays the scatter plots of the first two principal coordinates when $4N_0m = 4, 4N_0m = 20$, and there are 20 equally spaced sampled populations. In the scenario where $4N_0m = 20$, the occurrence of a barrier of gene flow in the middle of the range is visible in the MDS plot whereas it is much less visible when $4N_0m = 4$ even

with the perfectly regular sampling of populations. In the latter scenario, the MDS plot is above all influenced by the global IBD pattern that generates the inverted U-shaped pattern (Novembre and Stephens 2008). Patterns exactly similar to MDS were obtained with principal component analysis when using the population allele frequencies as raw data.

We additionally explore the running time of the algorithm. Once a pairwise matrix of F_{ST} or of other dissimilarity measures has been obtained, the most costly operations are the inversion of the matrix Ψ_{xx} required to compute the weight matrix τ_Ψ as well as the computation of a determinant required to evaluate the likelihood in the MCMC algorithm (see Appendix B). The computing cost of both operations is proportional to the cube of the number of sampled sites. We check this theoretical prediction in this example by increasing the total number of demes and the number of sampled demes. We find that the cubic prediction is quite accurate although a bit pessimistic because the running time of the algorithm actually grows as the number of sampled sites at the power 2.5 (Fig. S2).

Barriers in a two-dimensional model

We consider two examples of a two-dimensional model with genetic barriers. In the first two-dimensional example, the data are simulated using a stepping-stone model in a 10×10 grid. In each population, there are $N_0 = 1000$ diploid individuals per population and we sample all populations considering 20 chromosomes
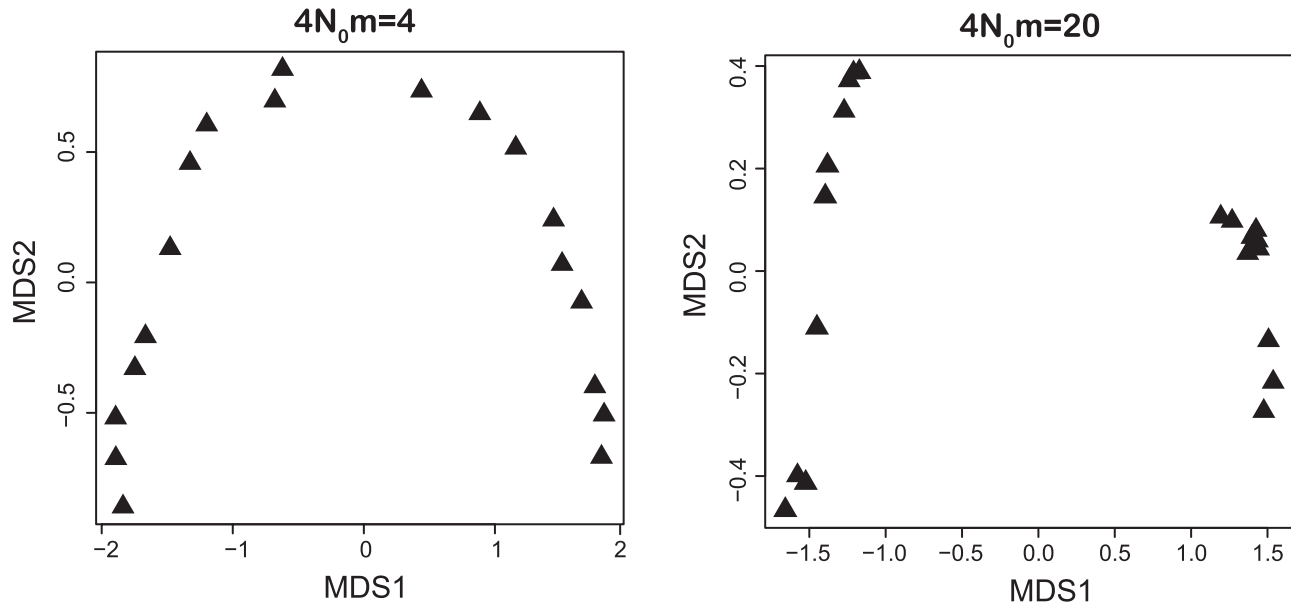
Figure 3. Multidimensional scaling plots in a one-dimensional range with a genetic barrier in the middle of the range.
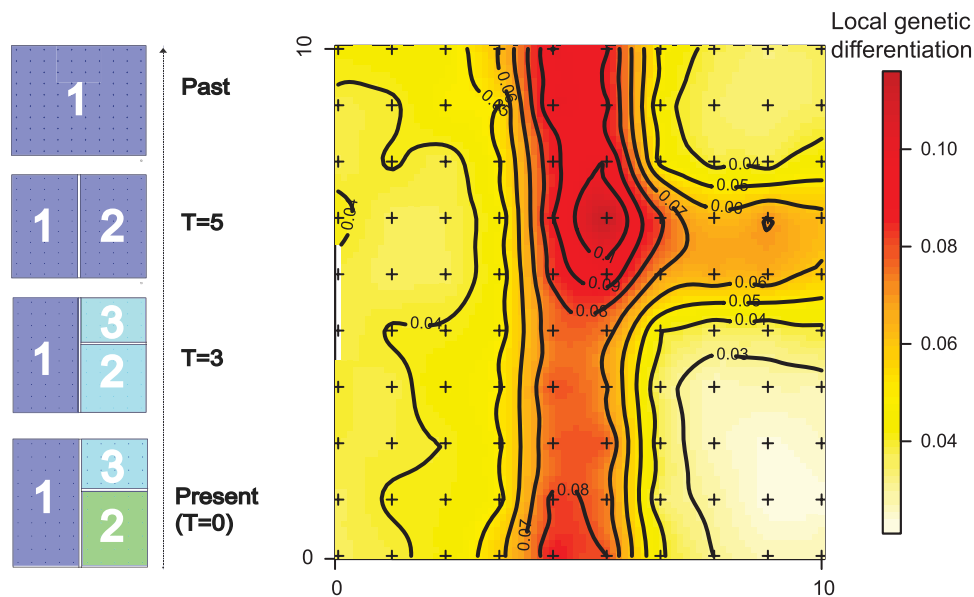

Figure 4. Map of local genetic differentiation in a two-dimensional stepping-stone model with two genetic barriers. On the left-hand side, the time line shows the time when the genetic barriers appeared. Local genetic differentiation was estimated using the pairwise correlation matrix, and it measures one minus the expected correlation between sampled populations and unsampled neighbors located at a distance of 0.1 from the sampled populations.

in each of them. The migration rate between neighboring populations is $4N_0m = 20$, where migration only occurs along horizontal and vertical lines but not along diagonals. We then assume that two barriers arose $T_1 = 5$ and $T_2 = 3$ units of time ago, where time is counted in units of $4N_0$ generations (see Fig. 4). In the second two-dimensional example, we specify explicitly the local decay of correlation using the nonstationary model of (Paciorek and Schervish 2006). We assume that there are three genetic barriers, which correspond to three different regions where the range parameter (r in equation (4)) is smaller (Fig. S3). In both examples, the input matrix of similarity is the correlation matrix between sampled sites, although the correlation is estimated based on simulated allele frequencies for the stepping stone example whereas it is obtained analytically using the convolution formula of Paciorek and Schervish (2006) for the second example.

Figures 4 and S3 show that estimated values of local genetic differentiation are larger around the genetic barriers as expected. For both examples, the relative importance of the barriers is

A. Local genetic differentiation and *barrier*

B. Multidimensional scaling
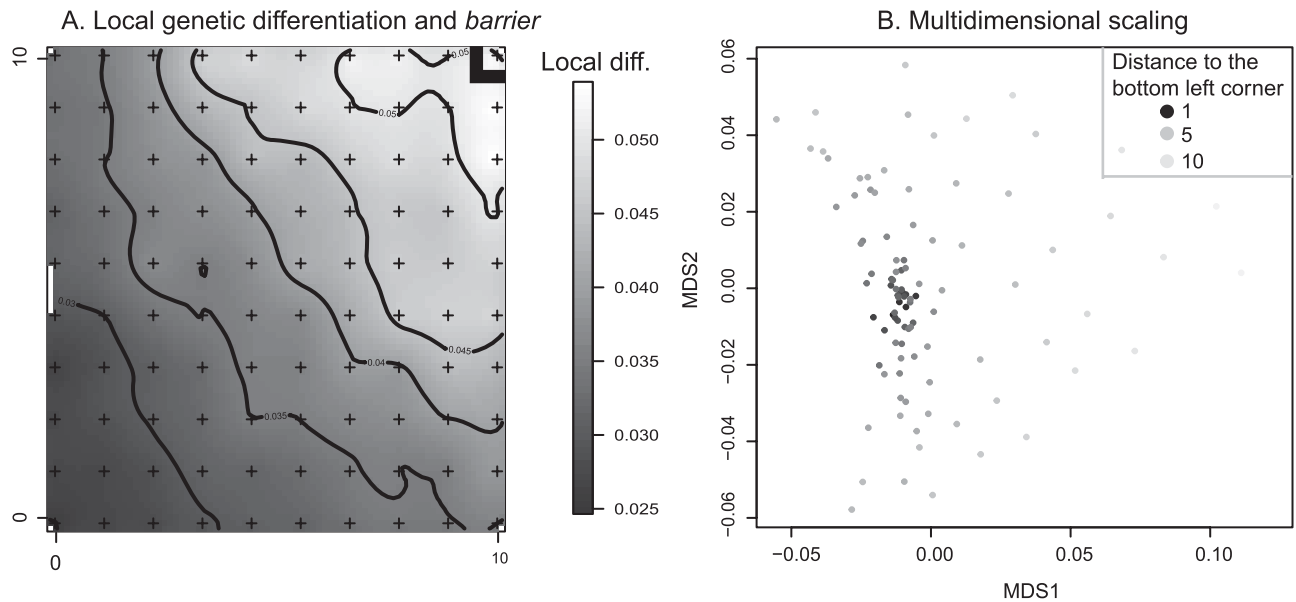
Figure 5. Investigating the pattern of population differentiation for a gradient of gene flow in a two-dimensional model. Gene flow is maximal at the lower left corner of the habitat and decreases quadratically to the distance from the lower left corner. (A) Estimation of local genetic differentiation values using the correlation with neighbors located at a distance of 0.1. The first genetic barrier that is found with the software *barrier* is shown with a thick black line. (B) Multidimensional scaling plot with a grayscale color scheme to represent the distance of each population to the lower left corner. Dark points correspond to populations that are close to the lower left corner whereas the lightest points are the farthest away from the lower left corner.

retrieved. The strongest barrier has the largest value of local genetic differentiation. For computing local genetic differentiation, we additionally consider F_{ST} pairwise values instead of correlation values in the stepping-stone model. In that case, the similarity matrix contains the pairwise $(1 - F_{ST})$ values that decay with increasing geographical distance as assumed by equation (4). The local genetic differentiation values now correspond to the expected F_{ST} between the sampled populations and their fictive neighbors. We find that the map of local genetic differentiation obtained with the F_{ST} measure is similar to the map obtained with the pairwise correlation matrix (Fig. S4). We also perform additional computations of local differentiation after incomplete sampling of the populations in the stepping-stone model. We sample, respectively, 50%, 33%, and 25% of the 100 populations present in the grid. We find that the oldest barrier is always recovered but not the most recent one, which is not detectable when sampling 33% or 25% of the populations (Fig. S5).

A gradient of gene flow

We also consider a different pattern of nonstationary IBD consisting of a two-dimensional stepping-stone model with a spatial gradient of gene flow. We assume that gene flow is maximum at the lower left corner of the habitat and decreases quadratically with distance from the lower left corner of the habitat (Fig. 5). As expected, we find that local genetic differentiation increases when moving away from the lower left corner of the habitat (Fig. 5).

When sampling 50%, 33%, or 25% of the 100 populations present in the grid, we also find a gradient of local genetic differentiation (Fig. S6).

This example is an instance of nonstationarity, which cannot be described with barriers to gene flow. For the previous examples, the software *barrier* (Manni et al. 2004), which detects zones of abrupt genetic change, is able to find barriers in both the one- and two-dimensional models (Fig. S7). However, for the gradient of gene flow, *barrier* incorrectly finds a barrier in the upper right corner of the habitat, which is nonetheless consistent with the fact that gene flow is minimal here (Fig. 5). Additionally, MDS provides a meaningful representation for the examples of barriers in the one-dimensional model with $4N_0m = 20$ (Fig. 3) and in the two-dimensional model (Fig. S8) because the populations that live on the same side of the barriers cluster together in the MDS plot (but see $4N_0m = 4$ in Fig. 3 where the clustering is less evident). However, interpreting the pattern obtained with MDS is much more difficult for the example of a gradient of gene flow (Fig. 5). The observed pattern found with MDS is consistent with the gradient of gene flow because populations living in regions of high gene flow (dark points in Fig. 5) are located more closely on the MDS plot than populations living in regions of low gene flow (clear points in Fig. 5). Although consistent with a gradient of gene flow, the MDS plot is not as easily interpretable as the map of local genetic differentiation for this example.
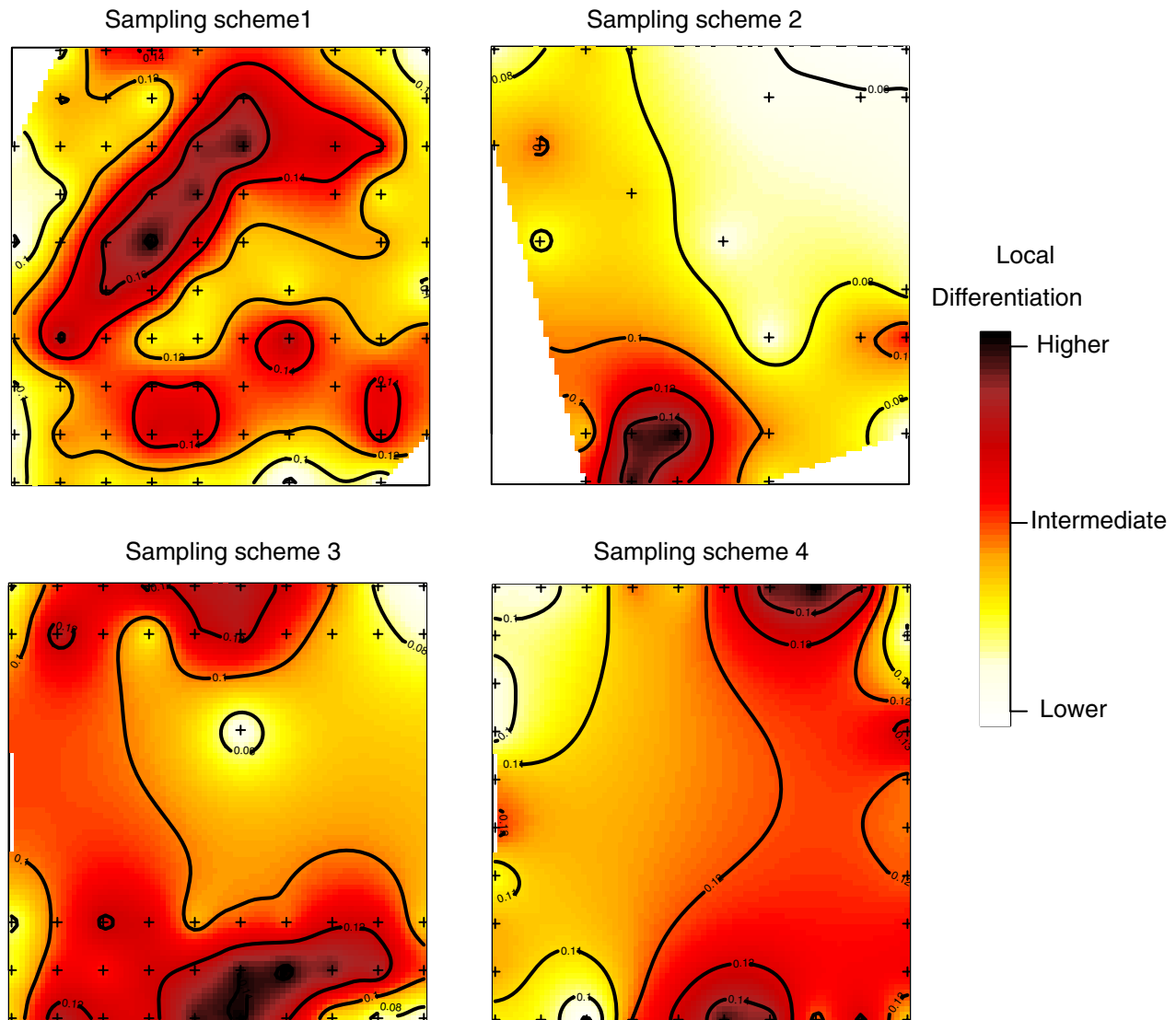
Figure 6. Effect of sampling scheme on local differentiation for data simulated with a stationary model of isolation-by-distance. Simulations are performed with *ms* under a stepping-stone model.

Testing nonstationary patterns of IBD

The estimates of local genetic differentiation may depend on the sampling scheme. Clustered or irregular sampling scheme in particular can be a matter of concern because they might generate false-positive patterns of nonstationarity. Here, we consider different sampling schemes in a two-dimensional stepping-stone model to study the risk of false positives. In the first and second sampling schemes, respectively, 25% and 75% of the total number of sites have been sampled (Fig. 6). The third sampling scheme consists of a clustered sampling scheme with two different geographic zones, which have been sampled, as well as an isolated sampled site between the two regions. In the last sampling scheme, only the perimeter of the two-dimensional square has been sampled. Figure 6 shows heat maps of local genetic differentiation for the stationary IBD simulations that have been generated with

ms. To evaluate whether the observed variations of local genetic differentiation are sufficient evidence for nonstationarity, we perform 100 simulations of stationary patterns for each sampling scheme. When simulations of the null models are performed with *ms*, stationarity is rejected for 3–7% of the simulations, which is consistent with the nominal 5% type I error we use. However if equation (4) is used as a null model, stationarity is rejected for all the simulations performed with *ms*. The stepping-stone simulations of stationary IBD show that equation (4) should not be used for hypothesis testing and we should instead resort to explicit simulation of IBD models for generating distributions of the test statistics under the null hypothesis of stationarity. In addition, for all the simulations of nonstationary processes considered so far (barriers in one- and two-dimensional models and gradient of gene flow), we reject stationarity as expected.

APPLICATIONS

### Nonstationary patterns of IBD among the swedish population

We first illustrate the kriging methodology using a human SNP dataset with a particularly dense geographic sampling. The data consist of genome-wide SNPs for 5174 Swedish individuals that cover all of the 21 Swedish counties Humphreys et al. (2011). To assign each individual to a county, Humphreys et al. (2011) used available geographic information with the following order of priority: city or village of birth, county of birth, municipality or city of residence, and county of residence if it is the only information available. They found strong differences between far northern counties and remaining counties, and also showed that northern counties are more clearly genetically differentiated from each other than southern counties are from each other.

Because our framework is an extension of IBD, we first check that population differentiation increases with increasing geographical distance. We confirm the prevalence of IBD in Sweden ($P < 10^{-7}$ for a Mantel test, see also Fig. S9). Then, we choose to quantify local genetic differentiation using the F_{ST} between a population living exactly in the barycentric center of the county and a putative neighboring population living 30 km away (see Fig. S10). We consider the pairwise $(1 - F_{ST})$ values between the counties as input matrix of pairwise similarities. Our hypothesis-testing procedure indicates significant nonstationary pattern of IBD in Sweden. We find that the northernmost counties (Nordbotten, Västerbotten, and Jämtlands) have the strongest values of local genetic differentiation (Fig. 7), whereas the smallest values are found in the regions around the Stockholm area (Östergötlands, Stockholms, Södermanlands, Västmanlands, and Jönköpings are the five counties with the lowest values of local differentiation). The fourth largest value is found for the Dalarna county (in north middle Sweden), which borders southern Norway, and counts more individuals with remote Finnish or Norwegian ancestry than other counties (Humphreys et al. 2011). As expected, the four counties with the largest local differentiation values are also the most differentiated from other counties even when controlling for geographic distance (Fig. S9).

In summary, we confirm the results of Humphreys et al. (2011) who found that there is more genetic differentiation within northern Sweden than within southern Sweden. The four counties with the largest values of local genetic differentiation are also the counties with the lowest population densities (Fig. S11) suggesting that low population density triggered population differentiation in northern Sweden.

### Nonstationary patterns of IBD for alpine plant species

We consider a set of 20 alpine plant species that have been sampled across the Alps (Gugerli et al. 2008; Alvarez et al. 2009); Jay et al. 2012a). The sampling is particularly dense with one to three indi-
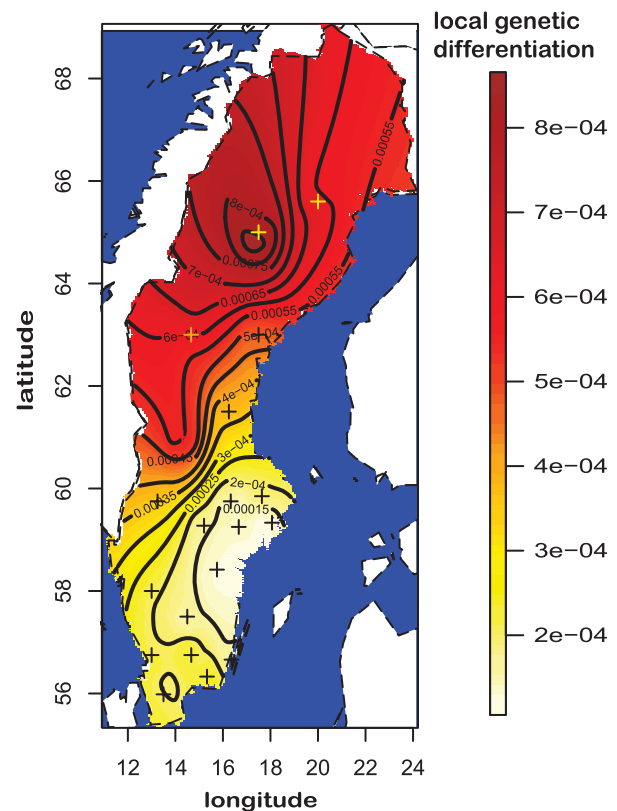
Figure 7. Map of local genetic differentiation computed for the 20 Swedish counties. Measures of local genetic differentiation correspond to the F_{ST} between the sampled counties and fictive neighboring populations, not shown, located at 30 km.

viduals per species collected for each cell of approximately 500 km^2. Individual genotypes consist of AFLPs. We compute allele frequencies at each sampling site—possibly using one individual only—and we consider the matrix of correlation between allele frequencies as input similarity matrix. Local genetic differentiation corresponds to one minus the expected correlation between sampled populations and neighboring populations located at 8 km.

The test for nonstationarity is significant for seven species with a type I error rate of 5% and it increases to nine species when accepting a type I error rate of 10% (Table S1). However, for the species with nonstationary IBD, the detailed pattern of local genetic differentiation is idiosyncratic to each species (Fig. S12). For instance, the alpine species *Phyteuma hemisphaericum* exhibits larger values of local genetic differentiation in a large region ranging from the central Alps to the southwestern Alps, whereas there are disconnected regions of larger genetic differentiation for the species *Arabis Alpina* all located in the South of the Alps (Fig. 8). To integrate the results found for all species with nonstationary IBD, we compute, for each species, normalized rank values between 0 and 1, where 0 corresponds to the site of
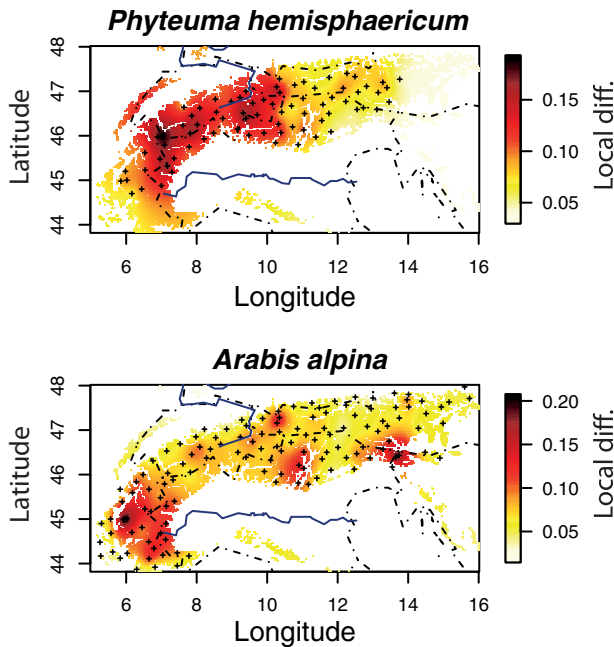
Figure 8. Heatmap of local genetic differentiation computed across the Alps for two alpine species: *Phyteuma hemisphaericum* and *Arabis alpina*. Local differentiation corresponds to one minus the correlation between sampled populations and fictive neighboring populations, not shown, located at 8 km around the sampling sites.

lowest local differentiation and 1 corresponds to the site of largest local differentiation. When averaging the normalized ranks across species, we found that a region of the western Alps encompassing the inner Alpine Aosta valley is the region with the larger values of local differentiation (Fig. S13). This region has already been found to be one of the two major break zones of allele distribution patterns for alpine plant species (Thiel-Egenter et al. 2011). Pleistocene glaciations are putative explanations for the occurrence of a break zone in this region: the populations of plants were initially fragmented into glacial refuges, then expanded via postglacial colonization routes, and a secondary contact zone finally arose where formerly allopatric populations admixed (Pawlowski 1970; Schönswetter et al. 2005; Thiel-Egenter et al. 2011).

Software

The software *LocalDiff* implementing the method is available at http://membres-timc.imag.fr/Michael.Blum/LocalDiff.html. It computes local genetic differentiation from a matrix of pairwise similarity score and can also handle raw genotype data that contain genotypes of individuals. In addition, it generates the *ms* command lines that are required for performing the stepping-stone simulations used in hypothesis testing.

Discussion

In this article, we present a new Bayesian method to characterize nonstationary patterns of IBD. Because the method aims at refining the description of IBD patterns, it should only be applied when IBD has already been detected. From *global* measures of pairwise similarity or dissimilarity, the method infers *local* measures of similarity or dissimilarity. Whatever is the exact measure of genetic (dis)similarity, we use the generic expression of *local genetic differentiation* when referring to the estimated local growth of genetic dissimilarity or differentiation. If considering for instance the F_{ST} pairwise matrix of genetic differentiation between populations, the inferred values correspond to the F_{ST} between the sampled populations and fictive neighboring populations located at a given distance. The method is not restricted to F_{ST} measures and can handle any type of measures of differentiation and is also valid at the individual scale. We consider for instance the correlation between the allelic types of individuals, but other measures would be valid such as identity by descent between individuals (Browning and Browning 2011) as well as coancestry measures (Lawson et al. 2012). Because the two latter measures are based on haplotypes instead of genotypes, they can provide information at a finer geographical scale (Gattepaille and Jakobsson 2012; Lawson et al. 2012).

GENETIC DIFFERENTIATION AND GENE FLOW

It is of course tempting to convert maps of local genetic differentiation into maps of gene flow or of dispersal distance. Assuming that differentiation occurs according to a stepping-stone model, such parameter estimates could be obtained using theoretical relationships between local F_{ST} and dispersal distance (Rousset 1997). However, relating F_{ST} or other measures of genetic differentiation to gene flow relies on many assumptions that may be unrealistic (Marko and Hart 2011). Although the estimation of gene flow with F_{ST}-based methods can be robust in some situations, such as temporal variation of gene flow (Leblois et al. 2004), there are other processes such as range expansion, local extinction, and recolonization that can modify drastically the pattern of genetic differentiation (Wade and McCauley 1988; Arenas et al. 2012). More generally, a map of local genetic differentiation is informative about the *pattern* of genetic differentiation but do not provide enough information to distinguish between the possible evolutionary *processes* that generated this pattern (for similar concerns about principal component analysis, see McVean 2009). To provide a concrete example, the same pattern, a zone of elevated local differentiation, can be interpreted as a barrier to gene flow in a equilibrium stepping-stone process (Figs. 2–4) or as a secondary contact zone following postglacial expansions in the case of the alpine plant species (Thiel-Egenter et al. 2011).

RELEVANCE TO LANDSCAPE GENETICS

Two key steps in landscape genetics are the detection of genetic discontinuities and the correlation of these discontinuities with landscape and environmental features such as barriers (Manel et al. 2003). Detection of genetic discontinuities is clearly provided by the proposed kriging method; for instance in the case of the alpine species *P. hemisphaericum*, we find genetic discontinuities, that is, larger local genetic differentiation, in a large region of the central and western Alps (Fig. 8). However, aiming at capture genetic discontinuities using barriers only might be too limited and the kriging method can reveal more complex patterns such as gradient of local differentiation across the species' range (Fig. 5). The second key step where the genetic discontinuities are correlated with landscape or environmental variable can also be obtained as a postprocessing step by correlating estimated local genetic differentiation with landscape variables. For instance, in the case of the human SNP Swedish data, we find that local genetic differentiation is correlated with population density. There are alternative and *integrative* approaches that account for both genetic data and landscape variables within the same statistical framework. Accounting for both sources of data can be performed either by a joint assessment of the pattern of population structure or differentiation and its correlation with landscape or environmental variable (Foll and Gaggiotti 2006; Jay et al. 2011), or by correlating genetic distances with distances based on landscape features (Cushman et al. 2006; McRae 2006). These integrative approaches are *hypothesis-driven* in the sense that each set of landscape features affecting population structure corresponds to one hypothesis that can be tested or compared to other ones. The proposed kriging approach is instead a technique of *exploratory data analysis*. It might be especially appropriate for large-scale conservation studies not focused on the underlying evolutionary processes but that should deal with reserve design and with the management of fragmented populations (Schwartz et al. 2007). To explore patterns of genetic differentiation, there are other statistical summaries of the data that can be computed. Population-specific F_{ST}'s based on the F-model can also provide local measures of differentiation by computing local values of F_{ST}'s (Gaggiotti and Foll 2010). However, compared to approaches based on the F-model, *LocalDiff* can also work with the individual-based sampling schemes often encountered in landscape genetics (Schwartz and McKelvey 2009). Moreover, F_{ST}'s based on the F-model relies on a parametric population-genetic model that may be sensitive to departures from the assumption of the F-model (Gaggiotti and Foll 2010). By contrast, the proposed approach relies on kriging (aka Gaussian process), which is a nonparametric approach that assumes a pattern of IBD only.

To compare or test the support of different evolutionary processes provided by the pattern of nonstationary IBD, we can rather resort to inference based on explicit simulations of evolutionary processes using for instance approximate Bayesian computation (Csilléry et al. 2010). Within this simulation framework, measures of local genetic differentiation can be included as statistical summaries of the data. When the number of sampling sites is large, the MCMC algorithm might be too slow to provide summary statistics for approximate Bayesian computation. To overcome this problem, we provide an option in *LocalDiff* where measures of local differentiation are computed by integrating the parameter triplet (α, λ, r) over the prior distribution instead of the posterior distribution. Although integration over the posterior with MCMC should be preferred when possible, integrating over the prior distribution provides relevant measures of local differentiation for the examples we investigated (see Fig. S14).

CAVEATS

Although, computing local differentiation is a descriptive technique of exploratory data analysis, we provide an hypothesis-testing procedure attached to it. This is desirable feature because it can prevent from overinterpreting maps of local differentiation (Fig. 6). The test of nonstationarity relies on two different test statistics and the null distribution is obtained from stepping-stone simulations performed with *ms*. To obtain the null distribution of the test statistic, we consider the same sampling as in the data, which is crucial because the sampling scheme can affect the inferred pattern of local differentiation (Fig. 6). For the simulations of the null model, we uniformly sample the parameter $4N_0m$ in a fixed range, but we acknowledge that using an estimated value of the effective migration parameter would also be a valid strategy. We also consider an alternative and approximate null model under which the correlation decays exponentially (equation (4)), but this approximation should not be used for hypothesis testing because it is much too liberal. However, even considering explicit stepping-stone simulations might have drawbacks because other processes such as range expansions can also generate IBD patterns, but they might induce different distributions of tests statistics (Edmonds et al. 2004). More generally, sampling scheme affects the different methods that characterize population differentiation and simulations can be informative about the effects of sampling scheme (McVean 2009; Schwartz and McKelvey 2009; Jay et al. 2012b).

Another concern about the kriging method concerns the choice of the distance between the sampled populations or individuals and the fictive neighboring populations or individuals. We consider various choices of distances for the Swedish dataset (10–150 km) and find that all these choices of distance provide similar patterns of local genetic differentiation although the patterns generated with the larger distances are smoother (Fig. S15). The last caveat to bear in mind concerns the choice of the pairwise dissimilarity or differentiation matrix. For instance, in a simulation study with five populations, Lawson and Falush (2012)

showed that the five populations were clearly distinguishable with some but not all pairwise dissimilarity matrices between individuals. Although a potential caveat, being able to choose the measure of dissimilarity also adds to the flexibility of the method and different measures can convey information about processes that occurred at different time periods.

PERSPECTIVES

When carefully addressing the aforementioned caveats, measures of local genetic differentiation can provide interpretable patterns for describing nonstationary patterns of IBD. Here, the expression *non-stationary* refers to spatial variations but note that if temporal data are available, a similar kriging framework can be used to study how genetic drift evolves as a function of time. The present method relies on a matrix of pairwise (dis)similarity between individuals or populations located on georeferenced sampling sites. An approach based on dissimilarity matrices is an appropriate methodology for the new genomic era where we have to deal with massive data. Computing pairwise dissimilarity matrix can be computationally efficient (e.g., Browning and Browning 2011) and can even be parallelized to compute different parts of the matrix (Lawson and Falush 2012). Having a statistical method able to scale with the dimension of the genetic data should make it a valuable tool for investigating patterns of genetic differentiation in a wide range of studies.

REFERENCES

Alvarez, N., C. Thiel-Egenter, A. Tribsch, R. Holderegger, S. Manel, P. Schönswetter, P. Taberlet, S. Brodbeck, M. Gaudeul, L. Gielly, et al. 2009. History or ecology? Substrate type as a major driver of spatial genetic structure in Alpine plants. Ecol. Lett. 12:632–640.

Arenas, M., N. Ray, M. Currat, and L. Excoffier. 2012. Consequences of range contractions and range shifts on molecular diversity. Mol. Biol. Evol. 29:207–218.

Arnaud, J. 2003. Metapopulation genetic structure and migration pathways in the land snail *Helix aspersa*: influence of landscape heterogeneity. Landscape Ecol. 18:333–346.

Barbujani, G., N. Oden, and R. Sokal. 1989. Detecting regions of abrupt change in maps of biological variables. Syst. Biol. 38: 376–389.

Bishop, C. 2006. Pattern recognition and machine learning, Vol. 4. Springer, New York.

Bocquet-Appel, J., and J. Bacro. 1994. Generalized wombling. Syst. Biol. 43: 442–448.

Brown, P., N. Le, and J. Zidek. 1994. Multivariate spatial interpolation and exposure to air pollutants. Can. J. Stat. 22:489–509.

Browning, B., and S. Browning. 2011. A fast, powerful method for detecting identity by descent. Am. J. Hum. Genet. 88:173–182.

Castella, V., M. Ruedi, L. Excoffier, C. Ibanez, R. Arlettaz, and J. Hausser. 2000. Is the Gibraltar strait a barrier to gene flow for the bat *Myotis myotis* (Chiroptera: Vespertilionidae)? Mol. Ecol. 9:1761–1772.

Cercueil, A., O. François, and S. Manel. 2007. The genetical bandwidth mapping: a spatial and graphical representation of population genetic structure based on the wombling method. Theor. Popul. Biol. 71:332–341.

Cressie, N. A. C. 1993. Statistics for spatial data (Wiley series in probability and statistics). Wiley-Interscience, New York.

Crida, A., and S. Manel. 2007. Wombsoft: an R package that implements the wombling method to identify genetic boundary. Mol. Ecol. Notes 7: 588–591.

Csilléry, K., M. G. B. Blum, O. E. Gaggiotti, and O. François. 2010. Approximate Bayesian computation (ABC) in practice. Trends Ecol. Evol. 25: 410–418.

Cushman, S., K. McKelvey, J. Hayden, and M. Schwartz. 2006. Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. Am. Nat. 168:486–499.

Dupanloup, I., S. Schneider, and L. Excoffier. 2002. A simulated annealing approach to define the genetic structure of populations. Mol. Ecol. 11: 2571–2581.

Edmonds, C. A., A. S. Lillie, and L. L. Cavalli-Sforza. 2004. Mutations arising in the wave front of an expanding population. Proc. Natl. Acad. Sci. USA 101:975–979.

Epps, C., P. Palsbøll, J. Wehausen, G. Roderick, R. Ramey II, and D. McCullough. 2005. Highways block gene flow and cause a rapid decline in genetic diversity of desert bighorn sheep. Ecol. Lett. 8:1029–1038.

Foll, M., and O. Gaggiotti. 2006. Identifying the environmental factors that determine the genetic structure of populations. Genetics 174: 875–891.

Gaggiotti, O. E., and M. Foll. 2010. Quantifying population structure using the F-model. Mol. Ecol. Res. 10:821–830.

Gattepaille, L. M., and M. Jakobsson. 2012. Combining markers into haplotypes can improve population structure inference. Genetics 190:159–174.

Gauffre, B., A. Estoup, V. Bretagnolle, and J.-F. Cosson. 2008. Spatial genetic structure of a small rodent in a heterogeneous landscape. Mol. Ecol. 17:4619–4629.

Gugerli, F., T. Englisch, H. Niklfeld, A. Tribsch, Z. Mirek, M. Ronikier, N. Zimmermann, R. Holderegger, and P. Taberlet. 2008. Relationships among levels of biodiversity and the relevance of intraspecific diversity in conservation—a project synopsis. Perspect. Plant Ecol. Evol. Syst. 10:259–281.

Handcock, M., and M. Stein. 1993. A Bayesian analysis of kriging. Technometrics 35:403–410.

Hardy, O., and X. Vekemans. 1999. Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. Heredity 83: 145–154.

Hardy, O. J., L. Maggia, E. Bandou, P. Breyne, H. Caron, M. H. Chevallier, A. Doligez, C. Dutech, A. Kremer, C. Latouche-hallé, et al. 2006. Fine-scale genetic structure and gene dispersal inferences in 10 neotropical tree species. Mol. Ecol. 15:559–571.

Hellberg, M. 2009. Gene flow and isolation among populations of marine animals. Annu. Rev. Ecol. Syst. 40:291–310.

Hudson, R. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics 18:337–338.

Humphreys, K., A. Grankvist, M. Leu, P. Hall, J. Liu, S. Ripatti, K. Rehnström, L. Groop, L. Klareskog, B. Ding, et al. 2011. The genetic structure of the Swedish population. PLoS One 6:e22547.

Jay, F., O. François, and M. G. B. Blum. 2011. Predictions of native American population structure using linguistic covariates in a hidden regression framework. PLoS One 6:e16227.

Jay, F., S. Manel, N. Alvarez, E. Y. Durand, W. Thuiller, R. Holderegger, P. Taberlet, and O. François. 2012a. Forecasting changes in population genetic structure of alpine plants in response to global warming. Mol. Ecol. 21:2354–2368.

Jay, F., P. Sjödin, M. Jakobsson, and M. G. B. Blum. 2012b. Anisotropic isolation by distance: the main orientations of human genetic differentiation. Mol. Biol. Evol. 30:513–525.

Kimura, M., and G. Weiss. 1964. The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics 49:561–576.

Lawson, D. J., and D. Falush. 2012. Population identification using genetic data. Annu. Rev. Genomics Hum. Genet. 13:337–361.

Lawson, D. J., G. Hellenthal, S. Myers, and D. Falush. 2012. Inference of population structure using dense haplotype data. PLoS Genet. 8:e1002453.

Leblois, R., F. Rousset, and A. Estoup. 2004. Influence of spatial and temporal heterogeneities on the estimation of demographic parameters in a continuous population using individual microsatellite data. Genetics 166:1081–1092.

Manel, S., and R. Holderegger. 2013. Ten years of landscape genetics. Trends Ecol. Evol. 28:614–621.

Manel, S., M. Schwartz, G. Luikart, and P. Taberlet. 2003. Landscape genetics: combining landscape ecology and population genetics. Trends Ecol. Evol. 18:189–197.

Manel, S., F. Berthoud, E. Bellemain, M. Gaudeul, G. Luikart, J. Swenson, L. Waits, P. Taberlet, et al. 2007. A new individual-based spatial approach for identifying genetic discontinuities in natural populations. Mol. Ecol. 16:2031–2043.

Manni, F., E. Guerard, and E. Heyer. 2004. Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using Monmonier's algorithm. Hum. Biol. 76:173–190.

Marko, P., and M. Hart. 2011. The complex analytical landscape of gene flow inference. Trends Ecol. Evol. 26:448–456.

McRae, B. 2006. Isolation by resistance. Evolution 60:1551–1561.

McRae, B., and P. Beier. 2007. Circuit theory predicts gene flow in plant and animal populations. Proc. Natl. Acad. Sci. USA 104:19885–19890.

McVean, G. 2009. A genealogical interpretation of principal components analysis. PLoS Genet. 5:e1000686.

Munshi-South, J. 2012. Urban landscape genetics: canopy cover predicts gene flow between white-footed mouse (*Peromyscus leucopus*) populations in New York City. Mol. Ecol. 21:1360–1378.

Nott, D. J., and W. T. Dunsmuir. 2002. Estimation of nonstationary spatial covariance structure. Biometrika 89:819–829.

Novembre, J., and M. Stephens. 2008. Interpreting principal component analyses of spatial population genetic variation. Nat. Genet. 40:646–649.

Paciorek, C. J., and M. J. Schervish. 2006. Spatial modelling using a new class of nonstationary covariance functions. Environmetrics 17:483–506.

Pawlowski, B. 1970. Remarques sur l'endémisme dans la flore des alpes et des carpates. Plant Ecol. 21:181–243.

Ramachandran, S., O. Deshpande, C. C. Roseman, N. A. Rosenberg, M. W. Feldman, and L. L. Cavalli-Sforza. 2005. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proc. Natl. Acad. Sci. USA 102:15942–15947.

Riley, S., J. Pollinger, R. Sauvajot, E. York, C. Bromley, T. Fuller, and R. Wayne. 2006. Fast-track: a southern California freeway is a physical and social barrier to gene flow in carnivores. Mol. Ecol. 15:1733–1741.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics 145:1219–1228.

Safner, T., M. Miller, B. McRae, M. Fortin, and S. Manel. 2011. Comparison of Bayesian clustering and edge detection methods for inferring boundaries in landscape genetics. Int. J. Mol. Sci. 12:865–889.

Schmidt, A., and A. O'Hagan. 2003. Bayesian inference for non-stationary spatial covariance structure via spatial deformations. J. R. Stat. Soc. Ser. B 65:743–758.

Schönswetter, P., I. Stehlik, R. Holderegger, and A. Tribsch. 2005. Molecular evidence for glacial refugia of mountain plants in the European Alps. Mol. Ecol. 14:3547–3555.

Schwartz, M., and K. McKelvey. 2009. Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. Conserv. Genet. 10:441–452.

Schwartz, M., G. Luikart, and R. Waples. 2007. Genetic monitoring as a promising tool for conservation and management. Trends Ecol. Evol. 22:25–33.

Serrouya, R., D. Paetkau, B. N. Mclellan, S. Boutin, M. Campbell, and D. A. Jenkins. 2012. Population size and major valleys explain microsatellite variation better than taxonomic units for caribou in western Canada. Mol. Ecol. 21:2588–2601.

Sharbel, T., B. Haubold, and T. Mitchell-Olds. 2000. Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. Mol. Ecol. 9:2109–2118.

Slatkin, M. 1985. Gene flow in natural populations. Annu. Rev. Ecol. Syst. 16:393–430.

———. 1993. Isolation by distance in equilibrium and non-equilibrium populations. Evolution 47:264–279.

Storfer, A., M. Murphy, S. Spear, R. Holderegger, and L. Waits. 2010. Landscape genetics: where are we now? Mol. Ecol. 19:3496–3514.

Székely, G. J., M. L. Rizzo, and N. K. Bakirov. 2007. Measuring and testing dependence by correlation of distances. Annals Stat. 35:2769–2794.

Thiel-Egenter, C., N. Alvarez, R. Holderegger, A. Tribsch, T. Englisch, T. Wohlgemuth, L. Colli, M. Gaudeul, L. Gielly, N. Jogan, et al. 2011. Break zones in the distributions of alleles and species in Alpine plants. J. Biogeogr. 38:772–782.

Wade, M., and D. McCauley. 1988. Extinction and recolonization: their effects on the genetic differentiation of local populations. Evolution 42:995–1005.

Wright, S. 1943. Isolation by distance. Genetics 28:114–138.

Zalewski, A., S. Piertney, H. Zalewska, and X. Lambin. 2009. Landscape barriers reduce gene flow in an invasive carnivore: geographical and local genetic structure of American mink in Scotland. Mol. Ecol. 18:1601–1615.

Associate Editor: J. Fordyce

Appendix A: Normalizing the Covariance Matrix

To renormalize equation (3) and compute the correlation matrix, we have to compute the diagonal elements of the variance–covariance matrix $\mathbf{Var}(\mathbf{Y})$ for the unsampled sites. Using the fact that the residuals of the regression equation (2) are of variance $\Psi_{yy} - \Psi_{xy}\Psi_{xx}^{-1}\Psi_{xy}$ (Bishop 2006), we can show that the variance–covariance matrix $\mathbf{Var}(\mathbf{Y})$ is given by

$$\mathbf{Var}(\mathbf{Y}) = \tau_\Psi \mathbf{Var}(\mathbf{X})\tau_\Psi^T + \Psi_{yy} - \Psi_{xy}\Psi_{xx}^{-1}\Psi_{xy}.$$

Appendix B: Gibbs Sampler

For the parameters of the correlogram model, we choose the following prior distributions

$$\alpha \sim \mathcal{U}(\min(\mathbf{Var}(\mathbf{X})) - \delta, \min(\mathbf{Var}(\mathbf{X})) + \delta) \qquad (5)$$

$$\log_{10} \lambda \sim \mathcal{U}(-4, -1) \qquad (6)$$

$$r \sim \mathcal{U}(\min D_{ij}, \max D_{ij}), \qquad (7)$$

where $\min(\mathbf{Var}(\mathbf{X}))$ is the smallest element of the variance–covariance matrix of $\mathbf{X}$ and D denotes the pairwise geographical distance between sampled sites. Here, we adopt an empirical Bayes approach, because we partly use the data when defining the prior distributions.

Using the Bayes formula we have

$$\log(p(\alpha, \lambda, r | \mathbf{Var}(\mathbf{X}))) = \log(p(\mathbf{Var}(\mathbf{X}) | \alpha, \lambda, r))$$
$$+ \log(p(\alpha, \lambda, r)) + C_1, \qquad (8)$$

where C_1 is a constant and the likelihood is given by the Wishart distribution Schmidt and O'Hagan (2003)

$$\log(p(\mathbf{Var}(\mathbf{X}) | \alpha, \lambda, r)) = -\frac{l-1}{2} \log(|\Psi_{\mathbf{xx}}|)$$
$$-\frac{l}{2} \mathrm{trace}(\Psi_{\mathbf{xx}}^{-1} \mathbf{Var}(\mathbf{X})) + C_2, \qquad (9)$$

where C_2 is another constant, and l is the number of loci. We then simulate a sample of the joint posterior probability using an hybrid algorithm with Gibbs and Metropolis–Hastings (MH) updating steps. To obtain replicates from the conditional distributions $p(\alpha | \lambda, r, \mathbf{Var}(\mathbf{X}))$ and $p(\lambda | r, \alpha, \mathbf{Var}(\mathbf{X}))$, we evaluate the likelihood function over a grid for the triplet (α, λ, r). The evaluations of the likelihood function are performed before running the MCMC algorithm and are saved to be subsequently used during the course of the algorithm. We evaluate the conditional densities for each point of the grid using equations (8) and (9) and then we use a sampling algorithm for simulating discrete random variables with known probability masses. The updates for the parameter r are performed with an MH algorithm using as proposal a discrete random walk over the grid. The resolution of the grid is more acute for the parameter r than for the other two parameters, which explains why we consider an MH step that was less time-consuming than evaluating the likelihood function for all the points of the grid.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website: