Analysing landscape effects on dispersal networks and gene flow with genetic graphs

Paul Savary, Jean-Christophe Foltête, Hervé Moal, Gilles Vuidel, Stéphane Garnier

Analysing landscape effects on dispersal networks and gene flow with genetic graphs

Savary, Paul*[1, 2, 3], Foltête, Jean-Christophe[2], Moal, Hervé[1], Vuidel, Gilles[2], and Garnier, Stéphane[3]

[1] *ARP-Astrance, 9 Avenue Percier, 75008 Paris, France*
[2] *ThéMA, UMR 6049 CNRS, Université Bourgogne-Franche-Comté, 32 Rue Mégevand, 25030 Besançon Cedex, France*
[3] *Biogéosciences, UMR 6282 CNRS, Université Bourgogne-Franche-Comté, 6 Boulevard Gabriel 21000 Dijon, France*

*Corresponding author: `paul.savary@univ-fcomte.fr`

Abstract

Graph-theoretic approaches have relevant applications in landscape genetic analyses. When species form populations in discrete habitat patches, genetic graphs can be used i) to identify direct dispersal paths followed by propagules or ii) to quantify landscape effects on multi-generational gene flow. However, the influence of their construction parameters remains to be explored. Using a simulation approach, we constructed genetic graphs using several pruning methods (geographical distance thresholds, topological constraints, statistical inference) and genetic distances to weight graph links (F_{ST}, D_{PS}, Euclidean genetic distances). We then compared the capacity of these different graphs to i) identify the precise topology of the dispersal network and ii) to infer landscape resistance to gene flow from the relationship between cost-distances and genetic distances. Although not always clear-cut, our results showed that methods based on geographical distance thresholds seem to better identify dispersal networks in most cases. More interestingly, our study demonstrates that a sub-selection of pairwise distances through graph pruning (thereby reducing the number of data points) can counter-intuitively lead to improved inferences of landscape effects on dispersal. Finally, we showed that genetic distances such as the D_{PS} or Euclidean genetic distances should be preferred over the F_{ST} for landscape effect inference as they respond faster to landscape changes.

Keywords: landscape genetics, ecological connectivity, graph theory, simulation, dispersal

First published online in *Molecular Ecology Resources* by the 18 january 2021

1 Introduction

Landscape connectivity is defined as the degree to which the landscape facilitates or impedes movement among resource patches (Taylor et al., 1993). Such dispersal events reduce metapopulation extinction risk (Den Boer, 1968; Hanski, 1998) and give rise to gene flow, thereby preventing inbreeding depression and maintaining local adaptation potential (but see Crispo et al. (2011), Richardson et al. (2016) and Lenormand (2002)). Therefore, understanding dispersal patterns is crucial for biodiversity conservation.

Landscape genetic approaches have been increasingly used to assess landscape influence on dispersal (Balkenhol et al., 2016; Dyer, 2015a; Manel et al., 2003; Storfer et al., 2007) because genetic data based inferences provide insights into effective movements that led to reproduction when inferences drawn from mark-recapture data or GPS tracks mostly identify current movements (Mateo-Sánchez et al., 2015; Zeller et al., 2018). Although advances have been achieved in landscape genetics in the last 15 years (Manel and Holderegger, 2013; Storfer et al., 2010), there are still methodological and theoretical challenges, to analysing and interpreting genetic data especially (Balkenhol et al., 2009a,b; Dyer, 2015a).

Graph-theoretic approaches are particularly relevant when dispersal events occur between patchy populations forming a network (Greenbaum and Fefferman, 2017). A genetic graph is made of i) a set of nodes corresponding to gene pools sampled from different sites, and ii) a set of links connecting them through gene flow. The graph is basically a pairwise adjacency matrix with 0 and 1 reflecting absence or presence of links between populations, but the links can also be weighted by measures of genetic differentiation. In this case, it is often recommended to prune the complete graph, in other words to remove links between some node pairs, e.g. indirectly connected through intermediate nodes, to make the topology easier to visualise and to keep only the most relevant links in light of the study aim.

Genetic graphs are flexible tools that can be used in multiple fashions in landscape genetic studies, offering a great potential for inferring models of network flow (Murphy et al., 2015). Indeed, a certain level of gene flow between two populations can result from direct exchanges of propagules and/or indirect exchanges through intervening populations in a stepwise way over several generations. Although considering only the genetic distance between two populations does not indicate whether gene flow occurred in a direct or indirect way, estimating genetic differentiation between a population pair conditionally upon other populations should make it possible to disentangle direct versus indirect gene flow between them (Dyer, 2015b). Hence, using the conditional independence principle (Magwene, 2001; Whittaker, 2009) can be a way to identify the precise topology of the dispersal network (i.e. the set of links depicting dispersal of propagules between populations), in other words identifying the set of edges that represents contributing connections among nodes (Murphy et al., 2015).

Alternatively, a genetic graph can be used for quantifying landscape feature resistance to gene flow through distance-based analyses (Garroway et al., 2011). Assessing the correlation between genetic distances and geographical or effective landscape distances is a way to identify the hypothesis that best fits the genetic data, and thus reflects landscape influence on gene flow, among several hypotheses of landscape feature resistance (Cushman et al., 2006; Khimoun et al., 2017; Peterman, 2018; Ruiz-Gonzalez et al., 2015). Although such inferences are usually based upon complete matrices of distance, several authors have suggested that reducing these matrices to a subset of population pairs may improve their robustness (Van Strien et al., 2015;

Van Strien, 2017; Wagner and Fortin, 2013; Zeller et al., 2016). Graph pruning precisely involves selecting a subset of population pairs which therefore makes genetic graphs particularly relevant in this context.

Reducing the dataset by removing population pairs in order to improve inferences of landscape resistance is somehow counter-intuitive, but it lies on the following rationale. Assuming that dispersal is generally spatially limited, several theoretical models of populations genetics predict that measures of genetic differentiation are linearly and positively correlated with geographical distance, provided enough time has elapsed for this equilibrium pattern to become established (Guillot et al., 2009; Kimura and Weiss, 1964; Slatkin, 1993; Wright, 1943). Models also predict that the spatial scale over which this pattern of Isolation by Distance (IBD) has reached its stationary state should increase with time following the establishment of populations (Slatkin, 1993). In other words, before complete equilibrium has been reached, IBD is only observed between nearby populations but not between more distant ones. Note that all these models assume that the landscape exerts an homogeneous effect on dispersal, and most of them exclude spatial variation in population density (Guillot et al., 2009). However, most real landscapes are heterogeneous, and a common way to consider landscape feature suitability for dispersal is to replace Euclidean distances by landscape distances (e.g. cost-distances or resistance distances) in the analysis of population genetic structure (Balkenhol et al., 2016; Coulon et al., 2004; Peterman, 2018). If the isolation by landscape resistance (IBLR) model extends the IBD model in heterogeneous landscapes, its theoretical expectations have been less strongly investigated. Nevertheless, a model developed by McRae (2006) predicts a linear positive relationship between genetic differentiation and landscape distances. Here again, some time is needed for patterns of differentiation to reflect the influence of landscape features on dispersal, and model assumptions are more likely to be verified at shorter landscape distances before the complete equilibrium has been reached (McRae, 2006). Hence, better inferences of landscape resistance to gene flow may be obtained when selecting the subset of populations pairs that are within a certain spatial distance. This issue is critical as landscape genetic studies are frequently performed in human-shaped landscapes which have undergone recent modifications potentially affecting demography (Manel and Holderegger, 2013; Storfer et al., 2010), but the relevance of the different graph pruning methods in this context has been rarely investigated.

In this study, we used a simulation approach to compare the relative efficiency of several graph pruning methods, genetic distances and analysis parameters of a genetic graph regarding two objectives in inferring network flow: i) identifying the precise topology of the dispersal network and ii) assessing the capacity of landscape distances to predict genetic distances. First, we assessed the efficiency of three kinds of criteria used for excluding graph links: geographical distance thresholds (leading to the exclusion of links corresponding to geographical distances larger than a threshold value), topology (involving topological constraints in graph pruning), and statistical inference of conditional independence based on genetic data (Dyer and Nason, 2004). Second, we compared some of the numerous genetic distances used to weight graph links (Murphy et al., 2015): F_{ST} (Keller et al., 2013; Munwes et al., 2010), D_{PS} (Naujokaitis-Lewis et al., 2013; Keller et al., 2013), genetic Euclidean distance (Excoffier et al., 1992). Finally, we compared two common practices in distance-based analyses. The first one relies on the correlation between genetic and landscape distances corresponding with population pairs that are directly connected in the genetic graph. The second one is based on the same correlation, but considering all pairwise genetic distances (between population pairs directly connected or not), by summing genetic distances along the shortest direct or indirect path between these populations on the graph. Dyer et al. (2010) revealed a higher correlation of this conditional

genetic distance (cGD) than pairwise F_{ST} with landscape distance. Yet, these two matrices (cGD *vs* complete F_{ST} matrix) involved two different genetic distances (Euclidean genetic distance and F_{ST}) and two kinds of links (paths made of direct dispersal paths only *vs* direct plus indirect paths) at the same time, thereby introducing a confounding factor in the comparison. Therefore, the ability of these practices in distance-based analyses to infer landscape effects on dispersal still needs investigation.

2 Material & Methods

2.1 Landscape data

We simulated 10 landscapes using spatially correlated Gaussian random fields models (autocorrelation range: 10)(Schlather et al., 2015) with NLMR package in R (Sciaini et al., 2018). Land cover proportions were close to those encountered in agricultural landscapes dominated by crops and grasslands with small remaining forest fragments. Cost values were assigned to five cover types to simulate the dispersal capacities of a forest specialist species. These cost values and land cover proportions were the following: crops (cost: 60, proportion: 35 %), grassland (40, 35 %), forest (1, 15 %), shrubland (5, 7.5 %) and artificial areas (1000, 7.5 %). We based these costs on values already employed to analyse ecological connectivity in forest species (Gurrutxaga et al., 2010; Schadt et al., 2002), and their range (1-1000) matches that inferred from field data in other empirical studies on a wide range of taxa with contrasted dispersal capacities (Khimoun et al., 2017; Pérez-Espona et al., 2008; Ruiz-González et al., 2014; Wang et al., 2008).

The resulting landscapes were square raster grids of 3600 square kilometres with a resolution of 100 m. We randomly selected 50 population locations within the forest patches, separated by a distance larger than 3 km from one another. Ten population location distributions were created for each landscape in order to vary the cost-distance value distribution. Each population contained 30 individuals during the simulation.

2.2 Gene flow simulation

We used CDPOP (Landguth and Cushman, 2010) to simulate gene flow. Population size and sex-ratio (equal to 1) remained constant throughout the simulation of 500 generations. At each generation, individuals mate in their own population and juveniles may disperse to establish in other populations. The number of offspring per female follows a Poisson distribution ($\lambda = 3$). Once every population is occupied by 30 native or dispersing individuals, following individuals immigrating die. Mating is done with replacement for males only, and generations are non-overlapping. Individual genotypes were simulated for 20 loci with 30 alleles per locus, thereby emulating the frequent use of microsatellites in landscape genetic studies (Storfer et al., 2010). Initial genotypes were assigned randomly at generation 0 as starting allele frequencies do not affect the overall final pattern of genetic differentiation (Graves et al., 2013). There was no selection but mutations could occur (k-alleles mutation model, $\mu = 0.0005$).

Gene flow depended on simulated landscape resistance. With respect to the second objective, i.e. assessing the capacity of landscape distances to predict genetic distances, we aimed at simulating contrasted patterns of genetic structure in terms of spatial scale at which IBLR was observed. We first explored several simulation settings before retaining the following one. For every 100 combinations of landscape and distribution of populations, a landscape graph with 50 nodes was built. Each node corresponded to a habitat patch occupied by one population. Cost-distances (CD) between habitat patches were calculated following Adriaensen et al. (2003)

as the accumulated cost along the least-cost path between each pair of habitat patches, using Dijkstra's algorithm on Graphab software (Foltête et al., 2012). Then, these CD values were used to weight the links of the graph, which initially had a complete topology. Using the edge-thinning method (Urban and Keitt, 2001), we removed links one by one in descending order of CD until we identified the link whose removal would have disconnected the graph into two components. The CD associated to this link was the "percolation threshold" (Rozenfeld et al., 2008). During gene flow simulations, dispersal probabilities associated with links whose CD values were above $1.1 \times percolation.threshold$ were set to 0. $1.1 \times percolation.threshold$ is therefore the maximum dispersal distance. The resulting population networks were made of the set of direct dispersal paths which could possibly be followed by individuals and thus represented the potential dispersal network. It had a single component, thereby preventing single populations from being totally isolated, which is theoretically necessary for populations to survive (Allendorf et al., 2007; Frankham et al., 2004).

The decrease of individual dispersal probability according to CD was modeled by a negative exponential function (Clobert et al., 2012; Hanski et al., 2000; Urban and Keitt, 2001), such that: $p(CD) = e^{-\beta CD}$. β values were calculated such that the CD associated with a dispersal probability of 0.01 was equivalent to 5 % of the percolation threshold. Preliminary tests revealed that these settings resulted in proportions of migrants akin to those empirically described by Bowne and Bowers (2004).

For each simulation scenario (*i.e.* combination of landscape and distribution of populations), gene flow was simulated 10 times (1000 simulations in total). We used genotypes from generations 50 and 500 to construct genetic graphs. After each simulation, a "realised dispersal graph" was built. Its links were all the links that had been followed by at least one individual during the simulation. Genetic graphs built in order to recover the topology of the dispersal network were supposed to reproduce the topology of this realised dispersal graph.

2.3 Genetic graphs

We constructed genetic graphs using several pruning methods and genetic distances to weight the links (see table 1 for the list of combinations).

2.3.1 Pruning method

We pruned the genetic graphs using nine pruning methods based upon three criteria: i) geographical distance thresholds, ii) topology and iii) statistical inference.

First, we pruned graphs by removing all the links between nodes separated by a geographical distance larger than a given threshold. We used 4 thresholds: 10, 15, 20 and 30 km (GEO-10, GEO-15, GEO-20 and GEO-30, respectively). We chose this range of values to keep most graphs connected (Naujokaitis-Lewis et al., 2013) and because above 30 km, the resulting graphs were complete graphs given the size of the landscapes. We used thresholds in geographical distance units instead of cost-distance units because in practice researchers are not supposed to have previous knowledge of cost values associated with land cover types.

The second pruning criterion aimed at constructing graphs with a specific topology, in agreement with the species dispersal pattern hypothesised *a priori*. Genetic graphs were first given the topology of a Gabriel graph (GAB)(Arnaud, 2003). This type of graph in which only neighbouring populations are connected assumes a stepping-stones migration model. We also created minimum spanning trees (MST)(Naujokaitis-Lewis et al., 2013) as they reflect the "backbone" of the dispersal network (Bunn et al., 2000). Here again, we assumed that cost values of landscape

features are unknown and GAB and MST connections were computed based on geographical distances as in Arnaud (2003) and Keller et al. (2013).

Third, graph pruning was based on a statistical procedure selecting the minimal set of links explaining population genetic structure. Based upon the conditional independence principle (Magwene, 2001; Whittaker, 2009), it is supposed to select links corresponding with direct dispersal paths and discard links associated with genetic similarities due to stepping-stones dispersal. We used the original method of Dyer and Nason (2004) but we also modified some of the calculation steps implemented in the popgraph package (cf. section 3 of supporting information). This method involves the calculation of a genetic covariance matrix from a genetic distance matrix that must have Euclidean properties, following Gower (1966). Therefore, we used a PCA-derived Euclidean genetic distance as well as the Euclidean genetic distance computed by default when using the `popgraph` package. From a strict mathematical point of view, the formula used to calculate the covariance c_{ij} from the distance d_{ij} between populations i and j is the following: $c_{ij} = -\frac{1}{2} \times (d_{ij}^2 - d_{i\bullet}^2 - d_{\bullet j}^2 + d_{\bullet\bullet}^2)$ (Everitt and Hothorn, 2011; Smouse and Peakall, 1999), although the formula implemented in `popgraph` package is: $c_{ij} = -\frac{1}{2} \times (d_{ij} - d_{i\bullet} - d_{\bullet j} + d_{\bullet\bullet})$ (Dyer and Nason, 2004)($d_{i\bullet}$ and $d_{\bullet j}$ correspond respectively to the sum of distances over a column/row of the distance matrix). In our modified version (CI), we used the former formula while we also implemented the latter for comparative purposes (CI2). We also added a p-value adjustment, following sequential Bonferroni procedure (Holm, 1979), to limit type-I errors. In sum, we constructed genetic independence graphs relying upon the conditional independence principle using either our modified method (CI) or the original method of Dyer and Nason (2004) (CI2), with either PCA-derived Euclidean distance (PCA, cf section 3 of supporting information) or popgraph derived Euclidean genetic distance (PG), and either adjusting (ADJ) p-values or not (Table 1).

Finally, we constructed complete genetic graphs (COMP) because graph topologies sometimes include all the potential links between nodes (Naujokaitis-Lewis et al., 2013). Besides, these complete graphs constituted a baseline to assess the relevance of graph pruning.

2.3.2 Genetic distance

Four genetic distances were used to weight the graph links. First, we used the linearised F_{ST} (*i.e.* $F_{ST}/(1\text{-}F_{ST})$), hereafter noted F_{ST} (Rousset, 1997). Second, we also used the "inter-population version" of D_{PS} (DPS), a genetic distance based on the dissimilarities of population allele pools computed as 1 - the proportion of shared alleles (Bowcock et al. (1994), cf. section 4 of supporting information). This commonly used genetic distance is supposed to reflect recent gene flow changes (Murphy et al., 2010b, 2015; Naujokaitis-Lewis et al., 2013). Third, we computed a Euclidean genetic distance by first performing a PCA of the matrix of allelic frequencies and then computing the Euclidean distance between populations in the space defined by all independent principal components to derive a PCA-based Euclidean genetic distance (PCA), following Paschou et al. (2014) and Shirk et al. (2017a). Finally, we used the Euclidean genetic distance computed by default in `popgraph` package (PG).

Genetic independence graph links were weighted only with the two Euclidean genetic distances. The links of the other genetic graphs were weighted using the F_{ST}, the D_{PS} and the PCA-derived genetic distance. Every genetic distance, including that computed with the original `popgraph` method, was used to weight the links of the complete graphs in order to provide a baseline for the comparison of all pruning methods. In sum, 30 genetic graphs were constructed at generations 50 and 500 for every simulation (Table 1).

2.4 Graphs analyses

The dispersal pattern of the simulated species is reflected by the realised dispersal graph topology, and simulated gene flow was driven by the cost-distance values between populations. Hence, a genetic graph can be considered accurate if i) its topology reflects well the direct paths of the realised dispersal graph or if ii) the genetic distances derived from its links are highly correlated to the cost-distance values between populations.

2.4.1 Topology similarity analyses

We assessed the topological similarity between realised dispersal graphs and genetic graphs. To that purpose, we created contingency tables classifying the potential links of both types of graphs into two categories: absence or presence (see Fletcher et al. (2011)). Then, we calculated the Matthews correlation coefficient (Matthews, 1975), considered as a reliable index of binary classification quality because it takes into account all the elements of the contingency table and is calculated with respect to a random baseline (Baldi et al., 2000). A Matthews correlation coefficient of 1 is reached when both graphs are identical, whereas a 0 value means that they are no more similar than if they were built by selecting links randomly. In our case, a large value indicates that a genetic graph recovers well the realised dispersal graph topology.

2.4.2 Distance-based analyses

We calculated the Mantel correlation coefficients r (Mantel, 1967) between genetic distances and CD values. For each simulated genetic dataset, we considered three sets of genetic distances: i) the subset of "raw" genetic distances associated with population pairs directly connected in genetic graphs (see Van Strien et al. (2015)), ii) the graph-based genetic distances between every population pair, calculated as the sum of link weights along the shortest path between nodes (an extended use of the "cGD" introduced by (Dyer et al., 2010) to other types of genetic graphs) and iii) the full set of "raw" genetic distances between every population pair derived from complete graphs. Large r values indicate that the set of genetic distances derived from genetic graphs reflects well the simulated landscape effects on gene flow. This approach is commonly used in landscape genetics (Graves et al., 2013; Shirk et al., 2017b; Van Strien et al., 2015; Zeller et al., 2016) as the use of Mantel correlation coefficients is relevant when the hypothesis can only be formulated in terms of distances (Legendre and Fortin, 2010). We focused on the correlation coefficient values rather than on statistical significance because it has been shown to provide reliable results when few hypotheses are compared (Shirk et al., 2017b; Zeller et al., 2016). Besides, type-I error rate is high with Mantel tests (Balkenhol et al., 2009b), which limits their relevance.

2.5 Simulation results ordination

We performed a large number of simulations by varying landscapes and population locations. Given our objectives, we intended to reproduce in our simulations the cases I and IV from the hypothetical classification of the relationship between genetic and geographical distances proposed by Hutchison and Templeton (1999). Although case I corresponds to an equilibrium between gene flow and drift over the whole region, case IV corresponds to a transient situation where this equilibrium has been reached at a smaller spatial scale because of the time lag of the genetic response. The case I is characterised by a linear increase of genetic differentiation with increasing geographical distances over the whole region considered. In contrast, the case IV depicts this positive correlation up to a certain geographical distance threshold above which the relationship flattens out. This distance threshold was defined by Van Strien et al. (2015) as the distance of maximum correlation (DMC), i.e. the geographical distance threshold below which the subset of population pairs maximises the linear correlation between genetic and geo-

graphical distances. The highest DMC occurs for case-I patterns of IBD as it should be equal to the maximum inter-population distance whereas it decreases when case-IV patterns of IBD are observed. Thereby, we used the DMC as a proxy of the spatial scale above which equilibrium has not been reached, and below which genetic structure depends both on gene flow and drift, which does not necessarily mean that equilibrium has been reached. Considering the linear positive relationship between genetic differentiation and landscape distances expected under an IBLR model (McRae, 2006), we aimed at reproducing the cases I and IV defined by Hutchinson and Templeton but considering cost-distances instead of geographic distances.

We determined the DMC by iteratively computing the Mantel correlation coefficients between i) the CD values driving the simulation and ii) the F_{ST} and the D_{PS}, using increasing threshold values. We also visualised scatter plots of the relationship between genetic distances and CD to identify the type of IBLR pattern corresponding to each simulation and time step. We could thus check for potential biases in Mantel r values. Most graph analyses were performed using `graph4lg` package in R (Savary et al., 2020).

To extract the main trend among the results of 1000 simulations, we applied a Principal Component Analysis to eight variables describing the simulation parameters (proportion of migrants per population, CD threshold used to build the potential dispersal graph, number of links in the realised dispersal graph, mean CD covered by migrants) and their genetic output (DMC computed at generation 50 and 500 for F_{ST} and D_{PS}). These variables were averaged over the 10 runs for each configuration combining a landscape and a population spatial distribution. We carried out a hierarchical clustering from the PCA factors in order to distinguish the main trend in the PCA results.

Graph name	Pruning method	Genetic distance
COMP-FST	No pruning	F_{ST}
COMP-DPS	No pruning	D_{PS}
COMP-PCA	No pruning	PCA-derived Eucl. dist.
COMP-PG	No pruning	Eucl. gen. dist. (from `popgraph`)
GEO-10-FST	Geo. dist. threshold (10-km)	F_{ST}
GEO-10-DPS	Geo. dist. threshold (10-km)	D_{PS}
GEO-10-PCA	Geo. dist. threshold (10-km)	PCA-derived Eucl. dist.
GEO-15-FST	Geo. dist. threshold (15-km)	F_{ST}
GEO-15-DPS	Geo. dist. threshold (15-km)	D_{PS}
GEO-15-PCA	Geo. dist. threshold (15-km)	PCA-derived Eucl. dist.
GEO-20-FST	Geo. dist. threshold (20-km)	F_{ST}
GEO-20-DPS	Geo. dist. threshold (20-km)	D_{PS}
GEO-20-PCA	Geo. dist. threshold (20-km)	PCA-derived Eucl. dist.
GEO-30-FST	Geo. dist. threshold (30-km)	F_{ST}
GEO-30-DPS	Geo. dist. threshold (30-km)	D_{PS}
GEO-30-PCA	Geo. dist. threshold (30-km)	PCA-derived Eucl. dist.
GAB-FST	Topological (Gabriel graph, geo. dist.)	F_{ST}
GAB-DPS	Topological (Gabriel graph, geo. dist.)	D_{PS}
GAB-PCA	Topological (Gabriel graph, geo. dist.)	PCA-derived Eucl. dist.
MST-FST	Topological (MST, geo. dist.)	F_{ST}
MST-DPS	Topological (MST, geo. dist.)	D_{PS}
MST-PCA	Topological (MST, geo. dist.)	PCA-derived dist.
CI-PCA	Condit. indep.	PCA-derived dist. (covar. from squared dist.)
CI-ADJ-PCA	Condit. indep.	PCA-derived dist. (covar. from squared dist.) with Holm-Bonferroni adjustment
CI-PG	Condit. indep.	Eucl. gen. dist. (from `popgraph`, covar. from squared dist.)
CI-ADJ-PG	Condit. indep.	Eucl. gen. dist. (from `popgraph`, covar. from squared dist.) with Holm-Bonferroni adjustment
CI2-PCA	Condit. indep.	PCA-derived Eucl. dist. (covar. from dist.)
CI2-ADJ-PCA	Condit. indep.	PCA-derived Eucl. dist. (covar. from dist.) with Holm-Bonferroni adjustment
CI2-PG	Condit. indep.	Original `popgraph` method
CI2-ADJ-PG	Condit. indep.	Original `popgraph` method with Holm-Bonferroni adjustment

Table 1: Genetic graph construction parameters. Cf. section 2 of supporting information for a glossary of the acronyms

3 Results

3.1 Simulation results

For each simulation, the realised dispersal graph was connected meaning that each population exchanged migrants with at least another population during the first 50 generations. The overall proportion of dispersing individuals over 500 generations ranged from 13.3 % to 24.1 %. Although all the landscapes were simulated with the same parameters and populations were located randomly in habitat patches, values of the maximum dispersal distance exhibited substantial variations (from 1321 to 3564 CD units). Consequently, the number of links in dispersal graphs ranged from 155 to 858 links (Figure 1), depicting a wide range of gene flow patterns.

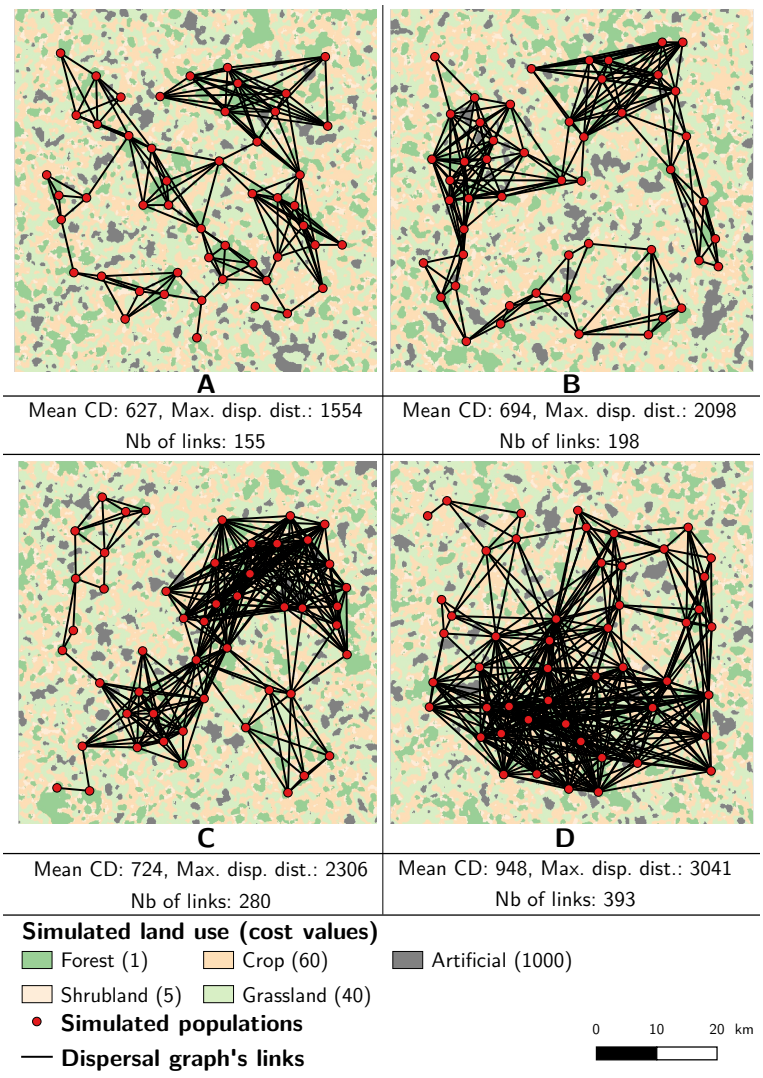

Figure 1: Four contrasted landscape/distribution of populations configurations exhibiting large differences in the number of links in the dispersal graph. Mean CD between populations, maximum dispersal distance in CD units and number of links followed by individuals are indicated in each landscape.

Although a case-IV pattern of IBLR was often observed at generation 50 (Figure 3), DMC values increased from generation 50 to 500 suggesting that genetic structure reached its stationary state at increasing spatial scale over time. Note that DMC values were always larger than

the maximum CD over which dispersal was possible. PCA results evidenced these variations (Figure 2). The first principal component (56.6 % of the variance) was positively correlated with the DMC (based on F_{ST} and D_{PS} values at generation 50 in particular), the maximum CD threshold, the number of links in the realised dispersal graph and to a lesser extent with the mean CD covered by individuals and the proportion of migrants. The second principal component explained a lower proportion of variance (24.9 %) and mainly reflected differences between simulations due to the interplay between the number of links, the proportion of dispersing individuals (negatively correlated) and the mean CD covered by dispersing individuals (positively correlated).

Three main clusters of landscape/distribution of populations configurations were identified through the hierarchical clustering of the PCA results (Figure 2). The first cluster is characterised by low numbers of links in dispersal graphs because of low maximum dispersal distances and by low DMC at generation 50 while the third cluster is characterised by high DMC, high numbers of links and high maximum dispersal distance. In the second cluster, dispersal graphs counted many links, the proportions of migrants were high and the DMC took intermediate values.

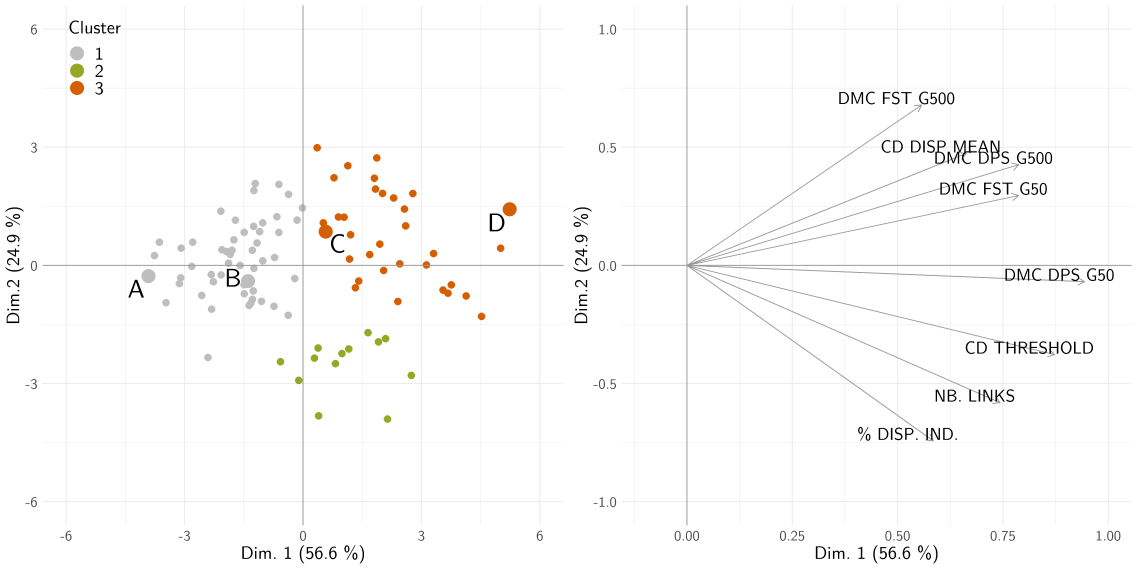

Figure 2: Principal Components Analysis of eight variables (100 observations) describing the simulation results. Configurations A to D are also displayed in figure 1.

The first and third clusters included configurations in which case-IV and case-I patterns of IBLR take form at generation 50, respectively. One objective of this study was to compare the usefulness of genetic graphs when gene flow influences genetic structure at the complete landscape scale (case I) or at smaller scale (case IV). In addition, the relative performance of graph construction and analysis methods exhibited marginal variation along the second principal component. Thus, for the sake of brevity, we chose to describe the results of the subsequent analyses based on four configurations (A, B, C, D; displayed on the figures 1, 3 and 2) along the first principal component which defines a gradient between these two opposite patterns. Configuration A was typical of a case-IV pattern of IBLR (at generation 50 in particular) and D of a case-I pattern (Figure 3). Configurations B and C corresponded to intermediate situations.

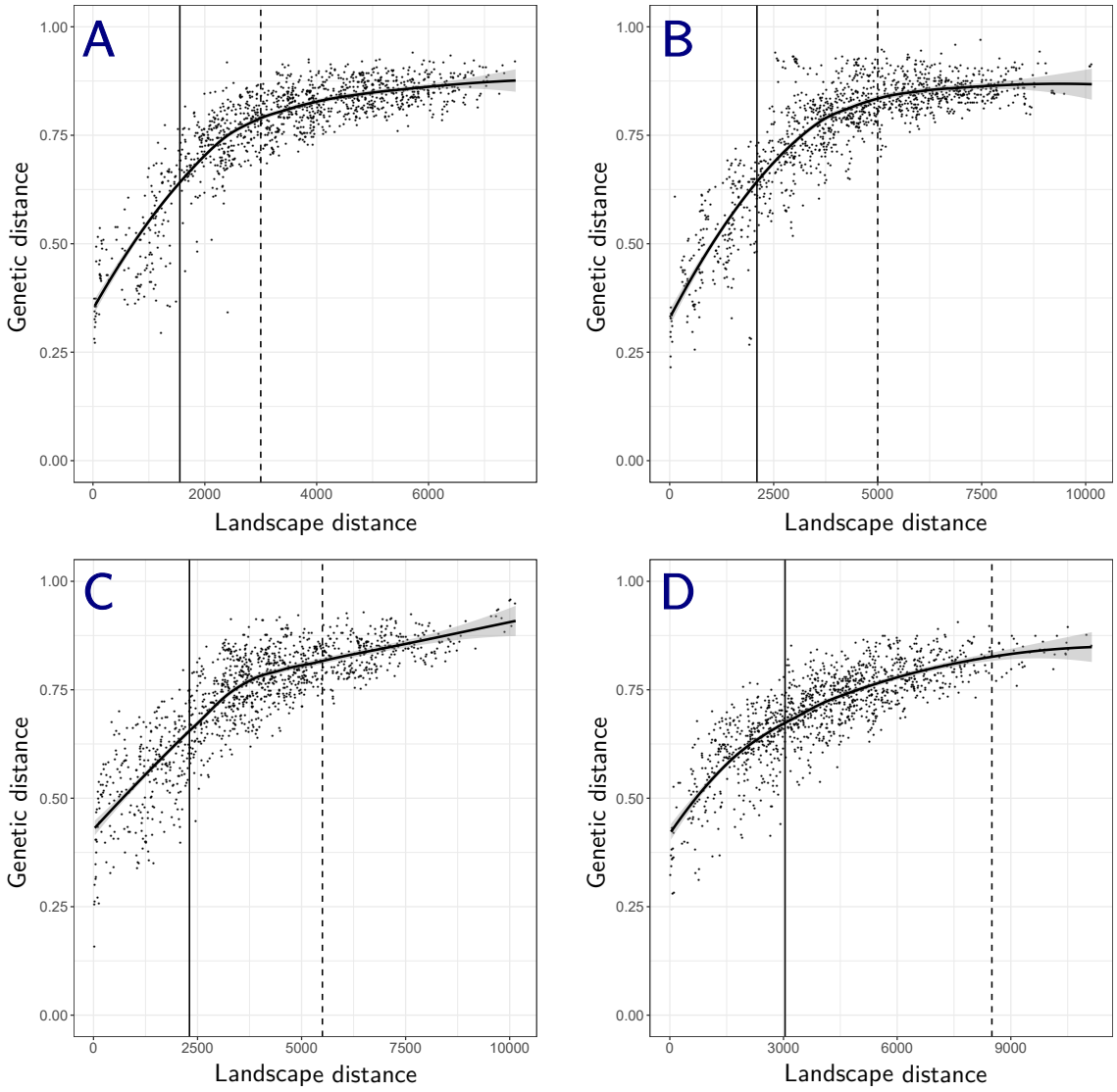

Figure 3: Scatter plots of the genetic distance (D_{PS}) plotted against cost-distance at generation 50. Cases A to D illustrate the gradient of IBD patterns (from type-IV to type-I). Solid vertical lines indicate the maximum dispersal distance, dashed lines indicate the DMC. See figure S1 for the same figure with the F_{ST}.

3.2 Genetic graphs

3.2.1 Topology similarity analyses

Depending on the pruning method used, the mean number of links in the genetic graphs was highly variable as it ranged from 49 (MST) to 802 (GEO-30)(Table 2 and figure S3). In contrast, the number of links in the realised dispersal graphs, which genetic graph topologies were supposed to reproduce, were 155, 198, 280 and 393 in the configurations A to D, respectively.

When the graphs were pruned with methods based on geographical distance thresholds or on topological constraints, their number of links was stable over generations given these methods do not rely on genetic data. The number of links of MST and Gabriel graphs (49 and around 90 links, respectively) was also highly stable among configurations and much lower than the number of links in realised dispersal graphs (Table 2). As a consequence, these topological pruning methods never performed well in reflecting realised dispersal graph topology (correlation values

from 0.29 to 0.54; Table 2).

On the contrary, as pruning based on conditional independence takes into account genetic data, the number of links in the genetic independence graphs varied strongly, from 58.5 to 519.7 in average. The number of links in these graphs tended to increase from generation 50 to generation 500 even if the number of realised direct dispersal paths was stable, but this trend was much lower when using genetic distances (CI2), as in the original `popgraph` method, than squared genetic distances (CI), as in our modified version. As a consequence, the ability of a genetic graph topology to reflect the topology of the realised dispersal graph was fairly stable between generations G50 and G500 when using genetic distances (CI2), whereas it decreases between G50 and G500 when using squared genetic distances (CI; Table 2). Adjusting the p-values to assess the significance of the partial correlations almost reduced the number of links by a factor of 2. When the covariance between allelic frequencies was calculated using squared genetic distances (CI), the number of links was consistently larger than when using genetic distances (CI2). In some cases, graphs obtained using the latter formula were not connected, especially when p-values were adjusted to assess the significance of partial correlation values.

Genetic graphs pruned with geographical distance thresholds presented the topology closest to that of realised dispersal graphs in all configurations (correlation values above 0.6) except the least connected one (*i.e.* configuration A)(Table 2). The closest the geographical distance threshold (GEO) from the maximum dispersal distance (CD threshold converted into Euclidean distance) used in the simulations, the better the genetic graph reflects the topology of the realised dispersal graph. In contrast, for the dispersal graphs created in configuration A, which counted fewer links (Figure S3), the highest correlation values were reached with pruning methods based on conditional independence. Correlation values above 0.6 were reached every time covariance was computed from genetic distances (CI2), and only at generation 50 with p-value adjustment when covariance was computed from squared genetic distances (CI) with our modified method. For the configuration B, correlation values above 0.6 were also reached when independence genetic graphs were pruned by computing the covariance from genetic distances (CI2) whatever the type of genetic distance used (PCA or PG). For the configuration C, the original `popgraph` method (CI2-PG) enabled to reach a correlation value of 0.6. Note that when computing the covariance from squared genetic distances (CI), the genetic graphs included links between population pairs not connected in the dispersal graph (Figure S3). p-value adjustment reduced the number of these false long-distance links.

Overall, genetic graphs which succeeded in accurately reproducing the topology of the realised dispersal graphs counted much the same number of links as the dispersal graph (Table 2). However, this condition is not sufficient to explain the correlation values given that in some cases, relatively low correlation values were obtained with a similar number of links to the realised dispersal graph (e.g. CI-ADJ-PG, configuration A at G500: Matthews correlation coefficient = 0.52, with a difference in the number of links between graphs equal to 2.4; Table 2).

3.2.2 Distance-based analyses

The correlation coefficients between genetic distances and CD separating population pairs which were directly connected in the genetic graphs were highly variable as they ranged from 0.47 to 0.86 in average at generation 50 (Figure 4, see figure S4 for generation 500). In all cases, Mantel correlation coefficients between genetic distances and geographical distances were lower than those between genetic distances and CD, showing that the isolation by landscape resistance model better explained genetic structure than the isolation by distance model did, as expected

from our simulations.

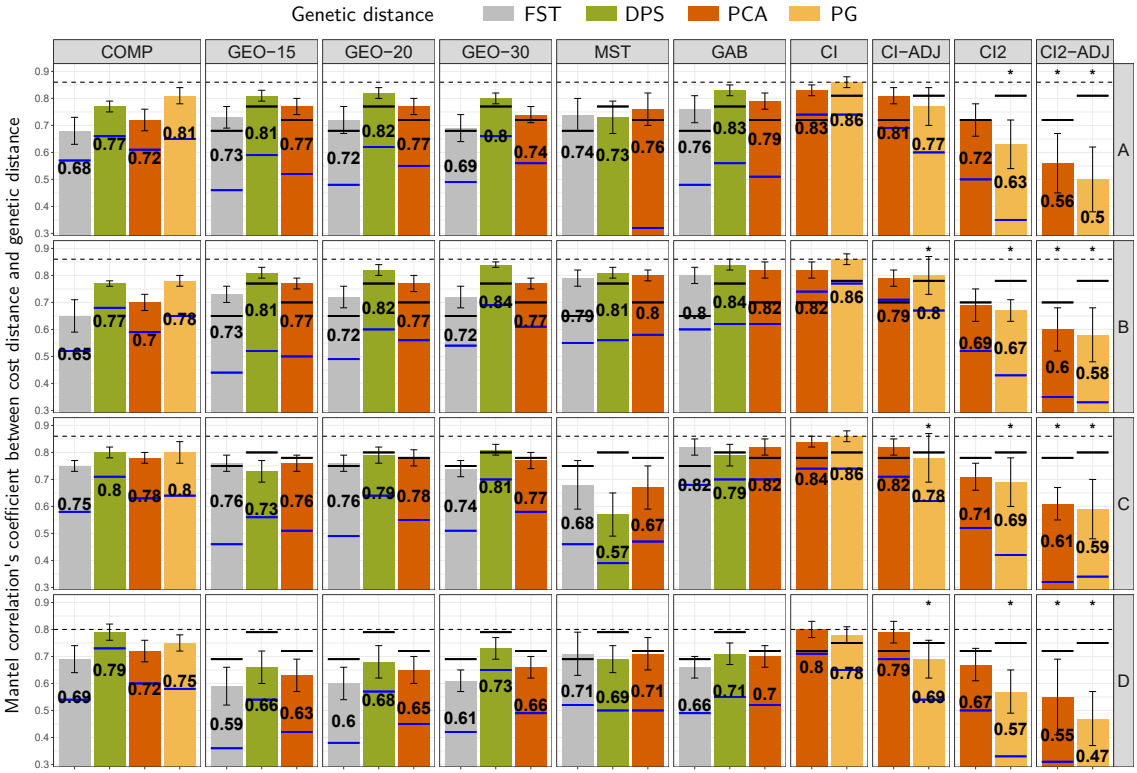

Figure 4: Mantel correlation between genetic distances and cost-distances separating nodes directly connected on the genetic graphs, according to the type of genetic distance and the pruning method at generation 50 (see table 1 for the graph names). Mean ± SD values were computed for the 10 runs simulated in each scenario. Blue bars refer to the correlation coefficient between genetic distance and geographical distance, when it is above 0.3. Black bars refer to the correlation coefficient obtained using every population pair to compute the correlation. When black and blue bars overlap, the bar is black. Stars indicate graphs counting several components. The dashed line indicates the maximum r value obtained for each configuration.

When a case-I pattern of IBLR was observed at generation 50 (configurations C and D), larger correlation coefficients were always those obtained when using genetic distance matrices derived from complete graphs instead of pruned graphs, except for genetic independence graphs built with our modified method (CI). Conversely, when a case-IV pattern of IBLR was observed at generation 50 (configurations A and B), correlation coefficients were almost always larger when genetic distance values were those associated with the links of a pruned graph, whatever the pruning method, than when they were associated with all population pairs (Figure 4). At generation 500, there were few differences between configurations (A to D) given that the relationship between genetic differentiation and cost-distance almost linearised over time in all cases, and higher correlation between genetic distances and CD were observed with a complete graph, compared with a pruned graph, except for genetic independence graphs based on squared genetic distance (CI) which still performed better (Figure S4).

The largest correlation coefficients were always reached when selecting genetic distance values from genetic independence graphs built without p-value adjustment and based on the computation of the covariance from squared genetic distances (CI). When computing covariance from genetic distances (CI2), as in the original `popgraph` method, correlation coefficients were much lower. This method never strengthened the correlation obtained with the corresponding complete

genetic distance matrices (COMP-PG or COMP-PCA), and it provided the lowest correlation values (Figure 4). For configurations A and B (case IV), correlation coefficients obtained when selecting population pairs from a Gabriel graph were slightly larger than correlation obtained selecting population pairs from an MST or by using geographical distance thresholds (Figure 4). In most cases, correlation coefficients between genetic distances and CD were lower when the genetic distance was the F_{ST} rather than the D_{PS} or Euclidean genetic distances.

When we computed the Mantel correlations between CD and graph-based genetic distances, correlation coefficients values ranged from 0.57 to 0.93 at generation 50 (Figure 5, see figure S5 for a similar variation at G500). However, differences between configurations were less pronounced when analysing the correlation this way, even at generation 50.

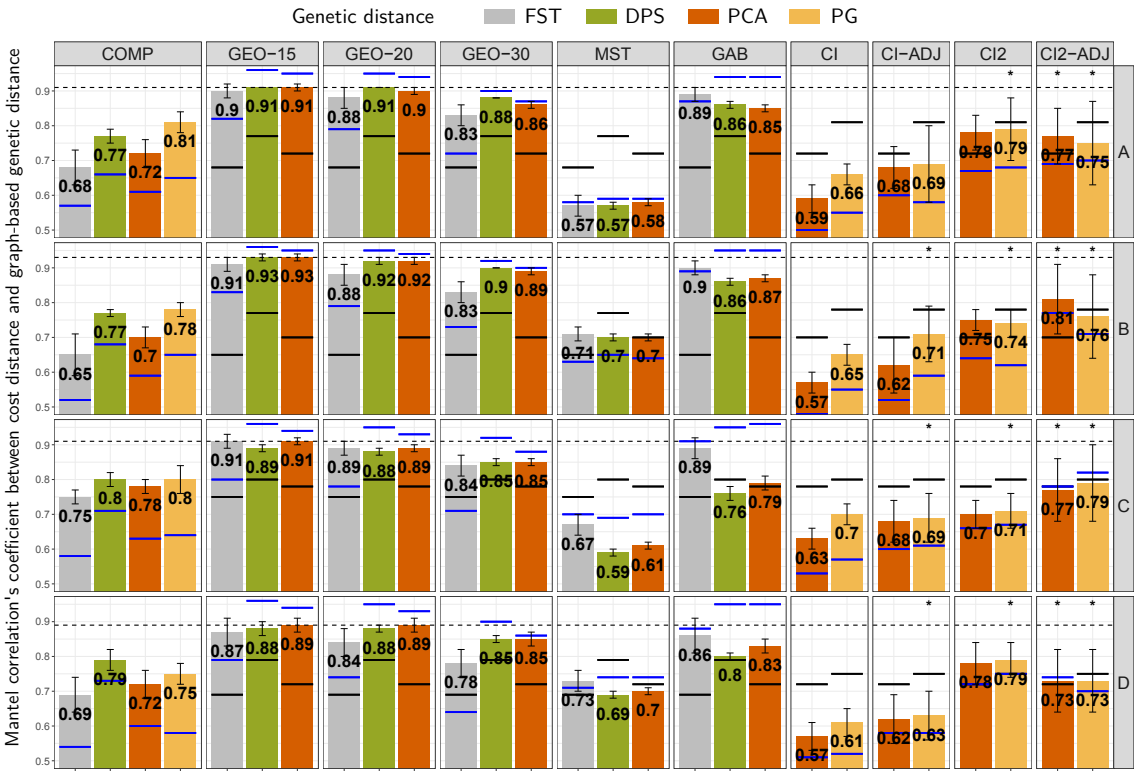

Figure 5: Mantel correlation between conditional genetic distances and cost-distances separating nodes on the genetic graphs, according to the type of genetic distance and the pruning method at generation 50 (see table 1 for the graph names). Mean $\pm SD$ values were computed for the 10 runs simulated in each scenario. Blue bars refer to the correlation coefficient between genetic distance and geographical distance, when it is above 0.5. Black bars refer to the correlation coefficient obtained using every population pair to compute the correlation. When black and blue bars overlap, the bar is black. Stars indicate graphs counting regularly several components. The dashed line indicates the maximum r value obtained for each configuration.

The correlation coefficient took its largest values when the graphs were pruned using a geographical distance threshold or a topological constraint. However, when computing graph-based genetic distances from these graphs, the correlation coefficients between these genetic distances and geographical distances were higher than those computed between the same genetic distances and CD values, supporting an IBD model over an IBLR model despite our simulation settings. The only exception was when this distance was computed from F_{ST} values (Figure 5).

Conversely, when computing graph-based genetic distances from independence graphs, these

distances were more correlated to CD than to geographical distances (Figure 5). The correlation coefficients were higher in that case when the pruning relied on the calculation of the covariance from genetic distances (CI2) rather than from squared genetic distances (CI). However, the correlation between Euclidean genetic distances and CD was higher when considering the complete graph instead of graph-based genetic distances, except when computing covariance from PCA-based genetic distances (CI2-PCA). Besides, we reproduced the result described by Dyer et al. (2010) who showed that landscape influence on gene flow was frequently better recovered when using these graph-based genetic distances derived from independence graphs (CI2) than when using the complete matrix of F_{ST}. Moreover, scatter plots created using graph-based genetic distance values revealed that summing genetic distances in case-IV pattern of IBLR tends to mask the fact that the relationship between genetic differentiation and CD flattens out beyond a CD threshold (Figure S2).

Combination	**A**		**B**		**C**		**D**	
a) Dispersal graphs								
Max. disp. dist.	1554		2098		2306		3041	
Max.disp.dist. (km)	13.1		15.1		18.2		20.8	
Nb. disp. paths	155		198		280		393	
b) Number of links								
Generation	G50	G500	G50	G500	G50	G500	G50	G500
GEO-10	**127.0**	***127.0***	**120.0**	***120.0***	120.0*	*120.0**	112.0*	*112.0**
GEO-15	274.0	*274.0*	**237.0**	***237.0***	**253.0**	***253.0***	**246.0**	***246.0***
GEO-20	455.0	*455.0*	**384.0**	***384.0***	**408.0**	***408.0***	**413.0**	***413.0***
GEO-30	802.0	*802.0*	693.0	*693.0*	746.0	*746.0*	757.0	*757.0*
GAB	92.0	*92.0*	89.0	*89.0*	96.0	*96.0*	92.0	*92.0*
MST	49.0	*49.0*	49.0	*49.0*	49.0	*49.0*	49.0	*49.0*
CI-PCA	274.4	*427.7*	269.7	*412.1*	283.8	*435.0*	305.7	*458.8*
CI-ADJ-PCA	**108.2**	*246.5*	107.7	*234.6*	109.0	*251.5*	114.8	*275.8*
CI-PG	325.5	*441.3*	330.9	*397.9*	342.8	*447.2*	375.3	*519.7*
CI-ADJ-PG	**96.8**	*152.6*	101.2*	*152.5*	95.7*	*149.6**	102.2*	*165.3*
CI2-PCA	**132.6**	***158.9***	**134.6**	***161.2***	141.8	*176.3**	146.9	*204.9*
CI2-ADJ-PCA	**76.7***	***85.8****	71.3*	*81.4**	65.0*	*81.5**	60.9*	*70.0**
CI2-PG	**135.4***	***148.1***	**134.3***	***149.2****	137.5*	***155.4***	140.5*	*153.9*
CI2-ADJ-PG	**77.8***	***80.5****	72.5*	*75.6**	63.7*	*72.9**	59.4*	*58.5**
c) MCC								
GEO-10	**0.64**	***0.64***	**0.68**	***0.68***	0.56*	*0.56**	0.44*	*0.44**
GEO-15	0.59	*0.59*	**0.73**	***0.73***	**0.68**	***0.68***	**0.62**	***0.62***
GEO-20	0.47	*0.47*	**0.60**	***0.60***	**0.68**	***0.68***	**0.69**	***0.69***
GEO-30	0.28	*0.28*	0.38	*0.38*	0.44	*0.44*	0.51	*0.51*
GAB	0.54	*0.54*	0.54	*0.54*	0.41	*0.41*	0.39	*0.39*
MST	0.49	*0.49*	0.43	*0.43*	0.36	*0.36*	0.29	*0.29*
CI-PCA	0.42	*0.25*	0.39	*0.24*	0.33	*0.19*	0.26	*0.14*
CI-ADJ-PCA	**0.60**	*0.34*	0.53	*0.33*	0.43	*0.25*	0.34	*0.18*
CI-PG	0.41	*0.34*	0.39	*0.37*	0.35	*0.32*	0.27	*0.21*
CI-ADJ-PG	**0.64**	*0.52*	0.55*	*0.48*	0.45*	*0.41**	0.35*	*0.29*
CI2-PCA	**0.77**	***0.69***	**0.69**	***0.65***	0.58	*0.55**	0.46	*0.40*
CI2-ADJ-PCA	**0.65***	***0.65****	0.55*	*0.57**	0.42*	*0.46**	0.33*	*0.33**
CI2-PG	**0.80***	***0.75***	**0.72***	***0.69****	0.59*	***0.60***	0.48*	*0.45*
CI2-ADJ-PG	**0.66***	***0.65****	0.56*	*0.56**	0.42*	*0.44**	0.32*	*0.31**

Table 2: Topologies of the dispersal graphs and genetic graphs. a) Topologies of the dispersal graphs. Maximum dispersal distances are given in CD units and in kilometres (conversion obtained after performing the linear regression of CD values against geographical distances values). b) Number of links in the genetic graphs. c) Matthews correlation coefficients assessing the topology similarity of both types of graphs (genetic and dispersal), according to the type of genetic distance and the pruning method in the four landscape/distribution of populations configurations and at two generations (see table 1 for the graph names). Mean values and standard deviations were computed for the 10 runs simulated in each scenario but standard deviations are not displayed because they were negligible. Matthews correlation coefficients above 0.6 and corresponding numbers of links in the genetic graphs are displayed in bold. Values referring to generation 500 are displayed in italics. Stars indicate that some of the ten graphs created for each combination were not connected.

4 Discussion

In this study, we demonstrated that the ability of different pruning methods to identify the precise topology of dispersal networks is highly variable, especially between methods based either on geographical distance or genetic independence criteria. In addition, we highlighted the importance of graph pruning for assessing landscape effects on gene flow in non-equilibrium situations. We provide users with rough guidelines that are schematically illustrated in Figure 6.

4.1 When and why to prune a genetic graph?

On the one hand, graph pruning is hardly avoidable when the objective is to identify the topology of the direct dispersal network followed by individuals (Figure 6). Indeed, except in very rare panmictic configurations or when the study area is very small, dispersal events are not expected between all population pairs (Kimura and Weiss, 1964). On the other hand, our results show that the relevance of graph pruning for inferring landscape resistance to gene flow depends on the scale at which gene flow effect on differentiation is detectable. When an IBLR pattern is observed at the scale of the entire landscape, graph pruning hardly ever improved the inference made from a complete graph, except when pruning relied on the conditional independence principle and squared genetic distances (CI). In contrast, when this pattern is observed up to a limited scale, graph pruning strengthened the linear correlation between genetic distances and cost-distance values driving the simulation, suggesting that graph pruning is useful to infer landscape resistance to gene flow in this situation.

Migration-drift equilibrium is less likely to be reached for the complete set of sampled population pairs when dispersal distances are short regarding the study area and/or landscapes have undergone recent modifications. Such non-equilibrium situations correspond to the case-IV pattern of IBD proposed by Hutchison and Templeton (1999). It has been observed in several theoretical (Slatkin, 1993) and empirical studies (Ciofi et al., 1999; Clegg and Phillimore, 2010; Hänfling and Weetman, 2006; Hutchison and Templeton, 1999; Kuehn et al., 2003; Méndez et al., 2011) and is expected to be frequent in landscape genetic studies dealing with dynamic human-shaped landscapes (Manel and Holderegger, 2013; Storfer et al., 2010). In such situations, not pruning a genetic graph might be problematic if the objective is to infer landscape resistance to gene flow. Indeed, such inferences may involve genetic distances that do not reflect the long-term effect of landscape on genetic structure. Wagner and Fortin (2013) suggested that considering a subset of population pairs could increase the power of distance-based analyses in landscape genetics. Indeed, a few studies reported stronger relationships between landscape structure and population genetic structure using this approach (Angelone et al., 2011; Coster et al., 2015; Jaquiéry et al., 2011; Keller et al., 2013; Van Strien et al., 2015). Most of them used geographical thresholds somehow linked to maximum dispersal abilities and they considered between populations distances while ignoring their spatial arrangement (but see Keller et al. (2013) for an explicit graph-based approach). However, Van Strien (2017) argued that population topology (*i.e.*, the arrangement of populations throughout a landscape) should be better incorporated in link-based landscape genetic studies. In this context, graph-theoretic methods offer great opportunities in link selection (Dyer, 2015b), but to date, the relative performance of the wide range of graph pruning methods had not been assessed.

4.2 How to prune a genetic graph to identify the precise topology of the dispersal network?

Our results show that pruning methods based on topological constraints are rarely suitable for recovering the topology of the dispersal network. Indeed, minimum spanning trees do not include any cycles and Gabriel graphs cannot take into account the presence of some long-distance dispersal events between population pairs. Because topological constraints generally impose a constant number of links given the number of populations, they lack ecological significance (Serrano et al., 2009).

When dispersal capacities of the study species are precisely known, pruning based on geographical distance thresholds always makes it possible to recover well the topology of the realised dispersal graphs, and this method is often the best one. Of course, estimating dispersal capacities is a difficult task (Schneider, 2003; Van Dyck and Baguette, 2005), and even if several thresholds can be tested in empirical studies, it is impossible to determine which genetic graph reflects best the true dispersal pattern. As expected, our results showed that a serious underestimation of maximum dispersal distance led to a disconnected graph, thus wrongly suggesting the existence of landscape barriers to dispersal. The similarity between genetic graphs pruned using distance thresholds and realised dispersal graphs may also depends on the correlation strength between cost distances (CD) and geographical distances, which is sometimes high (Marrotte and Bowman, 2017). However, geographical distance is not always a good proxy of CD (Balkenhol et al., 2009b), for instance when a barrier prevents dispersal between close populations or less frequently when large geographical distances are covered by dispersing individuals because they correspond to low CD values. Ignoring these rare long distance dispersal event may be problematic given their ecological and evolutionary consequences (Clobert et al., 2012; Greenbaum and Fefferman, 2017; Nathan et al., 2003).

Our results also suggest that building genetic independence graphs is a suitable option to recover dispersal network topology when dispersal distances are unknown, especially in less connected configurations (A and B, Table 2). In the latter case, these results can be as satisfactory as when dispersal distance is known. The topology of the dispersal graphs is better recovered when the covariance is computed with the original `popgraph` method from genetic distances (CI2) instead of squared genetic distances (CI). In the latter case, the presence of links in genetic graphs that were never followed by dispersing individuals during the simulations indicates that it does not identify direct dispersal paths reliably. The variability in the number of links among independence graphs was mainly due to the covariance formula, the genetic distance, the p-value adjustment, and to a lesser extent the generation. In contrast, the expected large difference in the number of links between very different connectivity configurations (A and D) was not observed.

It may appear puzzling to infer single-generation dispersal events from genetic structure shaped by multi-generational dispersal. However, it seems to be the promise behind the genetic independence graphs as the conditional independence is supposed to recover the actual route of propagules (Dyer, 2015b). Though our results seem to support this idea in some conditions, further research is needed on this pruning method. For instance, we expect this method to perform poorly when sampling is incomplete, which is often the rule in empirical studies, but the potential bias this introduces in the inferences remains to be estimated.

4.3 How to prune a genetic graph to infer landscape resistance to dispersal?

Genetic graphs reflecting precisely the dispersal network topology are not necessarily those that enable to quantify well landscape effects on dispersal. Indeed, the distance of maximum correlation (DMC) was always larger than the maximum dispersal distance in the simulation (Figure 3), suggesting that the set of genetic distance values to include in link-based analyses, should not be restricted to direct dispersal paths. Migration-drift equilibrium may become established between populations separated by distances beyond dispersal capacities, as expected under the stepping-stones model of Slatkin (1993), because they may exchange genes over several generations even if not connected by direct dispersal paths. We suggest including such population pairs in link-based inferences because their genetic divergence should reflect landscape influence on gene flow. Our view contrasts with the exclusive use of population pairs that are within migration range of each other recommended by others when assessing the effect of landscape on gene flow (Keller et al., 2013; Van Strien et al., 2015; Van Strien, 2017). Therefore, a reliable pruning method to estimate landscape resistance to gene flow should identify population pairs whose genetic differentiation reflects the long term gene flow between them.

In this context, we do not advise using pruning methods based on fixed criteria (*i.e.* geographical distances or topological constraints), even if they provided correlations between genetic distances and CD that were slightly lower than the maximum correlation obtained for a given configuration at generation 50, especially when using D_{PS} (Figure 4). Indeed, these methods seem inappropriate because the spatial scale of IBLR changes over time (McRae, 2006). Pruning methods relying on genetic data and statistical inference seem to provide the best inference of landscape resistance as they can account for the dynamic nature of IBLR. Indeed, in case-I and case-IV patterns of IBLR, the correlation between genetic distances associated with the genetic graph links and CD values was maximised when using pruning methods based upon the conditional independence principle. However, this result only holds when computing the covariance from squared genetic distances to stick with mathematical requirements (Everitt and Hothorn, 2011; Magwene, 2001; Smouse and Peakall, 1999). Although the original `popgraph` method reproduced the dispersal pattern quite well, it often produced the lowest correlation between genetic distances and CD. Nevertheless, these methods deserve further investigation because some connected population pairs in our independence graphs (using our modified method) were separated by CD values larger than the DMC. Even if the use of the DMC to determine the spatial scale at which genetic structure depends on both gene flow and drift needs stronger theoretical support, this suggests that genetic differentiation between these populations may still need time before stabilising.

We believe that an essential but tricky issue remains the identification of population pairs matching migration-drift equilibrium. Ciofi et al. (1999) developed a likelihood-based approach to assess whether population structure is best explained by a model of migration-drift equilibrium or by a model of pure drift. However, it seems that this approach fails to detect case-IV patterns of IBD (Hänfling and Weetman, 2006). Assuming that the DMC may be used as a proxy of the spatial scale of migration-drift equilibrium, a promising approach would consist in pruning the genetic graphs with a CD threshold equal to the DMC. This approach requires knowledge of the cost values associated with landscape features to estimate the CD between population pairs. However, assessing cost scenarios is often the aim of empirical link-based analyses (Balkenhol et al., 2016). Recent methods for optimisation of landscape resistance surfaces have been developed (Peterman, 2018), and a potential improvement would consist in using genetic graphs pruned with different CD thresholds in such optimisation procedures.

4.4 Which genetic distance to use and how?

Weighting graph links using the D_{PS} or Euclidean genetic distances always produced a better inference of landscape influence on gene flow than using F_{ST}, even if the difference in performance of these genetic distances decreased as pairwise genetic differentiation tends to reach its equilibrium level (*i.e.* from G50 to G500). Though F_{ST} is an excellent measure of genetic differentiation, using it for particular demographic inferences (e.g. the number of migrants entering a population every generation) requires assumptions of migration-drift equilibrium to be met (Neigel, 2002; Whitlock and Mccauley, 1999). Further theoretical work is required to assess the sensitivity of F_{ST}-based inferences of landscape resistance to equilibrium conditions. Besides, D_{PS} had already been shown to better reflect recent landscape changes than other genetic distances, including F_{ST} (Landguth et al., 2010; Robin et al., 2015). We would expect other genetic distances such as the Chord distance (Cavalli-Sforza and Edwards, 1967) to provide similar results as those obtained with D_{PS}.

Once the genetic graph has been pruned, we discourage summing genetic distances along shortest paths to create a complete matrix of graph-based genetic distances. This approach led to spurious conclusions by detecting an isolation by distance pattern instead of the true isolation by landscape resistance pattern when graph pruning was based on geographical thresholds or topological constraints (Figure 5). Interestingly, landscape influence on dispersal was frequently better recovered when using these graph-based genetic distances derived from independence graphs (CI2) than when using the complete matrix of F_{ST} (Figure 5). This result has been previously used to evidence the value of this graph-based genetic distance (Dyer et al., 2010). However, the correlation between genetic distances and the driver of dispersal (*i.e.* CD) was lower when considering these graph-based genetic distances than when using the complete matrix of corresponding raw genetic distances.

4.5 Limits and perspectives

Our simulations produced contrasted patterns of connectivity, but we acknowledge that our results are limited to cases where a single functional unit of populations that can somehow exchange migrants (i.e. a dispersal network made of a single component) is considered. We still need to investigate relative performances of graph-theoretic methods in landscapes with complete barriers isolating population clusters. The differences we detected between the compared methods were informative and promising, yet sometimes subtle. Although the migration rates we obtained were similar to those reported by Bowne and Bowers (2004), they were larger than those reported from other empirical data (Meirmans, 2014) or from simulated data reproducing case-IV patterns (Van Strien et al., 2015). Given that case-IV patterns of IBLR were observed in situations where dispersal was most constrained, repeating our simulations with more limited dispersal would have produced stronger contrasts between complete and pruned graphs in their ability to infer landscape resistance to gene flow and might have made the discrimination between pruning methods even more straightforward. Considering that low migration rates are probably the norm, this reinforces the relevance of genetic graph pruning in empirical studies. However, it remains to be determined whether there is a threshold below which dispersal has only a marginal effect on genetic differentiation as compared with genetic drift. If this case could reveal a complete barrier to dispersal, it would prevent from inferring the relative resistance of the different landscape features surrounding populations.

Our results challenge a common practice in landscape genetics consisting in using the complete matrix of genetic distance to infer the resistance of landscape features. Consequently, it must be further examined whether and how graph-theoretic methods may improve calibration

of resistance surfaces. Here, we did not compare different cost scenarios as we knew the "true" cost values driving the simulation. We therefore assumed that maximising the linear correlation (Mantel r) between genetic distances and landscape distances measured for a subset of population pairs allows for reliably identifying the best graph construction method, *i.e.* the one that selects the best subset of population pairs for the analysis (but see Graves et al. (2013)). As we did not aim at performing a fine-tuned calibration of cost values, we consider our approach as suitable (Shirk et al., 2017b; Zeller et al., 2016).

In our simulations, we assumed all populations of the study area were sampled. Such a sampling intensity is rarely achieved in practice although it is often recommended (Keller et al., 2013; Van Strien, 2017). Assessing how partial sampling of populations affects our conclusions needs further investigation (as in Koen et al. (2013) and Naujokaitis-Lewis et al. (2013)). In these situations, the complementarity between genetic graphs and landscape graphs needs to be explored, because the nodes of the latter are the exhaustive set of potential habitat patches in the study area (Foltête and Vuidel, 2017). In addition, a growing set of studies in landscape genetics now use individual-based sampling schemes. Though the conditional independence principle evaluated in our study is not applicable when nodes are individuals, a few studies have applied genetic graphs to individuals (Draheim et al., 2016; Greenbaum et al., 2016). This possibility offers a great potential and deserves further investigation.

Lastly, gene flow and drift were the main processes driving genetic differentiation in our simulations. Drift strength depends on population sizes, which were maintained equal and constant over generations. Although this choice allowed us to keep drift constant among our simulations in order to focus only on the effect of landscape on dispersal and to substantially reduce computation times, we acknowledge that it strongly simplifies the reality. Indeed, landscape changes also create spatial heterogeneity in effective population sizes, which can be a strong driver of genetic differentiation (Prunier et al., 2017). Besides, local features such as patch size or habitat quality can affect gene flow between populations (Pflüger and Balkenhol, 2014; Robertson et al., 2019; Weckworth et al., 2013), though we did not make it possible in our simulations. In this context, gravity models seem particularly relevant as they can be based on genetic graphs and additionally include local variables (Murphy et al., 2010a; Watts et al., 2015; Zero et al., 2017).

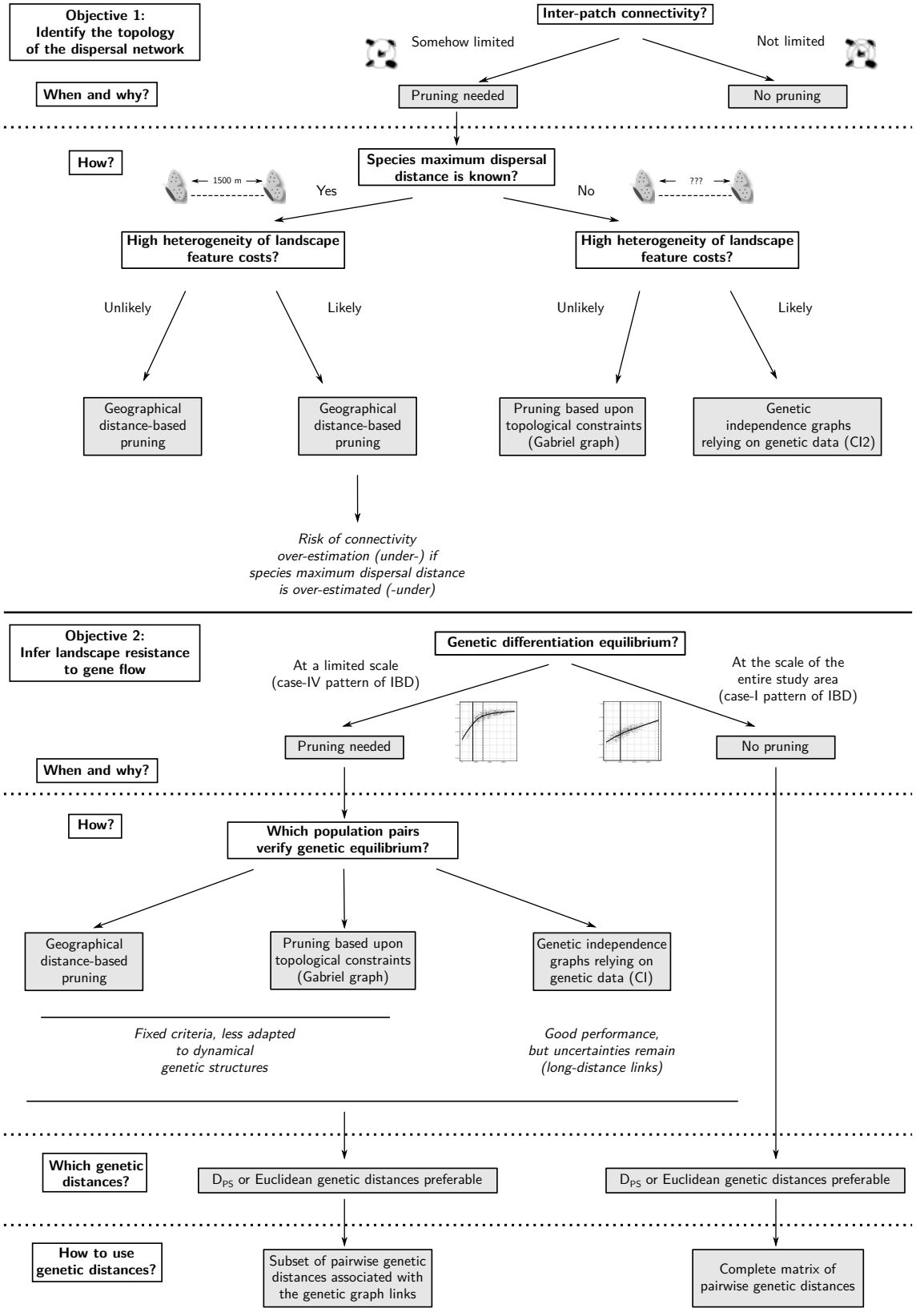

Figure 6: Guidelines based on our results to build and analyse genetic graphs for i) identifying the topology of a dispersal network and ii) inferring landscape resistance to gene flow. CI2 (CI): Conditional independence assessed with (squared) genetic distances.

5 Acknowledgements

We thank the editor and referees, as well as Aurélie Khimoun and Maarten van Strien, for their very constructive comments that improved our manuscript. This study is part of a PhD project supported by the ARP-Astrance company under a CIFRE contract supervised and partly funded by the ANRT (Association Nationale de la Recherche et de la Technologie). This work is also part of the project CANON that was supported by the French "Investissements d'Avenir" program, project ISITE-BFC (contract ANR-15-IDEX-0003). We are particularly grateful to ARP-Astrance team for its constant support along the project. We thank Ahmed Jebrane, Catherine Labruère and Catherine Larédo for their help on mathematical aspects. We thank Christopher Sutcliffe for revising the English manuscript. Simulations and analyses were carried out on the calculation "Mésocentre" facilities of the University of Bourgogne-Franche-Comté.

References

Adriaensen, F., Chardon, J., De Blust, G., Swinnen, E., Villalba, S., Gulinck, H., and Matthysen, E. (2003). The application of least-cost modelling as a functional landscape model. *Landscape and urban planning*, 64(4):233–247.

Allendorf, F. W., Luikart, G., and Aitken, S. N. (2007). Conservation and the genetics of populations. *mammalia*, 2007(2007):189–197.

Angelone, S., Kienast, F., and Holderegger, R. (2011). Where movement happens - scale-dependent landscape effects on genetic differentiation in the european tree frog. *Ecography*, 34(5):714–722.

Arnaud, J.-F. (2003). Metapopulation genetic structure and migration pathways in the land snail helix aspersa: influence of landscape heterogeneity. *Landscape Ecology*, 18(3):333–346.

Baldi, P., Brunak, S., Chauvin, Y., Andersen, C. A., and Nielsen, H. (2000). Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics*, 16(5):412–424.

Balkenhol, N., Cushman, S., Storfer, A., and Waits, L. (2016). *Landscape genetics: concepts, methods, applications*. John Wiley & Sons.

Balkenhol, N., Gugerli, F., Cushman, S. A., Waits, L. P., Coulon, A., Arntzen, J., Holderegger, R., Wagner, H. H., et al. (2009a). Identifying future research needs in landscape genetics: where to from here? *Landscape Ecology*, 24(4):455.

Balkenhol, N., Waits, L. P., and Dezzani, R. J. (2009b). Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. *Ecography*, 32(5):818–830.

Bowcock, A. M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J. R., and Cavalli-Sforza, L. L. (1994). High resolution of human evolutionary trees with polymorphic microsatellites. *nature*, 368(6470):455–457.

Bowne, D. R. and Bowers, M. A. (2004). Interpatch movements in spatially structured populations: a literature review. *Landscape ecology*, 19(1):1–20.

Bunn, A., Urban, D., and Keitt, T. (2000). Landscape connectivity: a conservation application of graph theory. *Journal of environmental management*, 59(4):265–278.

Cavalli-Sforza, L. L. and Edwards, A. W. (1967). Phylogenetic analysis: models and estimation procedures. *Evolution*, 21(3):550–570.

Ciofi, C., Beaumont, M. A., Swingland, I. R., and Bruford, M. W. (1999). Genetic divergence and units for conservation in the komodo dragon varanus komodoensis. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 266(1435):2269–2274.

Clegg, S. M. and Phillimore, A. B. (2010). The influence of gene flow and drift on genetic and phenotypic divergence in two species of zosterops in vanuatu. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1543):1077–1092.

Clobert, J., Baguette, M., Benton, T. G., and Bullock, J. M. (2012). *Dispersal ecology and evolution*. Oxford University Press.

Coster, S. S., Babbitt, K. J., Cooper, A., and Kovach, A. I. (2015). Limited influence of local and landscape factors on finescale gene flow in two pond-breeding amphibians. *Molecular Ecology*, 24(4):742–758.

Coulon, A., Cosson, J., Angibault, J., Cargnelutti, B., Galan, M., Morellet, N., Petit, E., Aulagnier, S., and Hewison, A. (2004). Landscape connectivity influences gene flow in a roe deer population inhabiting a fragmented landscape: an individual–based approach. *Molecular ecology*, 13(9):2841–2850.

Crispo, E., Moore, J.-S., Lee-Yaw, J. A., Gray, S. M., and Haller, B. C. (2011). Broken barriers: human-induced changes to gene flow and introgression in animals: an examination of the ways in which humans increase genetic exchange among populations and species and the consequences for biodiversity. *BioEssays*, 33(7):508–518.

Cushman, S. A., McKelvey, K. S., Hayden, J., and Schwartz, M. K. (2006). Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. *The American Naturalist*, 168(4):486–499.

Den Boer, P. J. (1968). Spreading of risk and stabilization of animal numbers. *Acta biotheoretica*, 18(1-4):165–194.

Draheim, H. M., Moore, J. A., Etter, D., Winterstein, S. R., and Scribner, K. T. (2016). Detecting black bear source–sink dynamics using individual-based genetic graphs. *Proc. R. Soc. B*, 283(1835):20161002.

Dyer, R. J. (2015a). Is there such a thing as landscape genetics? *Molecular ecology*, 24(14):3518–3528.

Dyer, R. J. (2015b). Population graphs and landscape genetics. *Annual Review of Ecology, Evolution, and Systematics*, 46:327–342.

Dyer, R. J. and Nason, J. D. (2004). Population graphs: the graph theoretic shape of genetic structure. *Molecular ecology*, 13(7):1713–1727.

Dyer, R. J., Nason, J. D., and Garrick, R. C. (2010). Landscape modelling of gene flow: improved power using conditional genetic distance derived from the topology of population networks. *Molecular Ecology*, 19(17):3746–3759.

Everitt, B. and Hothorn, T. (2011). *An introduction to applied multivariate analysis with R*. Springer Science & Business Media.

Excoffier, L., Smouse, P. E., and Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among dna haplotypes: application to human mitochondrial dna restriction data. *Genetics*, 131(2):479–491.

Fletcher, R. J., Acevedo, M. A., Reichert, B. E., Pias, K. E., and Kitchens, W. M. (2011). Social network models predict movement and connectivity in ecological landscapes. *Proceedings of the National Academy of Sciences*, 108(48):19282–19287.

Foltête, J.-C., Clauzel, C., and Vuidel, G. (2012). A software tool dedicated to the modelling of landscape networks. *Environmental Modelling & Software*, 38:316–327.

Foltête, J.-C. and Vuidel, G. (2017). Using landscape graphs to delineate ecologically functional areas. *Landscape Ecology*, 32(2):249–263.

Frankham, R., Ballou, J. D., and Briscoe, D. A. (2004). *A primer of conservation genetics*. Cambridge University Press.

Garroway, C. J., Bowman, J., and Wilson, P. J. (2011). Using a genetic network to parameterize a landscape resistance surface for fishers, martes pennanti. *Molecular ecology*, 20(19):3978–3988.

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53(3-4):325–338.

Graves, T. A., Beier, P., and Royle, J. A. (2013). Current approaches using genetic distances produce poor estimates of landscape resistance to interindividual dispersal. *Molecular Ecology*, 22(15):3888–3903.

Greenbaum, G. and Fefferman, N. H. (2017). Application of network methods for understanding evolutionary dynamics in discrete habitats. *Molecular Ecology*, 26(11):2850–2863.

Greenbaum, G., Templeton, A. R., and Bar-David, S. (2016). Inference and analysis of population structure using genetic data and network theory. *Genetics*, 202(4):1299–1312.

Guillot, G., Leblois, R., Coulon, A., and Frantz, A. C. (2009). Statistical methods in spatial genetics. *Molecular Ecology*, 18(23):4734–4756.

Gurrutxaga, M., Lozano, P. J., and del Barrio, G. (2010). Gis-based approach for incorporating the connectivity of ecological networks into regional planning. *Journal for Nature Conservation*, 18(4):318–326.

Hänfling, B. and Weetman, D. (2006). Concordant genetic estimators of migration reveal anthropogenically enhanced source-sink population structure in the river sculpin cottus gobio. *Genetics*, 173(3):1487–1501.

Hanski, I. (1998). Metapopulation dynamics. *Nature*, 396(6706):41.

Hanski, I., Alho, J., and Moilanen, A. (2000). Estimating the parameters of survival and migration of individuals in metapopulations. *Ecology*, 81(1):239–251.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70.

Hutchison, D. W. and Templeton, A. R. (1999). Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution*, 53(6):1898–1914.

Jaquiéry, J., Broquet, T., Hirzel, A. H., Yearsley, J., and Perrin, N. (2011). Inferring landscape effects on dispersal from genetic distances: how far can we go? *Molecular ecology*, 20(4):692–705.

Keller, D., Holderegger, R., and Strien, M. J. (2013). Spatial scale affects landscape genetic analysis of a wetland grasshopper. *Molecular Ecology*, 22(9):2467–2482.

Khimoun, A., Peterman, W., Eraud, C., Faivre, B., Navarro, N., and Garnier, S. (2017). Landscape genetic analyses reveal fine-scale effects of forest fragmentation in an insular tropical bird. *Molecular Ecology*.

Kimura, M. and Weiss, G. H. (1964). The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, 49(4):561.

Koen, E. L., Bowman, J., Garroway, C. J., and Wilson, P. J. (2013). The sensitivity of genetic connectivity measures to unsampled and under-sampled sites. *PloS one*, 8(2):e56204.

Kuehn, R., Schroeder, W., Pirchner, F., and Rottmann, O. (2003). Genetic diversity - gene flow and drift in bavarian red deer populations (cervus elaphus). *Conservation Genetics*, 4(2):157–166.

Landguth, E., Cushman, S., Schwartz, M., McKelvey, K., Murphy, M., and Luikart, G. (2010). Quantifying the lag time to detect barriers in landscape genetics. *Molecular ecology*, 19(19):4179–4191.

Landguth, E. L. and Cushman, S. (2010). Cdpop: a spatially explicit cost distance population genetics program. *Molecular Ecology Resources*, 10(1):156–161.

Legendre, P. and Fortin, M.-J. (2010). Comparison of the mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular ecology resources*, 10(5):831–844.

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, 17(4):183–189.

Magwene, P. M. (2001). New tools for studying integration and modularity. *Evolution*, 55(9):1734–1745.

Manel, S. and Holderegger, R. (2013). Ten years of landscape genetics. *Trends in ecology & evolution*, 28(10):614–621.

Manel, S., Schwartz, M. K., Luikart, G., and Taberlet, P. (2003). Landscape genetics: combining landscape ecology and population genetics. *Trends in ecology & evolution*, 18(4):189–197.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer research*, 27(2 Part 1):209–220.

Marrotte, R. R. and Bowman, J. (2017). The relationship between least-cost and resistance distance. *PloS one*, 12(3):e0174212.

Mateo-Sánchez, M. C., Balkenhol, N., Cushman, S., Pérez, T., Domínguez, A., and Saura, S. (2015). Estimating effective landscape distances and movement corridors: comparison of habitat and genetic data. *Ecosphere*, 6(4):1–16.

Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, 405(2):442–451.

McRae, B. H. (2006). Isolation by resistance. *Evolution*, 60(8):1551–1561.

Meirmans, P. G. (2014). Nonconvergence in bayesian estimation of migration rates. *Molecular ecology resources*, 14(4):726–733.

Méndez, M., Tella, J. L., and Godoy, J. A. (2011). Restricted gene flow and genetic drift in recently fragmented populations of an endangered steppe bird. *Biological Conservation*, 144(11):2615–2622.

Munwes, I., Geffen, E., Roll, U., Friedmann, A., Daya, A., Tikochinski, Y., and Gafny, S. (2010). The change in genetic diversity down the core-edge gradient in the eastern spadefoot toad (pelobates syriacus). *Molecular Ecology*, 19(13):2675–2689.

Murphy, M., Dyer, R., and Cushman, S. A. (2015). Graph theory and network models in landscape genetics. In Balkenhol, N., Cushman, S., Storfer, A., and Waits, L., editors, *Landscape genetics: Concepts, methods, applications*, pages 165–180. John Wiley & Sons, 1 edition.

Murphy, M. A., Dezzani, R., Pilliod, D. S., and Storfer, A. (2010a). Landscape genetics of high mountain frog metapopulations. *Molecular ecology*, 19(17):3634–3649.

Murphy, M. A., Evans, J. S., and Storfer, A. (2010b). Quantifying bufo boreas connectivity in yellowstone national park with landscape genetics. *Ecology*, 91(1):252–261.

Nathan, R., Perry, G., Cronin, J. T., Strand, A. E., and Cain, M. L. (2003). Methods for estimating long-distance dispersal. *Oikos*, 103(2):261–273.

Naujokaitis-Lewis, I. R., Rico, Y., Lovell, J., Fortin, M.-J., and Murphy, M. A. (2013). Implications of incomplete networks on estimation of landscape genetic connectivity. *Conservation genetics*, 14(2):287–298.

Neigel, J. E. (2002). Is fst obsolete? *Conservation Genetics*, 3(2):167–173.

Paschou, P., Drineas, P., Yannaki, E., Razou, A., Kanaki, K., Tsetsos, F., Padmanabhuni, S. S., Michalodimitrakis, M., Renda, M. C., Pavlovic, S., et al. (2014). Maritime route of colonization of europe. *Proceedings of the National Academy of Sciences*.

Pérez-Espona, S., Pérez-Barbería, F., McLeod, J., Jiggins, C., Gordon, I., and Pemberton, J. (2008). Landscape features affect gene flow of scottish highland red deer (cervus elaphus). *Molecular Ecology*, 17(4):981–996.

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

Pflüger, F. J. and Balkenhol, N. (2014). A plea for simultaneously considering matrix quality and local environmental conditions when analysing landscape impacts on effective dispersal. *Molecular Ecology*, 23(9):2146–2156.

Prunier, J. G., Dubut, V., Chikhi, L., and Blanchet, S. (2017). Contribution of spatial heterogeneity in effective population sizes to the variance in pairwise measures of genetic differentiation. *Methods in Ecology and Evolution*, 8(12):1866–1877.

Richardson, J. L., Brady, S. P., Wang, I. J., and Spear, S. F. (2016). Navigating the pitfalls and promise of landscape genetics. *Molecular ecology*, 25(4):849–863.

Robertson, E., Fletcher Jr, R., and Austin, J. (2019). The number of breeders explains genetic connectivity in an endangered bird. *Molecular ecology*.

Robin, V., Gupta, P., Thatte, P., and Ramakrishnan, U. (2015). Islands within islands: two montane palaeoendemic birds impacted by recent anthropogenic fragmentation. *Molecular ecology*, 24(14):3572–3584.

Rousset, F. (1997). Genetic differentiation and estimation of gene flow from f-statistics under isolation by distance. *Genetics*, 145(4):1219–1228.

Rozenfeld, A. F., Arnaud-Haond, S., Hernández-García, E., Eguíluz, V. M., Serrão, E. A., and Duarte, C. M. (2008). Network analysis identifies weak and strong links in a metapopulation system. *Proceedings of the National Academy of Sciences*, 105(48):18824–18829.

Ruiz-Gonzalez, A., Cushman, S. A., Madeira, M. J., Randi, E., and Gómez-Moliner, B. J. (2015). Isolation by distance, resistance and/or clusters? lessons learned from a forest-dwelling carnivore inhabiting a heterogeneous landscape. *Molecular ecology*, 24(20):5110–5129.

Ruiz-González, A., Gurrutxaga, M., Cushman, S. A., Madeira, M. J., Randi, E., and Gómez-Moliner, B. J. (2014). Landscape genetics for the empirical assessment of resistance surfaces: the european pine marten (martes martes) as a target-species of a regional ecological network. *PLoS One*, 9(10):e110552.

Savary, P., Foltête, J.-C., Moal, H., Vuidel, G., and Garnier, S. (2020). graph4lg: a package for constructing and analysing graphs for landscape genetics in r. *Methods in Ecology and Evolution*, pages 1–9.

Schadt, S., Knauer, F., Kaczensky, P., Revilla, E., Wiegand, T., and Trepl, L. (2002). Rule-based assessment of suitable habitat and patch connectivity for the eurasian lynx. *Ecological Applications*, 12(5):1469–1483.

Schlather, M., Malinowski, A., Menck, P. J., Oesting, M., Strokorb, K., et al. (2015). Analysis, simulation and prediction of multivariate random fields with package randomfields. *Journal of Statistical Software*, 63(8):1–25.

Schneider, C. (2003). The influence of spatial scale on quantifying insect dispersal: an analysis of butterfly data. *Ecological Entomology*, 28(2):252–256.

Sciaini, M., Fritsch, M., Scherer, C., and Simpkins, C. E. (2018). Nlmr and landscapetools: An integrated environment for simulating and modifying neutral landscape models in r. *Methods in Ecology and Evolution*, 9(1):2240–2248.

Serrano, M. Á., Boguná, M., and Vespignani, A. (2009). Extracting the multiscale backbone of complex weighted networks. *Proceedings of the national academy of sciences*, 106(16):6483–6488.

Shirk, A., Landguth, E., and Cushman, S. (2017a). A comparison of indiviudal-based genetic distance metrics for landscape genetics. *Molecular Ecology*, 17(6).

Shirk, A. J., Landguth, E. L., and Cushman, S. A. (2017b). A comparison of regression methods for model selection in individual-based landscape genetic analysis. *Molecular ecology resources*, 18(1):55–67.

Slatkin, M. (1993). Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, 47(1):264–279.

Smouse, P. E. and Peakall, R. (1999). Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity*, 82(5):561–573.

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

Storfer, A., Murphy, M. A., Spear, S. F., Holderegger, R., and Waits, L. P. (2010). Landscape genetics: where are we now? *Molecular ecology*, 19(17):3496–3514.

Taylor, P. D., Fahrig, L., Henein, K., and Merriam, G. (1993). Connectivity is a vital element of landscape structure. *Oikos*, pages 571–573.

Urban, D. and Keitt, T. (2001). Landscape connectivity: a graph-theoretic perspective. *Ecology*, 82(5):1205–1218.

Van Dyck, H. and Baguette, M. (2005). Dispersal behaviour in fragmented landscapes: routine or special movements? *Basic and Applied Ecology*, 6(6):535–545.

Van Strien, M. J. (2017). Consequences of population topology for studying gene flow using link-based landscape genetic methods. *Ecology and evolution*, 7(14):5070–5081.

Van Strien, M. J., Holderegger, R., and Van Heck, H. J. (2015). Isolation-by-distance in landscapes: considerations for landscape genetics. *Heredity*, 114(1):27.

Wagner, H. H. and Fortin, M.-J. (2013). A conceptual framework for the spatial analysis of landscape genetic data. *Conservation Genetics*, 14(2):253–261.

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

Watts, A. G., Schlichting, P. E., Billerman, S. M., Jesmer, B. R., Micheletti, S., Fortin, M.-J., Funk, W. C., Hapeman, P., Muths, E., and Murphy, M. A. (2015). How spatio-temporal habitat connectivity affects amphibian genetic structure. *Frontiers in genetics*, 6.

Weckworth, B. V., Musiani, M., DeCesare, N. J., McDevitt, A. D., Hebblewhite, M., and Mariani, S. (2013). Preferred habitat and effective population size drive landscape genetic patterns in an endangered species. *Proceedings of the Royal Society B: Biological Sciences*, 280(1769):20131756.

Whitlock, M. C. and Mccauley, D. E. (1999). Indirect measures of gene flow and migration: Fst≠ 1/(4nm+ 1). *Heredity*, 82(2):117–125.

Whittaker, J. (2009). *Graphical models in applied multivariate statistics*. Wiley Publishing.

Wright, S. (1943). Isolation by distance. *Genetics*, 28(2):114.

Zeller, K. A., Creech, T. G., Millette, K. L., Crowhurst, R. S., Long, R. A., Wagner, H. H., Balkenhol, N., and Landguth, E. L. (2016). Using simulations to evaluate mantel-based methods for assessing landscape resistance to gene flow. *Ecology and evolution*, 6(12):4115–4128.

Zeller, K. A., Jennings, M. K., Vickers, T. W., Ernest, H. B., Cushman, S. A., and Boyce, W. M. (2018). Are all data types and connectivity models created equal? validating common connectivity approaches with dispersal data. *Diversity and Distributions*, 0:1–12.

Zero, V. H., Barocas, A., Jochimsen, D. M., Pelletier, A., Giroux-Bougard, X., Trumbo, D. R., Castillo, J. A., Evans Mack, D., Linnell, M. A., Pigg, R. M., et al. (2017). Complementary network-based approaches for exploring genetic structure and functional connectivity in two vulnerable, endemic ground squirrels. *Frontiers in genetics*, 8:81.

6 Data accessibility

Simulated genotypes and R codes are available online at: Dryad DOI:10.5061/dryad.6q573n5xr

7 Author contributions

J.C.F., S.G. and H.M. designed the project and obtained the funding. P.S. and G.V. performed the simulations. P.S. designed the simulation study and analysed the data. P.S and S.G. wrote the manuscript with significant contributions and remarks from all co-authors.