the **genetics**society

ARTICLE

Check for updates

Assessing the influence of the amount of reachable habitat on genetic structure using landscape and genetic graphs

Paul Savary (ORCID)[1,2,3 ✉], Jean-Christophe Foltête[2], Maarten J. van Strien[4], Hervé Moal[1], Gilles Vuidel[2] and Stéphane Garnier[3]

Genetic structure, i.e. intra-population genetic diversity and inter-population genetic differentiation, is influenced by the amount and spatial configuration of habitat. Measuring the amount of reachable habitat (ARH) makes it possible to describe habitat patterns by considering intra-patch and inter-patch connectivity, dispersal capacities and matrix resistance. Complementary ARH metrics computed under various resistance scenarios are expected to reflect both drift and gene flow influence on genetic structure. Using an empirical genetic dataset concerning the large marsh grasshopper (*Stethophyma grossum*), we tested whether ARH metrics are good predictors of genetic structure. We further investigated (i) how the components of the ARH influence genetic structure and (ii) which resistance scenario best explains these relationships. We computed local genetic diversity and genetic differentiation indices in genetic graphs, and ARH metrics in the unified and flexible framework offered by landscape graphs, and we tested the relationships between these variables. ARH metrics were relevant predictors of the two components of genetic structure, providing an advantage over commonly used habitat metrics. Although allelic richness was significantly explained by three complementary ARH metrics in the best PLS regression model, private allelic richness and MIW indices were essentially related with the ARH measured outside the focal patch. Considering several matrix resistance scenarios was also key for explaining the different genetic responses. We thus call for further use of ARH metrics in landscape genetics to explain the influence of habitat patterns on the different components of genetic structure.

INTRODUCTION

The genetic structure of populations of the same species occupying subdivided habitat patches is characterised by two components: (i) the local genetic diversity within each population and (ii) the genetic differentiation between populations. Genetic drift and gene flow are the main processes influencing these two components when assessed from neutral genetic markers (Hedrick, 2011). Their combined effects depend on the habitat spatial pattern, i.e. the area and the configuration of habitat patches (DiLeo and Wagner, 2016, Keyghobadi, 2007). Indeed, on the one hand, when the effective size of a population is limited by the small area or poor quality of a habitat patch, genetic drift tends to erode its local genetic diversity (Frankham et al., 2004), thereby increasing the risk of inbreeding depression and local extinction (Frankham, 2005, Spielman et al., 2004). It also increases its genetic differentiation from other populations. On the other hand, if there are other habitat patches within dispersal distance, gene flow events due to dispersal from neighbouring populations can counterbalance this loss of local genetic diversity while limiting genetic differentiation between populations (Frankham, 2015, Ingvarsson, 2001, Lehnen et al., 2021). Understanding precisely how these two components of the genetic structure are influenced by the habitat spatial pattern is crucial in an era when habitat destruction is globally threatening all biodiversity levels (Díaz et al., 2019).

Describing the spatial pattern of habitat implies taking into account both habitat amount and configuration (Villard and Metzger, 2014), which are largely interdependent (Didham et al., 2012, Saura, 2021). For a given amount of habitat in the landscape, the configuration of habitat patches determines how much habitat is reachable from every patch (Saura, 2021, Villard and Metzger, 2014). The concept of habitat reachability integrates both habitat amount and configuration and extends that of habitat connectivity by considering both intra-patch and inter-patch connectivity (Pascual-Hortal and Saura, 2006, Saura and Rubio, 2010). The Amount of Reachable Habitat (ARH) computed for a patch is made of the area of the patch itself, and of the areas of its neighbouring patches according to species dispersal capacities. In addition, a patch may contribute to the ARH at a large scale by allowing "stepping-stone" dispersal over several generations between patches that are very distant from each other (Saura et al., 2014). To account for the latter situation when computing the ARH for a patch, one must consider the topology of the whole habitat network because it determines the role of that patch for indirect connections between distant patches (Saura and Rubio, 2010). Besides, as soon as the ARH includes habitat areas outside the focal patch, it should best include the resistance exerted by the landscape matrix on individual movements between patches (Andersson and Bodin, 2009, Joly et al.,

[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. [4]Planning of Landscape and Urban Systems, Institute for Spatial and Landscape Planning, ETH Zurich, Zurich, Switzerland. Associate editor: Sam Banks. ✉email: paul.savary@univ-fcomte.fr

2014). In sum, computing a set of complementary metrics makes it possible to measure the ARH from the species point of view and according to its dispersal capacities through the landscape matrix over large spatial scales and multiple generations (Saura and de la Fuente, 2017).

ARH metrics have been developed from landscape graphs, which represent habitat networks as sets of habitat patches (nodes) connected by potential dispersal paths (links) (Galpern et al., 2011, Saura and de la Fuente, 2017, Urban and Keitt, 2001). These graphs offer a unified framework for the computation of complementary habitat metrics in a more flexible way than commonly used metrics such as the distance to the nearest patch or the amount of habitat in a circular buffer area (see Fig. 1 for background information on habitat metrics). Accordingly, ARH metrics have proven helpful for explaining biological responses such as the composition of species communities (Awade et al., 2012, Mony et al., 2018) and are commonly used for conservation purposes (Bergès et al., 2020, Saura and de la Fuente, 2017). They have more rarely been used to explain the genetic structure of populations despite their potential relationships with both genetic drift and gene flow processes (but see Bertin et al. (2017), Flavenot et al. (2015) and Schoville et al. (2018)). Three metrics can be sufficient for describing the habitat pattern properties determining the ARH (Baranyi et al., 2011, Rayfield et al., 2011). These metrics should reflect the potential size of the population

occupying a patch and the contribution of a patch to dispersal fluxes and to long-distance dispersal events occurring through multiple generations over the whole habitat network. These properties have been named recruitment, flux and traversability by Urban and Keitt (2001), respectively.

In population genetics, the potential advantage of ARH metrics over other habitat metrics lies on the following rationale. Genetic drift depends on population size, which can be approximated by the capacity of a patch (i.e. recruitment component of the ARH, Fig. 1C). Besides, even if every suitable habitat patch in the landscape may not be systemically occupied by a population (Pasinelli et al., 2013), we can expect gene flow intensity between a given population and the others to increase with the flux component of the ARH. For a given patch, this component is measured by considering the potential connections to other habitat patches (e.g. with the F metric, Fig. 1D). Finally, the relative location of a patch in the topology of the whole network, taken into account in the traversability component of the ARH, is known to be a good predictor of multi-generational gene flow (Boulanger et al., 2020, van Strien, 2017, van Strien et al., 2014) (as reflected for example by the Betweenness Centrality (BC) metric, Fig. 1E). In contrast, while the distance to the nearest patch may only partially reflect the contribution of a patch to gene flow events (Fig. 1A), the amount of habitat in a buffer area (Fig. 1B) may not allow for distinguishing the effect of the habitat pattern on drift *versus* gene flow processes.
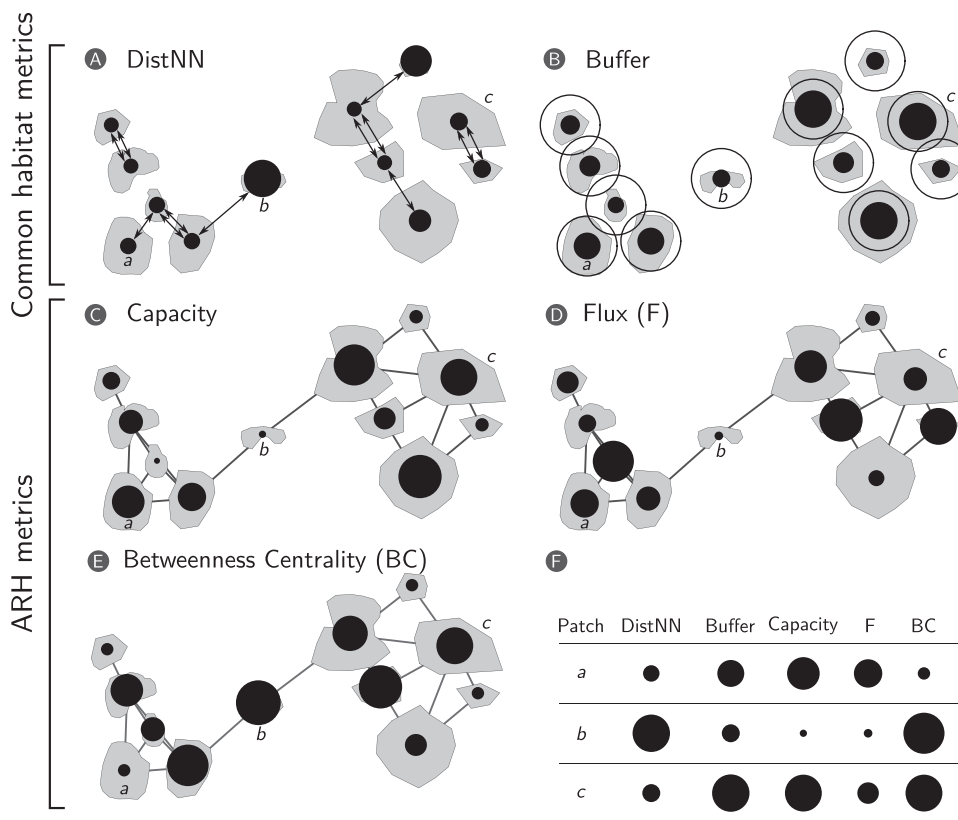

Fig. 1 Habitat metrics. Differences between common habitat metrics computed from a land cover map (**A**, **B**) and ARH metrics computed from a landscape graph (**C**, **D**, **E**). Grey areas correspond to habitat. The table (**F**) illustrates these differences by considering the metrics computed for three habitat patches (*a*, *b*, *c*). **A** The distance to the nearest habitat patch (DistNN) is computed for each habitat patch. **B** The amount of habitat in a circular buffer (so-called 'buffer' metric) is computed as the area of the pixels located within the black circle centred on each patch centroid. **C** The capacity is the area of each patch (node) of the landscape graph. **D** The Flux metric for patch *i* is the sum of the area of all the habitat patches *j* of the graph weighted by the dispersal probability between *i* and every patch *j*. **E** The Betweenness Centrality metric corresponds to the number of times every patch is located on a least-cost path between two other patches of the graph, weighted by the product of the connected patch areas and the dispersal probability between them. Brown lines on **B**, **D** and **E** correspond to landscape graph links.

The most frequent landscape genetic analyses focus on the relationship between genetic and landscape distances between patches (link-level, *sensu* Wagner and Fortin (2013)) to test for the effect of landscape structure on genetic differentiation (DiLeo and Wagner, 2016). In contrast, landscape influence on local genetic diversity or population-specific indices of genetic differentiation (node- or neighbourhood-level analyses, *sensu* Wagner and Fortin (2013)) have rarely been studied (DiLeo and Wagner, 2016)(see Barr et al. (2015), Millette and Keyghobadi (2015) or Toma et al. (2015) for examples). In addition, genetic diversity estimates tend to be taken as a result of genetic drift in empirical studies, while genetic differentiation is mainly explained by levels of gene flow. However, genetic diversity and differentiation are both influenced by the interaction of drift and gene flow. Furthermore, node-based studies mostly focus on either genetic diversity or differentiation (Flavenot et al., 2015, Toma et al., 2015) and consider simple habitat metrics such as habitat amount in circular neighbourhoods around populations and distances to nearest habitat patches (Hahn et al., 2013, Taylor and Hoffman, 2014). Because ARH metrics comprehensively reflect the drivers of both drift and gene flow, they could be relevant predictors of both genetic diversity and differentiation (Foltête et al., 2020). This would help understanding how each response is influenced by the habitat spatial pattern. Computing ARH metrics under different matrix resistance scenarios additionally offers the opportunity to assess the role of matrix resistance in these relationships.

ARH metrics are even more relevant for landscape genetics since the genetic structure of a set of populations can also be represented as a genetic graph in which nodes are sampled populations whereas links are weighted by genetic distances and represent substantial gene exchanges between populations (Dyer, 2015, Greenbaum and Fefferman, 2017, Savary et al., 2021a). Their nodes can be weighted by local genetic diversity indices (node-level) as well as indices considering genetic differentiation with other populations (neighbourhood-level) (Koen et al., 2016). In the latter case, the topology of the population network can be taken into account through graph pruning, which removes certain links between populations. It makes it possible to consider gene exchanges at different spatial scales when computing these genetic differentiation indices (Savary et al., 2021a). As evidenced by DiLeo and Wagner (2016), node- and neighbourhood-level approaches are the only landscape genetic approaches making it possible to study the relationships between (i) either genetic diversity or differentiation and (ii) either habitat amount or configuration.

Accordingly, in this study, we aimed at answering the following question: are ARH metrics better predictors of genetic structure than commonly used habitat metrics? To that purpose, we used an empirical genetic dataset concerning the large marsh grasshopper (*Stethophyma grossum*). This species has limited dispersal capacities and forms discrete populations in small habitat patches, making it a good model for understanding how the spatial patterns of habitat influence genetic structure. We computed local genetic diversity and genetic differentiation indices from genetic graphs. In parallel, we computed three ARH metrics (capacity, F, BC) at different scales in landscape graphs, while taking into account different matrix resistance scenarios. We also computed the distance to the nearest neighbouring patch (DistNN hereafter) and the amount of habitat in a circular buffer (buffer metric hereafter), two commonly used habitat metrics, for comparison purposes. We finally assessed the relationships between these genetic responses and landscape predictors through correlation analyses as well as partial least square regressions. These analyses also allowed us to compare the relationship between ARH metrics and either genetic diversity or differentiation, and the way the spatial scale and the resistance scenario influenced it.

MATERIAL AND METHODS
Study species and sampling area
We analysed an empirical dataset acquired and described by Keller et al. (2013) and van Strien et al. (2014). The large marsh grasshopper (*Stethophyma grossum*) is a specialist orthoptera species exhibiting a patchy distribution throughout most of Europe where it finds its habitat in periodically flooded grasslands and open wetlands (Bönsel and Sonneck, 2011, Reinhardt et al., 2005, Sonneck et al., 2008). In this species, dispersal seems possible even in suboptimal open areas such as dry grasslands (Marzelli, 1994) and the species is able to cross streams but suitable patches surrounded by trees cannot be reached (Reinhardt et al., 2005). Exceptionally, individuals can cover up to 1500 m, as observed by Griffioen (1996) in a permeable landscape.

Keller et al. (2013) modelled the potential habitat of the large marsh grasshopper in the surroundings of the city of Langenthal in the Oberaargau region of the Swiss plateau. This 180 km² area is characterised by intensive agriculture areas with forest patches and settlements. Across the potential habitat areas, 39 large marsh grasshopper populations were sampled exhaustively (Fig. 2) in July and August 2010. The tibia and tarsus of a mid-leg of each individual were sampled for genetic data analyses.

The genetic data analyses of eight microsatellite markers are described in Keller et al. (2013). Like those authors, we excluded the Sgr14 microsatellite marker from the analyses because of genotyping errors and high null allele frequency. This did not prevent us from detecting significant levels of genetic differentiation. Besides, we excluded two populations located on the eastern margin of the study area, as well as three other populations whose low numbers of individuals would have impaired our rarefied estimations of local genetic diversity (see below). In sum, we considered 34 populations with at least 12 individuals for a total of 886 individuals.

Genetic structure indices
At the intra-population level, we estimated the total ($_{AR}$) and private (Priv. $_{AR}$) allelic richness from rarefaction indices calculated using ADZE (Szpiech et al., 2008) to account for sample size differences. Note that the private allelic richness index indicates the number of alleles found in a given population while absent from all the others (Kalinowski, 2004). Thus, it can be considered as both a local genetic diversity index and a genetic differentiation index. For assessing genetic differentiation between pairs of populations, we computed the matrix of D_{PS} (calculated as 1—pairwise proportion of shared alleles) (Bowcock et al., 1994). This distance has been shown to respond quickly to recent landscape changes, making it relevant for estimating contemporary gene flow in landscape genetic analyses (Murphy et al., 2010). We also computed the matrix of pairwise F_{ST} (Weir and Cockerham, 1984), which is known to reflect historical gene flow (Latta, 2006, Murphy et al., 2010).

We then built genetic graphs whose nodes represented grasshopper populations. Links were weighted with either D_{PS} or F_{ST} values. In the complete graphs, every population was connected to every other population but we also created pruned graphs in which only a subset of links was included. In order to avoid any artefactual correlation between habitat metrics and graph-based genetic indices, we used a pruning method taking only genetic distances into account. To that purpose, we identified the so-called "percolation threshold" using an edge-thinning method (Urban and Keitt, 2001). Following Rozenfeld et al. (2008), we computed this threshold from genetic data, searching for the genetic distance associated with the graph link whose removal would break the graph into two components. All the links corresponding to genetic distances larger than this threshold were removed. Gene flow has been shown to be frequent but spatially limited in this area (Keller et al., 2013) and we therefore assumed that above this genetic threshold distance, genetic differentiation between populations poorly reflected landscape effects on gene flow. From these graphs, we computed the mean of the inverse weight of the links connected to each node (thereafter referred to as MIW-D_{PS} and MIW-F_{ST} according to the genetic distance used). High values of MIW indicate a high degree of genetic similarity of a population with the others. This metric has been shown to correlate well with the number of migrants (Koen et al., 2016) and other population-specific genetic differentiation indices have already been recommended and used for landscape genetic analyses (DiLeo and Wagner, 2016, Gaggiotti and Foll, 2010, Millette and Keyghobadi, 2015, Peterman et al., 2015). Genetic graphs were constructed and metrics were computed using the `graph4lg` package in R (Savary et al., 2021b).
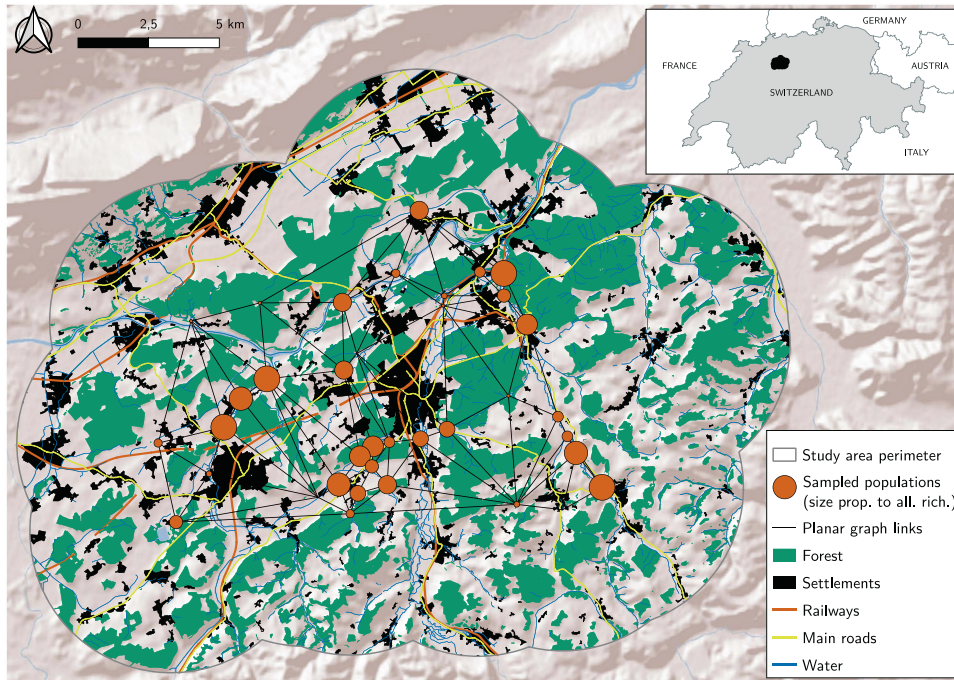
Fig. 2 Study area. Location of the sampled populations in the surroundings of Langental in the Oberaargau region.
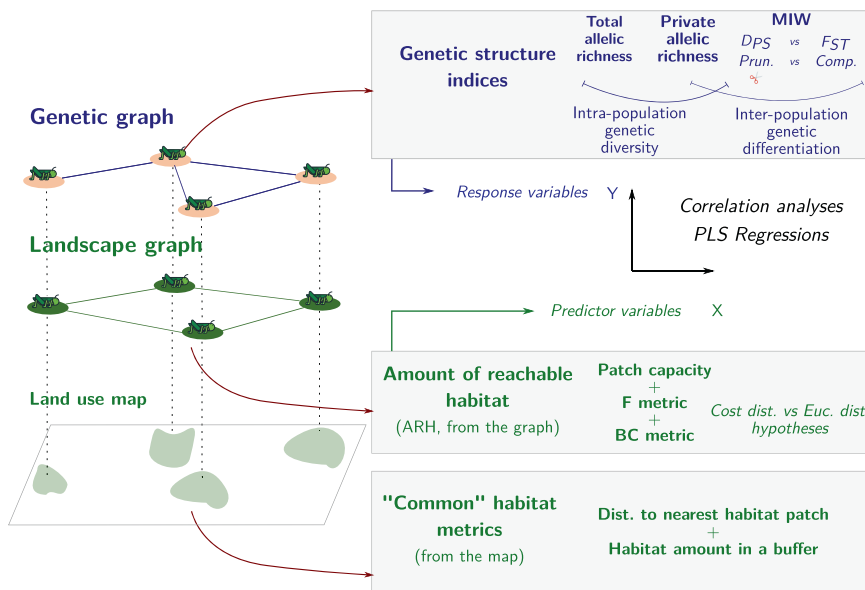

Fig. 3 Genetic indices and habitat metrics computed in both types of graphs to depict both genetic structure (diversity, differentiation) and the amount of reachable habitat and to perform correlation and partial least squares analyses. "Prun.": pruned graph, "Comp.": complete graph, "Cost dist.": cost-distances, "Euc. dist.": Euclidean distances.

Habitat metric calculations

We used rasterised (resolution: 10 m) land cover data from the sampling year in the area encompassing buffers of 5 km radius around each sampling site. In this area, we described the habitat spatial pattern by computing three ARH metrics (capacity, F, BC) from a landscape graph (Fig. 3).

Landscape graph construction. We considered six land cover types: (i) potential habitat areas, (ii) forest areas, (iii) settlements, (iv) agricultural areas, (v) wetlands and water areas, and (vi) railways and roads. Potential habitat areas corresponded to areas close to open water (≤500 m), within open agricultural areas and where water from the surroundings (500 m radius) can accumulate (Keller et al., 2013). We created a resistance surface

by combining these land cover data. The sampling of Keller et al. (2013) was exhaustive within the modelled potential habitat. Therefore, we built landscape graphs whose nodes were the 37 sampling sites in which several individuals were observed. The terms nodes, patches and populations are used interchangeably here. We used the resistance surface for computing the cost-distances between the nodes, which were used to weight the graph links.

We distinguished several "expert-based" scenarios of landscape matrix resistance when assigning a cost value to every land cover type on the resistance surface. In the first four scenarios, we set the cost values as indicated in Table 1. With these four scenarios, we varied the influence of wetlands and water areas (cost values: 50 or 1000, W50 and W1000 respectively) and of the roads and railways (cost values: 50 or 1000, R50

Table 1. Scenarios of matrix resistance considered for computing cost-distances between habitat patches.

Scenario	Potential habitat	Forest	Settlements	Agricultural areas	Wetlands, water	Roads, railways
W50-R50	1	1000	1000	50	50	50
W1000-R50	1	1000	1000	50	1000	50
W50-R1000	1	1000	1000	50	50	1000
W1000-R1000	1	1000	1000	50	1000	1000

and R1000, respectively) because we wanted to test for the respective influence of these potential linear barriers on gene flow. Cost values associated with other land cover types were set assuming that this species moves easily in potential habitat areas or open areas, whereas it is hardly able to move across forests and anthropogenic areas (Bönsel and Sonneck, 2011, Griffioen, 1996, Marzelli, 1994).

We computed the least-cost paths between every pair of habitat patches using Dijkstra's algorithm and weighted the links with the corresponding cost-distances. In a fifth scenario, we built a graph whose link weights were geodesic Euclidean distances between patches. As this species is assumed to disperse by stepping stones given its limited dispersal capacities, for every resistance scenario, the landscape graphs were pruned with a Delaunay triangulation resulting in a planar graph (Fig. 2).

Amount of reachable habitat metrics. To account for the influence of the ARH on genetic drift and gene flow, we took advantage of the spatial graph approach for computing three complementary ARH metrics. The graph nodes were located at the centroid pixel of every sampled habitat patch and we first computed their capacity as the area of potential habitat reachable at the patch scale. To that end, we assigned to every potential habitat cell surrounding the central pixel of the sampling site a weight that decreases with its cost-distance to this pixel. The weight of the potential habitat cell j located at a cost-distance d_{ij} from the central pixel of site i is equal to $e^{-ad_{ij}}$, such that the $Capacity_i$ of patch i is equal to:

$$Capacity_i = \sum_{j=1}^{N} e^{-ad_{ij}}$$

where N is the total number of potential habitat cells. We set a values such that $p = e^{-ad_{ij}} = 0.05$ at a cost-distance equivalent to 1500 m from the sampling site centroid, because distance-weighting exponential functions assuming that landscape effects on biological responses progressively decay with distance have been shown to outperform weighting functions based on fixed distance thresholds (Miguet et al., 2017). We converted this geodesic metric distance into cost-distance units using a log-log regression, following Tournant et al. (2013). After performing the same calculations for distances of 500 and 1000 m with very similar results, we retained 1500 m as the best scale because it is in the same order of magnitude as the maximum dispersal capacity of the species. Given that the large marsh grasshopper occupies small localised habitat patches, this metric reflects the amount of habitat reachable by individuals at the scale of the discrete patch occupied by their population. It is thus a suitable proxy for the effective population size driving genetic drift (DiLeo and Wagner, 2016). It was computed for each resistance surface and cost scenario.

As the capacity reflects the intra-patch component of the ARH, we computed two other metrics reflecting the ARH due to other patches:

- The Flux metric (F) represents the amount of habitat that is reachable when dispersing from a focal patch to other habitat patches. It can also be thought of as the amount of habitat from which migrants can originate. We computed the F using the following formula:

$$F_i = \sum_{j=1, j \neq i}^{n} Capacity_j^{\beta} e^{-ad_{ij}}$$

with i the index of the focal patch and j the index of all the other n patches, d_{ij} the distances (cost-distance or geodesic Euclidean distance) between patches i and j, $Capacity_j$ is the capacity of patch j and β indicates whether the patch capacity is taken into account ($\beta = 1$) or not ($\beta = 0$) in the calculation. Note that when $\beta = 0$, the F metric is essentially a topological metric reflecting the influence of the number and proximity of patches that are reachable. a was computed according to different dispersal kernels in order to test for the influence of the scale at which dispersal takes place. To that purpose, we set a values such that $p = e^{-ad_{ij}} = 0.05$ for

distances d_{ij} ranging from 1500 to 7500 m (with steps of 500 m). We thereby considered the ARH beyond the scale at which patch capacities were computed and until large scales as compared with the species dispersal capacities. For the sake of brevity, we refer to these distances d_{ij} at which $p_{d_{ij}} = 0.05$, either cost-distances or geodesic distances, as maximum dispersal distances (MDD) and express them in equivalent metric units after conversion.

- The Betweenness Centrality metric (BC) represents the number of times a focal patch (node/population) is a step on the indirect least-cost path from one patch to another when considering all possible patch pairs, excluding pairs involving the focal patch itself. It therefore reflects the role of that patch for potential dispersal movements at the scale of the whole habitat network and across several generations (traversability). Each term of this metric is weighted by the product of connected patch capacities (if $\beta = 1$) and dispersal probabilities associated with the inter-patch distance such that:

$$BC_i = \sum_{j} \sum_{k} Capacity_j^{\beta} Capacity_k^{\beta} e^{-ad_{jk}}$$

$$j, k \in \{1, \ldots, n\}, k < j, i \in P_{jk}$$

where P_{jk} represents the set of patches that are located along the least-cost path between patches j and k. We used the same a and β values as for the F index.

Because patches with large BC values may play a key role for dispersal between a large number of habitat patches ($\beta = 0$) and/or a great amount of habitat areas ($\beta = 1$), populations occupying these patches are expected to be genetically similar to the others and to have a high genetic diversity (Zetterberg et al., 2010).

As these three ARH metrics are complementary and make it possible to cover a large range of calculation parameters, other habitat metrics found in the literature (Capurucho et al., 2013, Peterman et al., 2015, Taylor and Hoffman, 2014) are particular cases of these metrics. Thus, although we aimed at assessing the relevance of the unified and flexible framework of the ARH metrics, we computed buffer metrics and the DistNN, two other habitat metrics, for comparative purposes. We first computed the buffer metrics, which measure the amount of potential habitat in circular neighbourhoods around each sampling sites considering similar scales as for the ARH metrics calculation. When considering small radius (from 100 to 500 m with steps of 100 m), "local buffer" metrics were akin to the capacity metric whereas "large buffer" metrics (from 1000 to 5000 m with steps of 500 m) more closely reflected the F metric calculation. We also computed the amount of potential habitat in non-circular neighbourhoods whose radius depended on cost-distance values according to every cost scenario. We use the terms "Local.Buffer" and "Large.Buffer" hereafter. In the Euclidean resistance scenario, the buffer is circular, and non-circular in the other scenarios. Finally, we computed the distance from each population to the nearest neighbour habitat patch occupied by a sampled population (DistNN) under every cost scenario. We built landscape graphs and computed metrics using Graphab 2.4 software (Foltête et al., 2012).

Analyses of the relationship between habitat metrics and genetic structure indices

Correlation analyses. We first assessed the correlations between the habitat metrics and the genetic indices (Fig. 3). Because all these variables were not normally distributed, we computed the Spearman rank correlation coefficient and tested for the significance of the correlations. We adjusted the p-values using the Benjamini and Hochberg (1995) method to control for the False Discovery Rate.

Partial Least Squares regressions. Simple correlation analyses allowed us to identify the habitat metrics, spatial scales and matrix resistance

scenarios most strongly related to each genetic response. However, they could not depict the complex relationships between genetic indices and our set of complementary ARH metrics. We therefore carried out Partial Least Squares regressions (PLS-R)(Carrascal et al., 2009) in which genetic indices were the response variables whilst ARH metrics were the predictor variables (Fig. 3). PLS regressions are an alternative to multiple linear regression and principal component regression (Roy et al., 2015, Wold et al., 2001), particularly adapted when predictor variables are collinear. The main difference with Principal Component Regression is that both the response and predictor variables are considered for creating a factorial space (Long, 2013). Response variables were rank-transformed because of departures from normal distributions. We assessed the complementarity of the ARH metrics through multivariate analyses, by testing for all combinations of three predictor variables involving a patch capacity, F and BC metric.

Following Tenenhaus (1998), we computed the Q^2 index to assess the role of every component in improving the prediction of the response variable when performing leave-one-out cross-validation. We only described the results obtained with models in which at least one component significantly improved the prediction of the response variable, i.e. when the Q^2 associated with this component is larger than 0.0975 (Supplementary information 2). We compared these models according to the Q^2 values associated with their significant components. Variable influences were assessed by computing their squared weights on the significant components. Variable weights were validated through bootstrap procedures following Pérez-Rodríguez et al. (2018). For every top model, the dataset was sampled with replacement 1000 times and the variable weights were estimated. If the 2.5–97.5% interval of the series of obtained values did not overlap zero, then we considered that the variable contributed significantly to the construction of the component.

RESULTS
Landscape and genetic graphs
The planar landscape graphs included 37 nodes and 95 links (Fig. 2) and the complete genetic graphs included 34 nodes connected by 561 links. The genetic graphs pruned using percolation thresholds computed from D_{PS} or F_{ST} values both included 412 links, although they had slightly different topologies (Figure S1).

Correlations between ARH metrics and genetic responses
The DistNN metric never significantly correlated with any genetic index (Table 2). Although the Local.Buffer metric consistently exhibited positive correlations with genetic indices (up to $r = 0.347$ with allelic richness), this correlation was never significant. Besides, the Large.Buffer metric was only significantly correlated to the allelic richness when considering a radius equivalent to 1000 m or 4500 m in the cost scenarios assigning water areas a low resistance (W50-R1000: $r = 0.482$ and W50-R50: $r = 0.432$, respectively). Overall, these commonly used habitat metrics performed poorly as compared with ARH metrics derived from landscape graphs.

Allelic richness was positively correlated with patch capacity ($r = 0.447$). This correlation was only significant when the capacity was computed under the cost scenario assigning a low resistance to water areas and a high resistance to roads and railways (W50-R1000). Thus, the local genetic diversity of a population is greater when this population occupies a patch with a large habitat surface reachable without crossing roads or railways. In contrast, patch capacities were not significantly correlated with any index of private allelic richness or relative genetic differentiation (MIW) derived from the genetic graphs, whatever the genetic distance and graph topology considered in the calculation.

All genetic indices tended to correlate more strongly with the Flux (F) metrics than they did with the Betweenness Centrality (BC) metrics (Table 2). Allelic richness and private allelic richness were respectively positively and negatively correlated with both metrics (Figs. S2 and S3). While allelic richness was more strongly correlated with the F metric computed considering cost-distances, especially when assigning roads and railways a high resistance

(W50-R1000: $r = 0.538$ or $r = 0.566$ when MDD = 1500 and $\beta = 0$ or $\beta = 1$ respectively, Table 2), private allelic richness was more strongly correlated with this metric when computed using Euclidean distances ($r = -0.609$ or $r = -0.593$ when MDD = 5500 or 2500 and $\beta = 0$ or $\beta = 1$, respectively, Table 2). The MDD did not have much influence on the correlation coefficients (Fig. S2) and we could not identify a scale of effect. Overall, the correlation values depended only slightly on the weight given to patch capacities (β value) when computing the metrics.

The MIW indices were positively and most often significantly correlated with the F and BC metrics (Table 2), indicating that populations located in habitat patches surrounded by large and nearby habitat patches tended to be genetically more similar to others than populations located in habitat patches isolated from large habitat patches (Fig. 4). Overall, the correlations were stronger when computing the MIW indices from pruned graphs rather than from complete graphs (Table 2). This was especially apparent when using the D_{PS} to weight the genetic graph links. However, these correlations were influenced by both the genetic distance used in the calculation and the type of distances (geodesic or cost-distances) used to compute the F and BC metrics. MIW-D_{PS} indices were more strongly correlated with F metrics computed using Euclidean distances whereas MIW-F_{ST} indices were more strongly correlated with F metrics computed using cost-distances under the scenarios W50-R50 or W50-R1000 which both assign a low resistance to water areas (Fig. 4). In both cases, correlation coefficients reached their highest values when the MDD was between 2000 and 3000 m.

Partial Least Squares regressions
Among all combinations of capacity, F and BC metrics, only one component had a significant effect in the PLS-R models explaining one of the genetic indices, except in one case where two components significantly explained the MIW-D_{PS} derived from a pruned graph. Among these combinations, the best models were very similar for a given response variable. Overall, the best model fits were obtained when patch capacities were not included in the calculation of the F metric ($\beta = 0$) and, except for the MIW-F_{ST}, included in the calculation of the BC metric ($\beta = 1$). Yet, these differences were most often subtle (Table 3). Accordingly, we only describe the results of the best models created with each response variable (Table 3).

The allelic richness was best explained when fitting a PLS-R model including the three following ARH metrics: capacity computed under the cost scenario W50-R1000, F metric computed under the same scenario with $\beta = 0$ and a MDD of 2000 m, and BC metric computed under the cost-scenario W1000-R1000 with $\beta = 1$ and a MDD of 7500 m. These variables were highly and positively correlated with the first component (Capacity: $r = 0.829$, F: $r = 0.860$, BC: $r = 0.787$ in the best model, Table 3 and Supplementary information 1, Fig. S5A). The R^2 associated with this component was equal to 0.325 in the best model, whereas the corresponding Q^2 was about 0.280 (Table 3).

In contrast, the only variable contributing significantly to the first component derived from the best PLS-R model explaining the private allelic richness was the F computed using Euclidean distances, with $\beta = 0$ and MDD of 1500 or 2000 m (Table 3). This variable was negatively correlated with the first component ($r = -0.816$) indicating that the private allelic richness is lower when habitat patches are surrounded by other nearby habitat patches (Supplementary information 1, Fig. S5B). Both the R^2 and the Q^2 values associated with this first component were larger than in the PLS-R models explaining allelic richness ($R^2 = 0.515$, $Q^2 = 0.415$).

The best PLS-R models explaining the MIW indices were obtained when computing them from pruned genetic graphs, the pruning step making the greatest differences in model fits when computing the MIW-D_{PS} (Table 3). Model goodness of fit

Table 2. Spearman correlation coefficients between genetic indices and ARH metrics according to the cost scenario used, the MDD considered and the weight given to patch capacities in the metric calculation (β value). The largest correlation coefficient obtained for each genetic index, habitat metric and β value are displayed. The 'Signif.' column indicates whether the correlation is still significant after p value adjustment ($p < 0.05$, $**p < 0.01$, $***p < 0.001$). For the cost scenarios, refer to Table 1. "DistNN" means "Distance to the Nearest Neighbour".

Genetic index	Habitat metric					Correlation	
	Metric	Cost sc.	MDD	β		$r_{Spearman}$	Signif.
AR	Capacity	w50-r1000	1500	$\varnothing$		0.447	*
AR	Local.Buffer	Euc.	200	$\varnothing$		0.347	
AR	DistNN	w50-r1000	$\varnothing$	$\varnothing$		−0.057	
AR	F	w50-r1000	1500	0		0.538	*
AR	F	w50-r1000	1500	1		0.566	*
AR	Large.Buffer	w50-r1000	1000	$\varnothing$		0.482	*
AR	BC	w1000-r1000	7500	0		0.382	
AR	BC	w1000-r1000	7500	1		0.387	
Priv. AR	Capacity	w50-r1000	1500	$\varnothing$		0.238	
Priv. AR	Local.Buffer	w50-r1000	300	$\varnothing$		0.289	
Priv. AR	DistNN	w50-r50	$\varnothing$	$\varnothing$		−0.181	
Priv. AR	F	Euc.	5500	0		−0.609	**
Priv. AR	F	Euc.	2500	1		−0.593	**
Priv. AR	Large.Buffer	w50-r1000	2000	$\varnothing$		0.391	
Priv. AR	BC	Euc.	3000	0		−0.437	*
Priv. AR	BC	Euc.	1500	1		−0.387	
MIW$_{comp.DPS}$	Capacity	w50-r1000	1500	$\varnothing$		0.203	
MIW$_{comp.DPS}$	Local.Buffer	w50-r50	500	$\varnothing$		0.274	
MIW$_{comp.DPS}$	DistNN	w50-r50	$\varnothing$	$\varnothing$		0.126	
MIW$_{comp.DPS}$	F	Euc.	3000	0		0.500	*
MIW$_{comp.DPS}$	F	Euc.	3500	1		0.467	*
MIW$_{comp.DPS}$	Large.Buffer	Euc.	5000	$\varnothing$		−0.348	
MIW$_{comp.DPS}$	BC	Euc.	1500	0		0.380	
MIW$_{comp.DPS}$	BC	Euc.	1500	1		0.335	
MIW$_{prun.DPS}$	Capacity	w50-r1000	1500	$\varnothing$		0.176	
MIW$_{prun.DPS}$	Local.Buffer	w50-r50	500	$\varnothing$		0.279	
MIW$_{prun.DPS}$	DistNN	w50-r50	$\varnothing$	$\varnothing$		0.113	
MIW$_{prun.DPS}$	F	Euc.	3000	0		0.632	**
MIW$_{prun.DPS}$	F	Euc.	3500	1		0.597	**
MIW$_{prun.DPS}$	Large.Buffer	Euc.	5000	$\varnothing$		−0.356	
MIW$_{prun.DPS}$	BC	Euc.	1500	0		0.453	*
MIW$_{prun.DPS}$	BC	Euc.	1500	1		0.411	
MIW$_{comp.FST}$	Capacity	w50-r1000	1500	$\varnothing$		0.329	
MIW$_{comp.FST}$	Local.Buffer	w50-r50	500	$\varnothing$		0.303	
MIW$_{comp.FST}$	DistNN	Euc.	$\varnothing$	$\varnothing$		-0.053	
MIW$_{comp.FST}$	F	w50-r50	3000	0		0.663	**
MIW$_{comp.FST}$	F	w50-r50	1500	1		0.602	**
MIW$_{comp.FST}$	Large.Buffer	w50-r1000	1000	$\varnothing$		0.370	
MIW$_{comp.FST}$	BC	w50-r1000	7000	0		0.472	*
MIW$_{comp.FST}$	BC	w1000-r1000	1500	1		0.441	*
MIW$_{prun.FST}$	Capacity	w50-r1000	1500	$\varnothing$		0.327	
MIW$_{prun.FST}$	Local.Buffer	w50-r50	500	$\varnothing$		0.310	
MIW$_{prun.FST}$	DistNN	w1000-r1000	$\varnothing$	$\varnothing$		0.059	
MIW$_{prun.FST}$	F	w50-r50	2000	0		0.686	**
MIW$_{prun.FST}$	F	w50-r50	1500	1		0.624	**
MIW$_{prun.FST}$	Large.Buffer	w50-r1000	1000	$\varnothing$		0.358	
MIW$_{prun.FST}$	BC	w50-r1000	7000	0		0.507	*
MIW$_{prun.FST}$	BC	w1000-r1000	1500	1		0.454	*

was overall better when modelling the MIW-F$_{ST}$ than the MIW-D$_{PS}$. The first component alone explained about 40% of the variance of the MIW index and up to 50% when modelling the MIW-F$_{ST}$ derived from a pruned graph (Table 3). This share was moderately reduced when performing the cross-validation (Q^2: from 0.314 to 0.443, Table 3). Here again, only the F contributed significantly to the first component, which was in most cases the only component explaining significantly the MIW (Supplementary information 1,
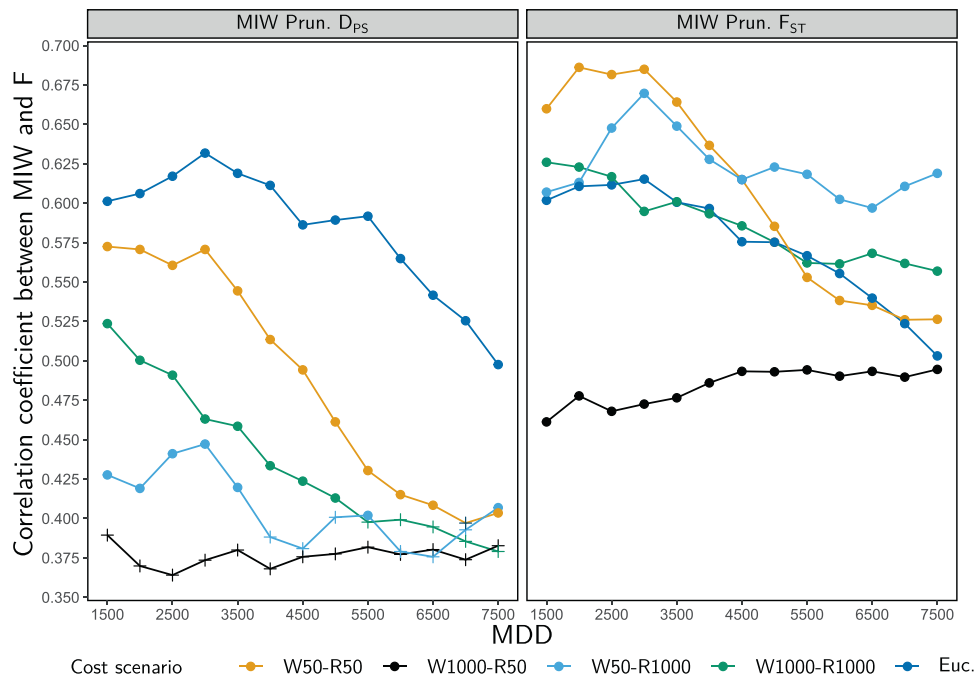
Fig. 4 Variation of the Spearman correlation coefficients between the relative genetic differentiation index MIW computed from the pruned genetic graphs and the F metric according to the genetic distance, cost scenarios and dispersal kernels used to compute these indices. The x axis indicates the dispersal kernels used to compute the metrics and corresponds to the MDD (maximum dispersal distances). In this figure, the F metric was computed without weighting patch capacities ($\beta = 0$). Point colours refer to the cost scenario used to compute cost-distances (see Table 1). The left and right panels display the variations observed when computing MIW from a genetic graph weighted with D_{PS} and F_{ST} values respectively. Crosses indicate that the correlation is not significant after p-value adjustment.

Figs. S5C and S5D). While the F was computed from Euclidean distances with $\beta = 0$ and MDD of 3500 m in the best model explaining the MIW-D_{PS}, it was computed with cost-distances under the scenario W50-R1000 with $\beta = 0$ and considering dispersal at a smaller scale (MDD = 2500 m) in the best model for MIW-F_{ST} (Table 3). In all cases, the correlation between the first component of the PLS models and the F was strong and positive (r about 0.97).

DISCUSSION

We assessed the advantage of using complementary metrics measuring the amount of reachable habitat (ARH) instead of two other commonly used habitat metrics for explaining population genetic structure. The three ARH metrics derived from the unified and flexible framework offered by landscape graphs, i.e. the patch capacity, Flux and Betweenness Centrality metrics, were relevant predictors of the two components of genetic structure, i.e. genetic diversity and genetic differentiation. They provided an advantage over the distance to the nearest neighbour patch (DistNN) and the amount of habitat in buffer areas (Local.Buffer or Large.Buffer) that were poor predictors in this study. Besides, although allelic richness was significantly explained by the three complementary ARH metrics in the best PLS-R model, private allelic richness and MIW indices were essentially related to the ARH measured outside the focal patch. Finally, considering several matrix resistance scenarios for computing ARH metrics was key for evidencing that local genetic diversity seemed to be negatively influenced by transport infrastructures and positively by water surfaces, whereas these landscape features did not influence genetic differentiation in the same way when measured with either the D_{PS} or the F_{ST}.

Are ARH metrics relevant predictors of genetic structure?

All the genetic indices describing the genetic structure of the grasshopper populations were significantly correlated with at least

one ARH metric and explained by these metrics in PLS models. In contrast, the two habitat metrics (DistNN and buffer metrics) previously used in landscape genetic analyses were hardly significantly correlated with the genetic indices, and in these rare cases, the correlation was much lower. Our results therefore confirm that the three ARH metrics here considered are relevant for describing the habitat pattern driving both genetic drift (capacity) and gene flow (Flux, BC) processes. Interestingly, our results match the results of Moilanen and Nieminen (2002) regarding the respective performance of several habitat metrics in predicting colonisation events. Although based on different biological responses, their results and ours provided similar evidence for the poor performance of metrics considering habitat amount in neighbourhoods delineated with fixed radius or distances to nearest patches, as compared with metrics considering dispersal probabilities to neighbouring patches.

As the computation of ARH metrics is very flexible, they include habitat metrics already computed in previous studies, as for example the amount of habitat in a circular neighbourhood with a radius of 15 km, identified by Capurucho et al. (2013) as the best predictor of genetic diversity in a tropical bird species (see Keyghobadi et al. (2005), Millette and Keyghobadi (2015) or Peterman et al. (2015) for other examples). Using complementary ARH metrics in this and similar study could thus have provided stronger statistical relationships and complementary insights into drift and gene flow processes driving genetic responses. In sum, although other metrics can explain genetic structure, landscape graphs offer a unified and flexible framework for understanding the influence of habitat patterns on genetic structure.

Including patch capacities in the calculation of the F and BC metrics only marginally influenced our results. Therefore, the number of reachable patches in a habitat network alone was often a good predictor of genetic structure. This recalls the results of Peterman et al. (2015) which have identified the isolation of a patch relative to others as the best predictor of population-

Table 3. Results of the Partial Least Squares regression (PLS-R) of the genetic indices by the capacity, flux and betweenness centrality metrics. For each genetic index and patch capacity weighting parameter for computing F and BC (β value), the best model according to the Q^2 associated with the first PLS component (Q^2.t1) is displayed (largest Q^2 value for each genetic index displayed in italics). When β is equal to 1, patch capacities are included in the metric calculation and not otherwise ($\beta = 0$). MDD indicates the distance at which the dispersal probability is set to 0.05 for the metric calculation. For the cost scenarios, refer to Table 1. The r.t1 column gives the Pearson correlation coefficient between the PLS components t1 and the habitat metrics. These values are displayed in bold when the metrics significantly contribute to the construction of the PLS components. R^2.t1 and Q^2.t1 values associated with the first component respectively indicate the proportion of the response variable variance and the cross-validated proportion of this variance explained by each PLS component. Q^2 values above 0.0975 indicate that the PLS component has a significant effect on the response variable and are displayed in bold. Q^2.t2 values associated with the second component are also displayed for information purposes. MIW$_{comp.DPS}$ and MIW$_{prun.DPS}$ refer to the MIW-D_{PS} computed from complete and pruned genetic graphs respectively (similar notation for the MIW-F_{ST}).

	Capacity				F				BC		Model fit		
Gen. index	Cost sc.	r.t1	β	MDD	Cost sc.	r.t1	β	MDD	Cost sc.	r.t1	Q^2.t1	R^2.t1	Q^2.t2
AR	w50-R1000	**0.848**	0	2000	w50-R1000	**0.885**	0	1500	w1000-R50	0.331	**0.273**	0.326	−0.035
ar	*w50-r1000*	*0.829*	*0*	*2000*	*w50-r1000*	*0.860*	*1*	*7500*	*w1000-r1000*	*0.787*	*0.280*	*0.325*	*−0.098*
AR	w50-R1000	**0.847**	1	2000	w50-r1000	**0.916**	0	5500	w1000-R1000	**0.870**	**0.214**	0.246	−0.095
AR	w50-R1000	**0.860**	1	2000	w50-R1000	**0.880**	1	7500	w1000-R1000	**0.807**	**0.230**	0.270	−0.136
Priv. AR	w50-r1000	0.201	0	1500	Euclid.	**−0.947**	0	5500	w1000-R1000	−0.396	**0.317**	0.443	−0.003
Priv.ar	*w50-r1000*	*0.427*	*0*	*2500*	*Euclid.*	*−0.813*	*1*	*6000*	*w1000-r1000*	*0.160*	*0.411*	*0.511*	*−0.147*
Priv. AR	w50-r1000	0.221	1	2000	Euclid.	**−0.943**	0	7000	w1000-R1000	−0.385	**0.309**	0.437	−0.006
Priv. AR	w50-r1000	0.448	1	2500	Euclid.	**−0.790**	1	7000	w1000-R1000	0.182	**0.410**	0.508	−0.141
MIW$_{comp.DPS}$	w1000-R50	−0.216	0	4000	Euclid.	**0.952**	0	2500	Euclid.	**0.880**	**0.148**	0.217	−0.076
miw$_{comp.dps}$	*w50-r1000*	*0.390*	*0*	*4500*	*Euclid.*	*0.929*	*1*	*6500*	*w1000-r1000*	*0.492*	*0.178*	*0.291*	*0.063*
MIW$_{comp.DPS}$	w1000-R50	−0.178	1	2500	Euclid.	**0.927**	0	3500	Euclid.	**0.905**	**0.110**	0.182	−0.077
MIW$_{comp.DPS}$	w50-R1000	0.491	1	3500	Euclid.	**0.904**	1	5500	w1000-R1000	0.542	**0.098**	0.211	0.025
MIW$_{prun.DPS}$	w1000-R50	−0.131	0	3000	Euclid.	**0.958**	0	2000	Euclid.	**0.901**	**0.312**	0.366	−0.081
miw$_{prun.dps}$	*w50-r1000*	*0.284*	*0*	*3500*	*Euclid.*	*0.974*	*1*	*7500*	*w1000-r1000*	*0.477*	*0.314*	*0.394*	*0.101*
MIW$_{prun.DPS}$	w1000-R50	−0.093	1	3000	Euclid.	**0.938**	0	1500	Euclid.	**0.899**	**0.270**	0.325	−0.077
MIW$_{prun.DPS}$	w1000-R50	−0.084	1	3500	Euclid.	**0.947**	1	1500	Euclid.	**0.891**	**0.241**	0.303	−0.065
miw$_{comp.fst}$	*w1000-r50*	*0.117*	*0*	*2000*	*w50-r1000*	*0.969*	*0*	*3500*	*Euclid.*	*0.618*	*0.381*	*0.445*	*−0.115*
MIW$_{comp.FST}$	w50-R1000	0.387	0	3500	Euclid.	**0.942**	1	2500	w1000-R50	0.091	**0.355**	0.448	−0.070
MIW$_{comp.FST}$	w1000-R50	0.228	1	2500	w50-r1000	**0.974**	0	3000	Euclid.	0.615	**0.324**	0.393	−0.091
MIW$_{comp.FST}$	w1000-R50	0.234	1	2500	w50-r1000	**0.969**	1	2000	Euclid.	0.701	**0.295**	0.365	−0.065
miw$_{prun.fst}$	*w1000-r50*	*0.147*	*0*	*2500*	*w50-r1000*	*0.972*	*0*	*3500*	*Euclid.*	*0.643*	*0.443*	*0.504*	*−0.103*
MIW$_{prun.FST}$	w1000-R50	0.164	0	2500	Euclid.	**0.965**	1	3000	Euclid.	0.714	**0.408**	0.474	−0.076
MIW$_{prun.FST}$	w1000-R50	0.226	1	2500	w50-r1000	**0.970**	0	3500	Euclid.	0.633	**0.368**	0.436	−0.083
MIW$_{prun.FST}$	w1000-R50	0.238	1	2500	w50-r1000	**0.965**	1	2500	Euclid.	0.717	**0.327**	0.402	−0.062

specific genetic differentiation indices. Thus, the advantage of the landscape graph approach for measuring the ARH could stem from their direct consideration of population topology, already recognised as an important driver of dispersal and gene flow patterns (Saura et al., 2014, van Strien, 2017).

Does the ARH influence genetic diversity and genetic differentiation to the same degree and at the same spatial scale?
It has previously been observed that genetic differentiation and local genetic diversity indices were not influenced to the same degree and at the same spatial scale by the habitat pattern (Balkenhol et al., 2013, Keyghobadi et al., 2005, Kierepka et al., 2020, Taylor and Hoffman, 2014). Our results confirm these previous results given that we used a common statistical approach for analysing these two components of genetic structure. On the one hand, allelic richness was significantly correlated with both the F metric and the patch capacity and was the only genetic index significantly explained by the capacity in the PLS models. On the other hand, private allelic richness and MIW indices appeared to be only related to F and BC metrics. Thus, local genetic diversity was influenced by the ARH at the scale of the focal patch and outside that patch, whereas genetic differentiation was influenced by the ARH outside the focal patch only. While genetic diversity and differentiation are expected to be driven by both gene flow and drift, DiLeo and Wagner (2016) suggested that a stronger effect of the local habitat amount on genetic diversity could stem from the close relationship between habitat amount and population size. In contrast, the effect of the large scale habitat pattern on migration rates seems to influence genetic differentiation more substantially than the effect of habitat area on drift does (Cushman et al., 2012).

The relative genetic differentiation among populations was better explained by the spatial pattern of habitats when computed from pruned genetic graphs. The relevance of graph pruning for landscape genetic analyses has already been suggested by Wagner and Fortin (2013) for link-level analyses and evidenced by Arnaud (2003), Angelone et al. (2011) and Savary et al. (2021a), among others. Besides, Shirk and Cushman (2011) have highlighted the importance of considering the spatial distribution of populations for computing genetic diversity indices in a genetic neighbourhood including several populations. Here, we further stress the relevance of reducing the set of population pairs considered for computing neighbourhood-level genetic indices from genetic graphs. The stronger relationship between the ARH and the relative genetic differentiation when considering only population pairs connected by frequent gene flow events confirms the result obtained by Keller et al. (2013) when analysing this dataset. They showed that the relationship between genetic differentiation and geodesic distance was positive only up to a

limited spatial scale, suggesting that the large marsh grasshopper is currently expanding. Indeed, although it has been negatively affected in the past by the reduction of wetland and grassland areas, intensive grassland management and river control reducing periodic flooding (Koschuh, 2004, Krause, 1996, Malkus, 1997, Reinhardt et al., 2005), the species has been recolonising new areas due to wetland conservation programmes and changes in grassland management practices, among others (Trautner and Hermann, 2008). Therefore, genetic differentiation at the scale of the entire study area might not have reached its equilibrium level, as expected from the IBD pattern dynamics theorised by Slatkin (1993). In this context (case-IV IBD *sensu* Hutchison and Templeton (1999)), the genetic differentiation pattern is best explained when considering only a subset of nearby population pairs, reinforcing the interest of genetic graph pruning. In summary, the spatial and temporal scales over which drift and gene flow influence population genetic structure could be identified by jointly using landscape and pruned genetic graphs for relating ARH metrics with genetic indices.

Does the resistance of the matrix affect genetic diversity and genetic differentiation in the same way?

The allelic richness and the relative genetic differentiation indices computed using the F_{ST} were most strongly correlated and best explained by ARH metrics computed with cost-distances. In contrast, considering geodesic Euclidean distances was the best option for explaining the private allelic richness and the relative genetic differentiation indices computed using the D_{PS}. In a context where the study species may be expanding due to landscape changes, these differences might result from i) the different time scales at which genetic diversity and differentiation respond to these changes and ii) the ability of genetic differentiation indices to reflect landscape influence on either historical or contemporary gene flow.

First, as expected from theory (Varvio et al., 1986), genetic differentiation reaches its equilibrium level faster than local genetic diversity does. For example, Keyghobadi et al. (2005) detected a positive influence of recent forests on genetic differentiation in a butterfly species dispersing through open areas and avoiding forests, while local genetic diversity was best explained by patch isolation metrics taking only geodesic Euclidean distances into account. Accordingly, the results we obtained can be interpreted from the following hypotheses. The closer relationship between local genetic diversity and ARH metrics considering cost-distances instead of geodesic Euclidean distances reflects the past influence of the matrix on dispersal. Second, the closer relationship between private allelic richness and MIW indices computed from graphs pruned with the D_{PS} and ARH metrics considering Euclidean distances instead of cost-distances points towards a lower influence of matrix resistance on contemporary dispersal. These hypotheses are consistent with the current expansion of this species.

Second, previous landscape genetic studies have shown that the D_{PS} reflects recent landscape effects on genetic structure while the F_{ST} should be preferred for reflecting past landscape effects (Holzhauer et al., 2006, Murphy et al., 2010, Storfer et al., 2010). This could explain why genetic differentiation indices computed using the F_{ST} were most correlated with ARH metrics taking into account the high resistance of some landscape features on dispersal. Although difficult to verify, this explanation would also mean that the landscape matrix have become more permeable for this species in recent years, thereby explaining its expansion.

On another note, as regards the nature of landscape feature effects on genetic structure, our results recall those obtained by Holzhauer et al. (2006), which observed that roads and railways might be barriers for the large marsh grasshopper while water areas are not. Indeed, the scenario in which roads and railways had a low resistance to movement and water areas a high resistance (W1000-R50) never provided the best fits when studying local genetic diversity and historical gene flow (F_{ST}). In contrast, the scenario in which transport infrastructures strongly limited dispersal and water areas were relatively permeable (W50-R1000) performed well in explaining these variables. This result is inconsistent with that of Keller et al. (2013) showing a positive effect of roads on dispersal in this species. However, these authors only considered a measure of genetic differentiation related to contemporary landscape influence on gene flow (mean assignment probabilities) as a response variable. Similarly, MIW indices based on the D_{PS} were best explained by ARH metrics without considering matrix resistance.

Limits and perspectives

The relationship between habitat structure and genetic structure is dynamic and takes time to reach an equilibrium (Slatkin, 1993). Besides, the topology of the habitat network has a strong influence on genetic structure, which may be related to the species dispersal pattern (van Strien, 2017). Even under the hypothesis where only the amount of habitat at a given scale drives diversity patterns (Fahrig, 2013), habitat configuration has been shown to affect them significantly (Saura, 2021). For example, different traversability properties of the habitat network may influence long-distance gene flow patterns over time, which would result in a different genetic structure. We also acknowledge that the relative effects of the ARH on the two components of genetic structure here observed may be specific to the habitat spatial pattern of our case study, but our results encourage using ARH metrics in empirical landscape genetic studies. These aspects could be further investigated using ARH metrics and performing gene flow simulations with varying population sizes, topologies, dispersal capacities, matrix resistances and habitat patterns.

Finally, our results are hardly comparable with previous ones distinguishing the effects of habitat amount and configuration on genetic structure (Cushman et al., 2012, Jackson and Fahrig, 2015, Millette and Keyghobadi, 2015). Most of these studies used link-level analyses (DiLeo and Wagner, 2016), whereas here we used a node- and neighbourhood-level approach. We may wonder whether it influences the detection of landscape genetic relationships. Indeed, the MIW index is based on genetic differentiation between one population and all links with neighbouring populations. It therefore averages landscape effects over all these links, which may preclude the possibility of precisely estimating the resistance of every type of landscape feature. Besides, in most previous studies, habitat configuration measures such as inter-patch distances or patch isolation were strongly correlated with habitat amount, which should have ruled out any conclusion that habitat configuration exerts a stronger influence than habitat amount does on genetic structure (Jackson and Fahrig, 2015). Accordingly, we focused here on complementary ARH metrics derived from spatial graphs because they account for the compounded effects of both habitat amount and configuration, which are highly interdependent (Didham et al., 2012). Their use has already been advocated (Saura, 2018) and we showed here that it makes it possible to understand how spatial habitat patterns influence both drift and gene flow at several spatial and temporal scales, while considering matrix resistance.

REFERENCES

Andersson E, Bodin Ö (2009) Practical tool for landscape planning? an empirical investigation of network-based models of habitat fragmentation. Ecography 32 (1):123–132

Angelone S, Kienast F, 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. Landsc Ecol 18 (3):333–346

Awade M, Boscolo D, Metzger JP (2012) Using binary and probabilistic habitat availability indices derived from graph theory to model bird occurrence in fragmented forests. Landsc Ecol 27(2):185–198

Balkenhol N, Pardini R, Cornelius C, Fernandes F, Sommer S (2013) Landscape-level comparison of genetic diversity and differentiation in a small mammal inhabiting different fragmented landscapes of the Brazilian Atlantic Forest. Conserv Genet 14(2):355–367

Baranyi G, Saura S, Podani J, Jordán F (2011) Contribution of habitat patches to network connectivity: redundancy and uniqueness of topological indices. Ecol Indic 11(5):1301–1310

Barr KR, Kus BE, Preston KL, Howell S, Perkins E, Vandergast AG (2015) Habitat fragmentation in coastal southern California disrupts genetic connectivity in the Cactus Wren (Campylorhynchus brunneicapillus). Mol Ecol 24(10):2349–2363

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soci Series B (Methodological) 57(1):289–300

Bergés A, Avon C, Bezombes L, Clauzel C, Duflot R, Foltête J-C, Gaucherand S, Girardet X, Spiegelberger T (2020) Environmental mitigation hierarchy and biodiversity offsets revisited through habitat connectivity modelling. J Environ Manag 256:1–10

Bertin A, Gouin N, Baumel A, Gianoli E, Serratosa J, Osorio R, Manel S (2017) Genetic variation of loci potentially under selection confounds species-genetic diversity correlations in a fragmented habitat. Mol Ecol 26(2):431–443

Bönsel AB, Sonneck A-G (2011) Habitat use and dispersal characteristic by Stethophyma grossum: the role of habitat isolation and stable habitat conditions towards low dispersal. J Insect Conserv 15(3):455–463

Boulanger E, Dalongeville A, Andrello M, Mouillot D, Manel S (2020) Spatial graphs highlight how multi-generational dispersal shapes landscape genetic patterns. Ecography 15(1):1–13

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368(6470):455–457

Capurucho JMG, Cornelius C, Borges SH, Cohn-Haft M, Aleixo A, Metzger JP, Ribas CC (2013) Combining phylogeography and landscape genetics of Xenopipo atronitens (Aves: Pipridae), a white sand campina specialist, to understand Pleistocene landscape evolution in Amazonia. Biol J Linnean Soc 110(1):60–76

Carrascal LM, Galván I, Gordo O (2009) Partial Least Squares regression as an alternative to current regression methods used in ecology. Oikos 118(5):681–690

Cushman SA, Shirk A, Landguth EL (2012) Separating the effects of habitat area, fragmentation and matrix resistance on genetic differentiation in complex landscapes. Landsc Ecol 27(3):369–380

Díaz SM, Settele J, Brondízio E, Ngo H, Guèze M, Agard J, Arneth A, Balvanera P, Brauman K, Butchart S, et al. (2019). The global assessment report on biodiversity and ecosystem services: summary for policy makers. Technical report, Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services

Didham RK, Kapos V, Ewers RM (2012) Rethinking the conceptual foundations of habitat fragmentation research. Oikos 121(2):161–170

DiLeo MF, Wagner HH (2016) A landscape ecologist's agenda for landscape genetics. Curr Landsc Ecol Rep 1(3):115–126

Dyer RJ (2015) Population graphs and landscape genetics. Annu Rev Ecol, Evolut Syst 46:327–342

Fahrig L (2013) Rethinking patch size and isolation effects: the habitat amount hypothesis. J Biogeogr 40(9):1649–1663

Flavenot T, Fellous S, Abdelkrim J, Baguette M, Coulon A (2015) Impact of quarrying on genetic diversity: an approach across landscapes and over time. Conserv Genet 16(1):181–194

Foltête J-C, Clauzel C, Vuidel G (2012) A software tool dedicated to the modelling of landscape networks. Environ Model Softw 38:316–327

Foltête J-C, Savary P, Clauzel C, Bourgeois M, Girardet X, Sahraoui Y, Vuidel G, Garnier S (2020) Coupling landscape graph modeling and biological data: a review. Landsc Ecol 35(5):1035–1052

Frankham R (2005) Genetics and extinction. Biol Conserv 126(2):131–140

Frankham R (2015) Genetic rescue of small inbred populations: Meta-analysis reveals large and consistent benefits of gene flow. Mol Ecol 24(11):2610–2618

Frankham R, Ballou JD, and Briscoe DA (2004) A primer of conservation genetics. Cambridge University Press

Gaggiotti OE, Foll M (2010) Quantifying population structure using the F-model. Mol Ecol Resour 10(5):821–830

Galpern P, Manseau M, Fall A (2011) Patch-based graphs of landscape connectivity: a guide to construction, analysis and application for conservation. Biol Conserv 144(1):44–55

Greenbaum G, Fefferman NH (2017) Application of network methods for understanding evolutionary dynamics in discrete habitats. Mol Ecol 26(11):2850–2863

Griffioen R (1996) Over het dispersievermogen van de moerassprinkhaan. Nieuwsbrief Saltabel 15(1):39–41

Hahn T, Kettle CJ, Ghazoul J, Hennig EI, Pluess AR (2013) Landscape composition has limited impact on local genetic structure in mountain clover Trifolium montanum L. J Heredity 104(6):842–852

Hedrick P (2011) Genetics of populations. Jones & Bartlett Learning

Holzhauer SI, Ekschmitt K, Sander A-C, Dauber J, Wolters V (2006) Effect of historic landscape change on the genetic structure of the bush-cricket Metrioptera roeseli. Landsc Ecol 21(6):891–899

Hutchison DW, Templeton AR (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

Ingvarsson PK (2001) Restoration of genetic variation lost-the genetic rescue hypothesis. Trends Ecol Evol 16(2):62–63

Jackson ND, Fahrig L (2015) Habitat amount—not habitat configuration—best predicts population genetic structure in fragmented landscapes. Landsc Ecol 31 (5):951–968

Joly CA, Metzger JP, Tabarelli M (2014) Experiences from the Brazilian atlantic forest: ecological findings and conservation initiatives. New Phytol 204(3):459–473

Kalinowski ST (2004) Counting alleles with rarefaction: private alleles and hierarchical sampling designs. Conserv Genet 5(4):539–543

Keller D, Holderegger R, Strien MJ (2013) Spatial scale affects landscape genetic analysis of a wetland grasshopper. Mol Ecol 22(9):2467–2482

Keyghobadi N (2007) The genetic implications of habitat fragmentation for animals. Can J Zool 85(10):1049–1064

Keyghobadi N, Roland J, Matter SF, Strobeck C (2005) Among- and within-patch components of genetic diversity respond at different rates to habitat fragmentation: an empirical demonstration. Proc R Soc B 272(1562):553–560

Kierepka EM, Anderson SJ, Swihart RK, Rhodes OE (2020) Differing, multiscale landscape effects on genetic diversity and differentiation in eastern chipmunks. Heredity 124(3):457–468

Koen EL, Bowman J, Wilson PJ (2016) Node-based measures of connectivity in genetic networks. Mol Ecol Resour 16(1):69–79

Koschuh A (2004) Verbreitung, lebensräume und gefährdung der sumpfschrecke (stethophyma grossum, l., 1758)(saltatoria) in der steiermark. Joannea, Zool 6:223–246

Krause S (1996) Populationsstruktur, Habitatbindung und Mobilität der Larven von Stethophyma grossum (Linné, 1758). Articulata 11(2):77–89

Latta RG (2006) Integrating patterns across multiple genetic markers to infer spatial processes. Landsc Ecol21(6):809–820

Lehnen L, Jan P-L, Besnard A-L, Fourcy D, Kerth G, Biedermann M, Nyssen P, Schorcht W, Petit E, and Puechmaille S (2021) Genetic diversity in a long-lived mammal is explained by the past's demographic shadow and current connectivity. Mol Ecol, 00(1)

Long FH (2013) Multivariate analysis for metabolomics and proteomics data. In Proteomic and metabolomic approaches to biomarker discovery, pages 299–311. Elsevier

Malkus J (1997) Habitatpräferenzen und mobilität der sumpfschrecke (stethophyma grossum l. 1758) unter besonderer berücksichtigung der mahd. Articulata 12 (1):1–18

Marzelli M (1994) Ausbreitung von mecostethus grossus auf einer ausgleichs-und renaturierungsfläche. Articulata 9(1):25–32

Miguet P, Fahrig L, Lavigne C (2017) How to quantify a distance-dependent landscape effect on a biological response. Methods Ecol Evol 8(12):1717–1724

Millette KL, Keyghobadi N (2015) The relative influence of habitat amount and configuration on genetic structure across multiple spatial scales. Ecol Evol 5 (1):73–86

Moilanen A, Nieminen M (2002) Simple connectivity measures in spatial ecology. Ecology 83(4):1131–1145

Mony C, Abadie J, Gil-Tena A, Burel F, Ernoult A (2018) Effects of connectivity on animal-dispersed forest plant communities in agriculture-dominated landscapes. J Veg. Sci. 29(2):167–178

Murphy MA, Evans JS, Storfer A (2010) Quantifying Bufo boreas connectivity in Yellowstone National Park with landscape genetics. Ecology 91(1):252–261

Pascual-Hortal L, Saura S (2006) Comparison and development of new graph-based landscape connectivity indices: towards the priorization of habitat patches and corridors for conservation. Landsc Ecol 21(7):959–967

Pasinelli G, Meichtry-Stier K, Birrer S, Baur B, Duss M (2013) Habitat quality and geometry affect patch occupancy of two Orthopteran species. PLoS One 8(5): e65850

Pérez-Rodríguez A, Khimoun A, Ollivier A, Eraud C, Faivre B, Garnier S (2018) Habitat fragmentation, not habitat loss, drives the prevalence of blood parasites in a Caribbean passerine. Ecography 41(11):1835–1849

Peterman WE, Anderson TL, Ousterhout BH, Drake DL, Semlitsch RD, Eggert LS (2015) Differential dispersal shapes population structure and patterns of genetic differentiation in two sympatric pond breeding salamanders. Conserv Genet 16 (1):59–69

Rayfield B, Fortin M-J, Fall A (2011) Connectivity for conservation: a framework to classify network measures. Ecology 92(4):847–858

Reinhardt K, Köhler G, Maas S, Detzel P (2005) Low dispersal ability and habitat specificity promote extinctions in rare but not in widespread species: the Orthoptera of Germany. Ecography 28(5):593–602

Roy K, Kar S, and Das RN (2015) Statistical methods in QSAR/QSPR. In A primer on QSAR/QSPR modeling, pages 37–59. Springer

Rozenfeld AF, Arnaud-Haond S, Hernández-Garcia E, Eguíluz VM, Serrão EA, Duarte CM (2008) Network analysis identifies weak and strong links in a metapopulation system. Proc Natl Acad Sci 105(48):18824–18829

Saura S (2018) The amount of reachable habitat—jointly measuring habitat amount and connectivity in space and time. In Proceedings of international conference of ecological sciences of the French Society for Ecology and Evolution

Saura S (2021) The Habitat Amount Hypothesis implies negative effects of habitat fragmentation on species richness. J Biogeogr 48(1):11–22

Saura S, Bodin Ö, Fortin M-J (2014) Stepping stones are crucial for species' long-distance dispersal and range expansion through habitat networks. J Appl Ecol 51(1):171–182

Saura S and de la Fuente B (2017) Connectivity as the amount of reachable habitat: conservation priorities and the roles of habitat patches in landscape networks. In Gergel, SE and Turner, MG, editors, Learning landscape ecology: a practical guide to concepts and techniques, pages 229–254. Springer

Saura S, Rubio L (2010) A common currency for the different ways in which patches and links can contribute to habitat availability and connectivity in the landscape. Ecography 33(3):523–537

Savary P, Foltête J-C, Moal H, Vuidel G, Garnier S (2021a) Analysing landscape effects on dispersal networks and gene flow with genetic graphs. Mol Ecol Resour 21 (4):1167–1185

Savary P, Foltête J-C, Moal H, Vuidel G, Garnier S (2021b) graph4lg: a package for constructing and analysing graphs for landscape genetics in R. Methods Ecol Evol 12(3):539–547

Schoville SD, Dalongeville A, Viennois G, Gugerli F, Taberlet P, Lequette B, Alvarez N, Manel S (2018) Preserving genetic connectivity in the European Alps protected area network. Biol Conserv 218:99–109

Shirk A, Cushman S (2011) sGD: software for estimating spatially explicit indices of genetic diversity. Mol Ecol Resour 11(5):922–934

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

Sonneck A-G, Bönsel A, Matthes J (2008) Der einfluss von landnutzung auf die habitate von stethophyma grossum (linnaeus, 1758) an beispielen aus mecklenburg-vorpommern. Articulata 23:15–30

Spielman D, Brook BW, Frankham R (2004) Most species are not driven to extinction before genetic factors impact them. Proc Natl Acad Sci 101(42):15261–15264

Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP (2010) Landscape genetics: where are we now? Mol Ecol 19(17):3496–3514

Szpiech ZA, Jakobsson M, Rosenberg NA (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. Bioinformatics 24 (21):2498–2504

Taylor Z, Hoffman S (2014) Landscape models for nuclear genetic diversity and genetic structure in white-footed mice (Peromyscus leucopus). Heredity 112 (6):588–595

Tenenhaus M (1998) La régression PLS: théorie et pratique. Editions TECHNIP

Toma Y, Imanishi J, Yokogawa M, Hashimoto H, Imanishi A, Morimoto Y, Hatanaka Y, Isagi Y, Shibata S (2015) Factors affecting the genetic diversity of a perennial herb Viola grypoceras A. Gray var. grypoceras in urban fragmented forests. Landsc Ecol 30(8):1435–1447

Tournant P, Afonso E, Roué S, Giraudoux P, Foltête J-C (2013) Evaluating the effect of habitat connectivity on the distribution of lesser horseshoe bat maternity roosts using landscape graphs. Biol Conserv 164:39–49

Trautner J, Hermann G (2008) Die Sumpfschrecke (Stethophyma grossum L., 1758) im Aufwind-Erkenntnisse aus dem zentralen Baden-Württemberg. Articulata 23 (2):37–52

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

van Strien MJ (2017) Consequences of population topology for studying gene flow using link-based landscape genetic methods. Ecol Evol 7(14): 5070–5081

van Strien MJ, Keller D, Holderegger R, Ghazoul J, Kienast F, Bolliger J (2014) Landscape genetics as a tool for conservation planning: predicting the effects of landscape change on gene flow Ecol Appl 24(2):327–339

Varvio S-L, Chakraborty R, Nei M (1986) Genetic variation in subdivided populations and conservation genetics. Heredity 57(2):189–198

Villard M-A, Metzger JP (2014) Beyond the fragmentation debate: a conceptual model to predict when habitat configuration really matters. J Appl Ecol 51 (2):309–318

Wagner HH, Fortin M-J (2013) A conceptual framework for the spatial analysis of landscape genetic data. Conserv Genet 14(2):253–261

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38(6):1358–1370

Wold S, Sjöström M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. Chemometr Intelig Lab Syst 58(2):109–130

Zetterberg A, Mörtberg UM, Balfors B (2010) Making graph theory operational for landscape ecological assessments, planning, and design. Landsc Urban Plan 95 (4):181–191

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

JCF, SG and HM obtained the funding for the project. MJVS provided the empirical and spatial data. GV and PS performed the statistical analyses. PS designed the study, wrote the first version of the article and answered to the reviewers. All co-authors contributed to the article writing.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41437-021-00495-w.

Correspondence and requests for materials should be addressed to Paul Savary.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.