



**HAL**  
open science

## **graph4lg: a package for constructing and analysing graphs for landscape genetics in R**

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

► **To cite this version:**

Paul Savary, Jean-Christophe Foltête, Hervé Moal, Gilles Vuidel, Stéphane Garnier. graph4lg: a package for constructing and analysing graphs for landscape genetics in R. *Methods in Ecology and Evolution*, 2021, 12, pp.539-547. 10.1111/2041-210X.13530 . hal-03030410

**HAL Id: hal-03030410**

**<https://hal.science/hal-03030410v1>**

Submitted on 30 Nov 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

graph4lg: a package for constructing and analysing graphs for  
landscape genetics in R

Savary, Paul<sup>\*1, 2, 3</sup>, Foltête, Jean-Christophe<sup>2</sup>, Moal, Hervé<sup>1</sup>, Vuidel, Gilles<sup>2</sup>, and  
Garnier, Stéphane<sup>3</sup>

<sup>1</sup>*ARP-Astrance, 9 Avenue Percier, 75008 Paris, France*

<sup>2</sup>*ThéMA, UMR 6049 CNRS, Université Bourgogne-Franche-Comté, 32 Rue Mégevand, 25030 Besançon  
Cedex, France*

<sup>3</sup>*Biogéosciences, UMR 6282 CNRS, Université Bourgogne-Franche-Comté, 6 Boulevard Gabriel 21000  
Dijon, France*

---

\*Corresponding author: [paul.savary@univ-fcomte.fr](mailto:paul.savary@univ-fcomte.fr)

## Abstract

1 - In landscape genetics, habitat connectivity and population genetic structure have been analysed using graph-theoretic approaches to understand how landscape features influence demography (i.e. dispersal and population size).

2 - Despite substantial advances in enhancing both genetic and landscape graph use, a software tool bringing together a large range of construction and analysis parameters for these two types of graphs was lacking in the landscape genetic toolbox. Moreover, although these two types of graphs appear complementary for answering landscape genetic questions, methods for comparing them have not been forthcoming.

3 - We have developed an R package to improve and encourage the use of these graphs. It includes functions for converting and importing genetic data and for genetic distance computing. It also implements time-efficient geodesic and cost-distance calculations from spatial data. A large range of parameters can be used to create genetic and landscape graphs from these data, including several graph pruning methods. We made available to R users the command-line facilities of Graphab software to easily model landscape graphs in R. The package functions perform preliminary analysis to adapt methodological choices to research questions. Landscape and genetic graphs created can be analysed with node-level metrics as well as link-level and modularity analyses. Users can compare and visualise these graphs and export them to shapefiles to facilitate interpretation and subsequent analyses.

4 - `graph4lg` contributes to expanding landscape and genetic graph potential for analysing ecological connectivity while encouraging further investigations on methodological implications related to these tools.

1 - En génétique du paysage, la connectivité des habitats et la structure génétique des populations ont été analysées à l'aide d'approches basées sur la théorie des graphes pour comprendre comment les éléments du paysage influencent la démographie (i.e. la taille des populations et la dispersion).

2 - Bien que des progrès conséquents aient amélioré l'utilisation des graphes génétiques et paysagers, un outil informatique réunissant une large gamme de paramètres de construction et d'analyse de ces deux types de graphes faisait défaut parmi les outils de la génétique du paysage. Par ailleurs, malgré l'intérêt potentiel de la complémentarité de ces deux types de graphes pour répondre aux questions de génétique du paysage, des méthodes permettant de les comparer n'ont pas encore été proposées.

3 - Nous avons développé un package R pour améliorer et encourager l'utilisation de ces graphes. Il intègre des fonctions de conversion et d'import de données génétiques et de calcul de distances génétiques. Il permet aussi de calculer des distances géodésiques et des distances-coût à partir de données spatiales. Une importante gamme de paramètres peut être utilisée pour créer des graphes génétiques et paysagers à partir de ces données, parmi laquelle on trouve plusieurs méthodes d'élagage. Nous avons rendu accessible l'utilisation du logiciel Graphab en lignes de commande aux utilisateurs du logiciel R pour faciliter la modélisation de graphes paysagers dans cet environnement. Les fonctions du package permettent des analyses préliminaires ayant pour but d'adapter les choix méthodologiques aux questions de recherche. Les graphes génétiques et paysagers créés peuvent être analysés avec des métriques calculées au niveau des nœuds ou avec des analyses des liens ou des modules de ces graphes. Les utilisateurs peuvent comparer et visualiser les graphes et les exporter sous forme de couches shapefile pour faciliter l'interprétation et les analyses ultérieures.

4 - `graph4lg` contribue à étendre le potentiel des graphes génétiques et paysagers pour l'analyse de la connectivité écologique tout en encourageant de futures recherches sur les aspects méthodologiques relatifs à ces outils.

**Keywords:** ecological connectivity, dispersal, graph theory, landscape genetics, R

# 1 Introduction

Landscape genetic studies aim at understanding how landscape characteristics such as habitat spatial distribution and matrix quality shape population genetic structure (Balkenhol et al., 2016; Manel et al., 2003). In this recent field, new methods have been developing at a sustained pace to describe landscape structure (e.g. Galpern et al. (2012)), population genetic structure (e.g. Al-Asadi et al. (2019); Prunier et al. (2017); Shirk and Cushman (2011)) and to bring landscape and genetic data together in multi-level analyses (Hall and Beissinger, 2014; Wagner and Fortin, 2013). Among these methods, graph-theoretic approaches were identified as particularly relevant (Manel and Holderegger, 2013) because they grasp interactions between sets of habitat patches or populations in a comprehensive way (Dale and Fortin, 2010; Dale, 2017; Fortin et al., 2012).

A graph is basically a set of nodes connected by a set of links. In landscape ecology, the use of landscape graphs dates back to the 2000s (Galpern et al., 2011; Urban and Keitt, 2001) and has developed through the availability of free software tools such as Graphab (Foltête et al., 2012) or Conefor (Saura and Torne, 2009). The nodes of landscape graphs are habitat patches and their links correspond to potential dispersal paths, e.g. identified by computing least-cost paths across resistance surfaces. From these graphs, a large range of connectivity metrics can be computed (Rayfield et al., 2011) and used for inference (Pereira et al., 2011) or conservation-oriented decision making (Foltête et al., 2014). Besides, landscape graph nodes can be partitioned through modularity analyses to identify management units or perform analyses at a coarser grain (Fletcher Jr et al., 2013; Foltête and Vuidel, 2017).

Although landscape graphs enable close investigation of habitat connectivity, their construction often relies upon expert-based opinion and combining them with biological data can improve this approach (Foltête et al., 2020). Landscape graph modelling software tools already make possible the import of biological data (Foltête et al., 2012) and genetic data are a relevant candidate for such a combination (Luque et al., 2012). Thus, this approach would benefit from being performed in a statistical software where genetic data processing and complex statistical analyses can both be done.

Similarly, migration models theorised in population genetics (Kimura and Weiss, 1964; Wright, 1931) often rely upon topological network representations and even if these models rarely reflect real situations (Greenbaum and Fefferman, 2017; Milligan et al., 2018), population geneticists have developed a large range of genetic graph construction methods that can potentially fit all the observed migration networks (Greenbaum and Fefferman, 2017). Thus, population genetic structure has been frequently represented as genetic graphs in which nodes correspond to sampled populations and links to substantial gene exchanges between them (Arnaud, 2003; Excoffier et al., 1992).

When building a genetic graph, the construction method should always be guided by the specific research question (Miele et al., 2019). For example, an important step in this process is graph pruning, which consists in removing some links and should be performed differently if the aim of the analysis is (i) to identify single generation (direct) dispersal paths (Boulanger et al., 2020; Dyer, 2015) or (ii) to infer landscape effects on dispersal from the genetic differentiation measurements between populations connected on the graph (Savary, P. et al., in correction; Van Strien (2017)). Indeed, in the first case, paths that are not within reach of individuals given their dispersal capacities should be removed in order to represent the dispersal network topology.

In the second case, links corresponding to multi-generational indirect dispersal can be conserved given that they reflect the genetic connectivity emerging over generations due to stepping-stone dispersal (Boulanger et al., 2020; Saura et al., 2014). In both cases, several graph pruning methods can be used and must be chosen accordingly (Greenbaum and Fefferman, 2017; Van Strien, 2017). Apart from these link-level analyses, once genetic graphs have been constructed, they can be analysed at the node- and boundary-levels (Wagner and Fortin, 2013), according to the research question. For example, node-level metric calculations and module partitions are possible to assess genetic diversity and relative genetic differentiation at the population level (e.g. Koen et al. (2016)) and to identify population clustering patterns (e.g. Fortuna et al. (2009)), respectively.

Then, comparing genetic graph characteristics such as node-level indices, graph topology, link weights and module partitions to the exact same characteristics derived from a landscape graph could contribute to a better understanding of the effect of habitat spatial patterns on population genetic structure. It has therefore been claimed that genetic graphs and landscape graphs should complement each other (Foltête and Vuidel, 2017; Galpern et al., 2011; Manel and Holderegger, 2013; Murphy et al., 2015). However, a practical tool for building, analysing and comparing these graphs was still lacking. Accordingly, we have developed the R package `graph4lg` to bridge all these gaps in the implementation of landscape and genetic graphs. It provides graph users with a software tool facilitating their choice and implementation of graph construction and analysis methods and builds on previous developments of R packages for landscape genetic and graph-theoretical analyses such as `igraph` (Csardi and Nepusz, 2006), `gstudio` (Dyer, 2014) or `adegenet` (Jombart, 2008).

## 2 Workflow

`graph4lg` package functions can be divided into four categories following the steps of landscape genetic analyses:

1. The package allows for genetic and spatial input data processing in preparation for graph construction, calculating intra-population genetic indices and inter-population genetic and landscape distances and performing preliminary analyses through diagnostic plots.
2. It provides users with functions for building genetic graphs and analysis tools for these graphs.
3. In parallel, some 'wrapper functions' run command-line functionalities of Graphab software (Foltête et al., 2012) directly from R to construct and analyse landscape graphs.
4. Finally, genetic and landscape graphs can be compared, plotted and exported to other formats.

This workflow is described in the following sections and depicted in Figure 1. All the package functions are also listed in Table S1 with their dependencies on other R packages.

### 2.1 Input data processing

#### 2.1.1 Genetic data

Studies of gene flow pattern and/or intensity and of landscape influence on it rely upon neutral genetic markers which reflect genetic variation due to demographic changes and are supposedly independent from adaptive processes (Holderegger et al., 2006). Microsatellite loci assumed or identified as being neutral are the most frequent type of markers used in landscape

genetics (Storfer et al., 2010). Besides, SNP markers are now becoming widespread and can also be used to perform these analyses provided that loci under selective pressures (identified as outliers) have been discarded (Cushman et al., 2018; Foll and Gaggiotti, 2008). Accordingly, `graph4lg` functions use genetic data with 2- or 3-digit allele coding to fit the common microsatellite coding. SNP data can also be used when loaded as `genind` object in R, because R objects with the `genind` class attribute from the `adegenet` package (Jombart, 2008) are the input of genetic data processing functions from `graph4lg`. Landscape genetic analyses performed with `graph4lg` are population-based and like most applications in this field rely on the a priori delimitation of populations (Milligan et al., 2018; Waits and Storfer, 2015). Populations are identified by the `pop` strata of `genind` objects and usually correspond to sampling units.

We included conversion functions (`gstud_to_genind`, `loci_to_genind`, `structure_to_genind`, `genepop_to_genind`) to easily get `genind` objects from formats used in other software tools such as `gstudio` (Dyer, 2014), `pegas` (Paradis, 2010), `STRUCTURE` (Pritchard et al., 2000) or `GENEPOP` (Raymond, 1995). We also made possible the creation of external text files in `GENEPOP` format from `genind` objects (`genind_to_genepop`) for users willing to perform analyses with this commonly used R package and executable software.

The `mat_gen_dist` function computes eight different inter-population genetic distances from `genind` objects (Table 1). However, 'external' genetic distance matrices or `dist` objects imported by users can be the input of the functions described in the next sections.

### 2.1.2 Spatial data

Two kinds of pairwise landscape distance matrices can be computed from point spatial coordinates and resistance surface raster layers:

- `mat_geo_dist` function computes geodesic distances from point sets with either projected or polar coordinate reference systems using Euclidean distance or great circle distance formulas, respectively.
- `mat_cost_dist` function computes pairwise cost-distance matrices from a point set, a categorical resistance surface raster layer and a `data.frame` indicating the cost associated with each cell value. This function depends on `gdistance` package (Van Etten, 2012), but can also use an external `.jar` file which substantially reduces computation times for large rasters (Table S2), providing R users with a time-efficient alternative to `gdistance` for cost-distance computing.

### 2.1.3 Preliminary analyses

When the study objective is to infer landscape effects on dispersal from the relationship between genetic and landscape distances associated with graph links, visualising a scatterplot using complete distance matrices (`scatter_dist`) can be a first step before graph construction. Isolation by distance patterns due to limited dispersal are common in population genetics (Wright, 1943). However, if the studied species has low dispersal capacities or after a decrease in landscape connectivity, the increase of genetic differentiation with distance is only observed at a small scale, which tends to expand over time (Slatkin, 1993). In that case, drift will be more important than migration as a driver of genetic differentiation between populations separated by large distances. This results in a non-linear relationship between landscape and genetic distances exhibiting a plateau beyond a given landscape distance threshold (Hänfling and Weetman, 2006; Hutchison and Templeton, 1999). Conversely, a linear relationship is expected when equilibrium is established at the scale of the study area. Because migration-drift equilibrium is a pre-requisite for genetic differentiation to reflect landscape effects on dispersal, inference ignoring it may be

biased (Bradbury and Bentzen, 2007; Van Strien et al., 2015). Genetic graph pruning method determines population pairs included in the inference and should therefore be chosen after consideration of the scale at which populations verify this equilibrium. Similarly, genetic distances based on fixation indices require equilibrium assumptions to be confirmed so that derived inferences are reliable (Neigel, 2002; Whitlock and Mccauley, 1999).

Van Strien et al. (2015) estimated the threshold distance between population pairs maximizing the correlation between genetic differentiation and landscape distance. This distance of maximum correlation (DMC) can be viewed as an estimate of the scale at which populations verify migration-drift equilibrium, i.e. their neighborhood size (Addicott et al., 1987; Kierepka et al., 2020). It is computed by the `dist_max_corr` function (Figure 2A), which can help choosing among pruning methods (cf. section 2.2.1).

When the objective is to recover direct dispersal paths by taking into account landscape resistance and maximum dispersal capacities of the study species, it is important to know how species dispersal distances expressed in geodesic distance units convert into cost-distance values, especially if geodesic distance thresholds are used to prune the graphs. To that purpose, the `convert_cd` function performs a linear or log-log linear regression of cost-distance values against geodesic distance values (Tournant et al., 2013).

## 2.2 Genetic graph construction and analysis

### 2.2.1 Genetic graph construction

The `graph4lg` package implements a wide range of pruning methods for constructing genetic graphs. Some of these methods can equally apply to landscape graphs. First, graphs can be pruned by removing the links between populations separated by genetic or landscape distances above or below a specified threshold value (`gen_graph_thr`). Such an approach can be efficient for identifying direct dispersal paths provided maximum dispersal distance is known. It has also been used to select population pairs to include in the inference of landscape effects on dispersal (Angelone et al., 2011; Keller et al., 2013). In that case, the distance threshold can be the DMC if a clear plateau is identified in the IBD pattern.

Second, when the study species is assumed to have stepping stone dispersal or when migration-drift equilibrium establishes at short distance, graphs can be pruned depending on topological constraints (`gen_graph_topo`), thereby constructing minimum spanning trees, planar graphs,  $k$ -nearest-neighbour graphs and Gabriel graphs (Arnaud, 2003; Bunn et al., 2000; Keller et al., 2013; Naujokaitis-Lewis et al., 2013). The topological constraints can be applied to matrices of landscape distances as well as genetic distances. Similarly, the edge-thinning method, linked to percolation theory, identifies the distance threshold above which graph thresholding breaks the graph into more than one connected component (Urban and Keitt, 2001; Rozenfeld et al., 2008) and creates a thresholded graph using this threshold.

Finally, the `gen_graph_indep` function creates genetic graphs directly from genetic data stored in `genind` objects, in the same way as the `popgraph` function from the `popgraph` package. This approach prunes graphs by conserving links between populations that are dependent on each other based on the covariance of their allelic frequencies, after having looked at the covariance with allelic frequencies from all the other populations. This use of the conditional independence principle (Whittaker, 2009) is supposed to conserve links between populations directly exchanging migrants through single generation dispersal events (Dyer and Nason, 2004). This function expands the original `popgraph` function by implementing  $p$ -value adjustments

(Benjamini and Hochberg, 1995; Holm, 1979), among other options compared by Savary et al. (in correction).

### 2.2.2 Genetic graph analyses

Once genetic graphs have been created, the `compute_node_metric` function computes graph-theoretic metrics such as the degree, closeness and betweenness centrality indices, which identify keystone hubs of genetic connectivity (Cross et al., 2018). It also computes the average and sum of the inverse genetic distance weighting the links. Koen et al. (2016) showed that these relative genetic differentiation indices can be good proxies of connectivity. Apart from these metrics depending on graph topology, population-level genetic diversity indices such as allelic richness and heterozygosity rates can be computed with the `pop_gen_index` function from `genind` objects. All these metrics can be added as node attributes with the `add_nodes_attr` function.

Link weights are used to partition nodes into modules (`compute_graph_modul`) and identify population clusters possibly delineated by sharp dispersal limitations (Fortuna et al., 2009; Garroway et al., 2008). We implement several modularity algorithms: `fast_greedy` (Clauset et al., 2004), `louvain` (Blondel et al., 2008), `optimal` (Brandes et al., 2008) and `walktrap` (Pons and Latapy, 2006) from the `igraph` package. Link weights can also be exported into data frames for subsequent link-level analyses using the `graph_to_df` function.

## 2.3 Landscape graph construction and analysis

The `graph4lg` package integrates the graph construction and analysis options of Graphab software (Foltête et al., 2012) by implementing its command-line functionalities. Thus, both the package documentation and Graphab software manual can be of substantial help for users.

First, the `graphab_project` function creates a Graphab project from a categorical raster layer. It defines habitat patches as contiguous cells with a given cell value and creates a directory containing this project in the user's machine. Then, the least cost paths between these habitat patches are computed (`graphab_link`). Once the Graphab project and link set have been created, users can create complete, thresholded or planar graphs (`graphab_graph`). A large range of connectivity metrics can be computed at the graph or node levels (`graphab_metric`). Delta-metrics can also be computed, e.g. for prioritisation analyses. These metrics have been extensively tested and compared in the literature (Baranyi et al., 2011; Rayfield et al., 2011).

Users can either import planar graphs created in Graphab as `igraph` objects (`graphab_to_igraph`) in order to compute metrics in R, or only import link weights or node-level metrics computed with Graphab (`get_graphab_metric`, `get_graphab_linkset`). In cases when users want to relate punctual field observations to connectivity metrics, they can get the metrics of the nearest habitat patches from a set of points (`graphab_pointset`). Finally, users can partition habitat patches through modularity analyses in Graphab (`graphab_modul`).

## 2.4 Landscape and genetic graph comparisons

Although landscape and genetic graphs have each been repeatedly used, their direct comparison has rarely been performed (Draheim et al., 2016; Schoville et al., 2018). To facilitate the interpretation of their respective topology and the formulation and test of hypotheses regarding their similarities, the `plot_graph_lg` function allows users to visualise the topology and connectivity of the created graphs (Figure 2D). It maps graphs in a spatially-explicit way or implements an attraction-repulsion algorithm based on link weights (Fruchterman and Reingold, 1991) to assess whether nodes cluster together independently from their spatial locations. Node metrics,



link weights and module partitioning can also be visualised with this function. Moreover, the link weight distribution can be plotted as a histogram (`plot_hist_w`) (Fig. 2C) and the pruning intensity can be visualised by representing population pairs in a different color on the scatter plot relating genetic distance with landscape distance (`scatter_dist_g`).

To test the formulated hypotheses, when landscape and genetic graphs share the same nodes, the correlation between population- or patch-based indices can be assessed (`graph_node_compar`) in order to understand the relationship between i) landscape attributes (habitat surface area or connectivity) and ii) genetic attributes (genetic differentiation or local diversity).

Similarly, the `graph_topo_compar` function compares the topologies of two graphs (link-based analysis) by creating a contingency table (based on the presence/absence of corresponding links in both graphs) and computing indices commonly used to assess classifications (e.g. Matthews' correlation coefficient, false discovery rate, accuracy). The congruence of two graph topologies can be visualised by plotting them on the same map while colouring their links to indicate whether they occur in both graphs or just one of them (`graph_plot_compar`). Mismatches between genetic and landscape graph topology can provide insights regarding the realised connectivity in the study area or the modelisation method itself.

Finally, quantifying how many node pairs classified in the same module in one graph are also classified together in the modules created from another graph indicates us how far these graphs reflect the same real-world situation. This boundary-based analysis is made possible by the `graph_modul_compar` function which computes the Adjusted Rand Index (Hubert and Arabie, 1985) to compare partitions.

### 3 Export facilities and included data

The graphs can be exported to shapefile layers for integration into a GIS (`graph_to_shp`) and analyses involving other types of geographical data. Included genetic and spatial data can be used to discover the functionalities of the package. We included a data set created from simulations with CDPOP (Landguth and Cushman, 2010) on a simulated landscape (`data_simul_genind`). It consists of 1500 individuals from 50 populations genotyped at 20 microsatellite loci. This dataset exhibits a typical type-IV pattern of IBD and was used to create Figure 2 and the tutorial presenting the package.

### 4 Limits and conclusion

Landscape and genetic graphs have great potential for analysing ecological connectivity and we do not claim to have compiled an exhaustive set of graph construction and analysis methods. Other pruning methods have been developing (Brooks, 2006; Greenbaum et al., 2016; Kininmonth et al., 2010; Milligan, prep; Peterson et al., 2018) and could be expanded to directed graphs for example. Besides, a large range of local metrics inspired from the metapopulation theory have been developed for landscape graphs and could similarly inspire genetic graph local metrics. Graph-based analyses could also benefit from significance testing approaches through permutations to enhance their robustness. Finally, although Graphab software can efficiently handle very large spatial data sets (Foltête et al., 2012), genetic graph modelling can involve higher computational costs, thereby limiting this approach to smaller spatial and genetic datasets. Further development of the package could introduce improvements.

Our first goal in developing `graph4lg` is to bring together and make accessible a large range of methods currently used in landscape genetics for constructing and analysing graphs. We hope this package will foster the use of genetic and landscape graphs as well as further investigation regarding theoretical as well as methodological aspects.

## 5 Acknowledgements

We thank the editor and referees, as well as the CRAN volunteers for their constructive comments. 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). We are particularly grateful to ARP-Astrance teams for their support throughout the project. We thank Ahmed Jebrane, Catherine Labruere and Catherine Laredo for their help with mathematical aspects.

## 6 Data availability

No empirical data were used in this article. The `graph4lg` package, all the R source codes, documentation files and vignettes can be downloaded by users under the GPL-2 license from the CRAN repository (<https://cran.r-project.org/web/packages/graph4lg/index.html>).

## 7 Authors' contributions

J.C.F., S.G. and H.M. obtained the funding for the project. P.S. and G.V. designed the package and developed the codes. P.S, J.C.F. and S.G. wrote the manuscript with significant contributions and remarks from all co-authors.

## References

- Addicott, J. F., Aho, J. M., Antolin, M. F., Padilla, D. K., Richardson, J. S., and Soluk, D. A. (1987). Ecological neighborhoods: scaling environmental patterns. *Oikos*, pages 340–346.
- Al-Asadi, H., Petkova, D., Stephens, M., and Novembre, J. (2019). Estimating recent migration and population-size surfaces. *PLoS genetics*, 15(1):e1007908.
- 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.
- Balkenhol, N., Cushman, S., Storfer, A., and Waits, L. (2016). *Landscape genetics: concepts, methods, applications*. John Wiley & Sons.
- Baranyi, G., Saura, S., Podani, J., and Jordán, F. (2011). Contribution of habitat patches to network connectivity: redundancy and uniqueness of topological indices. *Ecological Indicators*, 11(5):1301–1310.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, pages 289–300.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics - Theory and Experiment*, 10.
- Boulanger, E., Dalongeville, A., Andrello, M., Mouillot, D., and Manel, S. (2020). Spatial graphs highlight how multi-generational dispersal shapes landscape genetic patterns. *Ecography*, 15(1):1–13.

- 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.
- Bradbury, I. R. and Bentzen, P. (2007). Non-linear genetic isolation by distance: implications for dispersal estimation in anadromous and marine fish populations. *Marine Ecology Progress Series*, 340:245–257.
- Brandes, U., Delling, D., Gaertler, M., Gorke, R., Hofer, M., Nikoloski, Z., and Wagner, D. (2008). On modularity clustering. *IEEE transactions on knowledge and data engineering*, 20(2):172–188.
- Brooks, C. P. (2006). Quantifying population substructure: extending the graph-theoretic approach. *Ecology*, 87(4):864–872.
- Bunn, A., Urban, D., and Keitt, T. (2000). Landscape connectivity: a conservation application of graph theory. *Journal of environmental management*, 59(4):265–278.
- Clauset, A., Newman, M. E., and Moore, C. (2004). Finding community structure in very large networks. *Physical review E*, 70(6).
- Cross, T. B., Schwartz, M. K., Naugle, D. E., Fedy, B. C., Row, J. R., and Oyler-McCance, S. J. (2018). The genetic network of greater sage-grouse: Range-wide identification of keystone hubs of connectivity. *Ecology and Evolution*, 8(11):1–19.
- Csardi, G. and Nepusz, T. (2006). The igraph software package for complex network research. *International Journal of Complex Systems*, 1695(5):1–9.
- Cushman, S. A., Shirk, A., Howe, G. T., Dyer, R. J., Murphy, M. A., and Joost, S. (2018). The least cost path from landscape genetics to landscape genomics: challenges and opportunities to explore ngs data in a spatially explicit context. *Frontiers in Genetics*, 9:215.
- Dale, M. and Fortin, M.-J. (2010). From graphs to spatial graphs. *Annual Review of Ecology, Evolution, and Systematics*, 41.
- Dale, M. R. (2017). *Applying Graph Theory in Ecological Research*. Cambridge University Press.
- Dieringer, D. and Schlötterer, C. (2003). Microsatellite analyser (msa): a platform independent analysis tool for large microsatellite data sets. *Molecular ecology notes*, 3(1):167–169.
- 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. (2014). R package gstudio: analyses and functions related to the spatial analysis of genetic marker data. *R package version*, 1.
- Dyer, R. J. (2015). 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.
- 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 Jr, R. J., Revell, A., Reichert, B. E., Kitchens, W. M., Dixon, J. D., and Austin, J. D. (2013). Network modularity reveals critical scales for connectivity in ecology and evolution. *Nature communications*, 4:ncomms3572.
- Foll, M. and Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a bayesian perspective. *Genetics*, 180(2):977–993.
- 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., Girardet, X., and Clauzel, C. (2014). A methodological framework for the use of landscape graphs in land-use planning. *Landscape and Urban Planning*, 124:140–150.

- Foltête, J.-C., Savary, P., Clauzel, C., Bourgeois, M., Girardet, X., Saharoui, Y., Vuidel, G., and Garnier, S. (2020). Coupling landscape graph modeling and biological data: a review. *Landscape Ecology*, pages 1–18.
- Foltête, J.-C. and Vuidel, G. (2017). Using landscape graphs to delineate ecologically functional areas. *Landscape Ecology*, 32(2):249–263.
- Fortin, M.-J., James, P. M., MacKenzie, A., Melles, S. J., and Rayfield, B. (2012). Spatial statistics - spatial regression, and graph theory in ecology. *Spatial Statistics*, 1:100–109.
- Fortuna, M. A., Albaladejo, R. G., Fernández, L., Aparicio, A., and Bascompte, J. (2009). Networks of spatial genetic variation across species. *Proceedings of the National Academy of Sciences*, 106(45):19044–19049.
- Fruchterman, T. M. and Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and experience*, 21(11):1129–1164.
- Galpern, P., Manseau, M., and Fall, A. (2011). Patch-based graphs of landscape connectivity: a guide to construction, analysis and application for conservation. *Biological conservation*, 144(1):44–55.
- Galpern, P., Manseau, M., and Wilson, P. (2012). Grains of connectivity: analysis at multiple spatial scales in landscape genetics. *Molecular Ecology*, 21(16):3996–4009.
- Garroway, C. J., Bowman, J., Carr, D., and Wilson, P. J. (2008). Applications of graph theory to landscape genetics. *Evolutionary Applications*, 1(4):620–630.
- 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.
- Hall, L. A. and Beissinger, S. R. (2014). A practical toolbox for design and analysis of landscape genetics studies. *Landscape ecology*, 29(9):1487–1504.
- 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.
- Hedrick, P. W. (2005). A standardized genetic differentiation measure. *Evolution*, 59(8):1633–1638.
- Holderegger, R., Kamm, U., and Gugerli, F. (2006). Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecology*, 21(6):797–807.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70.
- Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of classification*, 2(1):193–218.
- 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.
- Jombart, T. (2008). adegenet: a r package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11):1403–1405.
- Jost, L. (2008). Gst and its relatives do not measure differentiation. *Molecular ecology*, 17(18):4015–4026.
- 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.
- Kierepka, E. M., Anderson, S. J., Swihart, R. K., and Rhodes, O. E. (2020). Differing, multiscale landscape effects on genetic diversity and differentiation in eastern chipmunks. *Heredity*, 124(3):457–468.
- 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.
- Kininmonth, S., van Oppen, M. J., and Possingham, H. P. (2010). Determining the community structure of the coral seriatopora hystrix from hydrodynamic and genetic networks. *Ecological Modelling*, 221(24):2870–2880.

- Koen, E. L., Bowman, J., and Wilson, P. J. (2016). Node-based measures of connectivity in genetic networks. *Molecular ecology resources*, 16(1):69–79.
- Landguth, E. L. and Cushman, S. (2010). Cdpop: a spatially explicit cost distance population genetics program. *Molecular Ecology Resources*, 10(1):156–161.
- Luque, S., Saura, S., and Fortin, M.-J. (2012). Landscape connectivity analysis for conservation: insights from combining new methods with ecological and genetic data. *Landscape ecology*, 27(2):153–157.
- 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.
- Miele, V., Matias, C., Robin, S., and Dray, S. (2019). Nine quick tips for analyzing network data. *PLOS Computational Biology*, 15(12):e1007434.
- Milligan, B. G. (in prep). Probabilistic graph models for landscape genetics. *PeerJ Preprints*.
- Milligan, B. G., Archer, F. I., Ferchaud, A.-L., Hand, B. K., Kierepka, E. M., and Waples, R. S. (2018). Disentangling genetic structure for genetic monitoring of complex populations. *Evolutionary Applications*.
- Murphy, M., Dyer, R., and Cushman, S. A. (2015). Graph theory and network models in landscape genetics. In Balkenhol, N., Cushman, S., Storer, A., and Waits, L., editors, *Landscape genetics: Concepts, methods, applications*, pages 165–180. John Wiley & Sons, 1 edition.
- 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.
- Paradis, E. (2010). pegas: an r package for population genetics with an integrated–modular approach. *Bioinformatics*, 26(3):419–420.
- 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*.
- Pereira, M., Segurado, P., and Neves, N. (2011). Using spatial network structure in landscape management and planning: a case study with pond turtles. *Landscape and urban planning*, 100(1):67–76.
- Peterson, E. E., Hanks, E. M., Hooten, M. B., Ver Hoef, J. M., and Fortin, M.-J. (2018). Spatially-structured statistical network models for landscape genetics. *Ecological Monographs*.
- Pons, P. and Latapy, M. (2006). Computing communities in large networks using random walks. *J. Graph Algorithms Appl.*, 10(2):191–218.
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–959.
- 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.
- Rayfield, B., Fortin, M.-J., and Fall, A. (2011). Connectivity for conservation: a framework to classify network measures. *Ecology*, 92(4):847–858.
- Raymond, M. (1995). Genepop: Population genetics software for exact tests and ecumenism. vers. 1.2. *Journal of Heredity*, 86:248–249.
- 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.
- Saura, S., Bodin, Ö., and Fortin, M.-J. (2014). Stepping stones are crucial for species' long-distance dispersal and range expansion through habitat networks. *Journal of Applied Ecology*, 51(1):171–182.
- Saura, S. and Torne, J. (2009). Conefor sensinode 2.2: a software package for quantifying the importance of habitat patches for landscape connectivity. *Environmental modelling & software*, 24(1):135–139.
- Schoville, S. D., Dalongeville, A., Viennois, G., Gugerli, F., Taberlet, P., Lequette, B., Alvarez, N., and Manel, S. (2018). Preserving genetic connectivity in the european alps protected area network. *Biological Conservation*, 218:99–109.
- Shirk, A. and Cushman, S. (2011). sgd: software for estimating spatially explicit indices of genetic diversity. *Molecular Ecology Resources*, 11(5):922–934.
- Shirk, A., Landguth, E., and Cushman, S. (2017). A comparison of individual-based genetic distance metrics for landscape genetics. *Molecular Ecology*, 17(6).
- Slatkin, M. (1993). Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, 47(1):264–279.
- 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.
- Tournant, P., Afonso, E., Roué, S., Giraudoux, P., and Foltête, J.-C. (2013). Evaluating the effect of habitat connectivity on the distribution of lesser horseshoe bat maternity roosts using landscape graphs. *Biological conservation*, 164:39–49.
- Urban, D. and Keitt, T. (2001). Landscape connectivity: a graph-theoretic perspective. *Ecology*, 82(5):1205–1218.
- Van Etten, J. (2012). R package gdistance: distances and routes on geographical grids (version 1.1-4). *Journal of Statistical Software*, 76:13.
- 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.
- Waits, L. P. and Storfer, A. (2015). Basics of population genetics: quantifying neutral and adaptive genetic variation for landscape genetic studies. In Balkenhol, N., Cushman, S., Storfer, A., and Waits, L., editors, *Landscape genetics: Concepts, methods, applications*, pages 35–57. John Wiley & Sons.
- Weir, B. S. and Cockerham, C. C. (1984). Estimating f-statistics for the analysis of population structure. *evolution*, 38(6):1358–1370.
- Whitlock, M. C. and Mccauley, D. E. (1999). Indirect measures of gene flow and migration:  $F_{st} \neq 1/(4nm + 1)$ . *Heredity*, 82(2):117–125.
- Whittaker, J. (2009). *Graphical models in applied multivariate statistics*. Wiley Publishing.
- Wright, S. (1931). Evolution in mendelian populations. *Genetics*, 16(2):97.
- Wright, S. (1943). Isolation by distance. *Genetics*, 28(2):114.

## 8 Tables and figures

Genetic distance	Description	Eq.	Depend.	Ref.
$F_{ST}$	Fixation index	Yes	diveRsity	Weir and Cockerham (1984)
Linearised $F_{ST}$	Linearised fixation index	Yes	diveRsity	Rousset (1997)
$G'_{ST}$	Standardised fixation index	Yes	diveRsity	Hedrick (2005)
$D_{Jost}$	Standardised fixation index	Yes	diveRsity	Jost (2008)
$D_{PS}$	1 - proportion of shared alleles	No	None	Bowcock et al. (1994), implementation of MSA software formula (Dieringer and Schlötterer, 2003)
Euclidean genetic distance	Computed from allelic frequencies differences	No	None	Excoffier et al. (1992)
Weighted Euclidean genetic distance	Computed from allelic frequencies differences giving more weights to rare alleles	No	None	Fortuna et al. (2009); Greenbaum et al. (2016)
PCA-derived Euclidean genetic distance	Inter-population distance in the space defined by the principal components obtained from a PCA of the allelic frequencies table	No	None	Inspired by the distances computed by Paschou et al. (2014) and Shirk et al. (2017)
popgraph-derived genetic distance	Inter-population distance in the space created after a PCA of the allelic frequencies table	No	None	Dyer and Nason (2004)

Table 1: Inter-population genetic distances computed with the `mat_gen_dist` function. The 'Eq.' column indicates whether the genetic distance implies that migration-drift equilibrium assumptions are made. The 'Depend.' column indicates the R packages on which the function depends for each genetic distance.

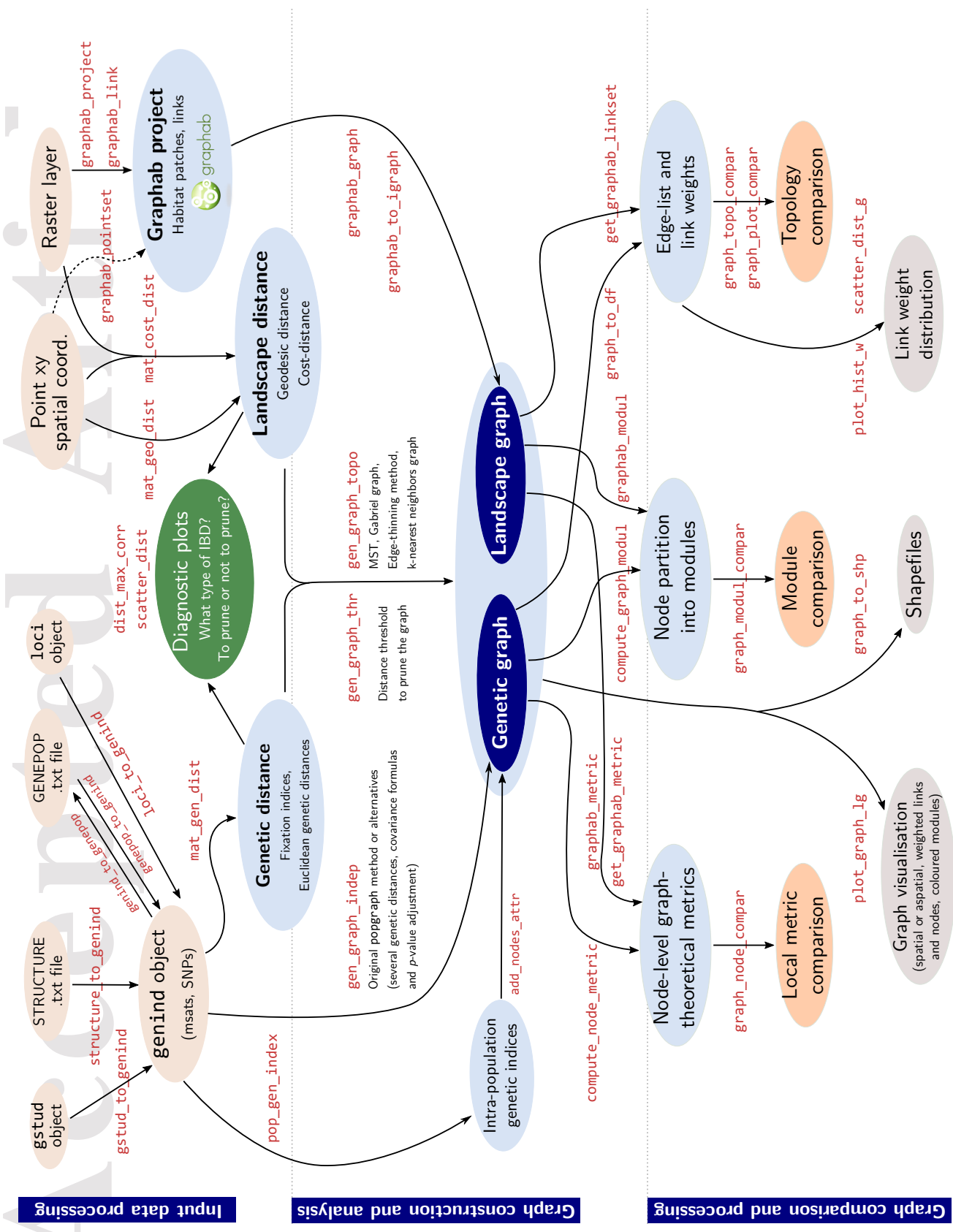


Figure 1: **graph4lg** workflow. Functions are represented by arrows linking their arguments to their output. Their names are written in red. Light beige boxes correspond to input data whereas grey boxes correspond to final output that can be exported to figures or files. Light blue boxes indicate analysis output that are used to construct graphs or derived from them. Dark blue boxes correspond to the two types of graphs that can be constructed with the package. The green box corresponds to diagnostic plots that can be assessed to help choose among the construction methods. Orange boxes indicate the analyses that can be carried out at several levels using both landscape and genetic graphs.



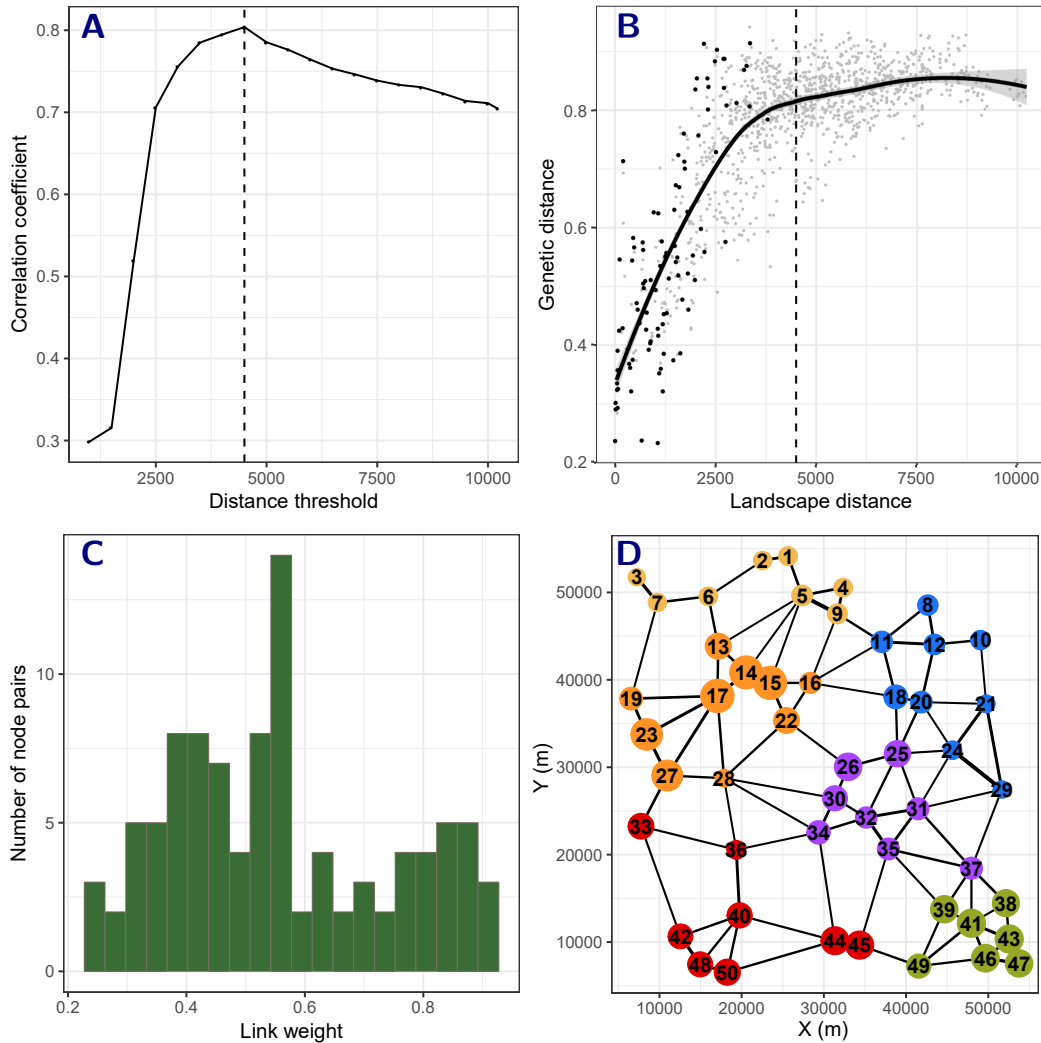


Figure 2: Diagnosis plots (A, B) and genetic graph analysis plots (C, D) produced with the datasets `data_simul_genind` and `pts_pop_simul` obtained after a simulation on CDPOP (Landguth and Cushman, 2010).

A) Identification of the distance of maximum correlation (DMC) (Van Strien et al., 2015) with the `dist_max_corr` function. Here, the DMC corresponds to the dashed vertical line on plots A and B. B) Scatter plot produced with the `scatter_dist_g` function representing the relationship between genetic distance ( $D_{PS}$ ) and cost-distance. The grey shaded region around the smoothed line corresponds to the 95 % confidence interval of the smoothing function. The black dots represent population pairs connected in the Gabriel graph. C) Histogram of the genetic distances separating the population pairs connected in the Gabriel graph produced with the `plot_w_hist` function. D) Gabriel graph mapped with the `plot_graph_lg` function. Link width is inversely proportional to the genetic distances weighting the links. Node size is proportional to the connectivity metric Flux derived from the corresponding landscape graph. Node color indicates the module to which pertains every node after a modularity analysis using `compute_graph_modul`.