



**HAL**  
open science

## Validating graph-based connectivity models with independent presence–absence and genetic data sets

Alexandrine Daniel, Paul Savary, Jean-Christophe Foltête, Aurélie Khimoun, Bruno Faivre, Anthony Ollivier, Cyril Éraud, Hervé Moal, Gilles Vuidel, Stéphane Garnier

### ► To cite this version:

Alexandrine Daniel, Paul Savary, Jean-Christophe Foltête, Aurélie Khimoun, Bruno Faivre, et al.. Validating graph-based connectivity models with independent presence–absence and genetic data sets. *Conservation Biology*, 2023, 37 (3), pp.e14047. 10.1111/cobi.14047 . hal-03982222

**HAL Id: hal-03982222**

**<https://hal.science/hal-03982222>**

Submitted on 24 Feb 2023

**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.



**HAL**  
open science

## Validating graph-based connectivity models with independent presence–absence and genetic data sets

Alexandrine Daniel, Paul Savary, Jean-christophe Foltête, Aurélie Khimoun, Bruno Faivre, Anthony Ollivier, Cyril Éraud, Hervé Moal, Gilles Vuidel, Stéphane Garnier

### ► To cite this version:

Alexandrine Daniel, Paul Savary, Jean-christophe Foltête, Aurélie Khimoun, Bruno Faivre, et al.. Validating graph-based connectivity models with independent presence–absence and genetic data sets. Conservation Biology, 2023, pp.e14047. 10.1111/cobi.14047 . hal-03982222

**HAL Id: hal-03982222**

**<https://hal.science/hal-03982222>**

Submitted on 10 Feb 2023

**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.

# Validating graph-based connectivity models with independent presence–absence and genetic data sets

Alexandrine Daniel<sup>1</sup> | Paul Savary<sup>1,2,3</sup> | Jean-Christophe Foltête<sup>2</sup> | Aurélie Khimoun<sup>1</sup> |  
 Bruno Faivre<sup>1</sup> | Anthony Ollivier<sup>1</sup> | Cyril Éraud<sup>4</sup> | Hervé Moal<sup>3</sup> | Gilles Vuidel<sup>2</sup> |  
 Stéphane Garnier<sup>1</sup>

<sup>1</sup>Biogéosciences, UMR 6282 CNRS, Université Bourgogne-Franche-Comté, Dijon, France

<sup>2</sup>ThéMA, UMR 6049 CNRS, Université de Franche-Comté, Besançon, France

<sup>3</sup>ARP-Astrance, Paris, France

<sup>4</sup>Office Français de la Biodiversité, Chizé, France

## Correspondence

Daniel, Alexandrine, Biogéosciences, UMR 6282 CNRS, Université Bourgogne-Franche-Comté, 6 Boulevard Gabriel 21000 Dijon, France.  
 Email: alexandrine.daniel@u-bourgogne.fr

**Article impact statement:** Graph-based connectivity models reflect genetic responses but the most complex methods are not necessarily leading to the best models.

## Funding information

European Regional Development Fund, Grant/Award Number: FEDER 2/2.4/-30528; Agence Nationale de la Recherche, Grant/Award Number: ANR-15-IDEX-0003; Association Nationale de la Recherche et de la Technologie; French Investissements d'Avenir; Direction de l'Environnement, de l'Aménagement et du Logement (DEAL) of Guadeloupe, Grant/Award Number: DEAL/RN971-2019-07-12-003;2009-053/SAPNSP/C; Parc National de la Guadeloupe

## Abstract

Habitat connectivity is a key objective of current conservation policies and is commonly modeled by landscape graphs (i.e., sets of habitat patches [nodes] connected by potential dispersal paths [links]). These graphs are often built based on expert opinion or species distribution models (SDMs) and therefore lack empirical validation from data more closely reflecting functional connectivity. Accordingly, we tested whether landscape graphs reflect how habitat connectivity influences gene flow, which is one of the main ecoevolutionary processes. To that purpose, we modeled the habitat network of a forest bird (plumbeous warbler [*Setophaga plumbea*]) on Guadeloupe with graphs based on expert opinion, Jacobs' specialization indices, and an SDM. We used genetic data (712 birds from 27 populations) to compute local genetic indices and pairwise genetic distances. Finally, we assessed the relationships between genetic distances or indices and cost distances or connectivity metrics with maximum-likelihood population-effects distance models and Spearman correlations between metrics. Overall, the landscape graphs reliably reflected the influence of connectivity on population genetic structure; validation  $R^2$  was up to 0.30 and correlation coefficients were up to 0.71. Yet, the relationship among graph ecological relevance, data requirements, and construction and analysis methods was not straightforward because the graph based on the most complex construction method (species distribution modeling) sometimes had less ecological relevance than the others. Cross-validation methods and sensitivity analyzes allowed us to make the advantages and limitations of each construction method spatially explicit. We confirmed the relevance of landscape graphs for conservation modeling but recommend a case-specific consideration of the cost-effectiveness of their construction methods. We hope the replication of independent validation approaches across species and landscapes will strengthen the ecological relevance of connectivity models.

## KEYWORDS

conservation modeling, habitat connectivity, landscape genetics, landscape graphs, species distribution models

## Resumen

La conectividad entre hábitats es un objetivo fundamental de las políticas de conservación actuales y con frecuencia se modela con grafos de paisaje (conjuntos de teselas de hábitat [nodos] conectados por vías potenciales de dispersión [enlaces]). Estos grafos se construyen a menudo con opiniones de expertos y modelos de distribución de especies (MDE), por lo que carecen de la validación empírica a partir de datos que reflejan de mejor

Alexandrine Daniel and Paul Savary contributed equally to this work.

manera la conectividad funcional. Por consiguiente, analizamos si los grafos de paisaje reflejan cómo la conectividad de hábitats influye sobre el flujo genético, que es uno de los principales procesos evolutivos. Con este propósito, modelamos la red de hábitats de un ave forestal (*Setophaga plumbea*) en Guadalupe con grafos basados en la opinión de un experto, en el índice de especialización de Jacobs o en un MDE. Usamos datos genéticos (712 aves de 27 poblaciones) para computar los índices genéticos locales y las distancias genéticas entre pares de poblaciones. Por último, analizamos las relaciones entre los índices o distancias genéticas y las distancias de costo o las métricas de conectividad con modelos de distancias de tipo maximum-likelihood-population-effect y correlaciones de Spearman entre las métricas e índices. En general, los grafos de paisaje reflejaron de manera confiable la influencia de la conectividad sobre la estructura genética de las poblaciones; el  $R^2$  de validación llegó hasta 0.30 y los coeficientes de correlación llegaron hasta 0.71. Aun así, la relación entre la pertinencia ecológica de los grafos, los requerimientos de datos y los métodos de construcción y análisis no fue directa porque los grafos basados en el método de construcción el más complejo (modelado a partir de la distribución de la especie) a veces tuvieron menos pertinencia ecológica que los otros. Los métodos de validación cruzada y el análisis de sensibilidad nos permitieron hacer espacialmente explícitas las ventajas y limitaciones de cada método de construcción. Así, confirmamos la pertinencia que tienen los grafos de paisaje para la conservación, aunque recomendamos se considere caso por caso el ratio entre la complejidad y la calidad de los métodos de construcción. Esperamos que la replicación de estrategias de validación independiente por varios paisajes y especies fortalezcan la pertinencia ecológica de los modelos de conectividad.

#### PALABRAS CLAVE

conectividad de hábitats, grafos de paisaje, genética de paisajes, modelos para la conservación, modelos de distribución de especies

## INTRODUCTION

The functional connectivity of habitat patches depends on their area and spatial configuration and on the degree to which the landscape matrix resists movements between them (Taylor et al., 2006). It determines dispersal and resulting ecoevolutionary processes, such as gene flow, which is key for long-term population persistence (Frankham, 2015). Connectivity is therefore a key objective of current conservation policies and several connectivity modeling approaches have been developed to support decision-making (Carroll et al., 2012) and to investigate biological responses to connectivity (Fletcher et al., 2016).

In this context, landscape graphs have been suggested as an optimal compromise for balancing data requirements, model complexity, and ecological relevance (Calabrese & Fagan, 2004). Landscape graphs make it possible to model a habitat network as a graph whose nodes are habitat patches and links are potential dispersal paths between the nodes (Urban & Keitt, 2001). Graph theoretical metrics quantify the role of every node and link in the connectivity of the habitat network (Galpern et al., 2011; Rayfield et al., 2011) and can be used to predict biological responses in subsequent analyzes (Mony et al., 2018; Ribeiro et al., 2011). The ability to identify landscape areas that make large contributions to connectivity from inexpensive data makes landscape graphs of strong operational interest for biodiversity conservation, as reflected by the 54 articles published from 2000 to 2013 alone (Correa Ayram et al., 2016).

Despite their strengths, landscape graphs have several limitations (Moilanen, 2011). To estimate functional connectivity, most graph-based connectivity models compute dispersal paths from cost surfaces. This requires making assumptions regarding the costs endured by the focal species when moving across the landscape matrix (Beier et al., 2008; Zeller et al., 2012). Most of these assumptions are based on expert opinion, and their proximity to ecological reality is rarely verified (Sawyer et al., 2011). Graph nodes are supposed to reflect the spatial distribution of the species habitat, but it is difficult to locate such habitat areas reliably from spatial environmental data alone, which brings the validity of the habitat delineation into question (Moilanen, 2011). Finally, a large range of connectivity metrics can be derived from landscape graphs (Baranyi et al., 2011; Rayfield et al., 2011), but the relationship between the connectivity pattern they quantify and the biological processes they are supposed to explain has rarely been tested.

To overcome these limitations and reinforce the ecological relevance of landscape graphs, several types of biological data have been used at various stages of modeling to define habitat patches and cost values associated with resistance to movement (Foltête et al., 2020). According to Beier et al. (2008), using genetic data, animal movement data, or interpatch movement measurements is the best option to estimate cost values. An advantage of genetic relatedness measures between populations over direct animal movement observations is that gene flow reflects the individual movements through time that are followed by successful reproduction (Robertson et al., 2018), and

such movements are key for dispersal to contribute to population persistence. Animal occurrence data, frequently used for deriving a species distribution model (SDM), are the second-best way to define habitat and cost values. Finally, relying on expert opinion or literature review should be reserved for cases where empirical data are not available (Beier et al., 2008). Yet, these assumptions have rarely been explicitly checked.

Genetic data reflect landscape connectivity well (Zeller et al., 2018) because they provide direct insights into dispersal-driven gene flow. Provided these data are not used in the graph modeling step, they could be a good candidate for assessing the ecological validity of graphs constructed using either presence data or expert opinion (Foltête et al., 2020).

Accordingly, we used 2 independent presence–absence and genetic data sets to answer the following question: Are landscape graphs built from expert opinion or presence–absence data close reflections of ecoevolutionary processes influenced by habitat connectivity? We modeled the habitat network of a forest bird (plumbeous warbler [*Setophaga plumbea*]) (Khimoun et al., 2016) in the Guadeloupe archipelago with expert opinion and presence–absence data and explored whether landscape graphs derived from these 2 approaches predicted genetic responses. We focused on the plumbeous warbler because of the importance of connectivity to forest bird species where forests have been reduced (Callens et al., 2011; Nevil Amos et al., 2014). We predicted that distances and connectivity metrics computed from the SDM-derived graph would provide the best predictions of the genetic responses given that its construction implies an additional modeling step relying on both empirical data and several types of spatial data.

## METHODS

We built 3 landscape graphs by delineating habitat patches and assigning dispersal costs based on expert opinion, Jacobs' indices measuring specialization, or an SDM (Figure 1), thereby reflecting common practitioner use. Graph links connected habitat patches (nodes) by minimizing dispersal costs between them and were weighted by the corresponding cost distances, that is, the accumulated cost along least-cost paths. Using patch areas and cost distances, we also computed node connectivity metrics. In parallel, we acquired genetic data in several bird populations and computed pairwise genetic distances and population-level genetic indices. We assessed the validity of landscape graphs regarding species genetic responses by focusing on the relationship between cost distances computed along graph links and pairwise genetic distances and the relationship between connectivity metrics computed at the graph node level and genetic indices computed at this same level.

### Study area and biological model

The Guadeloupe archipelago is in the Lesser Antilles (Figure 2) and is composed of 2 main islands (1440 km<sup>2</sup>) connected by a narrow isthmus: Grande-Terre (GT) and Basse-Terre (BT). The eastern part (i.e., GT) is more anthropized and has less

forest than the western part (i.e., BT), which hosts the Guadeloupe National Park. Entirely covered by forest before the 16th century, Guadeloupe experienced major deforestation episodes due to the agricultural activities of European settlers (Ricklefs & Bermingham, 2004).

The plumbeous warbler (*S. plumbea*) is endemic to Guadeloupe and Dominica. In this forest specialist species (Lovette et al., 1998), gene flow depends on forest cover and is influenced by landscape resistance (Khimoun et al., 2017). Thus, studying the genetic structure of this species could give hints about the influence of connectivity on genetic processes, thereby making it an ideal study species for validating connectivity models.

### Spatial and climatic data

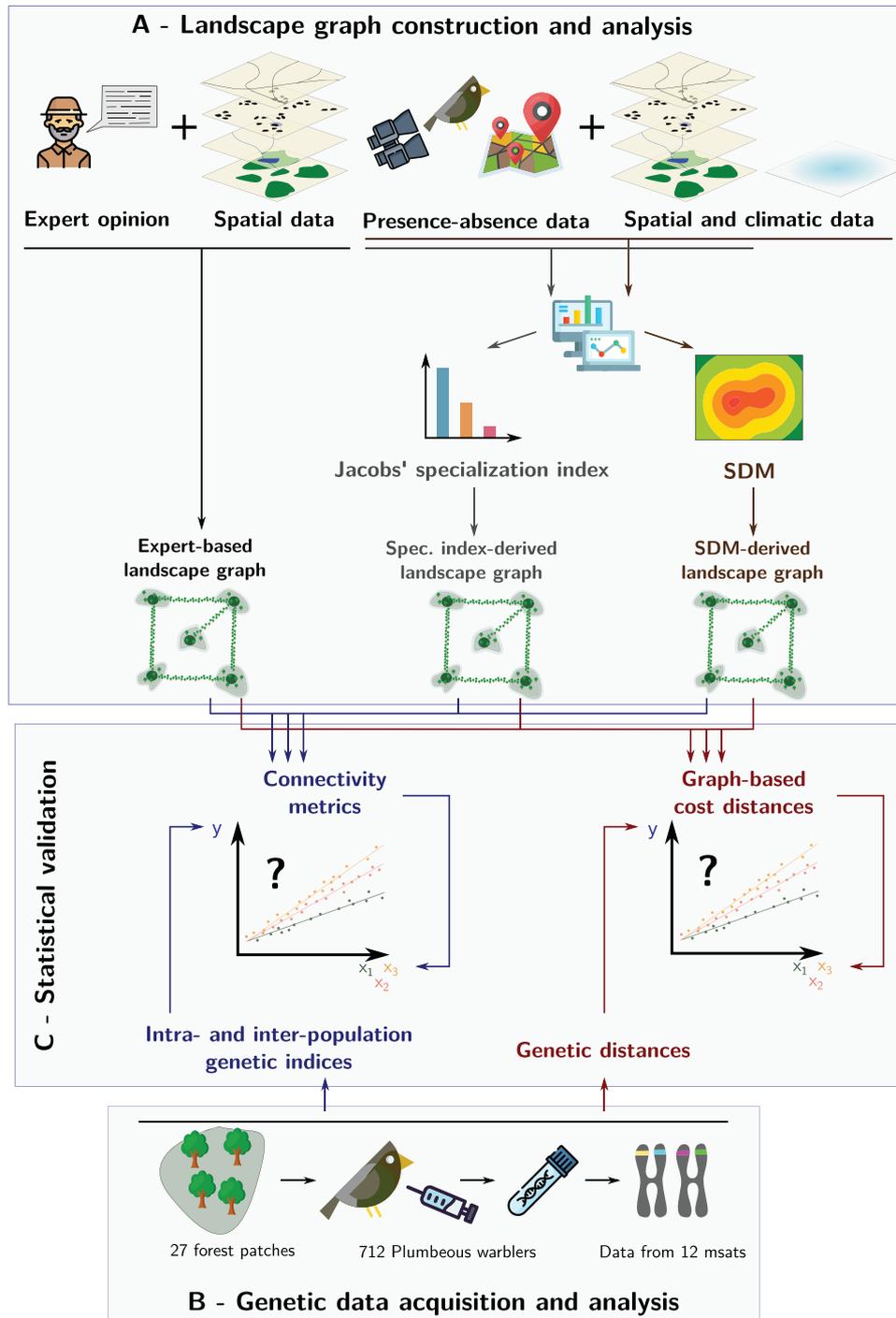
To build landscape graphs, we had to define 2 main parameters: habitat patches (graph nodes) and dispersal cost distances among pairs of nodes (weight of graph links) calculated from movement resistance surfaces relying on cost scenarios. We considered 3 methods for defining nodes and cost scenarios according to the input data (expert opinion, habitat specialization indices, and SDM) (Table 1).

We used land-cover and climatic spatial data to create the SDM and the landscape graphs. We used CORINE Land Cover data from 2012 to create a base map and the Institut National de l'Information Géographique et Forestière (IGN) BD TOPO database of 2012 to locate major roads (excluding dirt roads and tracks) and built elements. We used the 2010 map of 17 woody vegetation formations created by the Conseil Départemental of Guadeloupe, the IGN, and the Direction Régionale de l'Office National des Forêts (ONF) (Appendix 2). Annual mean precipitation and temperature raster layers with a resolution of 20 m were obtained by interpolation of punctual data of 61 Météo-France stations from 2012 to 2014 (Perrin et al., 2022).

We created a categorical raster surface (resolution 20 m) by classifying initial land-cover types into 10 types and intersecting the different spatial databases (Appendices S1 & S9). From this raster, we constructed continuous raster layers indicating the distance from each pixel to the closest road, forest, agricultural, or built elements. The proportions of these landscape features in circular landscapes of 500-m radius around each pixel were computed for species distribution modeling.

### Expert opinion

We asked a local ornithologist expert from Guadeloupe, who has been involved in conservation management as a member of the CSRPN (Conseil Scientifique Régional du Patrimoine Naturel) and in bird ringing campaigns, including the plumbeous warbler for the CRBPO (Centre de Recherche sur la Biologie des Populations d'Oiseaux), to identify the vegetation formations of the plumbeous warbler's habitat, which we used to create the habitat patches (graph nodes) of the expert-based

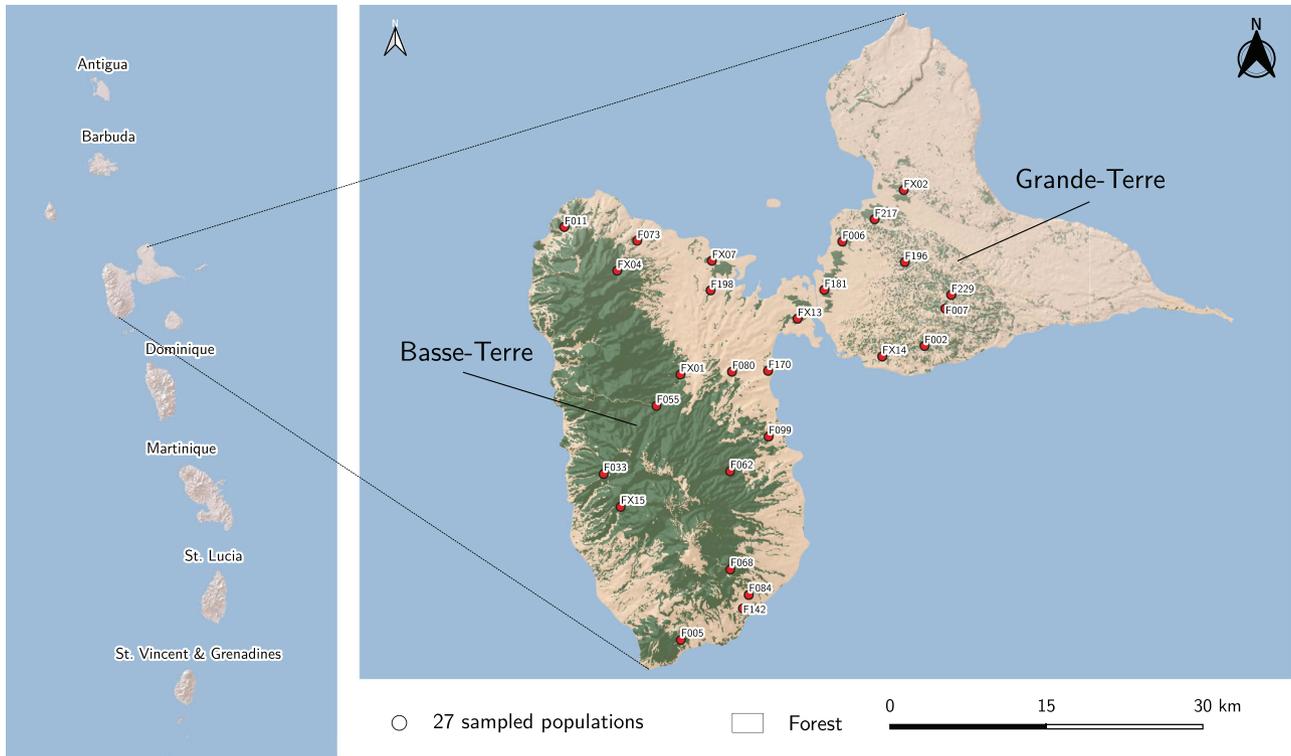


**FIGURE 1** Overall methodology in validating graph-based connectivity models with genetic data sets: (A) landscape graph construction and analysis from spatial data and expert opinion-only and from presence-absence, spatial, and climatic data. Graph-based cost distances between sampled patches and connectivity metrics (capacity, flux, betweenness centrality) are computed from the 3 different graphs (expert-based, habitat specialization index-derived [spec. index], and species distribution model-derived [SDM]); (B) genetic data acquisition and analysis in which 27 populations (712 individuals) are sampled in forest patches and 12 microsatellite markers (msats) are analyzed for every individual and local genetic structure indices (allelic richness, relative genetic differentiation) and pairwise genetic distances are computed at the population level; and (C) statistical validation through analyses of the relationships between cost distances and genetic distances and between connectivity metrics and genetic indices aiming at assessing the respective validity of the 3 landscape graphs.

landscape graph (Table 1). These areas represented a class of the cost surface. We also asked this expert to assign each of the 10 land-cover types a dispersal cost value for use in computing the least-cost paths between the habitat patches (graph links) of the expert-based landscape graph.

### Habitat specialization indices

We used 991 bird point counts acquired from 2009 to 2011 in different vegetation units (Khimoun et al., 2016). Detection and nondetection data recorded for the plumbeous warbler were



**FIGURE 2** Locations where 712 plumbeous warblers from 27 populations were sampled on the Guadeloupe archipelago in the Caribbean.

**TABLE 1** Landscape graph elements according to different graph construction methods

Landscape graph	Node delineation	Cost value for link weights
Expert based	Preferred vegetation formation according to expert opinion	Cost values following expert opinion
Specialization index derived	Land-cover type with the highest specialization index value	Cost values converted from specialization index values following Bourdouxhe et al. (2020) and assuming a cost of 1 in the preferred habitat
Species distribution model (SDM) derived	Areas with suitability scores above the suitability threshold maximizing the prediction performance of the SDM	Cost values converted from suitability scores following Bourdouxhe et al. (2020) and assuming a cost of 1 in the habitat

considered presence–absence data. From these point data, we computed a habitat specialization index for each of the 10 land-cover types following the formula of Jacobs (1974) (Appendix S4). This index varies from  $-1$  (avoidance) to  $1$  (exclusive use); an index value of  $0$  means the species is neutral relative to the land-cover type. To avoid pseudoreplication induced by the presence of points following linear and most accessible tracks in BT, we iteratively subsampled these points and kept only 2 points per set of linearly distributed points. The land-cover type with the maximum Jacobs’ index was considered habitat and was used to delineate habitat patches (nodes) of the specialization-index-derived graph (Table 1). Following the recommendation by Keeley et al. (2016), we assigned cost values to land-cover types with a negative exponential transformation. We used a formula adapted from Bourdouxhe et al. (2020) to specify the

range of cost values:

$$\text{cost}_i = e^{\frac{-\ln(\text{cost}_{\max}) \times \ln(\text{Jacobs}_i + 1)}{\text{Jacobs}_{\text{habitat}} + 1}} \times \text{cost}_{\max},$$

where  $\text{cost}_i$  and  $\text{Jacobs}_i$  are, respectively, the dispersal cost value and Jacobs’ index associated with land-cover type  $i$ ;  $\text{cost}_{\max}$  is the maximum cost value; and  $\text{Jacobs}_{\text{habitat}}$  is the maximum index value (i.e., associated with the species habitat), which ensures that the corresponding cost value is equal to 1. Because dispersal may occur through moderately suitable areas and is avoided only through the least suitable areas, we used a negative exponential distribution of the cost values. The maximum cost value was set to 1000, reflecting previous empirical results (Gurrutxaga et al., 2010).

## Species distribution model

We used logistic regressions to create the distribution model for plumbeous warbler because presence–absence methods should be used when both presence and absence data are available (Fletcher & Fortin, 2018), as in our case. This type of model is easy to use in the conservation-oriented modeling approaches we applied here. We used the same iterative subsampling approach as described above to avoid pseudoreplication.

After controlling for collinearity and explanatory power of the predictor variables, we retained the following variables in the model: percentage of agricultural and artificial areas at a radius of 500 m around the data point; distance from the data point to the closest road; distance from the data point to the closest forest; and mean annual precipitation. We computed the Matthews' (1975) correlation coefficient (MCC), which assesses the prediction quality of the model at different suitability thresholds. The threshold suitability value maximizing the MCC was selected as the probability threshold above which we considered a pixel to be a habitat pixel. We converted suitability scores into cost values using the same formula as for converting specialization indices into cost values:

$$\text{cost}_i = e^{\frac{-\ln(\text{cost}_{\max}) \times S_i}{S_{\text{thr}}}} \times \text{cost}_{\max},$$

where  $S_{\text{thr}}$  is the threshold probability used for delineating habitat patches and  $S_i$  is the suitability score of pixel  $i$ .

## Landscape graphs

We created landscape graphs whose nodes and links were defined in 3 different ways according to the information source (Table 1) with Graphab software (Foltête et al., 2012). The minimum patch size was set to 1 ha. The size of some habitat patches in BT exceeded the scale at which habitat amount is supposed to influence population dynamics, which decreases the explanatory power of graph-based connectivity metrics (Laroche et al., 2020). We thus controlled for the maximum size of habitat patches by fixing their maximum extent to 2500 m side length. The topology of the landscape graphs was planar because it is a good approximation of the complete graph. The links were weighted by pairwise cost distances.

From each landscape graph, we computed 3 complementary metrics reflecting the ways each patch contributes to the connectivity of the habitat network (Baranyi et al., 2011). First, we computed the capacity ( $\text{capa}$ ), a proxy for the carrying capacity of the patch representing its demographic potential and the intrapatch connectivity. It was equal to patch area.

We then computed the flux index ( $F$ ) to measure the contribution of each habitat patch to immigration and emigration flows. This metric is largely inspired by the incidence function by Hanski and Gilpin (1991). We used the following formula:

$$F_i = \sum_{j=1; j \neq i}^n \text{capa}_j \times e^{-\alpha \times d_{ij}}, \quad (3)$$

where  $i$  is the index of the focal patch,  $j$  is the index of all the other  $n$  patches, and  $d_{ij}$  is the cost distance between patches  $i$  and  $j$ . Alpha ( $\alpha$ ) was computed according to different dispersal kernels to test for the influence of the scale at which between-patch connections are assigned significant weights for computing the metrics. To that purpose, we set  $\alpha$  values such that  $p = e^{-\alpha \times d_{ij}} = 0.05$  for distances  $d_{ij}$  ranging from 500 to 15,000 m (with steps of 500 m), whereby we considered the amount of reachable habitat (Saura & de la Fuente, 2017) beyond the immediate neighborhood of a population.

Finally, we computed the betweenness centrality (BC) index to quantify the role of each patch for the traversability (Urban & Keitt, 2001) of the entire habitat network:

$$\text{BC}_i = \sum_j \sum_k \text{capa}_j \times \text{capa}_k \times e^{-\alpha \times d_{jk}}, \quad (4)$$

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

where  $P_{jk}$  is the set of crossed habitat patches along the least-cost paths between patches  $j$  and  $k$ , including patch  $i$ , so that only these paths are considered for computing the metric for patch  $i$ . Alpha ( $\alpha$ ) had the same range of values as in the  $F$  calculation. This index makes it possible to identify stepping stones in the habitat network. The BC index should have high values at patches that are key for maintaining high genetic diversity levels (Zetterberg et al., 2010), although this has rarely been proven.

## Field sampling and genotyping

We mist-netted 712 birds in 27 forest patches (9 and 18 from GT and BT, respectively) (Figure 2). Twenty sites were sampled in 2012–2013 and the other 7 in 2020 (protocol in Khimoun et al., 2017). Four sites were sampled twice, and temporal genetic differentiation was tested using GENEPOP (Raymond & Rousset, 1995). This temporal genetic differentiation was not significant for 3 of the 4 populations. The relative level of genetic differentiation between the only differentiated population and the other ones remained stable. We thus pooled the data. Permissions for bird sampling were granted by Direction de l'Environnement, de l'Aménagement et du Logement (DEAL) of Guadeloupe, Parc National de la Guadeloupe, Conservatoire du Littoral, and ONF de Guadeloupe.

Total DNA extraction was performed following either a standard phenol–chloroform protocol (2012–2013 data set) (Khimoun et al., 2017) or using commercial Blood Genomic DNA Mini-Preps Kits (BIO BASIC Inc., Markham, Canada) (2020 data set) from blood samples stored in Queen's lysis buffer. Individuals were genotyped at 12 microsatellite loci (Khimoun et al., 2016; Khimoun et al., 2017). Loci were amplified in simplex in a Dyad thermal cycler (Bio-Rad, Hercules, CA, USA), PCR products were multiloading for analyzes in an automated sequencer (ABI3730), and allele scoring was performed using GENEIOUS R.8 (Kearse et al., 2012).

## Genetic structure indices

After checking for Hardy–Weinberg equilibrium and for the absence of linkage disequilibrium, we assessed the genetic diversity of each population by its allelic richness ( $A_r$ ), computed using a rarefaction method implemented in ADZE 1.0 (Szpiech et al., 2008) to account for heterogeneous sample sizes. The relative genetic differentiation of each population was assessed by the mean of inverse weight metric (MIW) in the graph4lg R package (Savary et al., 2021b). The MIW of a population is the mean of the inverse pairwise genetic differentiation values between each population and the 26 others. Highest MIW values are obtained for the populations that are the least genetically different from the others. Genetic differentiation between populations (hereafter genetic distances) was estimated by pairwise  $F_{ST}$  (Weir & Cockerham, 1984).

## Validation of cost distances and connectivity metrics associated with landscape graph links and nodes

We assessed the ecological validity of landscape graphs by studying the relationship between cost distances computed along their links and genetic distances between populations connected by these links and the relationship between connectivity metrics and genetic indices computed at the level of patches occupied by sampled populations (Figure 1). In both cases, a strong relationship between these variables meant that landscape graphs correctly reflected reality.

Cost distances depend on cost scenarios and aim at reflecting dispersal constraints across landscapes more realistically than geographic distances. Therefore, we tested whether cost distances were better predictors of  $F_{ST}$  than geographic distances with Mantel and partial Mantel tests, following the causal modeling approach of Cushman et al. (2006). The cost distances that best explained pairwise genetic distances were the most reliable estimates of landscape connectivity among pairs of populations. For each graph, we thus modeled pairwise genetic distances as a function of cost distances between populations along the links of the landscape graphs with maximum-likelihood population-effects (MLPE) models (Clarke et al., 2002). These models include a population random effect to control for the nonindependence of the observations inherent to distance matrices.

Because the spatial scale and the topology of the links considered can influence the relationship between landscape distances and genetic distances (Savary et al., 2021a; Van Strien et al., 2015), we assessed the relationship between these distances by considering different sets of links: all population pairs; only the links between population pairs located on the same part of the island (intraisland pruning conserving only BT–BT or GT–GT links); and all population pairs separated by a cost distance lower than a threshold value (threshold pruning) (several thresholds were tested iteratively). The 2 latter approaches reflect the fact that a study area is sometimes divided in several parts for carrying out the analyzes (Angelone & Holderegger, 2009) and that

iterative thresholding can help determine the scale of landscape effects on genetic structure (Emaresi et al., 2011).

To assess the ecological validity of every cost scenario while controlling for potential model overfitting, we adapted the leave-one-out cross-validation method to our specific objectives and genetic distance data. When creating the MLPE models, we removed iteratively 1 of the 27 populations and all the links including it from the training data (i.e., 26 links for a complete topology). Then, we predicted the genetic distances involving this population with the calibrated model. The genetic distance associated with a given population pair was therefore predicted twice overall if it was included in the topology retained for the analysis. The mean of these 2 predicted genetic distances was compared with the corresponding observed genetic distances to assess the predictive quality of each cost scenario and link set. We computed a validation  $R^2$  to quantify the prediction error (Appendix S6). We also computed the root mean square of the errors (RMSE) associated with each population pair from the 2 corresponding predicted values and the mean of these RMSE at the population level. Finally, the predictive capacity of these cost distance models was compared with that of null models predicting genetic distances as a function of geographic distances with and without pruning.

To assess the ecological relevance of the connectivity metrics and identify metrics reflecting the best population genetic structure, we computed Spearman's rank correlations between the 2 genetic indices ( $A_r$ , MIW) and the connectivity metrics (capacity,  $F$ , BC). We only considered the 27 forest patches occupied by the sampled populations.

## RESULTS

### Graph construction parameters

According to the expert, the plumbeous warbler occupies all types of woody vegetation formations except coastal thickets and fallow lands with low-growth woody vegetation (Appendix 2). The assigned cost values ranged from 1 to 10; the smallest values were assigned to forests (Table 2). The ocean and artificial areas were the most resistant land-cover types (highest costs). For allowing comparisons with the other cost scenarios, we rescaled these cost values between 1 and 1000 (Appendices S5 & S13).

The maximum Jacobs' specialization index (0.73) was obtained for the land-cover type delineated as habitat areas by the expert, whereas the minimum was obtained for open areas (−0.83), agricultural areas (−0.68), and artificial areas (−0.26) (Table 2). Although there was no bird count in seminatural areas, these areas were mostly urban green spaces, and we assigned them the same index as for artificial areas. The cost values converted from these indices exhibited sharp contrasts as illustrated by the costs of 281 and 52, respectively, assigned to agricultural and artificial areas (Appendices S10 & S13).

The bird count point data set included 206 presence points and 785 absence points of the study species. The resulting SDM had good prediction accuracy (Area Under the Curve (AUC) =

**TABLE 2** Habitat specialization indices and raw expert-based cost values associated with every land-cover type and corresponding cost values according to the expert-based and specialization index-based cost scenarios for plumbeous warbler's dispersal

Land cover type	Initial values		Cost values	
	exp.	spec. index	exp. based	spec. index derived
Habitat	1	0.73	1	1
Suboptimal woody vegetation	2	0.46	112	3
Seminatural	4	-0.26	334	52
Semiopen	3	0.29	223	6
Open areas	4	-0.83	334	513
Agricultural	3	-0.68	223	281
Wet areas <sup>a</sup>	3	-1	223	1000
Water <sup>a</sup>	4	-1	334	1000
Ocean <sup>a</sup>	10	-1	1000	1000
Artificial	10	-0.26	1000	52

*Note:* Because the cost values derived from the species distribution models do not depend on land-cover types, they are not included in this table.

Abbreviations: exp., cost values deriving from expert opinion; spec. index, cost values obtained from the Jacobs' specialization indices.

<sup>a</sup>Given the absence of bird count points in water, wet areas, and the ocean, we set the corresponding indices at -1.

0.918) (Appendix S11). The suitability threshold maximizing the MCC above which we considered a pixel to be a habitat pixel was 0.328 (MCC = 0.547). The distance to the closest forest had the strongest effect on presence probability. It negatively influenced the suitability scores, as did the proportion of agricultural and artificial pixels in the 500 m surrounding a pixel. In contrast, an increase in distance to the closest road and in mean precipitation tended to increase the suitability (Appendix 12).

Cost values derived from the SDM were very different because the conversion of already contrasted presence probabilities assigned either very low or very high cost values to pixels (Appendix S10). In forest areas, mostly coinciding with habitat patches, pixels had low cost values, equal or close to 1 (Appendices S9 & S13). Because presence probabilities strongly decreased as distance to the closest forest increased and to a lesser extent as the proportion of agricultural and artificial areas increased, pixels located far from the main habitat patches were associated with high cost values (Appendix S13). This led to higher cost values in agricultural areas compared with the cost values associated with this land-cover type in the other cost scenarios (223 and 281 for the expert-based and specialization index-derived scenarios, respectively). Because artificial areas also took high values overall, areas dominated by artificial and agricultural areas (e.g., around the isthmus) were assigned much higher cost values in the SDM-derived scenario than in the others (Appendix S13).

## Graph element properties

The expert-based and specialization index-derived graphs shared the same nodes. They had more nodes with a smaller average area ( $n = 1653$ , mean area = 39 ha) than the SDM-derived graph ( $n = 621$ , mean = 102 ha). Overall, the cost distances between populations separated by the isthmus (BT–

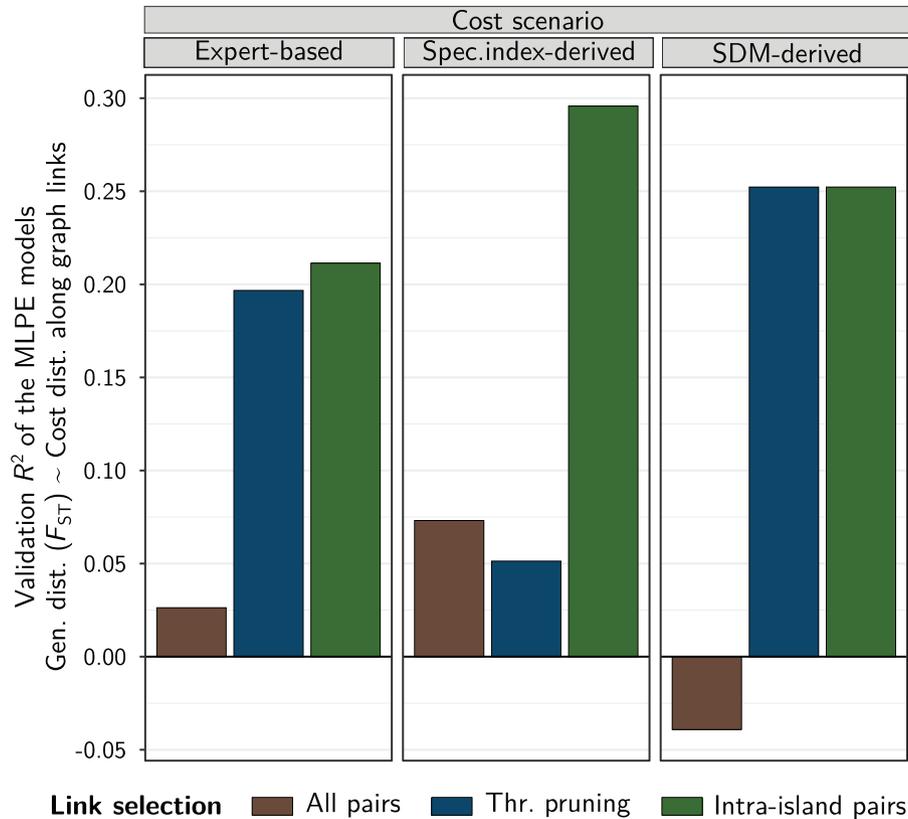
GT) were larger than the cost distances between population pairs from BT or GT (BT–BT or GT–GT). However, because of the high cost values assigned to both agricultural and artificial areas when building the graph from the SDM, the difference between the cost distances computed across the isthmus and the others was even larger when using the SDM-derived scenario rather than the other scenarios (Appendices S14 & S15).

## Genetic structure

On average, populations from BT and GT had similar levels of genetic diversity, although the allelic richness was more variable in BT (BT: 4.63 [SD = 0.34]; GT: 4.82 [SD = 0.13]) (Appendix S3). The main differences in allelic richness were observed in BT, where the populations inside the largest forest patches were more diverse than populations on their margins. Similarly, MIW exhibited a bimodal distribution on BT ( $\geq 50$  or  $\leq 26$ ) and had intermediary values on GT (26–50) (Appendix S3). The lowest MIW values, corresponding with the most genetically differentiated populations, were observed for BT populations on the margins of the main forest patches (Appendix S3). Overall, populations were highly differentiated genetically, with over 90% of population pairs significantly differentiated (Appendix S7). This differentiation was more significantly related to cost distances than to geographic distances according to the causal modeling approach (Appendix S8), suggesting a dominant pattern of isolation by landscape resistance rather than isolation by distance.

## Validation of cost distances associated with landscape graph links

The models explaining  $F_{ST}$  as a function of geographic distances had much less predictive power than the models based



**FIGURE 3** Validation  $R^2$  of the maximum-likelihood population-effects (MLPE) models explaining pairwise genetic distances ( $F_{ST}$ ) as a function of the cost distances between 27 plumbeous warbler populations along the links of the landscape graphs (vertical boxes, cost scenarios used for computing cost distances when creating landscape graphs; spec. index derived, cost values converted from Jacobs' specialization indices; SDM-derived, cost values converted from suitability scores of the species distribution model; matching colors, set of links [population pairs] considered in the models; All pairs, all population pairs; Thr. pruning, population pairs separated by a cost distance lower than a given cost distance threshold [only results of the best models obtained using a distance threshold reported for each cost scenario]; Intra-island, only pairs of populations located on the same part of the island are considered [Basse-Terre–Basse-Terre or Grande-Terre–Grande-Terre]).

on cost distances. The validation  $R^2$  values associated with the MLPE models based on geographic distances were negative when considering all population pairs ( $-0.05$ ), a pruning threshold ( $-0.03$ ), and on only intrainland links ( $-0.04$ ). The MLPE models with the cost distances along the graph links as predictor variables provided contrasting results according to the cost scenario and the set of links considered (Figure 3). Whatever the cost scenario, the models considering all population pairs had very low or negative validation  $R^2$ , especially when they included cost distances deriving from the SDM-derived landscape graph ( $R^2 = -0.04$ ). In contrast, when only a subset of links was considered in the models, the validation  $R^2$  was up to 0.30 under the cost scenario derived from specialization indices. Validation  $R^2$  reached maximum values of 0.25 and 0.21 with cost distances derived from the SDM and expert-based landscape graphs, respectively (Figure 3).

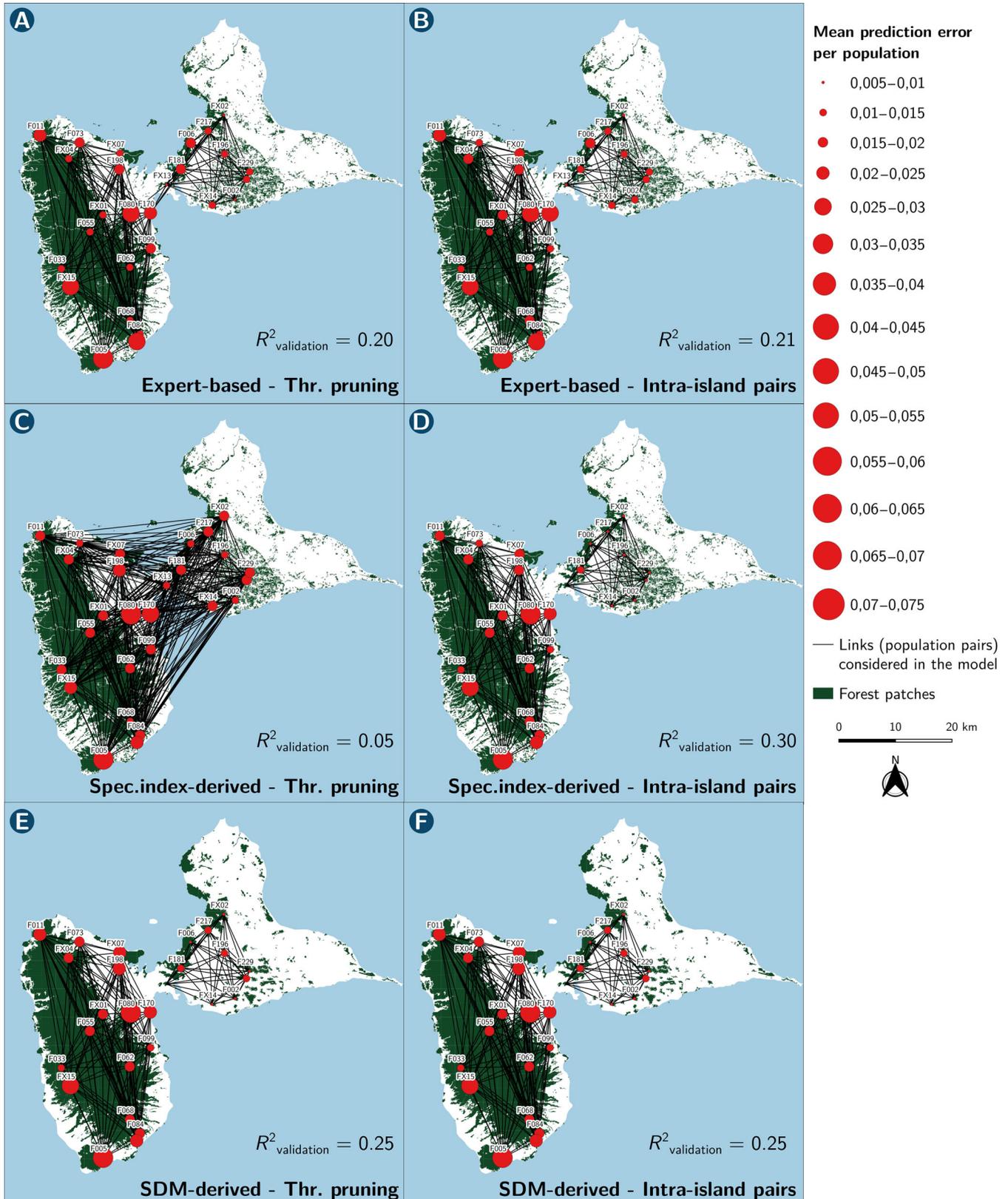
The set of links to consider for maximizing the predictive performance of the model was not the same depending on the cost distances used (Figures 3 & 4). For the specialization index-derived method, the best MLPE model relied on a set of intra-island links (BT–BT and GT–GT links; BT–GT links discarded) (Figure 4B), whereas the model relying on a cost distance threshold to prune the links had a poor predictive performance. In this latter model, BT–GT links were included in

the analyzes. For the expert-based or SDM-derived cost scenario, link selection based on the cost distance threshold and on the intra-island location led to similar topologies. Only few or no BT–GT links were conserved when the best pruning threshold was applied to cost distances based on the expert or SDM scenario, respectively. This led to similar validation  $R^2$  for a given scenario.

The different performances depending on the link set considered partly stemmed from the fact that the genetic distances between BT and GT populations were not predicted correctly by cost distances. This was especially the case with the SDM-derived cost distances, which led to a very negative validation  $R^2$  when a complete topology was considered. Besides, the prediction errors of the best models mainly concerned population pairs involving BT populations, especially those located on the margin of the largest forest patches (Figure 4).

### Validation of connectivity metrics associated with landscape graph nodes

The 2 genetic indices (Ar, MIW) were each significantly correlated to all 3 habitat connectivity metrics (capacity,  $F$ , BC). For each cost scenario, we only report the highest correlation



**FIGURE 4** Genetic distance prediction errors of the maximum-likelihood population-effects (MLPE) models (abbreviations defined in legend of Figure 3). The means of the prediction errors of the genetic distances involving a given population and the populations connected to it (26 for a complete topology) are computed for each population. The prediction error considered is the square root of the mean of the squared differences between the observed genetic distance and the genetic distance predicted by the MLPE models in which 1 of the 2 populations of a pair has been excluded from the training data set (2 values for each pair). The corresponding validation  $R^2$  of the MLPE model is reported. Results shown are for models including cost distances computed from cost scenarios deriving from (A, B) expert opinion, (C, D) Jacobs' specialization indices, or (E, F) species distribution model (SDM) and considering populations pairs either selected based on a threshold pruning (A, C, E) or corresponding to populations located on the same subpart of the island (Basse-Terre–Basse-Terre or Grande-Terre–Grande-Terre) (B, D, F).

**TABLE 3** Spearman's rank correlation coefficients ( $r$ ) between genetic indices and connectivity metrics ( $n = 27$  plumbeous warbler's populations, 712 individuals total) on the Guadeloupe archipelago.

Genetic index	Connectivity metric	Habitat delineation, cost scenario	$r_{sp. optim.}^a$	$d_{optim.}$	
MIW	Capacity	expert based, spec.ind.	0.50	*	
MIW	Capacity	SDM derived	0.57 <sup>b</sup>	**	
Ar	Capacity	expert based, spec.ind.	0.55 <sup>b</sup>	*	
Ar	Capacity	SDM derived	0.45		
MIW	F	expert based	0.68	***	1000
MIW	F	spec. ind. derived	0.71 <sup>b</sup>	***	500
MIW	F	SDM derived	0.58	**	3500
Ar	F	expert based	0.42	*	1000
Ar	F	spec. ind. derived	0.48 <sup>b</sup>	*	500
Ar	F	SDM derived	0.37		3500
MIW	BC	expert based	0.63	***	1000
MIW	BC	spec. ind. derived	0.68 <sup>b</sup>	***	500
MIW	BC	SDM derived	0.63	***	3500
Ar	BC	expert based	0.51 <sup>b</sup>	**	500
Ar	BC	spec. ind. derived	0.50	**	500
Ar	BC	SDM derived	0.42	*	1500

*Note.* Values are reported according to the genetic index (MIW, mean inverse weight; Ar, allelic richness), connectivity metric (capacity; F [flux metric]; BC [betweenness centrality metric]), method used for delineating habitat patches and fixing cost values (expert based, expert-based scenario; spec. ind. derived, habitat delineation and cost values deriving from Jacobs' specialization indices; SDM derived, habitat delineation and cost values deriving from suitability scores of the species distribution model), and distance at which dispersal probability is set to 0.05 for computing  $\alpha$  values. For each combination of a genetic index, a connectivity metric, and a construction scenario, the correlation coefficients ( $r_{sp. optim.}$ ) obtained when using the optimal distance ( $d_{optim.}$ ) and corresponding  $\alpha$  value for computing the connectivity metric are shown (spec., species; ind., index).

<sup>a</sup>Significance: \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

<sup>b</sup>The highest correlation coefficient for each combination of a genetic index and a connectivity metric.

coefficients obtained across the range of  $d$  values we used to compute the metrics (Table 3). These coefficients took large values (up to 0.71) and tended to be slightly higher between the connectivity metrics and the MIW rather than the allelic richness. The optimal  $d$  value used to compute the metric was always  $\leq 3500$  m.

The correlations obtained with metrics computed using the SDM-derived landscape graph were overall lower than those obtained with the 2 other graphs (Table 3). In all but 1 case (MIW vs capacity), the metrics computed from the landscape graphs based on expert opinion and on the specialization indices led to stronger correlations than metrics computed using the SDM-derived graph. Thus, we could not determine whether such differences of correlation strength stemmed from differences in terms of patch delineation or from cost distances associated with the graph links.

## DISCUSSION

Most connectivity models based on landscape graphs lack a posteriori empirical validation (Foltête et al., 2020; Kadoya, 2009). Using genetic data, we demonstrated that landscape graphs constructed using either expert opinion or presence-absence data reliably modeled the influence of habitat patch connectivity on population genetic structure. We thus confirmed their relevance for conservation modeling. Yet, the relation-

ship between graph ecological relevance, data requirements, and construction and analysis methods was not straightforward. The graph based on the most complex construction method (SDM) had a similar or even lower ecological relevance than the others.

## Empirical validation of landscape graphs by genetic data

The observed genetic differentiation pattern provides support for an isolation-by-landscape resistance hypothesis over the isolation-by-distance hypothesis. Furthermore, whatever the landscape graph construction method, the cost distances associated with graph links accurately predicted the genetic distances between populations sampled in graph nodes. Because genetic distance was partly due to the random and unpredictable effects of genetic drift, a validation  $R^2$  up to 0.30 seems satisfactory. The relationship between genetic distances and cost distances computed under different cost scenarios has most often been evidenced based on correlative approaches (Balbi et al., 2018; Creech et al., 2014). Introducing a novel cross-validation approach applied to linear mixed models, we showed that landscape graphs could be useful for implementing predictive approaches in landscape genetics. Predicting genetic distances between a whole set of habitat patches without implementing an

exhaustive genetic sampling protocol strengthens the usefulness of this approach in conservation modeling (Keller et al., 2015).

Although Moilanen (2011) raised concern about the uncertain ecological relevance of the multitude of metrics deriving from landscape graphs, a restricted set of these metrics sufficiently reflects how habitat connectivity influences ecoevolutionary processes (Baranyi et al., 2011; Urban & Keitt, 2001). Accordingly, we computed 3 complementary connectivity metrics (capacity,  $F$ , BC) and found that they were each significantly correlated with the local genetic diversity of each population and with their level of genetic differentiation with other populations.

Our results are consistent with previous studies showing the same landscape genetic relationships (Castillo et al., 2016; Creech et al., 2014), including studies using similar graph-based connectivity metrics and genetic indices (Bertin et al., 2017; Peterman et al., 2015). However, few studies have assessed these relationships with connectivity metrics completely independent of genetic data (but see Savary et al. [2022]). The independent validation approach we implemented therefore strengthens the robustness of landscape genetic inferences by showing that the metrics associated with the nodes and links of landscape graphs reliably reflect the drivers of both genetic drift and gene flow. This means that ranking habitat patches according to different connectivity metrics provides insight into the respective genetic responses of the populations occupying them and would help in the design of conservation measures. Nevertheless, our results are valid for a single species in a single landscape. Thus, to provide broad guidance to conservation scientists and planners, we call for studies spanning a wider range of species and landscapes while relying on an independent validation approach. In addition, when historical land-cover data are available, they should be used to test whether there is a mismatch between the observed genetic structure and the current landscape (Epps & Keyghobadi, 2015; Landguth et al., 2010; Robertson, Fletcher, & Austin, 2018) and to adjust the connectivity models accordingly.

## Relationships among graph ecological relevance, data requirements, and construction and analysis methods

Apart from validating the landscape graph approach, we aimed to compare the ecological relevance of graphs built using either expert-based information or empirical data. Our hypothesis that the graph deriving from an SDM would provide the most realistic outputs was only partly supported.

We used 991 presence–absence observations and both climatic and landscape variables for fitting the SDM; the same presence–absence data and only land-cover data were needed for computing habitat specialization indices. Only land-cover data and expert opinion were needed for the simplest expert-based approach. These graph construction methods thus differed in terms of both data requirement and complexity, which determine the overall modeling cost. Despite these differences, the connectivity metrics derived from the expert-based or specialization-index graphs were almost always the most

correlated to the genetic indices. The best model explaining genetic distances included cost distances computed with cost values derived from specialization indices, while cost distances obtained using the SDM-derived and expert-based cost scenarios led to similar predictive performances. This indicates that, compared with the SDM approach, simpler empirical approaches or expert-based approaches could be reliable and more cost-effective in many instances. Therefore, although our results confirm that integrating empirical data in landscape graph modeling leads to reliable models (Kadoya, 2009), they bring into question the cost-effectiveness of such an approach when it is based on an SDM.

The quality of the SDM cannot explain why the SDM-derived graph did not always appear as the most realistic. The SDM had a good accuracy ( $AUC > 0.9$ ), reflecting the quality of these models for specialist species such as the plumbeous warbler (McPherson & Jetz, 2007). Interestingly, the use of presence–absence data did not explain the main differences between the graphs. We used these same data to compute specialization indices, and the metrics derived from the graph built from these indices were more similar to the metrics derived from the expert-based graphs than those derived from the SDM-derived graph. Therefore, the choices made for delineating habitat patches and setting cost values seem to be equally important, if not more, than the data used for modeling landscape graphs.

Using suitability scores for delineating habitat patches produced fewer patches of larger sizes, partly because areas dominated by suitable pixels were bordered by areas with slightly lower suitability scores, sufficient for being integrated into habitat patches. This has already been observed (Bourdouxhe et al., 2020; Godet & Clauzel, 2021), although an opposite result can also be obtained when variables included in the SDM tend to decrease suitability in areas otherwise identified as habitat patches (Stevenson-Holt et al., 2014). The different correlations between genetic indices and connectivity metrics deriving from the graphs could be due to these differences in node delineation.

The methods used for setting the cost values substantially affected our results. The negative exponential function introduced by Keeley et al. (2016) for converting suitability scores into cost values assigned high cost values to pixels located far from forest areas. Consequently, we obtained large cost distance values between populations separated by the isthmus, around which forest areas were much less frequent than agricultural and artificial areas. This did not allow for predicting well the genetic distances between these populations. Yet, the cost scenario based on specialization indices obtained using the same conversion function led to better predictions of these same genetic distances (highest validation  $R^2$  for a complete topology). This means that this difference was not due to the negative exponential function in itself, but rather to the range of values to which it is applied. Because the transformation of expert opinion or empirical data into cost values is key to connectivity modeling, future studies might consider comparing different types of transformation for identifying best practices.

The scale at which the connections between populations or patches were considered largely affected the results. The set of links considered for predicting genetic distances from cost distances largely influenced the quality of the predictions, and in differing ways depending on the cost scenario. The inability of our models to predict the genetic distances separating populations from BT and GT may lie on the fact that at large scales, the influence of drift on genetic differentiation may be much larger than that of gene flow (Hutchison & Templeton, 1999), thereby preventing models based on cost distances from performing well. Although these complex results partly stem from the specific topography and spatial distribution of the Guadeloupe forests, they recall the importance of the topology of the population networks for the reliable assessment of landscape genetic relationships (Savary et al., 2021a; Van Strien, 2017).

### Implications for biodiversity conservation and landscape planning

Our results confirm the relevance of the wide range of graph-based methods specifically developed for conservation purposes (Foltête et al., 2014; Zetterberg et al., 2010). We further illustrated how landscape graphs can be used to address conservation issues whether genetic data are available or not.

Genetic data allowed us to assess the ecological validity of each graph element. First, our graph-based approach showed that the genetic diversity of plumbeous warbler populations was positively affected by the availability of forest areas at a limited radius around populations. The lower genetic diversity observed in forest patches located on the margins of the large BT forest patches indicated that this species is negatively affected by the breaking apart of forest patches, even when large forest areas are reachable nearby. Together, these results suggest the following conservation measures for maintaining genetic diversity in the plumbeous warbler populations of Guadeloupe: protect forest patches in northwestern GT; restore connectivity on BT between the small peripheral forest patches and the large central patches; and restore forest patches on the western part of the isthmus between BT and GT to allow long-distance gene flow (Frankham et al., 2017).

We also distinguished populations whose genetic differentiation mostly originated from the habitat pattern and matrix resistance from populations associated with high prediction errors where other factors could explain it. In the latter case, complementary field surveys should be carried out, for example, to identify other factors that could reduce population sizes (e.g., antagonistic interspecific interactions) or gene flow (unmodeled landscape barriers) and explain genetic structure. This illustrates that making the model limitations spatially explicit with predictive approaches when genetic data are available could expand the usefulness of the landscape graph approach for conservation modeling.

In contrast, when genetic data are not available, the construction and analysis parameters cannot be optimized, but we showed that the set of parameters we used still led to cost distances and metrics that significantly explained the genetic

responses. Thus, although landscape graph modeling should preferably be used for studying specialist species occupying discrete habitat patches (Urban & Keitt, 2001), it could continue to be a useful tool for addressing conservation issues in a wide range of species for which genetic data acquisition is difficult or expensive. The availability and the quality of either information from the literature or expert opinion or empirical data should determine the graph construction method because we found that empirical approaches were not necessarily the most cost-effective. In particular, using an SDM for constructing a landscape graph should be restricted to the cases where expert opinion and land-cover data do not allow for delineating habitat patches or setting cost values.

### ACKNOWLEDGMENTS

We thank the editor and reviewers for their constructive comments and suggestions. This work is part of the project CANON steered by S.G. and supported by the French Investissements d'Avenir program, project ISITE-BFC (contract ANR-15-IDEX-0003). This work also received financial support from the Direction de l'Environnement, de l'Aménagement et du Logement (DEAL) of Guadeloupe (DEAL/RN number 971-2019-07-12-003). P.S. was supported by the company ARP-Astrance under a CIFRE contract supervised and partly funded by the ANRT (Association Nationale de la Recherche et de la Technologie). The point count surveys carried out between 2009 and 2011 received support from the Fonds Européens de Développement Régional (FEDER number 2/2.4/-30528), the Direction de l'Environnement, de l'Aménagement et du Logement de la Guadeloupe (DEAL Guadeloupe; Convention number 2009-033/SAPNSP/C), and the Parc National de la Guadeloupe. We are grateful to all observers from several organizations (Parc National de la Guadeloupe, Office Français de la Biodiversité, AMAZONA, BIOS) for performing count surveys and to A. Charbonnel, E. Arnoux, H. Gadenne, A. Levesque, and L. Malglaive for compiling and managing count data. We also thank G. Mourgaud, T. Pagnon, C. Sineau, J. Abbou, C. Fagot, A. Vong, L. Pinto, and G. Leblond for providing assistance with bird mist-netting. We thank C. Serre for her contribution to the production of genetic data, and the Conseil Départemental de la Guadeloupe for providing the vegetation formation map.

### REFERENCES

- Angelone, S., & Holderegger, R. (2009). Population genetics suggests effectiveness of habitat connectivity measures for the European tree frog in Switzerland. *Journal of Applied Ecology*, 46(4), 879–887.
- Balbi, M., Ernoult, A., Poli, P., Madec, L., Guiller, A., Martin, M.-C., Nabucet, J., Beaujouan, V., & Petit, E. J. (2018). Functional connectivity in replicated urban landscapes in the land snail (*Cornu aspersum*). *Molecular Ecology*, 27(6), 1357–1370.
- Baranyi, G., Saura, S., Podani, J., & Jordán, F. (2011). Contribution of habitat patches to network connectivity: Redundancy and uniqueness of topological indices. *Ecological Indicators*, 11(5), 1301–1310.
- Beier, P., Majka, D. R., & Spencer, W. D. (2008). Forks in the road: Choices in procedures for designing wildland linkages. *Conservation Biology*, 22(4), 836–851.
- Bertin, A., Gouin, N., Baumel, A., Gianoli, E., Serratos, J., Osorio, R., & Manel, S. (2017). Genetic variation of loci potentially under selection confounds

- species–genetic diversity correlations in a fragmented habitat. *Molecular Ecology*, 26(2), 431–443.
- Bourdouxhe, A., Dufloot, R., Radoux, J., & Duffrène, M. (2020). Comparison of methods to model species habitat networks for decision-making in nature conservation: The case of the Wildcat in southern Belgium. *Journal for Nature Conservation*, 58, 125901.
- Calabrese, J. M., & Fagan, W. F. (2004). A comparison–shopper's guide to connectivity metrics. *Frontiers in Ecology and the Environment*, 2(10), 529–536.
- Callens, T., Galbusera, P., Matthysen, E., Durand, E. Y., Githiru, M., Huyghe, J. R., & Lens, L. (2011). Genetic signature of population fragmentation varies with mobility in seven bird species of a fragmented Kenyan cloud forest. *Molecular Ecology*, 20(9), 1829–1844.
- Carroll, C., McRae, B., & Brookes, A. (2012). Use of linkage mapping and centrality analysis across habitat gradients to conserve connectivity of gray wolf populations in western North America. *Conservation Biology*, 26(1), 78–87.
- Castillo, J. A., Epps, C. W., Jeffress, M. R., Ray, C., Rodhouse, T. J., & Schwalm, D. (2016). Replicated landscape genetic and network analyses reveal wide variation in functional connectivity for American pikas. *Ecological Applications*, 26(6), 1660–1676.
- Clarke, R. T., Rothery, P., & Raybould, A. F. (2002). Confidence limits for regression relationships between distance matrices: Estimating gene flow with distance. *Journal of Agricultural Biological and Environmental Statistics*, 7(3), 361–372.
- Correa Ayram, C. A., Mendoza, M. E., Etter, A., & Salicrup, D. R. P. (2016). Habitat connectivity in biodiversity conservation: A review of recent studies and applications. *Progress in Physical Geography*, 40(1), 7–37.
- Creech, T. G., Epps, C. W., Monello, R. J., & Wehausen, J. D. (2014). Using network theory to prioritize management in a desert bighorn sheep metapopulation. *Landscape Ecology*, 29(4), 605–619.
- Cushman, S. A., McKelvey, K. S., Hayden, J., & Schwartz, M. K. (2006). Gene flow in complex landscapes: Testing multiple hypotheses with causal modeling. *The American Naturalist*, 168(4), 486–499.
- Emaresi, G., Pellet, J., Dubey, S., Hirzel, A. H., & Fumagalli, L. (2011). Landscape genetics of the Alpine newt (*Mesotriton alpestris*) inferred from a strip-based approach. *Conservation Genetics*, 12(1), 41–50.
- Epps, C. W., & Keyghobadi, N. (2015). Landscape genetics in a changing world: Disentangling historical and contemporary influences and inferring change. *Molecular Ecology*, 24(24), 6021–6040.
- Fletcher, R., & Fortin, M.-J. (2018). *Spatial ecology and conservation modeling*. Springer.
- Fletcher, R. J., Burrell, N. S., Reichert, B. E., Vasudev, D., & Austin, J. D. (2016). Divergent perspectives on landscape connectivity reveal consistent effects from genes to communities. *Current Landscape Ecology Reports*, 1(2), 67–79.
- Foltête, J.-C., Clauzel, C., & Vuidel, G. (2012). A software tool dedicated to the modelling of landscape networks. *Environmental Modelling & Software*, 38, 316–327.
- Foltête, J.-C., Girardet, X., & 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., Sahraoui, Y., Vuidel, G., & Garnier, S. (2020). Coupling landscape graph modeling and biological data: A review. *Landscape Ecology*, 35(5), 1035–1052.
- Frankham, R. (2015). Genetic rescue of small inbred populations: Meta-analysis reveals large and consistent benefits of gene flow. *Molecular Ecology*, 24(11), 2610–2618.
- Frankham, R., Ballou, J. D., Ralls, K., Eldridge, M. D. B., Dudash, M. R., Fenster, C. B., Lacy, R. C., & Sunnucks, P. (2017). *Genetic management of fragmented animal and plant populations*. Oxford University Press.
- Galpern, P., Manseau, M., & Fall, A. (2011). Patch-based graphs of landscape connectivity: A guide to construction, analysis and application for conservation. *Biological Conservation*, 144(1), 44–55.
- Godet, C., & Clauzel, C. (2021). Comparison of landscape graph modelling methods for analysing pond network connectivity. *Landscape Ecology*, 36(3), 735–748.
- Gurrutxaga, M., Lozano, P. J., & del Barrio, G. (2010). GIS-based approach for incorporating the connectivity of ecological networks into regional planning. *Journal for Nature Conservation*, 18(4), 318–326.
- Hanski, I., & Gilpin, M. (1991). Metapopulation dynamics: Brief history and conceptual domain. *Biological Journal of the Linnean Society*, 42(1–2), 3–16.
- Hutchison, D. W., & 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.
- Jacobs, J. (1974). Quantitative measurement of food selection. *Oecologia*, 14(4), 413–417.
- Kadoya, T. (2009). Assessing functional connectivity using empirical data. *Population Ecology*, 51(1), 5–15.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647–1649.
- Keeley, A. T., Beier, P., & Gagnon, J. W. (2016). Estimating landscape resistance from habitat suitability: Effects of data source and nonlinearities. *Landscape Ecology*, 31(9), 2151–2162.
- Keller, D., Holderegger, R., van Strien, M. J., & Bolliger, J. (2015). How to make landscape genetics beneficial for conservation management? *Conservation Genetics*, 16(3), 503–512.
- Khimoun, A., Arnoux, E., Martel, G., Pot, A., Eraud, C., Condé, B., Loubon, M., Théron, F., Covas, R., Faivre, B., & Garnier, S. (2016). Contrasted patterns of genetic differentiation across eight bird species in the Lesser Antilles. *Genetica*, 144(1), 125–138.
- Khimoun, A., Eraud, C., Ollivier, A., Arnoux, E., Rocheteau, V., Bely, M., Lefol, E., Delpuech, M., Carpentier, M.-L., Leblond, G., Levesque, A., Charbonnel, A., Faivre, B., & Garnier, S. (2016). Habitat specialization predicts genetic response to fragmentation in tropical birds. *Molecular Ecology*, 25(16), 3831–3844.
- Khimoun, A., Peterman, W., Eraud, C., Faivre, B., Navarro, N., & Garnier, S. (2017). Landscape genetic analyses reveal fine-scale effects of forest fragmentation in an insular tropical bird. *Molecular Ecology*, 26(19), 4906–4919.
- Landguth, E., Cushman, S., Schwartz, M., McKelvey, K., Murphy, M., & Luikart, G. (2010). Quantifying the lag time to detect barriers in landscape genetics. *Molecular Ecology*, 19(19), 4179–4191.
- Laroche, F., Balbi, M., Grébert, T., Jabot, F., & Archaux, F. (2020). Three points of consideration before testing the effect of patch connectivity on local species richness: Patch delineation, scaling and variability of metrics. [bioRxiv. https://doi.org/10.1101/640995](https://doi.org/10.1101/640995)
- Lovette, I. J., Bermingham, E., Seutin, G., & Ricklefs, R. E. (1998). Evolutionary differentiation in three endemic West Indian warblers. *The Auk*, 115(4), 890–903.
- Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, 405(2), 442–451.
- McPherson, J. M., & Jetz, W. (2007). Effects of species' ecology on the accuracy of distribution models. *Ecography*, 30(1), 135–151.
- Moilanen, A. (2011). On the limitations of graph-theoretic connectivity in spatial ecology and conservation. *Journal of Applied Ecology*, 48(6), 1543–1547.
- Mony, C., Abadie, J., Gil-Tena, A., Burel, F., & Ernout, A. (2018). Effects of connectivity on animal-dispersed forest plant communities in agriculture-dominated landscapes. *Journal of Vegetation Science*, 29(2), 167–178.
- Nevil Amos, J., Harrisson, K. A., Radford, J. Q., White, M., Newell, G., Nally, R. M., Sunnucks, P., & Pavlova, A. (2014). Species- and sex-specific connectivity effects of habitat fragmentation in a suite of woodland birds. *Ecology*, 95(6), 1556–1568.
- Perrin, A., Khimoun, A., Ollivier, A., Richard, Y., Pérez-Rodríguez, A., Faivre, B., & Garnier, S. (2022). Habitat fragmentation matters more than habitat loss: The case of host-parasite interactions. *Molecular Ecology*, 1000, 1–19. <https://doi.org/10.1111/mec.16807>
- Peterman, W. E., Anderson, T. L., Ousterhout, B. H., Drake, D. L., Semlitsch, R. D., & Eggert, L. S. (2015). Differential dispersal shapes population structure and patterns of genetic differentiation in two sympatric pond breeding salamanders. *Conservation Genetics*, 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.

- Raymond, M., & Rousset, F. (1995). GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenism. *Journal of Heredity*, *86*, 248–249.
- Ribeiro, R., Carretero, M. A., Sillero, N., Alarcos, G., Ortiz-Santaliestra, M., Lizana, M., & Llorente, G. A. (2011). The pond network: Can structural connectivity reflect on (amphibian) biodiversity patterns? *Landscape Ecology*, *26*(5), 673–682.
- Ricklefs, R. E., & Bermingham, E. (2004). History and the species-area relationship in Lesser Antillean birds. *The American Naturalist*, *163*(2), 227–239.
- Robertson, E. P., Fletcher, R. J., & Austin, J. D. (2018). Microsatellite polymorphism in the endangered snail kite reveals a panmictic, low diversity population. *Conservation Genetics*, *19*(2), 337–348.
- Robertson, E. P., Fletcher, R. J., Cattau, C. E., Udell, B. J., Reichert, B. E., Austin, J. D., & Valle, D. (2018). Isolating the roles of movement and reproduction on effective connectivity alters conservation priorities for an endangered bird. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(34), 8591–8596.
- Saura, S., & de la Fuente, B. (2017). Connectivity as the amount of reachable habitat: Conservation priorities and the roles of habitat patches in landscape networks. In S. E. Gergel & M. G. Turner (Eds.), *Learning landscape ecology: A practical guide to concepts and techniques* (pp. 229–254). Springer.
- Savary, P., Foltête, J.-C., Moal, H., Vuidel, G., & Garnier, S. (2021a). Analysing landscape effects on dispersal networks and gene flow with genetic graphs. *Molecular Ecology Resources*, *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 in Ecology and Evolution*, *12*(3), 539–547.
- Savary, P., Foltête, J.-C., van Strien, M. J., Moal, H., Vuidel, G., & Garnier, S. (2022). Assessing the influence of the amount of reachable habitat on genetic structure using landscape and genetic graphs. *Heredity*, *128*(2), 120–131.
- Sawyer, S. C., Epps, C. W., & Brashares, J. S. (2011). Placing linkages among fragmented habitats: Do least-cost models reflect how animals use landscapes? *Journal of Applied Ecology*, *48*(3), 668–678.
- Stevenson-Holt, C. D., Watts, K., Bellamy, C. C., Nevin, O. T., & Ramsey, A. D. (2014). Defining landscape resistance values in least-cost connectivity models for the invasive grey squirrel: A comparison of approaches using expert-opinion and habitat suitability modelling. *PLoS ONE*, *9*(11), e112119.
- Szpiech, Z. A., Jakobsson, M., & Rosenberg, N. A. (2008). ADZE: A rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*, *24*(21), 2498–2504.
- Taylor, P. D., Fahrig, L., & With, K. A. (2006). Landscape connectivity: A return to the basics. In K. R. Crooks & M. Sanjayan (Eds.), *Connectivity conservation* (pp. 29–43). Cambridge University Press.
- Urban, D., & Keitt, T. (2001). Landscape connectivity: A graph-theoretic perspective. *Ecology*, *82*(5), 1205–1218.
- 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., & Van Heck, H. J. (2015). Isolation-by-distance in landscapes: Considerations for landscape genetics. *Heredity*, *114*(1), 27–37.
- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370.
- Zeller, K. A., Jennings, M. K., Vickers, T. W., Ernest, H. B., Cushman, S. A., & Boyce, W. M. (2018). Are all data types and connectivity models created equal? validating common connectivity approaches with dispersal data. *Diversity and Distributions*, *24*(7), 868–879.
- Zeller, K. A., McGarigal, K., & Whiteley, A. R. (2012). Estimating landscape resistance to movement: A review. *Landscape Ecology*, *27*(6), 777–797.
- Zetterberg, A., Mörtberg, U. M., & Balfors, B. (2010). Making graph theory operational for landscape ecological assessments, planning, and design. *Landscape and Urban Planning*, *95*(4), 181–191.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Daniel, A., Savary, P., Foltête, J.-C., Khimoun, A., Faivre, B., Ollivier, A., Éraud, C., Moal, H., Vuidel, G., & Garnier, S. (2023). Validating graph-based connectivity models with independent presence-absence and genetic data sets. *Conservation Biology*, e14047. <https://doi.org/10.1111/cobi.14047>