



**HAL**  
open science

# Évaluation de la validité des modèles de connectivité des habitats à partir de données génétiques

Alexandrine Daniel

► **To cite this version:**

Alexandrine Daniel. Évaluation de la validité des modèles de connectivité des habitats à partir de données génétiques. Environnement et Société. Université Bourgogne Franche-Comté, 2024. Français. NNT : 2024UBFCK061 . tel-04913394

**HAL Id: tel-04913394**

**<https://theses.hal.science/tel-04913394v1>**

Submitted on 27 Jan 2025

**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.

**UBFC**

UNIVERSITÉ  
BOURGOGNE FRANCHE-COMTÉ



**THESE DE DOCTORAT DE L'ETABLISSEMENT UNIVERSITE BOURGOGNE  
FRANCHE-COMTE**

**PREPAREE A L'UMR 6282 Biogéosciences**

Ecole doctorale n°554, Environnement - Santé



ÉCOLE DOCTORALE  
Environnements - Santé  
Bourgogne | Franche-Comté

Doctorat de Biologie des populations et écologie

# **Évaluation de la validité des modèles de connectivité des habitats à partir de données génétiques**

Par

**Alexandrine DANIEL**

Thèse présentée et soutenue à Dijon, le 27 septembre 2024.

Composition du Jury :

<b>Laurence Després</b>	Professeure des universités, Université Grenoble Alpes	Rapportrice - Présidente
<b>Marie-Josée Fortin</b>	<i>University Professor</i> , Université de Toronto	Rapportrice
<b>Bertrand Gauffre</b>	Chargé de recherche, INRAE PACA	Examineur
<b>Alain Frantz</b>	Conservateur de zoologie, MNHN Luxembourg	Examineur
<b>Bruno Faivre</b>	Professeur des universités, Université de Bourgogne	Directeur de thèse
<b>Aurélié Khimoun</b>	Maîtresse de conférences, Université de Bourgogne	Codirectrice de thèse
<b>Stéphane Garnier</b>	Maître de conférences, Université de Bourgogne	Invité



# TABLE DES MATIÈRES

Table des matières	iii
Liste des figures	xii
Liste des tables	xv
Résumés	xvii
Remerciements	xix
Avant-propos	xxiii

## 1. INTRODUCTION GÉNÉRALE ..... 1

### 1.1. Génétique du paysage, sa place dans notre compréhension de la biodiversité ..... 2

1.1.1. Origine et concepts de la génétique du paysage ..... 2

1.1.2. Interactions entre la structure du paysage et la variabilité génétique neutre et adaptative des populations ..... 3

1.1.2.1. Variabilité neutre, dérive et flux génétique, et connectivité ..... 5

1.1.2.2. Diversité adaptative, sélection naturelle et potentiel adaptatif ..... 6

1.1.3. Applications des inférences de génétique du paysage ..... 9

### 1.2. Modéliser les effets de la connectivité fonctionnelle des habitats sur la variabilité génétique neutre ..... 12

1.2.1. Les surfaces de résistance : un outil de modélisation des connectivités fonctionnelles. 12

1.2.1.1. Construction et paramétrage des surfaces de résistance ..... 12

1.2.1.2. Utilisation des surfaces de résistance ..... 16

1.2.2. Typologies de modèles en génétique du paysage ..... 17

1.2.2.1. La population ou l'individu comme unité d'analyse ..... 17

1.2.2.2. Les différents niveaux de modélisation en génétique du paysage ..... 17

1.2.3. Limites des modèles de génétique du paysage ..... 29

1.2.3.1. Limites liées au paramétrage et à la validité des modèles de génétique du paysage ..... 29

1.2.3.2. Limites liées au domaine de validité des modèles de génétique du paysage : la question de l'extrapolation.....	31
1.2.3.3. Limites liées à la traduction en mesure de conservation des inférences de génétique du paysage.....	33
<b>1.3. Objectifs de ce manuscrit de thèse.....</b>	<b>34</b>
<b>2. MÉTHODOLOGIE GÉNÉRALE.....</b>	<b>37</b>
<b>2.1. Cadre de recherche : projets FRAG&amp;BINV et CANON .....</b>	<b>38</b>
<b>2.2. Zone d'étude : la Guadeloupe .....</b>	<b>39</b>
<b>2.3. Modèles d'étude : cinq espèces d'oiseaux tropicaux.....</b>	<b>42</b>
<b>2.4. Échantillonnage .....</b>	<b>43</b>
<b>2.5. Analyses moléculaires .....</b>	<b>45</b>
<b>2.6. Génétique des populations – Analyses préliminaires.....</b>	<b>46</b>
2.6.1.1. Contrôle de l'évolution temporelle de la diversité génétique.....	46
2.6.1.2. Polymorphisme des loci.....	48
2.6.1.3. Equilibre de Hardy-Weinberg.....	48
2.6.1.4. Déséquilibre de liaison .....	49
<b>2.7. Données spatiales d'occupation du sol .....</b>	<b>50</b>
<b>3. VALIDITÉ DES MODÈLES DE L'EFFET DE LA CONNECTIVITÉ SUR LA VARIABILITÉ GÉNÉTIQUE.....</b>	<b>53</b>
<b>3.1. Validation par simulation d'une méthode d'optimisation de distances-coût à partir de données génétiques.....</b>	<b>55</b>
3.1.1. Introduction .....	55
3.1.2. Matériels et méthodes .....	55
3.1.3. Résultats principaux .....	56
3.1.4. Article : What can optimized cost distances based on genetic distances offer ?.....	57
3.1.4.1. Abstract.....	58
3.1.4.2. Introduction.....	59

3.1.4.3. Materials & Methods .....	63
3.1.4.3.1. Real landscape sampling .....	63
3.1.4.3.2. True resistance surfaces parametrization and cost-distances calculation....	63
3.1.4.3.3. Demo-genetics simulations and gene pool sampling .....	64
3.1.4.3.4. Spatial scale of landscape influence on genetic structure .....	64
3.1.4.3.5. RGA workflow and new implementations.....	65
3.1.4.3.6. Reliability of cost inferences from RGA optimization .....	66
3.1.4.4. Results.....	70
3.1.4.4.1. Landscape sampling and demo-genetic simulations .....	70
3.1.4.4.2. Optimization sensitivity and specificity to simulated genetic patterns .....	71
3.1.4.4.3. Assessment of RGA predictive performances.....	71
3.1.4.4.4. The congruence between optimized and true cost scenarios.....	76
3.1.4.5. Discussion.....	79
3.1.4.5.1. The spatial scale of landscape influence on genetic structure depends on the topology of the effective dispersal network.....	79
3.1.4.5.2. Pruning can enhance RGA predictive performance under strong dispersal limitation.....	80
3.1.4.5.3. Sampling design and landscape structure affect optimization performance	81
3.1.4.5.4. Optimization is not causation: good predictive abilities at the expense of accurate causal interpretation.....	82
3.1.4.5.5. The use of RGA in landscape genetics: conditions and prospects .....	86
3.1.4.6. Acknowledgements.....	86
<b>3.2. Validation empirique des modèles de graphes paysagers reflétant l'effet de la connectivité des habitats sur le flux génétique .....</b>	<b>87</b>
3.2.1. Introduction .....	87
3.2.2. Matériels et méthodes .....	87
3.2.3. Résultats principaux .....	88
3.2.4. Article : Validating graph-based connectivity models with independent presence-absence and genetic data sets.....	89
3.2.4.1. Abstract.....	90

3.2.4.2. Introduction.....	91
3.2.4.3. Methods .....	92
3.2.4.3.1. Study area and biological model .....	92
3.2.4.3.2. Spatial and climatic data .....	92
3.2.4.3.3. Expert opinion.....	92
3.2.4.3.4. Habitat specialization indices.....	93
3.2.4.3.5. Species distribution model .....	95
3.2.4.3.6. Landscape graphs .....	95
3.2.4.3.7. Fiel sampling and genotyping .....	95
3.2.4.3.8. Genetic structure indies.....	96
3.2.4.3.9. Validation of cost distances and connectivity metrics associated with landscape graph links and nodes.....	96
3.2.4.4. Results.....	96
3.2.4.4.1. Graph construction parameters.....	96
3.2.4.4.2. Graph element properties .....	97
3.2.4.4.3. Genetic structure .....	97
3.2.4.4.4. Validation of cost distances associated with landscape graph links.....	97
3.2.4.4.5. Validation of connectivity metrics associated with landscape graph nodes.....	98
3.2.4.5. Discussion .....	100
3.2.4.5.1. Empirical validation of landscape graphs by genetic data .....	100
3.2.4.5.2. Relationships among graph ecological relevance, data requirements, and construction and analysis methods.....	101
3.2.4.5.3. Implications for biodiversity conservation.....	102
3.2.4.6. Acknowledgments .....	103
3.2.4.7. References.....	103
3.2.4.8. Supporting information .....	104
<b>3.3. Conclusions .....</b>	<b>105</b>

<b>4. DOMAINE DE VALIDITÉ MULTI-SPÉCIFIQUE D’UN MODÈLE DE GÉNÉTIQUE DU PAYSAGE.....</b>	<b>109</b>
<b>4.1. Introduction.....</b>	<b>110</b>
<b>4.2. Matériels et méthodes .....</b>	<b>111</b>
<b>4.3. Résultats principaux .....</b>	<b>112</b>
<b>4.4. Article : Is landscape effect on gene flow shared across species? .....</b>	<b>113</b>
4.4.1. Abstract .....	114
4.4.2. Introduction .....	115
4.4.3. Materials & Methods.....	118
4.4.3.1. Species and study area .....	118
4.4.3.2. Sampling design and molecular analysis .....	118
4.4.3.3. Population genetic analyses .....	119
4.4.3.4. Spatial data and landcover map .....	121
4.4.3.5. Optimization of cost scenarios with ResistanceGA and computation of cost distances .....	121
4.4.3.6. Habitat connectivity models from landscape graphs and optimized cost scenarios .....	123
4.4.4. Results .....	124
4.4.4.1. Population genetic results .....	124
4.4.4.2. Optimized cost scenarios from the genetic data of the five studied species .....	129
4.4.4.3. Interspecies cross-validation for genetic distance modelling .....	130
4.4.4.3.1. Whole Guadeloupe scale.....	131
4.4.4.3.2. Within-island scale.....	131
4.4.4.4. Landscape graph properties and PCA analysis on Flux values .....	133
4.4.5. Discussion .....	135
4.4.5.1. Interspecific cross-validation of genetic structure predictability .....	135
4.4.5.2. Population genetic structure across species spanning a forest specialization gradient .....	135
4.4.5.3. The conservation ecology value of optimising resistance zones using genetic data .....	138



4.4.5.4. Landscape graph derived from optimized cost scenarios: implications for conservation .....	140
<b>4.5. Conclusions .....</b>	<b>141</b>
<b>5. DOMAINE DE VALIDITÉ TEMPOREL D'UN MODÈLE DE GÉNÉTIQUE DU PAYSAGE.....</b>	<b>145</b>
<b>5.1. Influence des effets phylogéographiques sur la distribution contemporaine de la variabilité génétique.....</b>	<b>147</b>
5.1.1. Origine historique et contemporaine de la structure génétique.....	147
5.1.2. Génétique du paysage et phylogéographie .....	148
5.1.3. Histoire de colonisation des Petites Antilles par les passereaux .....	150
5.1.4. Les outils de la phylogéographie : modèles de coalescence et approche 'ABC' .....	152
5.1.5. Inférer l'histoire évolutive des populations de Paruline caféïette en Guadeloupe .....	158
5.1.5.1. Paramètres populationnels .....	158
5.1.5.2. Echantillonnage, extraction d'ADN et séquençage RAD-seq .....	159
5.1.5.3. Travail en cours .....	160
<b>5.2. Équilibre temporel de l'interaction entre variabilité génétique et variabilité paysagère.....</b>	<b>162</b>
5.2.1. Introduction .....	162
5.2.2. Matériels et méthodes .....	163
5.2.3. Résultats principaux .....	164
5.2.4. Article : Disentangling past landscape legacies from the contemporary genetic structure of a forest bird using a historical map .....	164
5.2.4.1. Abstract.....	165
5.2.4.2. Introduction.....	166
5.2.4.3. Materials and Methods.....	169
5.2.4.3.1. Study area.....	169
5.2.4.3.2. Study model .....	170
5.2.4.3.3. Sampling, molecular analysis and genetic structure index.....	170
5.2.4.3.4. Construction of the historical and current land cover maps .....	171

5.2.4.3.5. Landscape description .....	172
5.2.4.3.6. Modelling landscape resistance: systematic screening of cost scenarios and resistance distances .....	172
5.2.4.3.7. Statistical analysis for current and historical model comparison .....	173
5.2.4.4. Results.....	175
5.2.4.4.1. Changes in the landscape between 1950 and 2012 .....	175
5.2.4.4.2. Population genetic structure .....	175
5.2.4.4.3. Identification of best cost scenarios for landscape resistance modelling with the whole-scale dataset .....	178
5.2.4.4.4. Identification of best cost scenarios for landscape resistance modelling with the intra-island-scale dataset .....	180
5.2.4.5. Discussion.....	180
5.2.4.6. Perspectives .....	184
<b>5.3. Conclusions .....</b>	<b>185</b>
<b>6. DISCUSSION GÉNÉRALE .....</b>	<b>187</b>
<b>6.1. Valider des modèles de connectivités, comment et pourquoi.....</b>	<b>190</b>
6.1.1. Validation de ResistanceGA à partir de données simulées .....	192
6.1.1.1. Apport méthodologique de notre approche de validation.....	192
6.1.1.2. Les critères de validité des inférences de RGA mis en perspective avec sa qualité d'apprentissage.....	194
6.1.1.3. Les limites de notre étude .....	200
6.1.2. Validation empirique des modèles de connectivité.....	202
6.1.2.1. Calibration du modèle : comment modéliser le processus de dispersion en interaction avec le paysage ?.....	204
6.1.2.2. Qualité des données de validation : les données empiriques, gage d'une validation indépendante reflétant la réalité du terrain.....	207
6.1.2.3. Conclusions et perspectives quant à la validité des graphes paysagers .....	208
<b>6.2. Domaine de validité spatial, temporel et multi-spécifique des modèles de connectivité .....</b>	<b>210</b>

6.2.1.Considération de l'échelle spatiale dans les modèles de connectivité .....	212
6.2.1.1. Résolution thématique .....	212
6.2.1.2. 'Scale of effect' et seuil de distances.....	212
6.2.1.3. Niveaux d'analyse .....	214
6.2.1.4. L'apport des graphes paysagers pour la prise en compte de l'échelle spatiale dans les inférences de génétique du paysage.....	214
6.2.2.2.Considération de l'échelle temporelle dans les modèles des connectivité .....	218
6.2.2.1. Dynamique temporelle des paysages et des déplacements d'individus.....	218
6.2.2.2. Délais de réponse de la variabilité génétique à une perturbation paysagère et héritage des paysages passés.....	218
6.2.2.3. Intégrer les effets phylogéographiques dans les modèles de génétique du paysage .....	220
6.2.3.3.Modèles de connectivité multi-spécifiques .....	221
<b>6.3. Perspectives pour des modèles de connectivité valides.....</b>	<b>223</b>
6.3.1.Vers une vision dynamique des modèles de connectivité .....	225
6.3.2.Vers une diversification des variables explicatives de la connectivité.....	226
6.3.3.Modéliser l'influence du paysage sur la diversité adaptative.....	227
6.3.3.1. Interactions entre flux génétique et adaptation locale.....	227
6.3.3.2. Comparaison de la variabilité neutre et adaptative .....	228
6.3.3.3. Association entre gradient environnemental et variabilité adaptative .....	228
6.3.3.4. Perspectives de recherche en Guadeloupe .....	229
6.3.4.Traduction des inférences en mesures de conservation.....	230
6.3.4.1. Les apports de cette thèse à la conservation .....	230
6.3.4.2. La génétique du paysage face aux changements globaux.....	233
<b>7. CONCLUSIONS GÉNÉRALES.....</b>	<b>237</b>
<b>8. RÉFÉRENCES.....</b>	<b>243</b>

<b>9. ANNEXES.....</b>	<b>269</b>
9.1. Annexe 1 : Analyses moléculaires .....	270
9.2. Annexe 2 : Supplemental Information for: “ <i>What can optimized cost distances based on genetic distances offer? A simulation study on the use and misuse of ResistanceGA</i> ” .....	275
9.3. Annexe 3 : Supplemental Information for: “ <i>Is landscape effect on gene flow shared across species? A multispecies cross-validation landscape genetic study</i> ” .....	291
9.4. Annexe 4 : Supplemental Information for: “ <i>Disentangling past landscape legacies from the contemporary genetic structure of a forest bird using a historical map</i> ” .....	299
9.5. Annexe 5: Rapport « Etude de la connectivité des habitats forestiers entre la Grande-Terre et la Basse-Terre » .....	305

# LISTE DES FIGURES

## 1. Chapitre 1

- 1.1. **Figure 1.1** : Interactions entre la dérive, le flux génétique, et le patron d'hétérogénéité du paysage, et leurs effets sur le patron de variabilité génétique neutre. .... 4
- 1.2. **Figure 1.2** : Les différentes étapes de la paramétrisation d'une surface de résistance à partir de données biologiques ..... 13
- 1.3. **Figure 1.3** : Deux niveaux de modélisation en génétique du paysage (d'après la Figure 1 de DiLeo et Wagner (2016)) ..... 18
- 1.4. **Figure 1.4** : Modèle des distances de moindre coût..... 20
- 1.5. **Figure 1.5** : Modèle des distances de résistance, d'après McRae et al. (2009)..... 22
- 1.6. **Figure 1.6** : Construction des graphes paysagers, et métriques de connectivité en dérivant..... 28

## 2. Chapitre 2

- 2.1. **Figure 2.1** : Cartographie des îles de la Guadeloupe, et des fragments forestiers échantillonnés en 2012-2013 et en 2020..... 40
- 2.2. **Figure 2.2** : Description des cinq espèces de passereaux échantillonnées ..... 44

## 3. Chapitre 3

- 3.1. **Figure 3.1.1**: Overall methodology for assessing the ability of optimized cost distances to reliably reflect landscape effect on genetic structure. .... 62
- 3.2. **Figure 3.1.2**: Diagram summarizing the successive stages of data filtering throughout the analyses. . 66
- 3.3. **Figure 3.1.3**: Comparison of predictive performance of RGA models based either on the sampling-scale or the effective-dispersal-scale datasets, as a function of dispersal capacities and specialization levels. .... 72
- 3.4. **Figure 3.1.4**: Performance criteria as a function of dispersal capacities and specialization levels. Only sampling-scale datasets coming from true positive runs have been considered here. .... 74
- 3.5. **Figure 3.1.5**: Evaluation of the interpolation and the extrapolation capacities of the optimized IBR models for a specialist species. .... 75
- 3.6. **Figure 3.1.6**: Comparison of predictive performances (assessed by the validation R<sup>2</sup> of IBR models) between the true and the optimized cost scenarios for a specialist species. .... 76

- 3.7. **Figure 3.1.7:** Assignment of each land cover type to the 4 ranks of relative resistance in the optimized cost scenarios for A) a generalist species, B) a specialist species, C &D) virtual species with ecological profiles derived from variations in the specialist profile (the order of land use resistance has been reversed as compared with the specialist scenario). ..... 78
- 3.8. **Figure 3.1.8:** A) Characterization of the optimized cost scenarios according to their contrast as a function of the ecological profile, B) Distribution of the ranking consistency with true cost scenarios of optimized runs (i.e., Spearman correlation between the ranking of land covers in the optimized cost scenario and in the true cost scenario) according to the ecological profile ..... 78

## 4. Chapitre 4

- 4.1. **Figure 4.1:** General methodology for the multi-species cross-validation approach. .... 117
- 4.2. **Figure 4.2:** Map of the 9 landcover types considered in this study, with a 100m resolution ..... 120
- 4.3. **Figure 4.3:** Graphical visualization of Isolation-by-Distance (IBD) models explaining linearized FST by the logarithm of geographical distances. .... 125
- 4.4. **Figure 4.4:** Spatialization of the results from the clustering analyses with STRUCTURE for each of the five species ..... 127
- 4.5. **Figure 4.5:** Graphical representation of the results of the multi-species cross-validation models..... 132
- 4.6. **Figure 4.6:** A.) Biplot representation of the PCA of the Flux connectivity metrics associated with the five species, computed for each forest habitat patch. B-C.) Spatialization of these results ..... 134

## 5. Chapitre 5

- 5.1. **Figure 5.1.1 :** Formule du théorème de Bayes pour le calcul de la probabilité postérieure d'un scénario évolutif (Scenario) conditionné aux données génétiques observées (data) ..... 157
- 5.2. **Figure 5.1.2 :** Principes de l'ABC résumé en 3 étapes, d'après Lopes and Beaumont (2010) et Bertorelle, Benazzo, and Mona (2010) ..... 157
- 5.3. **Figure 5.1.3 :** Scénarios évolutifs pressentis pour l'analyse phylogéographique. La probabilité postérieure de ces scénarios sera calculée par méthode ABC avec le logiciel DIYABC. .... 161
- 5.4. **Figure 5.2.1:** General methodology for the diachronic landscape genetics approach. .... 168
- 5.5. **Figure 5.2.2:** Land cover map of Guadeloupe in A) 1950 and B) 2008. The dark dots correspond to the localisations the 27 populations of plumbeus warbler that were sampled in 2012-2013 and 2020. .... 169
- 5.6. **Figure 5.2.3:** Values of FRAGSTAT landscape metrics characterizing both the current and the historical landscape ..... 176
- 5.7. **Figure 5.2.4:** Spatial visualization of genetic clusters inferred by STRUCTURE ..... 177

## 6. Chapitre 6

6.1. <b>Figure 6.1</b> : Schéma général synthétisant les deux approches développées dans cette thèse, concernant l'exploration de l'interaction entre patron et processus en génétique du paysage. ....	189
6.2. <b>Figure 6.2</b> : Focus depuis la figure 6.1 sur l'approche d'inférence inductive .....	191
6.3. <b>Figure 6.3</b> : Focus depuis la figure 6.1 sur l'approche de modélisation. ....	203
6.4. <b>Figure 6.4</b> : Focus depuis la figure 6.1 sur les dynamiques spatiales, temporelles et spécifiques de l'interaction entre patron et processus, contraignant le domaine de validité des modèles de connectivité. ....	211
6.5. <b>Figure 6.5</b> : Focus depuis la figure 6.1 sur les perspectives d'amélioration et d'application des modèles de connectivité en génétique du paysage.....	224

# LISTE DES TABLES

---

## 2. Chapitre 2

- 2.1. **Table 2.1** : Résultats des tests de différenciation génétique dans le temps au sein des 4 - 5 populations échantillonnées en 2012-2013 et en 2020, pour les cinq espèces étudiées ..... 47
- 2.2. **Table 2.2** : Résultats concernant le niveau de variabilité des *loci* génotypés pour les cinq espèces considérées. Nombre moyen d'allèles : moyenne de nombre d'allèles observés à chaque *locus*. ..... 48
- 2.3. **Table 2.3** : Résultats des tests de déviation par rapport à l'équilibre de Hardy-Weinberg, après correction de Bonferroni, pour chaque *locus*, dans chaque population, et pour les cinq espèces considérées ..... 49
- 2.4. **Table 2.4** : Résultats des tests de déséquilibre de liaison entre toutes les paires de *loci* dans chaque population, après correction de Bonferroni. .... 49

## 3. Chapitre 3

- 3.1. **Table 3.1.1**: Results of statistical model explaining the predictive performance of optimized IBR models from the sampling-scale dataset (assessed by the validation R2 index) as a function of the level of specialization, the dispersal capacity and two landscape structure metrics. .... 73

## 4. Chapitre 4

- 4.1. **Table 4.1**: Species specialization index (SSI) based on occurrence data across forest and non-forest patches. .... 117
- 4.2. **Table 4.2**: Graphical visualization of Isolation-by-Distance (IBD) models explaining linearized FST by the logarithm of geographical distances ..... 126
- 4.3. **Table 4.3**: Summary table of the optimization results at the two spatial scales considered and for the five species. .... 130

## 5. Chapitre 5

- 5.1. **Table 5.2.1**: Statistical description of genetic differentiation values between sampled populations of plumbeous warbler ..... 178
- 5.2. **Table 5.2.2**: Table of cost scenarios resulting in the best resistance models according to the AIC criterion for the *whole-scale models*. .... 179
- 5.3. **Table 5.2.3**: Table of cost scenarios resulting in the best resistance models according to the AIC criterion for the *within-island models*. .... 179





## **Titre : Évaluation de la validité des modèles de connectivité des habitats à partir de données génétiques**

**Mots clés :** dispersion, surface de résistance, conservation, oiseaux

**Résumé :** Préserver la connectivité des habitats est un enjeu majeur pour la conservation de la biodiversité. Modéliser la connectivité permet d'améliorer la compréhension de l'interaction entre patrons et processus éco-évolutifs et d'orienter, à terme, les mesures de conservation. Les modèles de connectivité en génétique du paysage traduisent une hypothèse concernant l'action d'un processus évolutif (*e.g.*, le flux génétique) en interaction avec un patron paysager sur la distribution de variabilité génétique. Explorer la validité de ces modèles est nécessaire pour assurer l'efficacité des mesures de conservation en dérivant. L'objectif de cette thèse était d'explorer les bases de cette validité, à savoir : i) la validité des estimations des paramètres de l'interaction patron-processus ; ii) la validité de l'hypothèse sur l'action du processus. Tout d'abord, nous avons validé par simulation l'estimation des paramètres de résistance du paysage au flux génétique réalisée par la méthode d'inférence

ResistanceGA, et avons identifié leur faible pouvoir d'extrapolation. Ensuite, nous avons validé empiriquement la capacité du modèle des graphes paysagers à représenter l'influence de la connectivité sur le patron de variabilité génétique d'une espèce de passereau forestier en contexte insulaire et fragmenté. Enfin, nous avons étudié les domaines de validité spatio-temporels et multi-spécifiques de ces modèles de connectivité. Nous avons ainsi i) confronté un patron empirique contemporain de différenciation génétique à des résultats de modèles de connectivité diachroniques, et ii) démontré que les espèces présentant des habitats similaires partageaient une perception similaire de l'effet de la connectivité sur le flux génétique. Les résultats de cette thèse contribuent ainsi à améliorer certaines approches méthodologiques en génétique du paysage et à affirmer son apport à des fins de conservation de la biodiversité et de la connectivité fonctionnelle des habitats.

## **Title: Evaluating the validity of habitat connectivity models from genetic data**

**Keywords:** dispersal, resistance surface, conservation, birds

Maintaining habitat connectivity is a major challenge for biodiversity conservation. Modelling connectivity allows for a better understanding of the interaction between eco-evolutionary patterns and processes, and ultimately guides conservation actions. Connectivity models in landscape genetics reflect a hypothesis about the effect of an evolutionary process (*e.g.* gene flow) interacting with a landscape pattern on the distribution of genetic variability. Testing the validity of these models is necessary to ensure the relevance of derived conservation actions. The aim of this work was to investigate the foundations of this validity, namely: i) the validity of the parameter estimates of the pattern-process interaction; ii) the validity of the hypothesis about the process under study. First, we validated by simulation the estimation of landscape resistance parameters to gene flow carried out by the

ResistanceGA inference method, and pointed out its low extrapolation power. We then empirically validated the ability of landscape graph models to reflect the influence of connectivity on the genetic structure of a forest passerine species in an insular and fragmented context. Finally, we investigated the spatio-temporal and multi-species validity ranges of these connectivity models. We thus i) compared a contemporary empirical pattern of genetic differentiation with the results of diachronic connectivity models, and ii) demonstrated that species with similar habitats share a similar perception of the effect of connectivity on gene flow. Thus, the results of this thesis contribute to the improvement of methodological approaches in landscape genetics and confirm the contribution of this field to the conservation of biodiversity and functional connectivity of habitats.



# REMERCIEMENTS

---

Je souhaiterais remercier Laurence Després, Marie-Josée Fortin, Alain Frantz et Bertrand Gauffre pour avoir accepté d'évaluer mon travail de thèse en faisant partie de mon jury. Merci à Bertrand Gauffre et Christophe Thébaud d'avoir jalonné ces quatre années de recherche par de précieux conseils et pertinentes discussions dans le cadre de mes comités de thèse.

Aurélie et Stéphane, je vous remercie de m'avoir offert l'opportunité de vivre cette belle aventure de thèse. La génétique des populations était encore pour moi un domaine assez obscur il y a quatre ans, et je vous suis reconnaissante d'avoir cru en moi et de m'avoir accompagné dans mon apprentissage. J'ai découvert une discipline passionnante dans laquelle je n'aspire aujourd'hui qu'à progresser. Aurélie, merci pour ta présence bienveillante et tes conseils, toujours pertinents et efficaces. Merci infiniment d'avoir été dans le temps de la thèse et du doctorant. Stéphane, les longues heures passées dans le bureau N322 à discuter de ResistanceGA m'ont fait appréhender la rigueur scientifique, la précision et le choix des mots. Cet apprentissage m'est précieux et m'accompagnera tout au long de ma vie de chercheuse. Tu as pour cela toute ma gratitude. Bruno, merci beaucoup pour ta grande disponibilité et tes remarques toujours justes et avisés. Je te suis très reconnaissante d'avoir pu compter sur toi, notamment en cette fin de thèse. Anthony, merci beaucoup de m'avoir formé et accompagné dans les manip de biologie moléculaire. Merci pour ta gentillesse et ta patience, à l'épreuve de ma maladresse ! Merci beaucoup également à Coline, pour avoir partagé la paillasse avec moi durant ma première année. Merci d'avoir apporté le rire et le soleil en ces temps sombres de confinement !

Paul, je ne te remercierais jamais assez pour tout ce que tu m'as apporté durant ces quatre années. Tu allies l'extrême bienveillance, l'attention, le bouillonnement d'idées, la force de travail, la pertinence et l'efficacité. Sans toi, cette thèse se serait parée d'une autre couleur. Si je reste encore souvent scotchée derrière ta longueur d'avance, je me suis vue progresser à tes côtés et j'espère du fond du cœur que nos chemins continueront de se croiser.

Jean-Christophe et Gilles, je vous remercie énormément pour votre aide précieuse tout au long de cette thèse. Merci pour votre disponibilité et l'accueil toujours très chaleureux à Théma ! Je remercie également toute l'équipe Graphab, Marc, Céline, François et Lise pour les moments de science et de convivialité partagés.

Je remercie aussi tous les membres de l'équipe BioME pour leurs retours pertinents sur mon travail, ainsi que tous les membres du laboratoire Biogéosciences. Merci à Claire, pour les sourires, l'enthousiasme et les balades avec Puky et Vent d'Ouest.

Mille mercis à toute l'équipe des doctorants et jeunes chercheurs de Biogéosciences. En regardant en arrière, je suis émue de notre histoire, et émue que vous continuiez à l'écrire.

L'éloignement géographique de nos bureaux comme de nos sujets de thèse n'ont pas eu raison des liens d'entraide et d'amitié qui se sont noués. Merci à Antoine de m'avoir accueillie et d'avoir suivi mon travail. Merci à Gabrielle et Agathe, pour cette si belle première année. Merci particulièrement à Agathe, binôme de thèse en or et Tata pouic. Merci à Louis (chasse neige du courage à jamais dans mon cœur), Chris, Pierre, Esteban, Gaétan, et Robin. La chute du mur de Gabriel fut un si bel accomplissement. Merci pour les randos dans un jura un caniculaire, l'épopée à ski et les soirées au Blues et à l'American. Robin, merci pour ton amphytrionnage et ton soutien sans faille en ces moments de fin de thèse. Sarah, Lola, Léa et Mélissa, merci d'avoir lancé la phase d'approche. Camille, merci pour ton engagement. Et pour le concert de Pomme. Myriam, merci d'être le petit soleil du fond du couloir. Batistin, merci d'être cet écrivain maudit, pourvoyeur de café et de réconfort. Merci à Jeanne, co-bureau de folie et de douceur. Merci à Luc, roi des biscuits de Noël et des spaetzle. Merci à William, tonton de Nesquik. Tessie, merci de faire vire la montagne, l'escalade et le yoga au labo. Quentin, merci pour ta gentillesse et ton humour ! Yann, merci infiniment pour ton amitié, pour l'aide et le soutien apporté. Grâce à toi, j'ai appris l'art de la bibliographie. Nathan, go Fourmilière dans un an ! Heureuse de t'avoir retrouvé à Dijon. Constance et Marie, merci d'être vous, pleines de folie douce et d'attention ! Merci à Alice et Ophélie, de la team Pompon Bibliothèque. Merci à Mathilde, je suis heureuse que nos routes continuent à se croiser. Camille-Sophie, merci pour ta gentillesse et ton soutien, notamment durant ce mémorable colloque à Metz. Merci à Alexandre et Jeanne-Emma, à Majd et Justine, merci à Marlisa, Célia, Erwan, Julien, Clément, Morgan, et Ivan pour tous les moments partagés. Une page se tourne sur ce microcosme dijonnais où il faisait bon vivre, mais je sais que nous resterons liés !

A tous mes amis loin de Dijon, merci du fond du cœur. Je suis touchée des liens solides qui nous unissent. Merci pour votre soutien dans cette aventure de thèse, et dans la vie !

A la fin de cet accomplissement, j'ai une pensée pour mes grands-parents. A Papi Henri, grand marin, intellectuel et musicien.

A mes parents et à Hortense, merci pour tout. Avec amour.





# AVANT-PROPOS

---

Ce manuscrit de thèse rassemble les résultats des recherches menées durant trois années en contrat doctoral et une année en contrat d'Attachée Temporaire d'Enseignement et de Recherche au laboratoire Biogéosciences (UMR 6282 Université de Bourgogne – CNRS - EPHE). Ces résultats ont été synthétisés sous la forme de quatre articles scientifiques rédigés en anglais. Un de ces articles a été publié, et un second est en révision en vue d'une publication. Les deux derniers articles sont en cours de finalisation. Faute de temps, certaines des données acquises dans le cadre de cette thèse n'ont pas pu être analysées selon les méthodes envisagées pour répondre à une problématique préalablement identifiée. Un des chapitres de la thèse, rédigé en français, présente donc ces données, les analyses pressenties, et le contexte bibliographique dans lequel elles se positionneront.

Ce manuscrit de thèse s'articule de la façon suivante :

- Une introduction générale replace le travail de thèse dans le contexte scientifique de la modélisation en génétique du paysage.
- Une présentation générale du matériel et des méthodes mobilisées positionne ce travail de thèse dans le contexte plus restreint du projet de l'équipe de recherche qui m'a accueillie. Il décrit dans les grandes lignes les données génétiques et spatiales utilisées, davantage explicitées dans les parties 'matériel et méthodes' de chaque chapitre, par soucis de concision.
- Suivent trois chapitres, construits chacun autour des articles scientifiques soumis ou en préparation, avec une courte introduction et un résumé étendu des méthodes et les principaux résultats. L'article en lui-même, rédigé en anglais, fait partie de ces chapitres, et est suivi d'une courte discussion en français.
- Une discussion générale et une conclusion font enfin la synthèse des résultats majeurs de ce travail de thèse. Ils soulignent sa contribution à la progression du champ de la modélisation de la connectivité en génétique du paysage.





# CHAPITRE 1

## INTRODUCTION GÉNÉRALE

---

1. INTRODUCTION GÉNÉRALE .....	1
1.1. Génétique du paysage, sa place dans notre compréhension de la biodiversité.....	2
1.1.1. Origine et concepts de la génétique du paysage.....	2
1.1.2. Interactions entre la structure du paysage et la variabilité génétique neutre et adaptative des populations .....	3
1.1.3. Applications des inférences de génétique du paysage.....	9
1.2. Modéliser les effets de la connectivité fonctionnelle des habitats sur la variabilité génétique neutre .....	12
1.2.1. Les surfaces de résistance : un outil de modélisation des connectivités fonctionnelles .....	12
1.2.2. Typologies de modèles en génétique du paysage.....	17
1.2.3. Limites des modèles de génétique du paysage.....	29
1.3. Objectifs de ce manuscrit de thèse .....	34

## 1.1. Génétique du paysage, sa place dans notre compréhension de la biodiversité

### 1.1.1. Origine et concepts de la génétique du paysage

La biodiversité regroupe l'ensemble de la diversité du vivant : « ce qui vit, ce qui est organisé pour vivre, et ce qui concourt à la vie » (Ollagnon, 2006). Elle se structure en trois niveaux d'organisation : la diversité des écosystèmes, des espèces et des gènes. Ces trois niveaux de diversité sont animés par des processus écologiques et évolutifs, tels que les grands cycles climatiques, les interactions inter et intraspécifiques, ou les forces d'évolution génétique, qui les structurent de façon hétérogène à la surface du globe. La biodiversité est aujourd'hui à l'aube d'un 6<sup>e</sup> événement d'extinction de masse, sous l'effet de l'activité humaine (Ceballos *et al.*, 2015; Cowie *et al.*, 2022). Le rythme actuel d'érosion de la biodiversité est sans précédent, 100 fois plus rapide que le taux d'extinction de fond naturel<sup>1</sup> (Ceballos *et al.*, 2015). Cette érosion résulte de plusieurs facteurs anthropiques : la surexploitation des ressources, l'expansion d'espèces exotiques envahissantes, la pollution des milieux, le changement climatique et principalement, pour les milieux terrestres, la destruction et la fragmentation des habitats<sup>2</sup> (Butchart *et al.*, 2010; IPBES, 2019). Si le point d'orgue de l'érosion spécifique correspond à la disparition définitive d'une espèce de la surface du globe, les extinctions locales de populations surviennent bien avant que l'extinction complète d'une espèce (Ceballos *et al.*, 2017). Ces extinctions locales sont étroitement liées à l'érosion du niveau d'organisation le plus fin et le moins perceptible de la biodiversité : celui de la diversité génétique. Certaines populations d'espèces menacées présentent à l'approche de leur extinction une diversité génétique appauvrie (Spielman *et al.*, 2004). Cet appauvrissement entraîne une sensibilité accrue des populations à l'impact de facteurs stochastiques environnementaux et démographiques, les rendant vulnérables au phénomène de vortex d'extinction (Fagan and Holmes, 2006; Gilpin, 1986).

Comment ces perturbations environnementales et paysagères affectent-elles la diversité génétique des populations ? Et quelle est la conséquence de cette érosion sur la

---

<sup>1</sup> Le taux d'extinction de fond désigne le nombre d'espèces amenées à disparaître au cours d'une période donnée, sur la base de facteurs non anthropiques et sans la survenue de catastrophes. Il s'élèverait à 2 espèces pour 10000 en 100 ans chez les mammifères, par exemple.

<sup>2</sup> Nous considérons, dans ce manuscrit, l'habitat d'une espèce comme la zone réunissant l'ensemble des ressources et conditions nécessaires à la réalisation complète de son cycle de vie.

persistance des populations ? Les espèces vont-elles s'éteindre ou s'adapter au changement climatique ? Telles sont les problématiques adressées par la génétique du paysage.

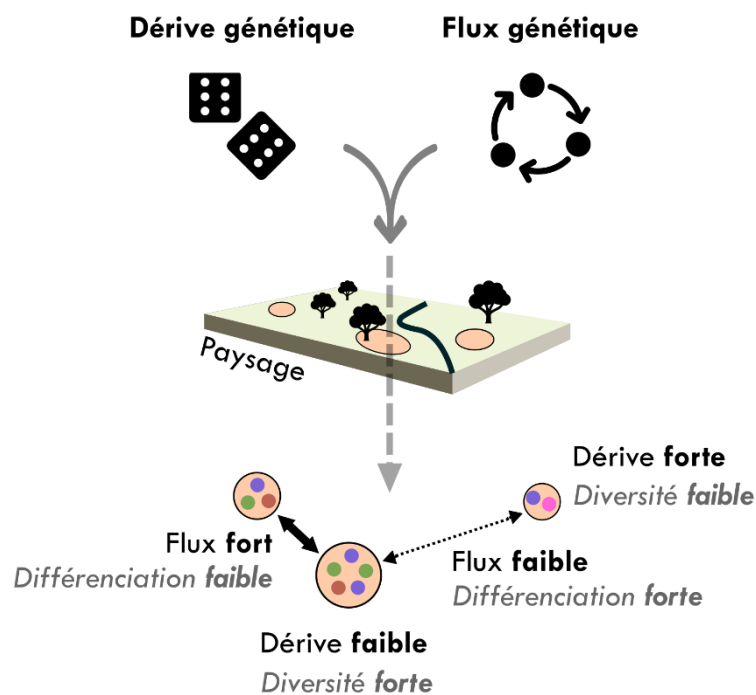
La génétique du paysage est une discipline née de l'alliance formalisée en 2003 (Manel *et al.*, 2003) entre l'écologie du paysage (Encadré n°1) et la génétique des populations. L'objectif de cette alliance est de comprendre comment les facteurs géographiques, paysagers et environnementaux structurent la variabilité génétique à l'échelle des individus ou des populations, et influent sur les processus écologiques et évolutifs (Balkenhol *et al.*, 2016; Manel, 2013; Storfer *et al.*, 2010). La portée des inférences de génétique du paysage est ainsi autant fondamentale, pour améliorer les connaissances sur les dynamiques évolutives des espèces et des populations, que patrimoniale, pour participer à la conservation de la biodiversité (Manel, 2013; Segelbacher *et al.*, 2010).

### 1.1.2. Interactions entre la structure du paysage et la variabilité génétique neutre et adaptative des populations

Face aux pressions anthropiques, la persistance des espèces suit deux stratégies principales (Manel, 2013; Pauls *et al.*, 2013). La première repose sur des épisodes de migration et de dispersion pour suivre le déplacement des conditions biotiques et abiotiques favorables. La seconde est évolutive. Elle nécessite une plasticité phénotypique et une diversité génétique suffisante pour garantir une adaptation aux nouvelles conditions locales. Ces deux stratégies sont généralement complémentaires et concomitantes. La génétique du paysage explore l'effet des facteurs environnementaux et paysagers sur la variabilité génétique traduisant ces processus démographiques et évolutifs.

L'évolution de cette variabilité génétique repose sur l'action de quatre forces : la dérive génétique, le flux génétique, les mutations et la sélection naturelle. La dérive et le flux génétique sont deux forces déterminées par des facteurs démographiques et agissant à l'échelle du génome en limitant et favorisant respectivement la variabilité génétique. A ces effets s'ajoutent ceux des mutations, correspondant à des changements aléatoires des séquences génétiques apparaissant lors des étapes de réplication de l'ADN, et susceptibles de conduire à l'apparition de nouveaux allèles. L'introduction de variabilité par mutations peut conférer un avantage ou un désavantage sélectif aux individus, ou rester silencieuse (Kimura

*et al.*, 1968; Ohta, 1992). Les variants génétiques avantageux ou désavantageux constituent la variabilité adaptative. Cette variabilité est soumise à une 4<sup>e</sup> force évolutive, la sélection naturelle, qui détermine l'évolution des fréquences de ces variants adaptatifs. Les variants silencieux, à l'inverse, ne sont pas soumis à la sélection naturelle. Etudier l'influence de mécanismes démographiques sur la distribution de la variabilité génétique, nécessite donc de considérer l'action conjuguée du flux et de la dérive génétique sur la variabilité neutre (Figure 1.1). En revanche, l'étude du potentiel adaptatif d'une population se fonde généralement sur l'évaluation des effets de la sélection naturelle et des mutations sur la diversité adaptative.



**FIGURE 1.1 :** Interactions entre la dérive, le flux génétique, et le patron d'hétérogénéité du paysage, et leurs effets sur le patron de variabilité génétique neutre. Lorsque les populations sont séparées par de grandes distances ou par des éléments barrières (*e.g.*, route, cours d'eau), le flux génétique est contraint et ne contrebalance pas les effets de la dérive génétique, médiés par la taille des taches d'habitat et l'effectif des populations. Les populations ont alors tendance à se différencier génétiquement, tandis que leur diversité génétique diminue. A l'inverse, si les populations sont proches et non séparées par des barrières, le flux génétique entre elles est important et l'effet de la dérive est réduit. Dans ce cas, les populations tendent à s'homogénéiser génétiquement.

### 1.1.2.1. Variabilité neutre, dérive et flux génétique, et connectivité

La variabilité neutre correspond donc à la variabilité allélique de *loci* non soumis à la sélection naturelle et mais influencée par le flux génétique et la dérive génétique (Figure 1.1).

La dérive génétique correspond à l'échantillonnage aléatoire des allèles dans le pool gamétique à chaque reproduction, déterminant la transmission des allèles d'une génération à la suivante. Suite à ce brassage aléatoire, certains allèles ne sont pas transmis à la descendance et disparaissent. La dérive conduit donc à l'amointrissement de la variabilité génétique. Cet effet est inversement proportionnel à la taille des populations sur lesquelles la dérive agit, les populations de faible effectif souffrant ainsi plus rapidement d'un appauvrissement génétique (Frankham *et al.*, 2017). Dans le cas d'une population complètement isolée, la dérive finit par conduire à la perte totale de variabilité. Chaque gène n'est alors représenté que par une unique version dans toute la population, et cet allèle est dit fixé. A l'échelle d'un ensemble de populations isolées, la dérive génétique conduit à la différenciation génétique des populations, en raison de leurs trajectoires stochastiques indépendantes aboutissant parfois à la fixation aléatoire d'allèles potentiellement différents au sein de chacune d'elles.

Toutefois, dès lors que la population n'est pas complètement isolée, cette perte de diversité par dérive génétique peut être compensée par l'arrivée de migrants, potentiellement porteurs de nouveaux allèles (Mills and Allendorf, 1996). La reproduction des migrants dans la population d'arrivée conduit alors à un échange de matériel génétique, restaurant la diversité perdue sous l'effet de la dérive. Ce transfert de gènes, lié à la dispersion d'individus parvenant à se reproduire et à transmettre leurs gamètes, correspond au flux génétique. Ce flux maintient la diversité génétique d'une population, et limite sa différenciation génétique avec les autres populations (Figure 1.1).

Le flux génétique est donc étroitement lié au processus de dispersion, correspondant au déplacement d'un individu depuis sa population de naissance vers une autre population dans laquelle il se reproduira (Clobert, 2012). La part de variabilité génétique influencée par le flux génétique reflète donc la capacité des espèces à se déplacer à travers un paysage hétérogène et fournit une estimation de la connectivité fonctionnelle des habitats (Encadré 1.1) (Holderegger and Wagner, 2008; Manel, 2013; Segelbacher *et al.*, 2010). Cette connectivité est d'une importance cruciale pour la conservation de la biodiversité, puisqu'elle conditionne le déplacement des individus et donc des gènes. Ces échanges et déplacements

sont en effet fondamentaux pour limiter l'accumulation de mutations délétères, favoriser la propagation d'allèles avantageux, et assurer aux espèces la possibilité de changer d'aire de distribution pour suivre le déplacement des conditions de leur niche écologique. Étudier l'intensité du flux génétique à travers un paysage *via* l'étude de la variabilité génétique neutre permet donc d'évaluer la réponse démographique d'une espèce, en particulier face à des changements globaux.

Toutefois, ces processus de dérive et de flux génétique sont difficilement observables directement, puisqu'ils se rapportent au niveau d'organisation de la biodiversité le moins perceptible. Ils sont donc généralement estimés à partir de *proxy* écologiques, tels que le suivi de déplacement des individus, des mesures de distances génétiques et des effectifs de populations. De plus, leurs effets sur l'évolution des fréquences alléliques d'une population sont souvent conjoints et difficilement distinguables à partir d'un patron de diversité observé. L'effet aléatoire de la dérive génétique est ainsi souvent considéré comme faisant partie du bruit des modèle étudiant l'influence du flux génétique sur un patron observé.

#### 1.1.2.2. Diversité adaptative, sélection naturelle et potentiel adaptatif

La diversité génétique adaptative correspond à la variabilité allélique de gènes impliqués dans l'expression de traits phénotypiques conférant sous certaines conditions un avantage ou un désavantage sélectif. Cette diversité est donc à la fois impactée par la dérive et le flux génétique, mais également par la sélection naturelle et les mutations. Contrairement à la dérive et au flux génétique impactant l'ensemble du génome, la sélection naturelle est un processus spécifique agissant à l'échelle de gènes dits 'sous sélection'. Il existe trois types de sélection naturelle, impactant différemment les niveaux de variabilité génétique. La sélection directionnelle favorise un trait extrême au détriment des autres, tandis que la sélection stabilisante favorise l'expression moyenne du trait. La sélection diversifiante favorise quant à elle plusieurs expressions extrêmes d'un trait. A l'échelle d'une population et sous pression de sélection constante, seule la sélection diversifiante produit de la variabilité génétique au niveau des *loci* sous sélection. A l'échelle d'un réseau de populations en revanche, une distribution hétérogène des pressions sélectives peut conduire à des sélections directionnelles divergentes entre populations, et ainsi à une différenciation génétique inter-populations au niveau des *loci* sous sélection.

Etudier la variabilité adaptative en relation avec les conditions environnementales permet de détecter les facteurs dirigeant les processus d'adaptation, et de comprendre comment se distribuent spatialement les variants génétiques présentant un potentiel adaptatif (Holderegger and Wagner, 2008; Manel, 2013; Segelbacher *et al.*, 2010). De plus, évaluer l'effet de la structure paysagère sur le flux génétique (estimé à partir de la distribution de la variabilité génétique neutre) renseigne sur le potentiel de propagation de ces gènes adaptatifs à travers le paysage. Un flux élevé garantit une bonne circulation des allèles, contrebalance les effets de la dérive, et favorise les chances qu'une population inclue un variant adapté aux conditions locales. Combiner des modèles de distribution des flux génétiques et de la variation génétique adaptative permet donc de prédire la future distribution spatiale de variants adaptatifs selon divers scénarios de changements environnementaux (Manel, 2013).

### **Encadré 1.1 : Hétérogénéité du paysage, écologie du paysage et connectivité**

Contrairement aux études classiques de génétique des populations qui considèrent le paysage comme un espace théorique homogène, temporellement stable et uniquement caractérisé par les distances euclidiennes, la génétique du paysage évalue la distribution de la variabilité dans un paysage hétérogène. Le concept *d'hétérogénéité du paysage* a été introduit par Forman et Godron (1986) avec le '*paradigme de la mosaïque de taches*'. Le paysage y est décrit comme une mosaïque de taches d'habitats discrets, incluse dans une matrice et reliées par des corridors. Deux principaux modèles de structure du paysage sont issus de ce paradigme : le modèle statique de biogéographie insulaire inspiré de la théorie éponyme de MacArthur et Wilson, (1967), et le modèle dynamique du paysage mosaïque (voir Cushman *et al.*, 2010 pour une revue détaillée). Dans le modèle de biogéographie insulaire, les taches d'habitats sont perçues comme des îles entourées d'un océan inhospitalier. L'océan représente la matrice, qui est considérée comme une surface uniforme de non-habitat. Le paysage est ainsi modélisé de façon binaire et statique : les taches d'habitats peuvent disparaître, mais elles sont alors remplacées par la matrice de façon immuable. Dans le modèle de paysage mosaïque, la conception statique et binaire habitat/matrice est remplacée par une vision plus complexe et dynamique du paysage, caractérisée par l'arrangement spatial de différents types d'occupation du sol. Ces types d'occupation se répartissent sur un gradient d'adéquation selon la perception d'un organisme ciblé. Cette mosaïque de taches d'habitats est dynamique dans le temps, et évolue selon les processus écologiques (e.g., succession) et la survenue de perturbations, ainsi que selon l'action humaine.



Ces modèles d'hétérogénéité des paysages sont les fondements de *l'écologie du paysage*. Cette discipline considère l'hétérogénéité et la dynamique des systèmes paysagers comme principaux déterminants des processus écologiques qui s'y déroulent et des patrons de diversité qu'on y observe à différentes échelles spatio-temporelles (Burel, 2003; Forman and Godron, 1986; Turner, 1989). En particulier, pour expliquer la distribution des patrons de diversité dans les paysages hétérogènes, l'écologie du paysage étudie l'influence de la structure des paysages sur le déplacement des individus. Cette propriété de facilitation ou de limitation du mouvement des organismes correspond à la notion de *connectivité des habitats* (Taylor *et al.*, 1993). Cette notion peut se diviser selon deux composantes structurelles et fonctionnelles. La connectivité structurelle des habitats décrit les liens structuraux entre les différentes taches d'habitats, et mesure le niveau de contiguïté spatiale de ces taches, indépendamment des attributs des organismes. À l'inverse, la connectivité fonctionnelle des habitats décrit la réponse d'un organisme aux facteurs paysagers, en prenant en compte ses capacités de dispersion et sa perception de la résistance au déplacement exercée par les types d'occupation du sol. La connectivité fonctionnelle des habitats est donc une propriété émergente de l'interaction entre les caractéristiques d'un individu et l'hétérogénéité d'un paysage. Elle peut être définie comme la quantité d'habitat atteignable lors du déplacement des individus au sein de leurs taches d'habitats et vers les autres taches d'habitat (Savary, 2021). Cette propriété émergente de connectivité fonctionnelle fait elle-même émerger des processus à l'échelle du paysage, tels que la dispersion en milieu hétérogène. Les mesures évaluant l'intensité des flux génétiques entre les populations sont régulièrement utilisées en génétique du paysage pour estimer cette connectivité fonctionnelle (Lowe and Allendorf, 2010).

### 1.1.3. Application des inférences de génétique du paysage

La génétique du paysage étudie donc l'influence de l'hétérogénéité du paysage sur la distribution de la variabilité génétique neutre et adaptative. Ses inférences contribuent à la compréhension des patrons de diversité, ainsi qu'à la prédiction de la persistance des populations en milieux changeants. La discipline a donc des implications en écologie, en biologie évolutive et en biologie de la conservation.

#### 1.1.3.1. Améliorer la compréhension des processus écologiques et évolutifs

La génétique du paysage permet en particulier d'améliorer la compréhension des processus évolutifs que sont le flux génétique et l'adaptation locale, et de leurs impacts sur les patrons de diversité. La détection des discontinuités génétiques spatiales causées par des facteurs environnementaux, c'est-à-dire des zones géographiques associées à un changement brutal de la variabilité génétique, a été un des premiers objectifs de la discipline (Manel *et al.*, 2003). Ces inférences sont précieuses en évolution, puisqu'elles permettent de comprendre comment la dispersion, en interaction avec les facteurs paysagers, façonne la structure génétique d'une population. Ces inférences renseignent également sur les facteurs impliqués dans l'adaptation génétique locale et sur les dynamiques de propagation des mutations avantageuses, permettant ainsi d'évaluer la trajectoire évolutive d'une population (Manel *et al.*, 2003). Les inférences de génétique du paysage trouvent leurs applications dans les grandes théories d'écologie modélisant la dynamique des populations, telle que la théorie des métapopulations (Levins, 1969; Murphy *et al.*, 2010). La génétique du paysage prend également part aux grands débats théoriques contemporains, comme celui du '*Single Large Or Several Small*' en écologie du paysage (Fahrig, 1997; Fahrig *et al.*, 2022; Fletcher *et al.*, 2018; Saura, 2020) pour lequel des études ont évalué l'influence relative de la quantité d'habitat et de la fragmentation *per se* des habitats sur la variabilité génétique (Cushman *et al.*, 2012; Millette and Keyghobadi, 2015; Weckworth *et al.*, 2013). Le champ disciplinaire s'attache ainsi à valider des modèles conceptuels représentant l'influence des processus éco-évolutifs sur la structuration des patrons de biodiversité observés (Keeley *et al.*, 2017; Keyghobadi *et al.*, 2005; Zeller *et al.*, 2018), et à comprendre les effets d'échelle associés

(Anderson *et al.*, 2010). Enfin, la génétique du paysage œuvre également à l'inférence de paramètres populationnels (Epps *et al.*, 2013) et à la prédiction des dynamiques des patrons de diversité selon des approches '*data-driven*'<sup>3</sup>, indépendantes des hypothèses concernant les processus à l'œuvre (Kittlein *et al.*, 2022; Vanhove and Launey, 2023).

### 1.1.3.2. Contribuer à la conservation des connectivités fonctionnelles des habitats et de la biodiversité

La fragmentation et la perte des habitats liées aux activités anthropiques peuvent conduire à une chute de la connectivité fonctionnelle, et donc du flux génétique entre populations. Lorsque les populations ont de faibles effectifs, un tel isolement fonctionnel peut conduire à une réduction à long-terme de la fitness des individus et par conséquent de la résilience des populations face aux perturbations futures (Frankham *et al.*, 2017). La première étape pour atténuer les effets de la destruction et fragmentation des habitats sur l'érosion de la biodiversité réside dans l'identification des populations vulnérables. La génétique du paysage, en décrivant l'influence des éléments paysagers et de leurs dynamiques spatio-temporelles sur la structure génétique des populations, apporte des clefs essentielles pour des prédictions plus exactes du statut de conservation futur d'une population à partir de son patrimoine génétique (Segelbacher *et al.*, 2010). Une fois les populations vulnérables identifiées, l'objectif peut alors être de restaurer leurs connectivités fonctionnelles afin de promouvoir les échanges génétiques, par exemple *via* la mise en place de corridors de dispersion (Langhammer *et al.*, 2024). Albert *et al.* (2017) ont ainsi montré qu'imposer la prise en compte d'un critère de connectivité lors de l'établissement d'un plan de conservation, en plus d'un critère sur la préservation des habitats, assure la robustesse de ces mesures de conservation face à l'incertitude des changements climatiques. Ces corridors doivent être localisés de façon optimale et être constitués d'éléments paysagers favorables au déplacement des individus. Or, pour de nombreuses espèces, l'observation directe des mouvements de dispersion est difficilement réalisable. Dans ce contexte, les inférences de génétique du paysage peuvent permettre d'identifier les éléments paysagers facilitant ou entravant la dispersion, et ainsi

---

<sup>3</sup> Les approches '*data-driven*' ou '*pilotées par la donnée*' sont des approches exploratoires fondées sur l'analyse de données pour produire des inférences inductives. Cette approche s'oppose à celle pilotée par la théorie, qui repose sur un schéma hypothético-déductif classique (élaboration d'hypothèses, collecte et analyse de données, tests d'hypothèses et conclusions théoriques sur la base des résultats).

faciliter la construction efficace de corridors de dispersion (Keller *et al.*, 2015). Les inférences de génétiques du paysages sont également prometteuses pour prédire les effets de changements paysagers, liés à l'action humaine directe ou aux conséquences des changements climatiques, sur les flux génétiques entre populations (van Strien *et al.*, 2014). En effet, elles peuvent permettre aux gestionnaires d'évaluer et de comparer des scénarios d'aménagement du territoire en fonction de leur impact sur les flux génétiques, pour implémenter ceux dont les effets négatifs sur la biodiversité sont les plus limités. Ces inférences à visée prédictive sont notamment précieuses pour évaluer les impacts de différentes projections de changement climatique, et estimer le potentiel d'atténuation de mesures de conservation.

Toutefois, favoriser le flux génétique n'est pas universellement bénéfique. En effet, bien qu'il puisse convoier des allèles avantageux, le flux génétique est également susceptible de restreindre l'adaptation locale, par l'apport régulier de variants non adaptés (Lenormand, 2002). Les modèles de génétique du paysage peuvent donc permettre d'identifier les populations sensibles au flux génétique et œuvrer à la conservation de leur patrimoine génétique adaptatif. Les objectifs de conservation peuvent également nécessiter la suppression de liens de connectivité, par exemple pour contrôler la pullulation d'espèces ravageuses (Foltête and Giraudoux, 2012). Les études de génétique du paysage sont donc prometteuses pour prédire et contrôler la propagation d'espèces exotiques envahissantes (Sacks *et al.*, 2016) ou de maladies (Biek and Real, 2010), évaluer la dynamique des interactions hôte-parasite (Kozakiewicz *et al.*, 2020) et limiter les risques d'introgession génétique entre des variants domestiques et des variétés sauvages (Bohling, 2016).

Face à la multitude et l'importance de ces enjeux, la communauté scientifique s'est largement saisie de la problématique des connectivités fonctionnelles des populations et des habitats. Il s'en est suivi une multiplication des définitions, des interprétations et des méthodes de modélisation de la connectivité, sans émergence de consensus (Calabrese and Fagan, 2004; Fagan and Calabrese, 2006). Il est donc délicat pour les gestionnaires de naviguer parmi tous les modèles et les métriques alternatives, d'autant que leur validité écologique est parfois remise en question (Moilanen, 2011). Il existe donc actuellement un réel enjeu méthodologique. Pour être à la hauteur des attendus mentionnés ci-dessus en termes de compréhension des processus impliqués, de prédictions des trajectoires des populations et d'opérationnalité en matière de conservation, il est crucial d'identifier nettement des modèles de connectivités fonctionnelles robustes, valides et non redondant.

## 1.2. Modéliser les effets de la connectivité fonctionnelle des habitats sur la variabilité génétique neutre

Dans ce manuscrit, nous nous intéresserons exclusivement à l'effet de la structure du paysage sur la distribution de la variabilité génétique neutre. Ainsi, les modèles focalisés sur la diversité neutre constituent une forme de modèles nuls, permettant d'étudier dans un second temps les effets du paysage sur la variabilité adaptative tout en contrôlant les effets de la dérive et du flux génétique. Ces modèles apportent donc un éclairage sur la capacité des individus à se déplacer dans un environnement hétérogène, et sur le taux de circulation des gènes à travers un réseau de populations. Dans cette partie, nous dresserons le panorama de ces différents types de modèles représentant de l'effet connectivités paysagères sur la distribution spatiale de la variabilité génétique neutre.

### 1.2.1. Les surfaces de résistance : un outil de modélisation des connectivités fonctionnelles

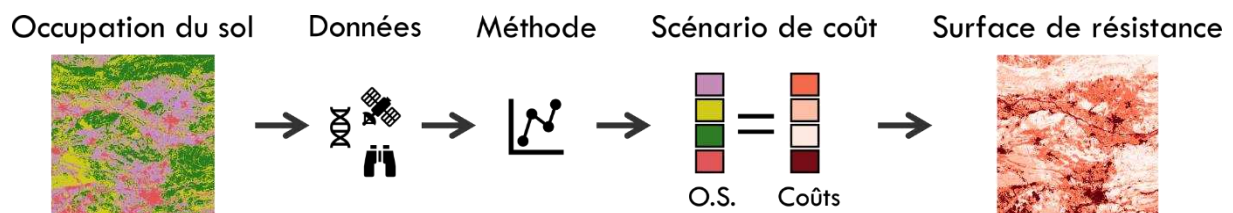
#### 1.2.1.1. Construction et paramétrage des surfaces de résistance

La modélisation des connectivités est un prérequis souvent nécessaire à leur restauration efficace. Ainsi, de nombreuses approches ont été développées pour cartographier les éléments paysagers facilitant le déplacement des organismes et modéliser la connectivité des populations (Cushman *et al.*, 2013). Ces approches partagent généralement un fondement commun : la représentation du paysage selon une surface de résistance au déplacement (Figure 1.2, Encadré 1.2). Ces surfaces de résistance modélisent la propension au mouvement, le coût physiologique et le risque de mortalité encouru par les individus qui se dispersent dans des environnements hétérogènes (Diniz *et al.*, 2020 ; Zeller *et al.*, 2012). Elles prennent la forme d'un raster<sup>4</sup> où à chaque pixel est associé une valeur de coût, représentant la résistance au déplacement exercé par l'élément paysager ou environnemental sous-jacent au pixel (Figure 1.2, Spear *et al.*, 2015, 2010). Ces traits peuvent correspondre à des caractéristiques

---

<sup>4</sup> Dans les Système d'Information Géographique (SIG), un raster correspond à une grille de pixels, où chaque pixel est porteur d'une valeur.

discrètes du paysage (classiquement, un type d'occupation du sol) ou à des facteurs environnementaux continus (*e.g.*, altitude, température). Une surface de résistance est donc une représentation spatiale explicite d'une hypothèse sur le comportement de dispersion d'un organisme en interaction avec le paysage traversé. L'attribution des valeurs de coût à chaque caractéristique paysagère est une étape essentielle de la modélisation, conditionnant la fiabilité des analyses. Dans ce manuscrit, nous désignerons cette étape comme celle du *paramétrage de la surface de résistance* par un *scénario de coût* (*i.e.*, l'association entre les traits paysagers et les valeurs de coût). De nombreuses méthodes ont été développées pour tendre vers un paramétrage fiable et biologiquement valide des surfaces de résistance (Dutta *et al.*, 2022; Spear *et al.*, 2010; Zeller *et al.*, 2012). Ces méthodes se distinguent notamment en fonction du type de données sur lesquelles se fonde l'estimation des valeurs de coût. Ces données peuvent correspondre à des représentations directes du déplacement des individus (*e.g.*, des données GPS, de télémétries ou de marquage-recapture), ou à des représentations indirectes (*e.g.*, des données de présence ou des données génétiques reflétant les flux de gènes). Lorsqu'aucune donnée biologique n'est disponible, de nombreuses études ont recours à l'opinion d'expert.e.s et aux données bibliographiques pour construire les scénarios de coût. Ces méthodes possèdent chacune leurs intérêts et leurs inconvénients, notamment en termes de coût d'acquisition et de pertinence écologique. Le choix de l'une d'entre elles doit donc avant tout se fonder sur les objectifs de l'étude et les aspects biologiques pertinents à leur égard (Spear *et al.* (2010), Encadré 1.2). De plus, plusieurs auteurs recommandent d'adopter des méthodes d'optimisation ou des études de sensibilité pour comparer les performances de plusieurs surfaces candidates, ainsi que des approches de validation pour vérifier la fiabilité du paramétrage sélectionné (Dutta *et al.*, 2022; Spear *et al.*, 2010; Zeller *et al.*, 2012).



**FIGURE 1.2** : Les différentes étapes de la paramétrisation d'une surface de résistance à partir de données biologiques. Il s'agit tout d'abord de cartographier les éléments paysagers retenus comme variables explicatives, en choisissant les résolutions spatiales et thématiques. Des données biologiques reflétant la dispersion sont ensuite collectées, puis transformées par un outil analytique pour conduire à l'estimation de paramètres de coût liés à la dispersion. Ces paramètres sont associés aux éléments paysagers (*i.e.*, sous la forme d'un 'scénario de coût') et cette association est transcrite spatialement sous la forme d'une surface de résistance. O.S. : occupation du sol.

### **Encadré 1.2 : Construction d'une surface de connectivité**

#### **Étape 1 : Définir les objectifs de la modélisation**

Des objectifs de la modélisation découlent les choix méthodologiques. En effet, une meilleure compréhension de l'influence du type d'occupation du sol sur la connectivité génétique ne s'acquiert pas en réalisant les mêmes inférences que lorsqu'il s'agit de déterminer la localisation optimale d'un passage à faune.

#### **Étape 2 : Choisir des variables environnementales et des résolutions thématiques, temporelles et spatiales**

Les variables environnementales constituent les prédicteurs du modèle de connectivité. Leur choix doit donc être méticuleux. Les variables d'occupation du sol sont les plus classiquement utilisées, bien que les variables mesurant l'altitude, le réseau routier ou le réseau hydrologique par exemple alimentent aussi couramment les modèles de connectivité (Zeller *et al.*, 2012). Ces variables peuvent être catégorielles et présenter divers niveaux d'agrégation potentiels, ou être continues et issues de diverses transformations permettant de représenter des relations non linéaires entre l'espèce et l'élément paysager (Keeley *et al.*, 2016). Elles sont idéalement choisies en fonction des connaissances que l'on a *a priori* sur la perception du paysage par l'espèce focale. Toutefois, ces informations sont souvent manquantes. Une surface peut être multifactorielle, ou monofactorielle si un seul facteur paysager est pressenti comme important. Par ailleurs, les variables environnementales peuvent affecter les connectivités à différentes échelles spatiales et temporelles (e.g., variation saisonnière du réseau hydrographique) (voir Anderson *et al.* (2010) et Cushman and Landguth (2010) pour des revues sur le sujet). Le choix des résolutions thématique<sup>5</sup>, spatiale et temporelle peut faire l'objet d'études de sensibilité préalables (Zeller *et al.*, 2017).

#### **Étape 3 : Choisir des données biologiques**

Les données biologiques utilisées pour paramétrer les surfaces de résistances doivent refléter les connectivités fonctionnelles. Elles peuvent être de plusieurs types. Les données mesurant le déplacement des individus de manière directe, par télémétrie, marquage recapture ou GPS apportent des informations fiables et précises sur la spatialisation des chemins de dispersion,

---

<sup>5</sup> La résolution thématique correspond au nombre de catégories d'occupation du sol modélisées dans le cas d'une surface catégorielle, ou à la forme de la fonction reliant une variable environnementale continue à son effet de résistance sur le flux génétique dans le cas d'une surface continue.

mais sont rares et onéreuses à obtenir. Les données de (télé-)détection sont plus accessibles, mais leur pouvoir explicatif peut être faible, car la présence n'est pas forcément synonyme de mouvement et parce que les exigences écologiques pour le mouvement ne sont pas forcément identiques aux exigences pour les autres étapes du cycle de vie (reproduction par exemple). Toutefois, il est assez courant de dériver une surface de résistance à partir d'un modèle de distribution d'espèce (SDM) fondé sur des données de présence, en transformant les probabilités de présence en valeurs de résistance (Keeley *et al.*, 2016). Les données génétiques sont également utilisées régulièrement, car elles reflètent pour partie l'effet des flux génétiques, eux-mêmes déterminés par la dispersion des individus. Toutefois, les données de mouvement et de flux génétique ne sont pas non plus synonymes, puisque le flux génétique traduit non seulement un épisode de dispersion, mais aussi l'installation de l'individu, sa reproduction et la survie de sa descendance. De plus, le flux génétique ne reflète que les mouvements de dispersion et non pas les mouvements réalisés au sein même du domaine vital. Un point d'attention concerne donc l'adéquation entre les objectifs poursuivis, les données biologiques utilisées et les conclusions tirées. Pour le design de corridors, des données de suivi GPS peuvent être très pertinentes. En revanche, pour évaluer la persistance à long terme d'une métapopulation, les données génétiques peuvent se révéler plus appropriées. Enfin, de nombreuses études paramètrent les surfaces de résistance à partir du savoir d'expert.e.s. Cette source de données est la moins coûteuse et la plus rapidement disponible, mais sa subjectivité est parfois questionnée. Beier *et al.* (2008) propose ainsi que les inférences de coût à partir de données génétiques ou de données de déplacement améliorent la pertinence écologique des modèles de connectivité, par rapports à celles fondées sur des dires d'expert.e.s.

#### Étape 4 : Choisir un modèle analytique

Il s'agit ensuite de choisir une approche analytique permettant de dériver des valeurs de coûts à partir des données biologiques utilisées. Lorsque le paramétrage se fonde sur des dires d'expert.e.s, cette étape n'a pas lieu d'être, les valeurs de coût étant directement estimées. Les fonctions de sélection de ressources, permettant d'établir la probabilité d'utilisation d'un habitat pour une espèce donnée, sont souvent utilisées à cette étape. Une autre approche courante en génétique du paysage consiste à optimiser des valeurs de coût pour maximiser l'ajustement entre des distances écologiques (basées sur les coûts à optimiser) et des distances génétiques, par exemple grâce à l'algorithme d'optimisation ResistanceGA que nous étudierons dans le chapitre 3.1 (Peterman, 2018).



### Étape 5 : Validation de la surface

Enfin, Spear *et al.* (2010) insistent sur l'intérêt de valider la capacité des surfaces de résistance à prédire correctement les connectivités, par exemple en utilisant des méthodes de validations croisées à partir de données empiriques indépendantes ou simulées (Ch 3.1).

#### 1.2.1.2. Utilisation des surfaces de résistance

Ces surfaces de résistance peuvent ensuite être utilisées de différentes façon pour modéliser les connectivités fonctionnelles. La représentation cartographique des surfaces en elles-mêmes peut tout d'abord servir à identifier les éléments paysagers influençant les connectivités fonctionnelles ainsi que la magnitude de leur influence. Cette approche est courante dans les études de génétique du paysage. Certaines études évaluent ainsi la perméabilité des facteurs paysagers au flux génétique en interprétant les scénarios de coûts optimisés à partir de données génétiques (Peterman, 2018). Nous reviendrons néanmoins sur le risque associé à un tel usage dans le chapitre 3.1. Les surfaces de résistance peuvent également servir de données d'entrée dans d'autres modélisations visant à représenter les chemins de dispersion ou distances écologiques entre les taches d'habitats (Adriaensen *et al.*, 2003; McRae, 2006), le réseau d'habitat ou de dispersion des individus (voir Encadré n°3 sur les graphes paysagers, Dyer, 2015; Urban and Keitt, 2001) ou encore à calculer des métriques de connectivité à différentes échelles spatiales à partir des liens estimés (Keeley *et al.*, 2021). Ces modèles de connectivité peuvent ensuite répondre à différents objectifs, tels que la détermination des processus écologiques influençant la connectivité génétique des populations, la prédiction des impacts de futurs changements environnementaux, ou la construction de couloirs de dispersion dans le cadre de mesures de conservation.

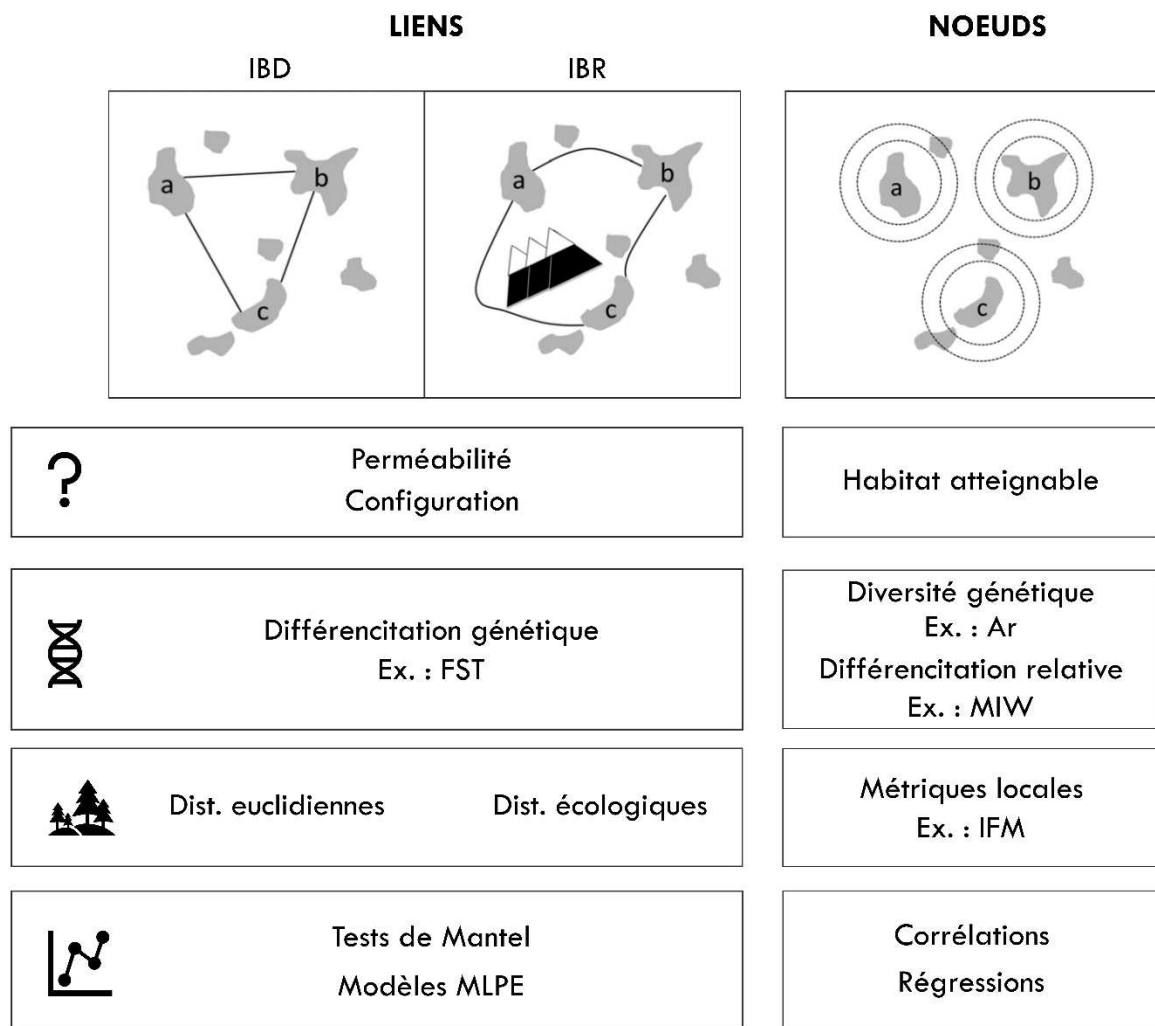
## 1.2.2. Typologies de modèles en génétique du paysage

### 1.2.2.1. La population ou l'individu comme unité d'analyse

Les patrons génétiques et paysagers peuvent se modéliser en considérant soit des individus soit des populations comme entités de base. De façon théorique, il est recommandé de conduire une analyse au niveau des populations pour les espèces présentant une distribution spatiale morcelée, et au niveau des individus pour les espèces distribuées de façon plus continue à travers le paysage (Waits and Storfer, 2015). Bien que des travaux de simulation et des analyses empiriques aient montré que le niveau individuel pouvait se révéler pertinent pour des organismes distribués en taches (Coulon *et al.*, 2004; Murphy *et al.*, 2008; Prunier *et al.*, 2013), ce niveau est moins fréquemment utilisé que celui des populations dans les travaux de génétique du paysage et manque de support théorique (DiLeo and Wagner, 2016; Richardson, 2012). Le choix de l'unité statistique d'étude est important car il influence la stratégie d'échantillonnage, le choix des métriques génétiques et paysagères, et les méthodes d'analyses utilisées. Ici, nous avons choisi le niveau populationnel, en partie en raison de la fragmentation importante de l'habitat forestier des zones échantillonnées.

### 1.2.2.2. Les différents niveaux de modélisation en génétique du paysage

Les modèles de génétique du paysage consistent à mettre en relation des variables génétiques avec des variables paysagères. Ces mises en relation peuvent s'opérer à différents niveaux, tel que cela a été formalisé par DiLeo et Wagner (2016) et Wagner et Fortin (2013) (Figure 1.3).



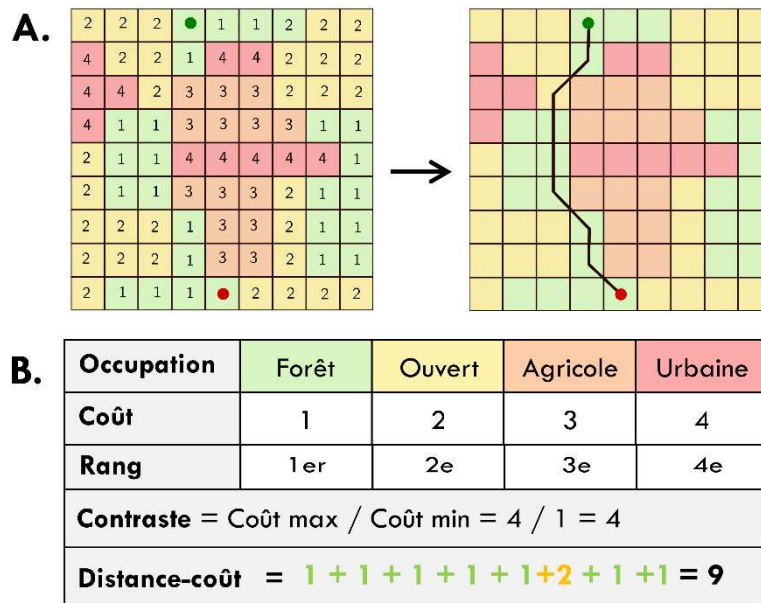
**FIGURE 1.3** : Deux niveaux de modélisation en génétique du paysage (d'après la Figure 1 de DiLeo et Wagner (2016)). Le niveau des liens étudie l'effet de la perméabilité de la matrice et de la configuration du paysage sur la différenciation génétique entre les populations, en mettant en relation des distances paysagères et génétiques *via* des tests de Mantels et des modèles MLPE. La matrice peut être considérée comme homogène (utilisation de distances euclidiennes) ou hétérogène (utilisation de distances écologiques) en fonction de l'objectif d'étude (étude des patrons d'IBD et de la configuration des habitats *versus* étude des patrons d'IBR et de l'effet de la matrice). Le niveau des nœuds étudie l'effet de la quantité d'habitat atteignable sur la diversité et la différenciation génétique relative d'une population *via* des approches corrélatives.

### Niveau des liens

Une majorité des analyses de génétique du paysage se font au niveau des liens entre populations (Figure 1.3). Elles mettent principalement en relation des distances euclidiennes ou écologiques avec des distances génétiques entre deux populations (DiLeo and Wagner, 2016). Ce niveau d'étude est particulièrement adapté pour évaluer la perméabilité de la matrice paysagère au flux génétique, tester des hypothèses sur la configuration et la résistance

des éléments paysagers et identifier des corridors pour répondre à des enjeux de conservation. Les prédicteurs correspondent à des distances écologiques qui modélisent à la fois la longueur et la perméabilité au mouvement des voies de dispersion, deux paramètres censés conditionner le flux génétique. L'effet de la dérive génétique sur la différenciation génétique constitue une source de stochasticité que l'on peut rarement modéliser et qui se retrouve donc dans le « bruit » des résidus du modèle. En effet, les prédicteurs sous forme de distances ne modélisent pas la quantité d'habitat, qui reflète l'effectif des populations fortement liée à la dérive. Cette variable est très rarement modélisée au niveau des liens entre populations (voir Prunier *et al.*, 2017). Quand elle est prise en compte, ce n'est pas la « direction » de l'effet aléatoire de la dérive qui est modélisée, mais plutôt l'hétérogénéité de l'intensité de la dérive, qui dépend de façon plus prédictible de la taille respective des deux populations entre lesquelles on mesure la différenciation génétique (Prunier *et al.*, 2017; Savary *et al.*, 2023; Serrouya *et al.*, 2012; Weckworth *et al.*, 2013). DiLeo et Wagner (2016) soulignent toutefois que la réplication d'analyses de types *liens* sur différents paysages caractérisés par des quantités d'habitat variables permettrait d'étudier les interactions entre perméabilité de la matrice, configuration et quantité d'habitat. Enfin, l'influence de ces variables paysagères sur la diversité génétique est rarement évaluée à ce niveau, davantage adaptée à la modélisation des différenciations génétiques.

A ce niveau d'analyse, les métriques de distances génétiques se fondent sur la comparaison des fréquences alléliques entre populations. Une métrique très couramment utilisée est le  $F_{ST}$  développée par Sewall Wright pour décrire le partitionnement de la variation génétique entre populations (Wright, 1931). Cet indice quantifie la divergence des fréquences alléliques entre deux populations, en évaluant l'écart entre les taux d'hétérozygotie observés dans ces populations par rapport au taux attendu dans une unique population panmictique de taille équivalente. D'autres métriques se sont également développées pour répondre à différents enjeux méthodologiques, tels que le  $G_{ST}$  prenant en compte une multiplicité d'allèles pour un unique locus, le  $G'_{ST}$  indépendant du niveau de polymorphisme des *loci* ou encore le  $D_{PS}$  basé sur la proportion d'allèles partagés entre individus (Bowcock *et al.*, 1994; Hedrick, 2005; Meirmans and Hedrick, 2011).

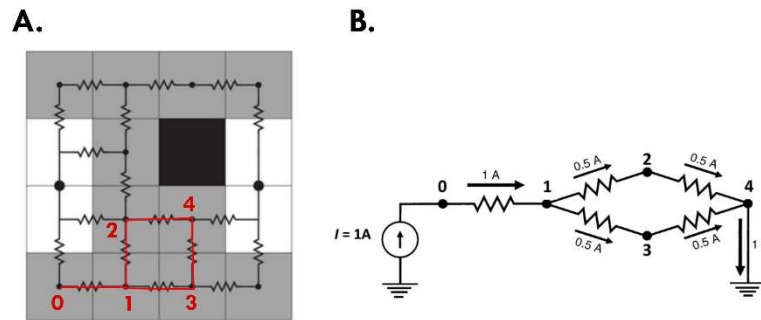


**FIGURE 1.4 :** Modèle des distances de moindre coût. A) La construction du modèle repose sur une surface de résistance, représentant le paysage par un raster dont chaque pixel est associé à une valeur de coût liée au déplacement d’un individu à travers ce pixel. Le chemin de moindre coût entre un point A (rond vert) et un point B (rond rouge) correspond au chemin traversant les pixels les moins coûteux pour relier ces deux points (d’après la Figure 10 de Savary (2021)). B) La surface de résistance est paramétrée par un scénario de coût. En considérant l’occupation du sol comme variable paysagère explicative, un scénario associe chaque occupation du sol à une valeur de coût représentant la résistance de cette occupation au déplacement. Un scénario de coût se caractérise par le rang de résistance de chaque occupation du sol, et son contraste (*i.e.*, ratio entre le coût minimal et le coût maximal). La distance-coût enfin, correspond au coût cumulé des pixels traversés par le chemin de moindre coût.

Ces distances génétiques entre paires de populations sont mises en relation avec des distances écologiques entre les taches d’habitat associées à ces populations. Parmi les types de distances écologiques, nous pouvons dissocier les distances euclidiennes reliant en ligne droite les taches d’habitats indépendamment du paysage traversé, des distances-coût et des distances de résistance qui, elles, considèrent la résistance exercée par la matrice paysagère (Figure 1.3). Les distances euclidiennes constituent un modèle nul permettant d’évaluer l’effet de la configuration spatiale des habitats sur les distances génétiques, indépendamment des effets de l’hétérogénéité de la matrice paysagère (van Strien *et al.*, 2015). Ces distances sont classiquement utilisées pour détecter et caractériser les patrons d’Isolement par la Distance (IBD), caractérisés par une augmentation de la différenciation génétique des populations avec les distances euclidiennes en raison de la diminution des flux génétiques à longue distance (Hutchison and Templeton, 1999; Slatkin, 1993). Toutefois, il est fréquent que ces distances expliquent moins bien les distances génétiques que d’autres distances écologiques considérant à la fois cet effet de la distribution des habitats, mais également celui de l’hétérogénéité de la matrice. Les distances de moindre coût (Adriaensen *et al.*, 2003) et les distances de résistance dérivées de la théorie des circuits (McRae, 2006) constituent les deux approches les plus populaires en génétique du paysage. Le calcul de ces distances nécessite le paramétrage

préalable d'une surface de résistance. Les distances de moindre coût reposent sur le concept du chemin de moindre coût (Figure 1.4). Ce chemin est celui qui permet de relier deux taches d'habitats en traversant les pixels les moins coûteux sur une surface de résistance. Ces chemins de moindre coût sont couramment utilisés dans les études de conservation pour planifier des corridors de dispersion (Beier *et al.*, 2009; Carroll *et al.*, 2012; Sawyer *et al.*, 2011). Les distances de moindre coût (ou distances-coût) correspondent au coût cumulé le long de ce chemin. Elles sont donc en quelque sorte une somme pondérée de la distance parcourue pondérée par le coût de chaque segment du déplacement. Ce modèle de dispersion implique que les organismes aient une connaissance parfaite du territoire et ne se déplacent qu'en suivant l'unique chemin de moindre coût. L'absence de considération de chemins alternatifs est une limite de cette méthode, qui est abordée par la méthode des distances de résistance (Figure 1.5).

Cette seconde méthode, dérivée de la théorie des circuits, modélise une exploration aléatoire du paysage. Selon ce modèle, le paysage est découpé par une grille, dont les nœuds réguliers sont connectés par des résistances, à l'image d'un circuit électrique (Figure 1.5). La surface de résistance permet de définir la valeur de la résistance exercée entre deux nœuds. La distance de résistance entre deux nœuds résume l'ensemble des chemins parallèles empruntés par un courant pour relier ces deux nœuds. Si ces distances de résistances considèrent ainsi de nombreux chemins alternatifs, les calculs associés sont nettement plus longs que ceux des distances de moindre coût. Enfin, les distances écologiques entre deux populations peuvent être évaluées non pas directement sur une surface de résistance, mais le long des liens d'un graphe paysager (Encadré n°3). Lorsque le graphe présente une topologie complète (i.e., que l'ensemble des taches d'habitats sont reliées les unes aux autres sur le graphe), les distances écologiques le long des liens du graphe sont équivalentes à celles calculées sur une surface de résistance. En revanche, si le graphe est élagué (i.e., si les taches d'habitat ne sont reliées qu'à certaines taches, selon un critère de proximité par exemple), les distances écologiques refléteront alors un processus de dispersion en *stepping-stone* suivant les taches connectées. Les distances écologiques, qu'elles dérivent directement d'une surface de résistance ou d'un graphe paysager, sont largement utilisées pour détecter des patrons d'isolement par la Résistance du Paysage (IBR) et évaluer l'influence de l'hétérogénéité du paysage sur le flux génétique. Ces patrons sont les pendants des patrons d'IBD avec des distances écologiques : ils témoignent d'une augmentation de la différenciation génétique entre populations avec l'augmentation des distances écologiques entre ces mêmes populations (Figure 1.3).



**FIGURE 1.5 :** Modèle des distances de résistance, d'après McRae *et al.* (2009). A) Un raster paysager est converti en circuit électrique : chaque pixel devient un nœud (représenté par un point), et chaque nœud est connecté à ses voisins (pixels adjacents) par des résistances. La couleur d'un pixel représente sa contrainte à la dispersion. Les valeurs de résistances joignant deux pixels peuvent par exemple correspondre à la valeur moyenne de résistance entre deux pixels adjacents. B) Représentation schématique sous forme de circuit électrique des deux chemins alternatifs permettant de relier le nœud 0 au nœud 4, représentés en rouge sur la figure A.

La détection de ces patrons d'IBD et d'IBR nécessite la mise en relation des données génétiques et paysagères (Figure 1.3). Une des approches les plus fréquentes en génétique du paysage repose sur les tests de Mantel et des tests de Mantel partiels, qui évaluent la corrélation entre deux matrices de distances génétiques et paysagères conditionnellement à une troisième matrice de distances (dans le cas d'un test partiel) souvent euclidiennes (Manel, 2013). Cushman *et al.* (2006) se sont basés sur cette approche analytique pour définir le 'causal modeling' permettant de tester plusieurs hypothèses concernant les facteurs paysagers limitant le flux génétique. Des approches de régression ont également été développées, notamment pour modéliser simultanément l'effet de plusieurs variables de distance sur la structure génétique (méthode de la régression multiple des matrices de distances, Legendre and Fortin, 2010). Ces approches ont toutefois été critiquées, en raison notamment de leur faible capacité à détecter une relation entre distances génétiques et paysagères lorsque celle-ci est présente (i.e., faible puissance statistique, voir Legendre and Fortin, 2010; Raufaste and Rousset, 2001). Par ailleurs, le format des données au niveau des *liens*, impliquant des paires de populations, conduit à des problèmes d'indépendance rendant invalides les tests statistiques associés aux modèles. L'utilisation de modèles linéaires mixtes permet de résoudre ce problème, en considérant comme effet aléatoire l'identifiant des populations impliquées dans la distance considérée, et en modélisant ainsi la covariance des résidus. Cette approche, nommée MLPE ('Maximum Likelihood Population Effect') (Clarke *et al.*, 2002) a été introduite en génétique du paysage par (Van Strien *et al.*, 2012) et semble prometteuse (Shirk *et al.*, 2018).

### Niveau des nœuds

Les données génétiques peuvent également être mises en relation avec les caractéristiques des taches d'habitats et de leur contexte paysager local : on parle alors d'analyses au niveau des nœuds (DiLeo and Wagner, 2016; Wagner and Fortin, 2013) (Figure 1.3). L'unité statistique de ce niveau d'analyse n'est pas une paire de taches d'habitats ou de populations, mais une tache d'habitat ou une population focale. Ce niveau d'étude caractérise la connectivité d'une tache d'habitat, et son influence sur la structure génétique de la population associée à cette tache. Contrairement aux analyses fondées sur les liens, les analyses au niveau des nœuds permettent d'évaluer explicitement l'effet de la quantité d'habitat environnante sur la diversité génétique ou la différenciation relative d'une population par rapport aux autres, modélisant ainsi conjointement les effets de la dérive génétique et du flux génétique. Des travaux empiriques et de simulation ont mis en évidence une relation forte entre taille des populations et diversité génétique (Capurcho *et al.*, 2013; Jackson and Fahrig, 2014). Une particularité des analyses basées sur les nœuds est donc la prise en compte de la diversité génétique locale d'une population, non réalisable dans le cadre d'une modélisation focalisées sur les liens entre populations, ainsi que celle de la quantité d'habitat disponible. Ce niveau d'étude étant très flexible, il permet également d'étudier les effets de la configuration du paysage et de la résistance de la matrice sur la diversité génétique ainsi que sur des indices de différenciation génétiques population-spécifiques (Pflüger and Balkenhol, 2014).

Au niveau des nœuds, il est donc possible de décrire la diversité génétique d'une population à travers des indices de richesse allélique, ou d'hétérozygotie observée ou attendue sous l'équilibre d'Hardy-Weinberg. D'autres indices permettent de caractériser la particularité du *pool* allélique d'une population vis-à-vis de celui de l'ensemble des populations. Ainsi, la richesse allélique privée évalue le nombre d'allèles propres à une population et absents de toutes les autres, tandis que la richesse allélique manquante correspond au nombre d'allèles présents dans toutes les populations hormis la population focale. Enfin, il existe des indices qualifiant la différenciation génétique d'une population relativement à toutes les autres, tels que le '*population-specific  $F_{ST}$* ' dérivé d'une approche bayésienne (Gaggiotti and Foll, 2010) ou le '*Mean of Inverse Weight*' (MIW) issu de la théorie des graphes (Koen and Bowman, 2015).



Ces indices génétiques sont mis en relation avec des métriques paysagères, calculées soit dans un rayon fixe autour de la tache d'habitats échantillonnée, soit dans un voisinage défini par un rayon mouvant, selon la distance écologique des taches d'habitats voisines et la capacité de dispersion de l'organisme étudié (Figure 1.3). On peut distinguer deux types de métriques de connectivité au niveau des nœuds (Calabrese and Fagan, 2004; Keeley *et al.*, 2021): les métriques structurelles et les métriques fonctionnelles. Les métriques structurelles décrivent les attributs de composition et de configuration du paysage, indépendamment des caractéristiques de l'espèce étudiée. On retrouve ainsi classiquement des métriques telles que la distance à la tache voisine la plus proche, le nombre de taches d'habitats dans le voisinage ou la quantité d'habitat dans un rayon défini. Le programme FRAGSTATS (McGarigal and Marks, 1995) permet de calculer une grande diversité de métriques structurelles. Plus récemment, l'introduction des graphes paysagers dans les études de génétique du paysage a conduit à l'émergence de nouvelles métriques, fondées sur la topologie du réseau d'habitat (Encadré 1.3) (*e.g.*, degré des nœuds ou métriques de centralité, Rayfield *et al.*, 2011). Les métriques de connectivité fonctionnelle considèrent l'interaction entre les traits de l'espèce étudiée et la résistance du paysage. Ces métriques peuvent dériver de cartes binaires, représentant les taches d'habitats incluses au sein d'une matrice homogène, ou dériver de surfaces de résistances prenant en compte l'effet de cette matrice décrite selon une résolution plus fine, basée notamment sur l'occupation du sol. Dans le premier cas, les métriques intègrent des distances euclidiennes, et dans le second des distances-coût ou des distances de résistance issues de la théorie des circuits. Une métrique de connectivité fonctionnelle canonique est la fonction d'incidence ('IFM') définie par Hanski (Hanski, 1994) et dérivée de la théorie des métapopulations. Cette fonction permet d'estimer le flux de migrants au niveau d'une tache d'habitat, en fonction de sa distance écologique aux taches voisines et de la surface de ces taches. La métrique de Flux issue des graphes paysagers est une forme d'héritage de cette fonction d'incidence, qui modélise la quantité d'habitat atteignable au niveau d'une tache pour un organisme donné (Urban and Keitt, 2001). De nombreuses métriques fonctionnelles peuvent être dérivées des graphes paysagers. Leur intérêt réside notamment dans leur capacité à modéliser la contribution de chaque nœud à plusieurs composantes de la connectivité, telles que la capacité de charge des habitats, la quantité d'habitat atteignable dans leur voisinage et leur importance pour la traversabilité du réseau (Baranyi *et al.*, 2011).

Les données génétiques et paysagères, ainsi décrites par ces variables au niveau des nœuds, sont alors mises en relation par des analyses de corrélations et des régressions multiples. Il est également possible d'utiliser l'extension directe de ces approches sous forme d'analyses canoniques de redondance (RDA) pour évaluer l'effet des variables paysagères sur les fréquences alléliques d'une population (Legendre and Legendre, 2012).

#### *Autres niveaux d'analyse*

La relation entre structure paysagère et génétique peut s'analyser à d'autres niveaux, bien que cela soit nettement moins courant. On peut citer les études de types '*boundary-level*' (Wagner and Fortin, 2013) qui s'attachent à détecter les éléments paysagers constituant des barrières au flux génétique. Ce type d'analyse consiste à détecter des clusters génétiques (dans lesquels les individus partagent un *pool* d'allèle commun), par exemple *via* l'algorithme de *clustering* bayésien STRUCTURE (Pritchard *et al.*, 2000), et à la mettre en relation avec des éléments paysagers.

Enfin, certains modèles de génétique du paysage allient plusieurs niveaux d'étude. C'est notamment le cas des modèles gravitaires, intégrant à la fois des variables liées aux liens et aux nœuds du réseau de populations, pour modéliser les effets de l'environnement local et de la perméabilité de la matrice sur le flux génétique (Murphy *et al.*, 2010; Zero *et al.*, 2017).

### **Encadré 1.3 : Les graphes paysagers**

L'application de la théorie des graphes en écologie a été formalisée par l'article fondateur d'Urban et Keitt, (2001). Les graphes paysagers sont des graphes spatiaux constitués de *nœuds*, correspondant aux taches d'habitat, reliés par des *liens*, correspondant aux chemins de dispersion potentiels entre ces taches d'habitat (Figure 1.6.A). Les nœuds sont généralement pondérés par une variable censée approximer leur capacité de charge (e.g., leur surface), tandis que les liens sont généralement pondérés par les distances écologiques reflétant la résistance du paysage entre deux taches d'habitat.

Les graphes paysagers peuvent être perçus comme une modélisation de la connectivité atteignant un compromis optimal entre coût d'acquisition des données spatiales et biologiques, complexité de leur traitement et pertinence écologique du modèle en résultant (Calabrese and Fagan, 2004). Ils sont en effet relativement faciles à mettre en œuvre et

proposent une représentation exhaustive du réseau d'habitat et des enjeux de connectivité associés.

Les graphes paysagers ont une portée fondamentale dans la mesure où ils contribuent à la compréhension des déplacements des individus au sein d'un réseau d'habitat dans un paysage hétérogène, et de leurs conséquences sur la dynamique des populations. Ils trouvent leur origine dans la théorie des métapopulations, tout autant qu'ils en poursuivent les applications théoriques. Urban et Keitt (2001) ont ainsi montré que la persistance d'une métapopulation était en partie déterminée par la persistance des connexions du graphe d'habitat associé. Les graphes paysagers ont également une portée opérationnelle et sont souvent utilisés à des fins de conservation. En effet, ils permettent de représenter spatialement les enjeux de connectivité, d'évaluer la contribution de taches d'habitat à la connectivité du réseau, d'identifier les zones vulnérables ou d'intérêt pour établir des plans de priorisation, ou encore de prédire l'impact d'aménagements futurs sur la connectivité d'un réseau (Foltête *et al.*, 2014; Galpern, 2011). Le succès des graphes paysagers dans le domaine de la conservation repose notamment sur le développement de logiciels gratuits permettant une construction et une analyse aisée, à destination de la communauté scientifique mais également des gestionnaires, tel que le logiciel Graphab (Foltête *et al.*, 2021, 2012; Girardet and Clauzel, 2018).

La construction d'un graphe paysager repose sur l'identification des nœuds d'une part, et sur le choix des liens et de leur pondération d'autre part (Figure 1.6.B). Les seules données nécessaires sont donc des données géographiques, support de la délimitation des taches d'habitats sur la base d'hypothèses concernant la distribution de l'espèce étudiée et ses mouvements. Les liens peuvent éventuellement être pondérés par les distances euclidiennes entre les taches d'habitats lorsqu'à défaut d'informations sur l'espèce étudiée, il est préférable de ne pas faire d'hypothèses plus précises concernant la perméabilité de la matrice. À l'inverse, pour modéliser explicitement l'effet de la matrice paysagère sur la dispersion, il devient nécessaire de créer une surface de résistance à partir des données géographiques, qui permettra de pondérer les liens par des distances-coûts.

Une caractéristique de la modélisation par les graphes paysagers est la possibilité de choisir les connexions que l'on souhaite modéliser entre les taches d'habitat (Figure 1.6.C). En effet, un graphe complet modélise les liens qu'établit chaque tache d'habitat avec l'ensemble des autres taches, mais il est possible d'*élaguer* ce graphe pour ne conserver qu'un sous-ensemble de liens, représentant par exemple les connexions les plus fortes, c'est-à-dire

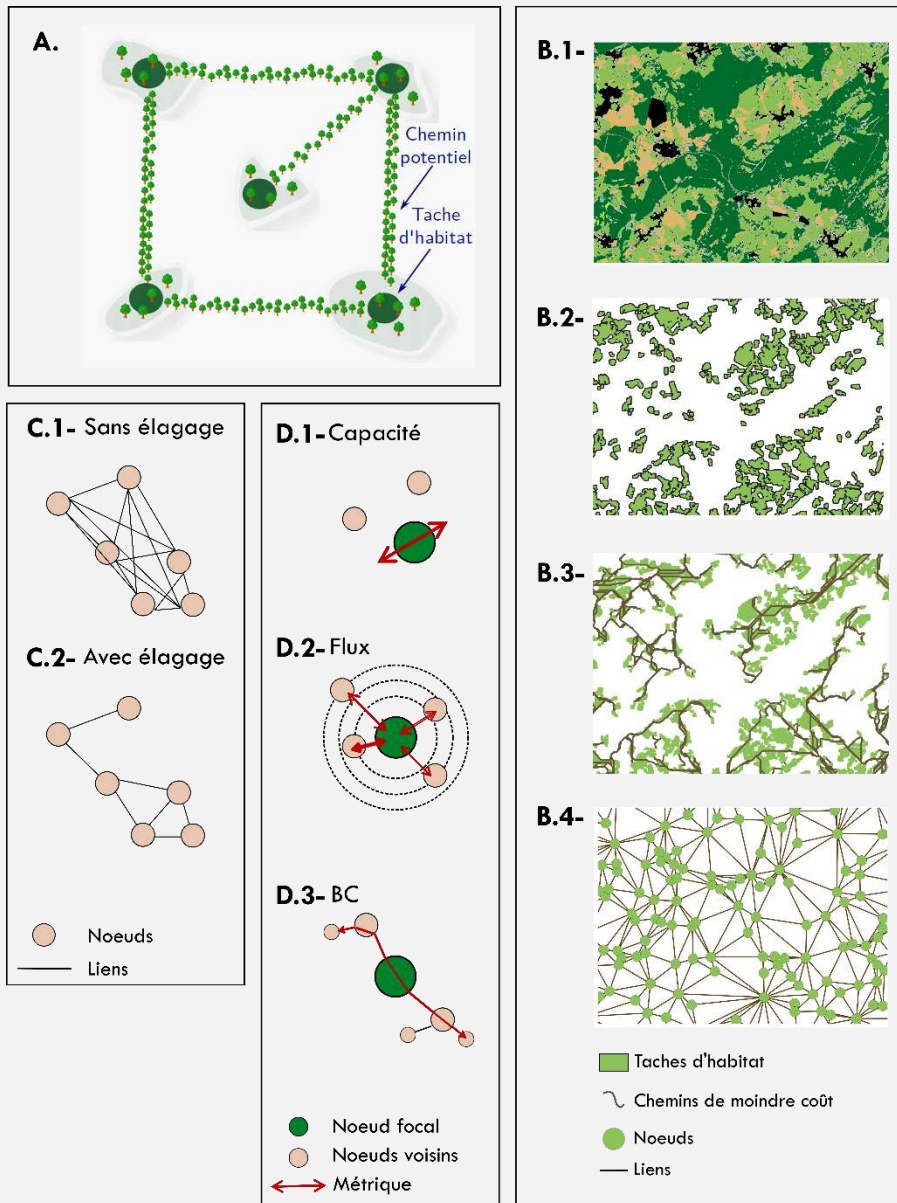
les plus propices aux déplacements compte-tenu de leur localisation et du paysage qu'elles traversent. L'élagage d'un graphe peut reposer sur des critères écologiques (e.g., choisir de ne conserver que les liens associés à une distance inférieure à la capacité de dispersion de l'espèce focale), mais également permettre une meilleure lisibilité des cartographies issues du modèle et une diminution des temps de calcul (Bunn *et al.*, 2000; Foltête *et al.*, 2012).

Les graphes paysagers offrent la possibilité de modéliser la connectivité comme la quantité d'habitat atteignable, à différentes échelles spatiales et niveaux d'analyse. Au niveau des nœuds, les métriques de connectivité permettent d'évaluer la quantité d'habitat atteignable depuis la tache focale et la contribution de cette tache à la connectivité du réseau. Bien qu'une multitude de métriques puissent être calculées au niveau des nœuds d'un graphe, on en distingue trois groupes qui reflètent de façon complémentaire les processus écologiques d'extinction, de colonisation et de dispersion à longue distance :

- Les métriques traduisant le potentiel de *recrutement* d'une tache d'habitat, et reflétant donc son potentiel démographique (e.g., la Capacité, souvent assimilée à la surface de la tache faute de données supplémentaires) (Figure 1.6.D-1).
- Les métriques traduisant le *flux* potentiel de migrants depuis et vers la tache focale (e.g., le Flux, Figure 1.6.D-2)
- Les métriques traduisant la contribution d'une tache à la *traversabilité* du réseau d'habitat, et reflétant donc son rôle dans les événements de dispersion à longue distance. Ces taches permettent que des individus relient facilement des taches d'habitat ou ensembles de taches (e.g., massifs forestiers) éloignés spatialement et/ou séparés de barrières naturelles et anthropiques. La métrique la plus caractéristique de cette catégorie est la métrique de centralité intermédiaire ('*Betweenness Centrality*', BC, Figure 1.6.D-3).

Au niveau des liens, il est possible d'extraire des métriques similaires à celles calculées pour les nœuds, caractérisant par exemple la contribution du lien à la traversabilité du réseau. On peut également exploiter les distances écologiques pondérant les liens. La distance entre deux nœuds modélise un déplacement suivant les liens du graphe, et correspond à la distance cumulée des poids associés à chacun des liens formant le chemin le plus court entre deux nœuds. Cette distance 'sur le graphe' peut ensuite être utilisée dans des analyses de connectivité de la même façon que des distances-coût classiques.

Enfin, les graphes paysagers permettent de caractériser la connectivité à l'échelle d'un réseau d'habitats dans son ensemble. Ces métriques considèrent de façon conjointe la capacité de l'ensemble de taches et leur degré de connexion, et résume ces information en une unique métrique synthétique (Laita *et al.*, 2011).



**FIGURE 1.6 :** Construction des graphes paysagers, et métriques de connectivité en dérivant. A) Représentation schématique des nœuds (tache d'habitat) et des liens (chemin de dispersion) constituant un graphe paysager. B) Étapes de construction d'un graphe paysagers : 1- création de la carte d'occupation du sol, 2- identification des taches d'habitat, 3- détermination des chemins de moindre coût reliant les taches d'habitat, représentation spatialisée, 4- représentation topologique du graphe paysager. C) Représentation schématique de la propriété d'élagage des graphes paysagers. D) Représentation schématique de trois métriques de connectivité complémentaires calculées au niveau des nœuds d'un graphe paysager : 1) la Capacité, 2), le Flux, 3) la 'Betweenness Centrality'(BC).

### 1.2.3. Limites des modèles de génétique du paysage

Le champ disciplinaire de la génétique du paysage est relativement récent et capitalise 21 années de développement depuis la publication de Manel *et al.* (2003). Alors que dans un contexte de changement global, les attentes vis-à-vis de cette discipline sont très grandes, les méthodes d'acquisition de données et les outils analytiques se multiplient rapidement sans qu'on ait assez d'évaluations et de recul concernant leur utilisation (Richardson *et al.*, 2016). La communauté scientifique donc naviguer parmi de nombreuses méthodes de modélisation et prendre le risque de mal les utiliser ou d'extrapoler leurs conclusions hors du domaine de validité de ces approches. Plusieurs études se sont ainsi attelées à répertorier les limites et perspectives des modélisations en génétique du paysage (DiLeo and Wagner, 2016; Manel and Holderegger, 2013; Richardson *et al.*, 2016; Storfer *et al.*, 2010).

#### 1.2.3.1. Limites liées au paramétrage et à la validité des modèles de génétique du paysage

Les premiers écueils surviennent dès les choix de construction du plan d'échantillonnage, s'étendent à l'interprétation des variables génétiques utilisées comme réponses et peuvent compromettre la validité des analyses ultérieures lorsque l'outil analytique utilisé n'a pas été complètement évalué.

##### *Construction du plan d'échantillonnage*

La stratégie d'échantillonnage conditionne la portée et la validité des inférences qui pourront dériver des modèles de génétique du paysage. Short Bull *et al.* (2011) soulignent ainsi que, comme dans toute démarche scientifique, l'absence de détection de l'effet d'un élément paysager ne signifie pas automatiquement que cet élément n'a pas d'importance sur la dispersion des espèces, mais peut aussi résulter d'un manque de puissance statistique. Dans le cadre d'analyses corrélatives, et sans protocole expérimental de type effet-contrôle dans lequel les paramètres peuvent être davantage contrôlés, l'absence de détection de l'effet d'une variable peut simplement s'expliquer par un manque de variabilité de celle-ci à l'échelle du paysage échantillonné. Une attention particulière doit donc être donnée au choix de la zone d'étude et aux caractéristiques du paysage devant permettre de tester les hypothèses et de

répondre aux objectifs de l'étude. Cette considération amène à la question du choix de l'étendue spatiale échantillonnée (Anderson *et al.*, 2010). La génétique du paysage est adaptée à l'étude de l'effet de l'hétérogénéité paysagère sur la distribution de la variation génétique à petite échelle spatiale. Si le paysage échantillonné s'étend à large échelle, la probabilité que les individus échantillonnés proviennent de différentes lignées évolutives augmente. Les effets du paysage contemporain sur la structure génétique sont alors susceptibles d'interagir avec des facteurs phylogéographiques (Richardson *et al.*, 2016; Vandergast *et al.*, 2007), et cette interaction doit être prise en compte lors de l'interprétation du modèle. Enfin, des travaux alertent sur l'influence de la distribution spatiale des populations échantillonnées sur la détection des relations de génétique du paysage (Naujokaitis-Lewis *et al.*, 2013; van Strien *et al.*, 2014). Le protocole d'échantillonnage semble donc être source de facteurs confondants, qu'il conviendrait de contrôler pour améliorer la fiabilité des inférences.

#### *Interprétations du signal de structuration génétique*

Il est assez courant en génétique du paysage d'interpréter un patron de différenciation génétique comme résultant d'une restriction du flux génétique, et de chercher les variables paysagères pouvant être associées à cette potentielle restriction. Pourtant, la différenciation génétique entre deux populations résulte souvent d'une interaction entre de nombreux processus, parmi lesquels le flux génétique n'est pas systématiquement dominant (Richardson *et al.*, 2016; Wagner and Fortin, 2013). En effet, une divergence génétique peut également découler de l'effet de la dérive génétique, notamment dans un réseau de populations de tailles hétérogènes (Prunier *et al.*, 2017), de facteurs phylogéographiques (Dyer *et al.*, 2010) ou encore de la sélection *via* des phénomènes de '*genetic hitchhiking*' (Bertin *et al.*, 2017; Storfer *et al.*, 2015). Compte-tenu de ces multiples sources de structuration génétique, Richardson *et al.* (2016) appellent à considérer les processus évolutifs susceptibles de contribuer à cette structuration dans les modèles de génétique du paysage. Sans cette prise en compte, les résultats pourraient être erronés.

#### *Validité des outils analytiques*

DiLeo et Wagner (2016) ont mis en évidence la grande prépondérance dans les études de génétique du paysage des modèles focalisés sur les connexions entre populations (i.e., au niveau des *liens*) et évaluant l'influence des distances paysagères sur les patrons de différenciation génétique. Ainsi, les effets de la configuration spatiale des habitats et de la perméabilité de la matrice sur la structure génétique sont sur-représentés par rapport aux effets de la quantité d'habitat ou de la composition du paysage autour des populations. De plus, très

peu d'études évaluent la réponse de la diversité génétique à ces facteurs paysagers. Or, la quantité et la configuration des habitats agissent conjointement sur la diversité et la différenciation génétique, car elles affectent toutes deux à la fois la dérive et le flux génétique (Wagner and Fortin, 2013). DiLeo et Wagner (2016) appellent donc au développement d'analyses évaluant simultanément l'effet des deux composantes de la structure du paysage sur les deux composantes de la structure génétique. Des études basées sur les *nœuds* seraient très adaptées à ce type d'analyse et permettraient de modéliser la relation complexe entre patrons paysagers et processus évolutifs.

Par ailleurs, les outils d'analyses classiquement employés pour les études de génétique du paysage au niveau des liens souffrent de nombreux biais, notamment en raison de la non indépendance des valeurs de distance, mentionnée précédemment (Legendre and Fortin, 2010). De nouveaux développements méthodologiques, fondés par exemple sur des analyses multivariées prenant les populations comme individus statistiques (*i.e.*, au niveau des *nœuds*) sont donc nécessaires à la progression du champ disciplinaire. Certaines approches semblent prometteuses, notamment concernant les problématiques de dépendance statistique (Galpern *et al.*, 2014). Par ailleurs, certaines approches déjà adoptées par la communauté scientifique manquent d'étude de validation (Manel and Holderegger, 2013; Richardson *et al.*, 2016). Manel and Holderegger (2013) soulignent donc l'enjeu crucial de la validation des modèles de génétique du paysage, en particulier compte-tenu de leur implication dans l'élaboration de stratégies de conservation. Bien que cette étape de validation ait été mentionnée par Spear *et al.*, (2010) comme une étape à part entière du processus de modélisation par des surfaces de résistance, relativement peu d'études s'y sont jusqu'alors intéressées.

### 1.2.3.2. Limites liées au domaine de validité des modèles de génétique du paysage : la question de l'extrapolation

Les études de génétique du paysage sont généralement centrées sur un couple espèce-paysage unique. Le coût d'acquisition des données limite en effet la réplification large et aisée des modélisations pour un éventail d'espèces et de paysages. Cette limite a très tôt été identifiée (Holderegger and Wagner, 2008; Manel and Holderegger, 2013; Segelbacher *et al.*, 2010; Wagner and Fortin, 2013) et des appels ont été lancés pour orienter les recherches vers des études comparatives multi-espèces et multi-paysages. La progression de la discipline et



l'extension de sa portée, notamment dans le champ de la conservation, repose sur la généralisation des principales conclusions au-delà des zones et modèles d'études considérés jusqu'alors. Il est à l'heure actuelle difficile d'estimer le domaine de validité des conclusions tirées d'une étude : l'espèce focale présentera-t-elle le même comportement de dispersion dans un autre paysage ? Le paysage étudié exerce-t-il la même résistance sur le déplacement d'une autre espèce ? Face à ces questionnements et au besoin des gestionnaires de disposer de recommandations claires à l'échelle de leur territoire d'action et des communautés d'espèces y résidant (Beier *et al.*, 2006), il est parfois tentant d'extrapoler les résultats obtenus pour une espèce dans un paysage particulier. Richardson *et al.* (2016) mettent en garde contre ce risque de « surinterprétation ». En effet, de premières études comparatives ont montré qu'une espèce n'était pas forcément influencée de la même façon sur l'ensemble de son aire de répartition par un même élément paysager (Burgess and Garrick, 2020; Short Bull *et al.*, 2011; Vergara *et al.*, 2017). D'autres études ont mis en évidence que des espèces phylogénétiquement proches et co-occurant sur un même territoire pouvaient présenter des réponses différentes à la connectivité des habitats (Burkhart *et al.*, 2017; Engler *et al.*, 2014; Kebäili *et al.*, 2023). Ces résultats soulignent l'interaction complexe entre les patrons paysagers, les processus évolutifs, et les traits écologiques des espèces. Ils démontrent par ailleurs la nécessité d'un examen attentif des limites inférentielles des modèles de génétique du paysage et de la validité de leurs projections.

Les domaines de validité des modèles doivent donc être évalués dans un espace multifactoriel, considérant bien sûr les espèces, mais également l'espace et le temps. La question de la réplification à travers différents paysages est une première forme de considération du facteur spatial en génétique du paysage. Des études ont également mis en évidence l'impact des résolutions spatiale et thématique, ou encore de l'étendue spatiale à laquelle on évalue les indices de connectivité, sur la détection des relations de génétique du paysage (Anderson *et al.*, 2010; Richardson *et al.*, 2016). Par ailleurs, le postulat d'une correspondance temporelle entre les structures génétiques et paysagère est couramment admis mais rarement vérifié. Or, plusieurs études soulignent l'existence d'un délai de réponse de la structure génétique à un changement paysager (Epps and Keyghobadi, 2015). Une prise en compte des différences d'échelles temporelles entre dynamiques paysagères et dynamiques évolutives dans les modèles de génétique du paysage serait ainsi très pertinente pour garantir la validité de leurs inférences.

### 1.2.3.3. Limites liées à la traduction en mesure de conservation des inférences de génétique du paysage

En 2016, Bowman *et al.* (2016) indiquaient que seules 4% des études de génétique du paysage étaient à visée appliquée pour la conservation de la biodiversité. Pourtant, cette discipline est appelée à jouer un rôle important dans la gestion et la conservation des écosystèmes, en particulier face aux pressions anthropiques croissantes (Manel and Holderegger, 2013; Segelbacher *et al.*, 2010). Pour remplir ce rôle, la génétique du paysage doit opérer un tournant, depuis des approches essentiellement descriptives vers des approches prédictives (Storfer *et al.*, 2007; van Strien *et al.*, 2014). En effet, pouvoir prédire l'impact du changement climatique comme celui d'un aménagement local sur le flux génétique serait précieux pour les décisions publiques en matière de gestion de la biodiversité. Par ailleurs, rendre opérationnelles les inférences de génétique du paysage nécessite de fournir à ces gestionnaires des recommandations claires et des outils facilement exploitables. Bien que ces enjeux soient aujourd'hui clairement identifiés (Foltête *et al.*, 2014; Keeley *et al.*, 2021; Keller *et al.*, 2015), la marge de progression dans ce domaine reste importante. Enfin, développer des approches pour vérifier la validité des modèles de génétique du paysage avant leur opérationnalisation contribuerait certainement à une plus large traduction de leurs résultats au domaine de la conservation.

### 1.3. Objectifs de ce manuscrit de thèse

Les travaux de cette thèse gravitent autour de la question de la validité des modèles de connectivité, et en particulier de ceux dont la construction repose sur les récents apports de la génétique du paysage.

Notre objectif est d'évaluer la validité et les conditions d'application de modèles classiquement utilisés dans le cadre de la génétique du paysage pour inférer l'effet du paysage sur la variabilité génétique neutre, et par suite, les paramètres démographiques de populations situées dans des paysages hétérogènes. Nous souhaitons par ce biais contribuer à la progression de ce champ disciplinaire sur le plan théorique et méthodologique, mais aussi à ses applications à des fins de conservation de la biodiversité.

Plus précisément, nous nous intéresserons à deux outils de modélisation : l'algorithme d'optimisation des surfaces de résistance ResistanceGA, et les modèles de connectivité par graphes paysagers. Notre premier objectif est de proposer des approches basées sur des simulations (Ch. 3.1) et empiriques (Ch. 3.2), permettant de vérifier la validité écologique de ces outils. Un deuxième objectif est de caractériser le domaine de cette validité dans un espace multifactoriel défini par les espèces (Ch. 4), l'espace (Ch. 3, 4) et le temps (Ch. 5). Nous visons ainsi contribuer à la compréhension du potentiel d'extrapolation et de généralisation des inférences de génétique du paysage. Enfin, notre dernier objectif est de proposer des perspectives d'amélioration de ces modèles, pour aller vers des inférences plus robustes et traductibles en mesures de conservation de la biodiversité.





# CHAPITRE 2

## MÉTHODOLOGIE GÉNÉRALE

---

2. MÉTHODOLOGIE GÉNÉRALE .....	37
2.1. Cadre de recherche : projets FRAG&BINV et CANON.....	38
2.2. Zone d'étude : la Guadeloupe.....	39
2.3. Modèles d'étude : cinq espèces d'oiseaux tropicaux.....	42
2.4. Échantillonnage .....	43
2.5. Analyses moléculaires .....	45
2.6. Génétique des populations – Analyses préliminaires .....	46
2.6.1. Contrôle de l'évolution temporelle de la diversité génétique .....	46
2.6.2. Polymorphisme des <i>loci</i> .....	48
2.6.3. Equilibre de Hardy-Weinberg .....	48
2.6.4. Déséquilibre de liaison .....	49
2.7. Données spatiales d'occupation du sol.....	50

## 2.1. Cadre de recherche : projets FRAG&BINV et CANON

Ce projet de recherche doctoral s'inscrit dans la continuité du projet de recherche européen FRAG&BINV (FRAGmentation & Biological INVasion), financé par l'ANR dans le cadre de l'appel à projet ERA-NET Net-Biome 2010. Le projet FRAG&BINV avait pour objectif d'étudier les conséquences de la fragmentation des forêts sur les populations de plusieurs espèces d'oiseaux de la région Caraïbes. En particulier, il s'agissait de comparer les réponses génétiques, phénotypiques et d'interactions parasitaires à la fragmentation des habitats pour un cortège d'espèces distribuées le long d'un gradient de spécialisation pour l'habitat forestier. Ce projet, officiellement terminé depuis 2016, a conduit à l'obtention de plusieurs résultats intéressants. Arnoux *et al.* (2014) et Khimoun *et al.* (2016) ont démontré l'existence d'une structuration génétique pour neuf espèces d'oiseaux à l'échelle de l'archipel des Petites Antilles mais également, et de façon plus étonnante, à l'échelle plus fine de la Guadeloupe pour sept d'entre elles. Les espèces les plus spécialistes de l'habitat forestier se sont révélées les plus sensibles aux effets de la fragmentation (Khimoun *et al.*, 2016b). En particulier, une étude a démontré que la structure du paysage était la principale source de différenciation génétique chez la Paruline caféïette (passereau forestier) en Guadeloupe (Khimoun *et al.*, 2017). Perrin *et al.*, (2021) ont mis en évidence une augmentation de la prévalence parasitaire en milieu fragmenté, associée à une diminution de la diversité génétique neutre et à une détérioration de la condition corporelle chez une espèce d'oiseau hôte en Guyane française. Enfin, Perrin *et al.*, (2023) ont montré que la fragmentation forestière expliquait deux fois mieux la variance de prévalence d'haemosporidés que la perte forestière ou l'hétérogénéité du paysage. La démarche et les thématiques de recherche du projet FRAG&BINV ont ensuite été poursuivies en 2019 par le projet CANON (Landscape Connectivity and Biodiversity: Networks of Networks), financé par le programme 'Investissement d'Avenir' dans le cadre d'un financement ISITE-BFC (ANR-15-IDEX-0003). Les deux objectifs majeurs du projet CANON sont de déterminer comment la structure du paysage influence i) la connectivité fonctionnelle des populations et des communautés d'oiseaux, et ii) la circulation des communautés de parasites au sein des communautés d'oiseaux sauvages en Guadeloupe. Des développements méthodologiques fondés sur la combinaison de graphes paysagers, génétiques et d'interactions hôte-parasite sont depuis ses débuts au cœur de ce projet. Les projets FRAG&BINV et CANON ont d'ores et déjà conduit à la soutenance de deux thèses de doctorat, apportant des éclairages sur les interactions hôte-parasite en milieu fragmenté (Perrin, 2021) et sur l'intérêt de l'utilisation conjointe des

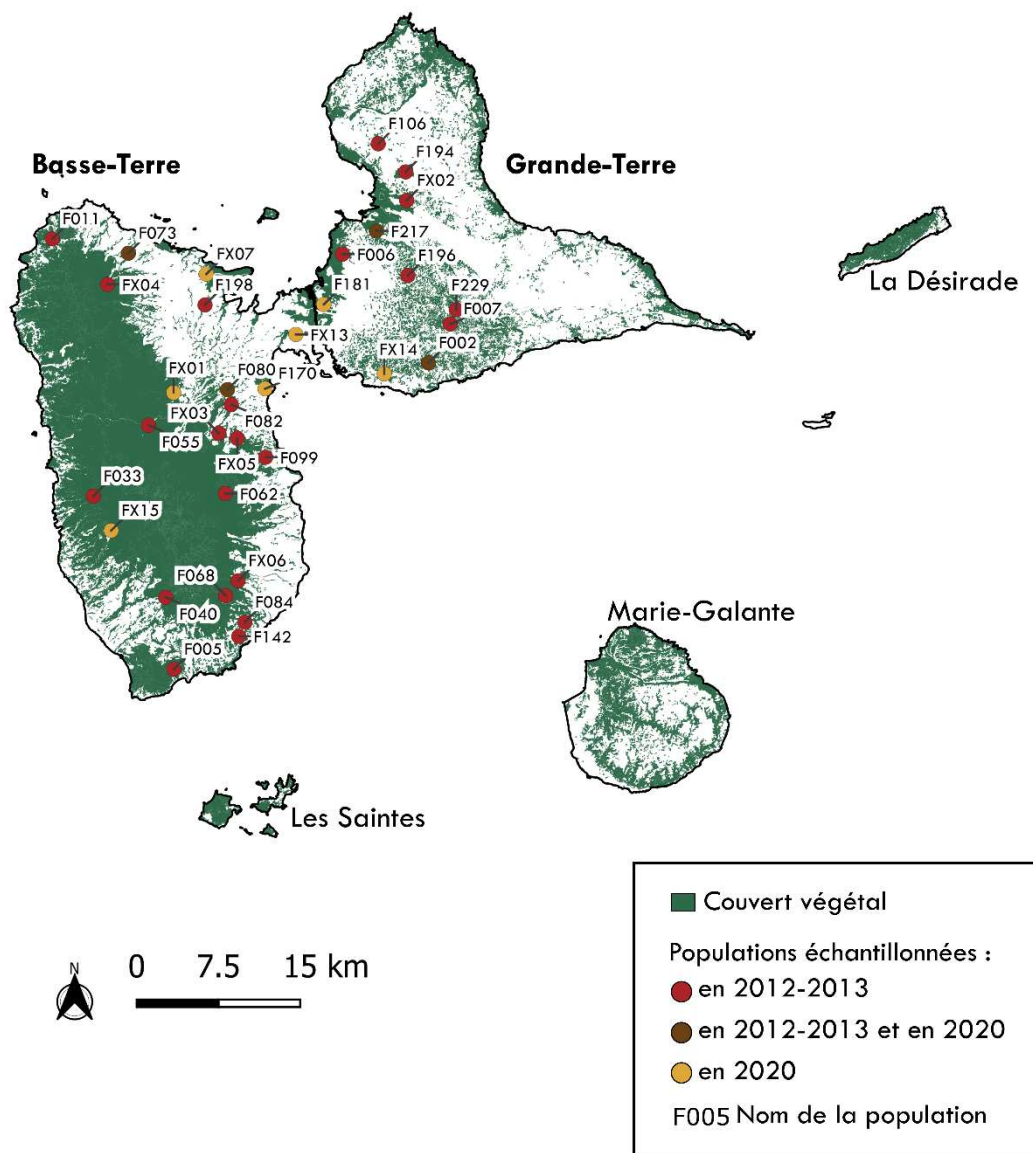
graphes génétiques et paysagers pour l'analyse de la connectivité des habitats (Savary, 2021). Ce manuscrit de thèse se positionne ainsi dans la continuité de ces travaux, qu'il poursuit en évaluant la validité écologique des modèles de connectivité, en particulier dans un contexte paysager insulaire et fragmenté.

## 2.2. Zone d'étude : la Guadeloupe

La Guadeloupe est une région française (1713 km<sup>2</sup>) faisant partie de l'archipel des Petites Antilles situé dans la mer Caraïbes (Figure 2.1). Elle est constituée de 5 îles, dont deux îles majeures formant la 'Guadeloupe continentale' et représentant 89% de sa surface terrestre. Ces deux îles, reliées par un isthme étroit et urbanisé, correspondent à la Grande-Terre (GT) et à la Basse-Terre (BT). Dans cette thèse, seules ces deux îles principales ont fait l'objet d'échantillonnages et d'analyses.

Le climat de la Guadeloupe est tropical, chaud et humide, et l'habitat forestier couvre 43% de l'île. Des variations climatiques et paysagères existent toutefois entre la GT et la BT. En effet, la GT est une île calcaire de faible altitude, balayée par les alizés. Le climat y est donc plus sec que sur BT, comme en témoigne la végétation principalement xérophile. Ce climat a rendu GT historiquement plus propice à l'agriculture et aux implantations humaines. Le couvert forestier y est moins dense que sur BT (22% de couverture forestière, contre 60% sur BT) et très fragmenté, notamment au profit des activités agricoles (cannes à sucre, et élevage bovin et caprin). BT présente quant à elle un relief marqué, comprenant notamment le volcan de la Soufrière, point culminant des Antilles (1467m). Le couvert forestier y est plus dense, cohésif et humide que sur GT (Figure 2.1). Les activités agricoles se concentrent sur les étages de faibles altitudes, et s'orientent vers la culture de bananes, de cannes à sucre, d'agrumes et de café (Eraud *et al.*, 2012; Levesque *et al.*, 2008). La destruction et la fragmentation de l'habitat forestier en Guadeloupe s'est principalement déroulée au cours du XVIIIe siècle, suite à d'intenses épisodes de déforestation visant à accroître la production agricole bananière et sucrière. Elle s'est aujourd'hui stabilisée (Ricklefs and Bermingham, 2008).





**FIGURE 2.1 :** Cartographie des îles de la Guadeloupe, et des fragments forestiers échantillonnés en 2012-2013 et en 2020. 34 sites forestiers ont été échantillonnés : 11 sur Grande-Terre (9 en 2012-2013 et 2 en 2020), et 23 sur Basse-Terre (18 en 2012-2013 et 5 en 2020).

La Guadeloupe est une des îles les plus peuplées des petites Antilles. La pression anthropique sur les zones naturelles y est forte, en partie en raison de la forte croissance démographique. L'INSEE estime ainsi une augmentation de 16% de la densité de populations d'ici l'horizon 2030, pour atteindre 321 hab./km<sup>2</sup> et 548000 habitants, alors qu'elle n'en comptait que 305 000 en 1968. Cette hausse démographique, associée au développement de l'activité touristique et au desserrement des ménages, se traduit par une importante

expansion urbaine. Ces zones urbaines se concentrent majoritairement sur les zones de faibles altitudes, sur GT et au niveau de l'isthme (Eraud *et al.*, 2012; Levesque *et al.*, 2008).

Les milieux insulaires abritent généralement de nombreuses espèces endémiques. Ainsi, BirdLife International a défini les Petites Antilles comme 'Grande Aire d'Endémisme des Oiseaux'. On peut noter que l'ensemble des territoires ultra-marins français rassemblent près de 60 fois plus d'oiseaux endémiques que la métropole (Eraud *et al.*, 2012). Toutefois, la taille réduite de la Guadeloupe et son éloignement aux grandes sources de dispersion de la faune et de la flore du continent américain ont restreint l'endémisme de l'île, qui ne compte aujourd'hui qu'une seule espèce endémique stricte : le Pic de Guadeloupe *Melanerpes herminieri*. La Guadeloupe abrite toutefois 258 espèces d'oiseaux, dont 72 nicheuses et 13 endémiques des Petites Antilles (Levesque *et al.*, 2008).

Ce patrimoine écologique est toutefois menacé. Les espèces endémiques sont en effet caractérisées par des faibles abondances locales et des aires de répartition morcelées, les rendant vulnérables au risque d'extinction local. La France occupe ainsi le 6e rang mondial des pays présentant le plus grand nombre d'espèces d'oiseaux menacées d'extinction (Eraud *et al.*, 2012). A l'échelle de la Guadeloupe, de nombreuses mesures sont mises en place pour œuvrer à la conservation de cette diversité. L'île compte notamment un Parc National couvrant 33500 ha au cœur de BT, une réserve mondiale de biosphère, une Zone Humide d'Importance Internationale Ramsar et deux Réserves Naturelles Nationales.

La Guadeloupe est donc une zone d'étude particulièrement intéressante pour évaluer l'influence de la connectivité des habitats sur la structure génétique des populations aviaires. En effet, l'habitat forestier présente une distribution très contrastée entre BT et GT, tant en termes de surface que de fragmentation. Par ailleurs, les espèces insulaires ont généralement des capacités de dispersion inférieures à celles des espèces du continent, les rendant plus sensibles à la fragmentation de leur habitat (Adler and Levins, 1994; Haig *et al.*, 2011). Ces conditions sont donc favorables à la détection d'un effet de la connectivité de l'habitat forestier sur la structure génétique des populations. De plus, l'insularité de la Guadeloupe, sa petite taille et les interactions limitées des espèces avec les autres îles de l'archipel lui confèrent un statut d'unité paysagère indépendante et exhaustive. Enfin, l'endémisme de cette île et le statut 'vulnérable' de certaines espèces d'après l'Union Internationale pour la Conservation de la Nature (UICN) ajoutent à l'étude de ce territoire des enjeux patrimoniaux et de conservation importants. Manel et Holderegger (2013) reconnaissent les enjeux liés à la préservation des endémismes des

régions tropicales, et déplorent que seul 10 % des études de génétique du paysages publiées avant 2010 portent sur ces régions (Storfer *et al.*, 2010). Ils appellent donc à orienter davantage les recherches vers la préservation des espèces tropicales menacées et nous espérons contribuer à cela dans le cadre de cette thèse.

### 2.3. Modèles d'étude : cinq espèces d'oiseaux tropicaux

Durant les jeunes années de la génétique du paysage, les oiseaux n'étaient pas des modèles d'étude fréquents. Sur les 665 articles considérés dans la synthèse bibliographique de Storfer *et al.* (2010), seuls 10 étudiaient l'effet du paysage sur la structure génétique de populations d'oiseaux (Haig *et al.*, 2011). En effet, la capacité de vol amène facilement à penser que les oiseaux sont capables de contourner ou traverser rapidement des habitats inhospitaliers. Pourtant, leur capacité de dispersion peut être largement restreinte par des barrières, notamment comportementales (Harris and Reed, 2002). De plus, il existe un gradient latitudinal dans leurs capacités de vol, les oiseaux tropicaux étant davantage contraints dans leurs déplacements que les espèces continentales (Haig *et al.*, 2011). Enfin, les zones tropicales ayant été moins impactées que les milieux tempérés par les grands cycles et perturbations climatiques, les espèces de ces régions présentent une plus faible capacité d'adaptation et une plus forte sensibilité à la fragmentation de leur habitat (Betts *et al.*, 2019; Stutchbury and Morton, 2022). Les oiseaux, et en particulier ceux rencontrés dans des climats tropicaux, sont donc des modèles d'études pertinents pour traiter les problématiques actuelles liées à la connectivité fonctionnelle. Ils sont ainsi de plus en plus représentés dans les études de génétique du paysage (Amos *et al.*, 2014; Miller *et al.*, 2013; Mulvaney *et al.*, 2021; Radford *et al.*, 2021).

Dans cette thèse, nous nous sommes intéressés à cinq espèces de passereaux endémiques des Petites Antilles et présents sur l'île de la Guadeloupe : le Sporophile rouge-gorge (*Loxigilla noctis*, ci-après dénommé SR), le Saltator gros-bec (*Saltator albicollis*, SB), le Moqueur grivotte (*Allenia fusca*, MG), la Grive à pieds jaunes (*Turdus lherminieri*, GJ) et la Paruline caféïette (*Setophaga plumbea*, PC) (Figure 2.2).

Ces cinq espèces se positionnent le long d'un gradient de spécialisation pour l'habitat forestier (Figure 2.2). Le SB et le SR sont les espèces les plus ubiquistes (Eraud *et al.*, 2012; Khimoun *et al.*, 2016c). Elles se distinguent par leur abondance importante, notamment au

niveau des Grands Fonds sur GT, et par leur grande aire de répartition au sein de la Guadeloupe et de l'archipel des Petites Antilles. Le MG, bien que plus spécialiste de l'habitat forestier, peut s'observer dans les milieux semi-ouverts, et ne présente donc pas de très fortes exigences écologiques pour l'habitat. La PC et la GJ, en revanche, sont deux espèces fortement spécialistes inféodées à l'habitat forestier. Elles présentent une niche écologique restreinte, conditionnée par la densité et la cohésion du couvert forestier ainsi que l'abondance des précipitations (Eraud *et al.*, 2012; Khimoun *et al.*, 2017). Contrairement au SB et SR, le MG, la PC et la GJ présentent des abondances plus élevées sur BT que sur GT, particulièrement concentrées au niveau du Parc National. Ces trois espèces ont fait l'objet d'une estimation démographique (Eraud *et al.*, 2012) : les effectifs de la PC, du MG et de la GJ en Guadeloupe s'élevaient en 2012 à environ 100 558, 87 711 et 46 920 individus respectivement. La GJ présente donc les plus faibles effectifs et est identifiée comme espèce 'Vulnérable' sur la liste rouge de l'UICN. Les quatre autres espèces de notre étude ont actuellement le statut de 'préoccupation mineure'. Ces cinq espèces de passereaux ont fait l'objet d'études de génétique des populations et du paysage dans le cadre des projets de recherche FRAG&BINV et CANON (Arnoux *et al.*, 2014; Khimoun *et al.*, 2016c, 2016a, 2017). Les résultats de ces études ont mis en évidence l'existence d'une structuration génétique des populations à l'échelle de la Guadeloupe, ainsi qu'une augmentation de l'intensité de la réponse génétique à la connectivité des habitats avec le niveau de spécialisation des espèces.

## 2.4. Échantillonnage <sup>1</sup>

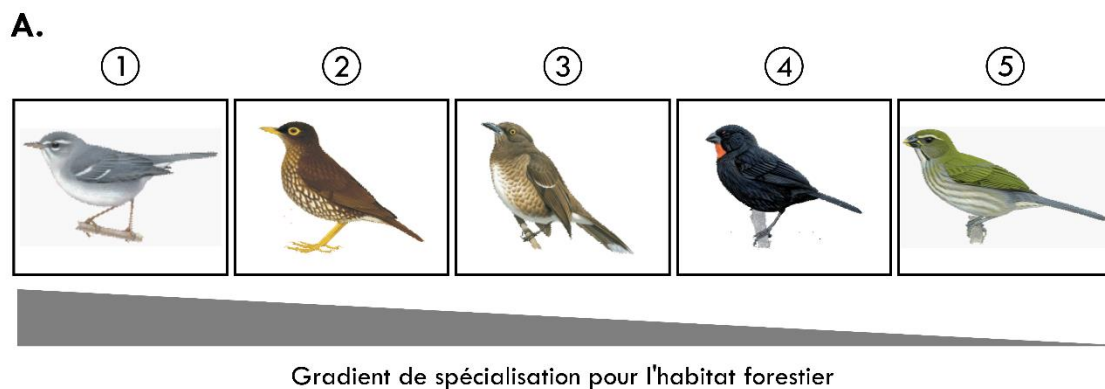
Les oiseaux ( $n = 2917$ ) ont été échantillonnés à l'aide de filets maillants dans 34 taches d'habitat forestier (23 sites sur BT et 11 sites sur GT) (Figure 2.1). L'échantillonnage s'est déroulé entre les mois de mars et juillet, en 2012-2013 pour 27 sites et en 2020 pour 7 autres sites. La plupart du temps, un site était échantillonné en dispersant une dizaine de filets pendant cinq à dix matinées non nécessairement consécutives, et en déplaçant les filets à chaque session de capture. Toutes les espèces n'ont pas été observées sur l'ensemble des 34 sites, et le nombre de population<sup>2</sup> par espèce varie ainsi de 17 à 29 (Figure 2.2). Environ 20  $\mu$ L de sang ont été

---

<sup>1</sup>Je n'ai pas participé à ce terrain, cette campagne d'échantillonnage ayant été réalisée avant le début de cette thèse, dans le cadre des projets CANON et FRAG&BINV. Voir Khimoun *et al.* (2016b).

<sup>2</sup>Nous considérons les individus d'une même espèce échantillonnés sur un même site, et vivant ainsi suffisamment proches pour pouvoir potentiellement se reproduire entre eux, comme une population.

prélevés au niveau de la veine brachiale sur chaque oiseau capturé, à l'aide d'aiguilles stériles et de microcapillaires héparinés, conformément aux exigences légales en France. Les échantillons de sang ont été conservés dans 500  $\mu$ L de *Queen's lysis buffer*. Les oiseaux ont enfin été bagués pour éviter un éventuel rééchantillonnage, et relâchés sur leur lieu de capture. Les autorisations pour l'échantillonnage des oiseaux ont été accordées par la Direction de l'Environnement, de l'Aménagement et du Logement (DEAL) de Guadeloupe, le Parc National de la Guadeloupe, le Conservatoire du Littoral, et l'ONF de Guadeloupe.



**B.**

	Espèces	Individus	Populations	Marqueurs microsatellites
①	Paruline Caféïette	712	27	12
②	Grive à Pieds Jaunes	503	22	10
③	Moqueur Grivotte	322	17	10
④	Sporophile Rouge-Gorge	817	29	10
⑤	Saltator Gros-Bec	563	24	14

**FIGURE 2.2 :** Description des cinq espèces de passereaux échantillonnées. A) Illustrations de chacune de ces espèces d'après le *'Handbook of birds of the worlds'*. Les cinq espèces sont classées selon un gradient de spécialisation pour l'habitat forestier, la Paruline Caféïette (1) étant la plus spécialiste pour l'habitat forestier, et le Saltator Gros-Bec (5) étant le moins spécialisé. B) Table indiquant le nombre d'individus et de populations échantillonnés, ainsi que le nombre de marqueurs microsatellites génotypés, pour chacune des espèces.

## 2.5. Analyses moléculaires

L'ADN des échantillons collectés en 2012 été extrait selon un protocole phénol-chloroforme standard (optimisé à partir de Hillis *et al.* 1996), à partir de 240 µL de sang stocké dans du QLB. Cette étape d'extraction a été initiée par une première étape de digestion, consistant en une période d'une nuit d'incubation à 55 °C avec 10 µL de protéinase K. L'ADN des échantillons de 2020 a été extrait à l'aide des kits *Blood Genomic DNA Mini-Preps (BIO BASIC INC., Markham, Canada)*.<sup>3</sup>

Selon l'espèce, les individus ont été génotypés pour 10 à 14 *loci* microsatellites (Figure 2.2, Annexe 1). Les *loci* ont été amplifiés par PCR, en utilisant des amorces avec une queue M13 (5- CAC GAC GTT GTA AAA CGA C -3) marquées par des fluorochromes (FAM, PET, NED and VIC) (Schuelke, 2000). Les amplifications ont été réalisées à partir de mix PCR de 10 µL contenant l'ADN dilué au 1/10<sup>e</sup>, un tampon GoTaq Flexi 5X, les amorces Forward et Reverse (20 µM), les amorces M13 marquées par fluorescence (20 µM), des dNTPs<sup>4</sup> (5mM), du MgCl<sub>2</sub> (25mM) et l'enzyme Taq polymérase (Promega, Madison, WI, USA). Les concentrations en MgCl<sub>2</sub> et les températures d'hybridation ont été optimisées lors d'analyses préliminaires (Annexe 1). Les *loci* ont été amplifiés en simplex dans un thermocycleur Dyad (Bio-Rad, Hercules, CA, USA) selon le protocole suivant : un premier cycle de dénaturation (3 min), 33 cycles consistant en une phase dénaturation à 94°C (30 s), une phase d'hybridation à la température optimale (1 min 30 s) et une phase d'élongation à 72°C (1 min), suivis de 10 cycles constitués d'une phase de dénaturation à 94 °C (30 s), d'hybridation à 50 °C (1 min 30) et d'élongation à 72 °C (30 s), complétés par une extension finale à 72 °C (5 min). Les produits PCR ont ensuite été multiplexés, ont migré sur un séquenceur automatique (ABI3730), et l'identification des allèles a été réalisée par interprétation de chromatogrammes à l'aide du logiciel GENEIOUS R.8 (Kearse *et al.*, 2012).

<sup>3</sup> J'ai participé à l'extraction d'ADN et au génotypage des échantillons récoltés en 2020 uniquement.

<sup>4</sup> désoxyribonucléoside triphosphate, monomères précurseurs de l'ADN.

## 2.6. Génétique des populations – Analyses préliminaires

Nous présentons dans cette partie les analyses préliminaires de génétique des populations permettant de vérifier la qualité de l'échantillonnage et du génotypage, et d'attester ainsi de la compatibilité du jeu de données avec les analyses de génétique du paysage réalisées.

### 2.6.1. Contrôle de l'évolution temporelle de la diversité génétique

Les jeux de données génétiques de chaque espèce proviennent du regroupement de deux campagnes d'échantillonnage réalisées en Guadeloupe en 2012-2013 et en 2020. Le regroupement de ces données nécessitait de vérifier en amont que durant cette période de temps, les populations ré-échantillonnées n'avaient pas évolué génétiquement au point de rendre significative la différenciation génétique entre les génotypes des individus situés dans une même population en 2012 et en 2020. Dans cet objectif, entre trois et quatre populations échantillonnées en 2012 ont été ré-échantillonnées en 2020 pour chaque espèce. La différenciation génétique au sein des populations entre ces deux pas de temps a ensuite été testée avec le logiciel GENEPOP (Rousset, 2008) par un G-test basé sur la log-vraisemblance.

Les résultats (Table 2.1) témoignent d'une absence de différenciation génétique entre 2012-2013 et 2020 pour toutes les populations de SB, de SR et de MG. Les jeux de données de 2012-2013 et de 2020 ont donc été fusionnés pour ces trois espèces. Pour la GJ et la PC, 2/3 et 3/4 des populations respectivement ne présentaient pas de différenciation génétique temporelle. En revanche, une population chez chacune de ces deux espèces (F080 pour la GJ et F073 pour la PC) était significativement différenciée entre 2012 et 2020.

Nous avons donc réalisé un test de corrélation pour évaluer l'évolution du niveau de différenciation génétique entre ces deux populations concernées par une différenciation temporelle et l'ensemble des autres populations échantillonnées entre 2012. Plus précisément, nous avons étudié la corrélation entre X et Y, avec X correspondant aux valeurs de différenciation génétique entre la population F080 pour la GJ (ou F073 pour la PC) échantillonnée en 2012-2013 et l'ensemble des 21 autres populations échantillonnées en 2012-13, et Y correspondant aux valeurs de différenciation entre la population F080 pour la GJ (ou F073 pour la PC) échantillonnée en 2020 et l'ensemble des 21 autres populations échantillonnées en 2012-13. Une corrélation élevée traduit une stabilité temporelle des niveaux

de différenciation relatif entre populations, indépendamment de l'évolution des fréquences alléliques dans la population concernée par une divergence temporelle, ce qui permettrait de fusionner les données des différentes années. En effet, cela signifierait que bien que les fréquences alléliques de ces populations aient changé au cours du temps, le niveau de différenciation génétique de cette population par rapport aux autres, principale variable réponse dans nos études, serait lui resté stable.

Espèce	Population	Significativité ( <i>pvalue</i> )
Grive à Pieds Jaunes	F073	0.39
	F080	0.047 *
	F217	0.298
Moqueur Grivotte	F073	0.935
	F080	0.955
	F217	0.103
Saltator Gros-Bec	F002	0.483
	F073	0.978
	F080	0.341
	F217	0.829
Sporophile Rouge-Gorge	F002	0.497
	F080	0.338
	F217	0.664
Paruline Caféïette	F073	0.010 *
	F002	0.187
	F080	0.512
	F217	0.729

**TABLE 2.1 :** Résultats des tests de différenciation génétique dans le temps au sein des 4 - 5 populations échantillonnées en 2012-2013 et en 2020, pour les cinq espèces étudiées. Les tests significatifs sont indiqués par un astérisque au niveau de la probabilité du test, dans la colonne 'significativité'.

Les résultats ont montré une corrélation élevée et significative pour les deux espèces (GJ :  $r=0.94$ ,  $p=3,3.10^{-10}$  ; PC :  $r=0.91$ ,  $p=6,2.10^{-11}$ ). La proximité génétique entre la population distinguée par le test de différenciation temporelle et les autres populations chez l'espèce considérée était donc quasiment identique en 2012 et 2020. Ce résultat est soutenu par l'absence de déviation à l'équilibre d'Hardy-Weinberg (cf II.-F. 3) des *loci* dans chacune de ces deux populations (F080-GJ et F073-PC) constituées d'individus échantillonnés en 2012-2013 et



2020. Nous avons donc également fusionné les données génétiques issues de ces deux campagnes d'échantillonnage pour la PC et la GJ.

### 2.6.2. Polymorphisme des *loci*

Interpréter la structure génétique des populations à partir de fréquences alléliques nécessite d'observer plusieurs allèles associés à des fréquences variables sur chaque *locus*. Pour vérifier si les *loci* microsatellites séquencés étaient bien informatifs, nous avons évalué leur polymorphisme en calculant le nombre d'allèles observés à chaque *locus*. Les résultats ont montré que le nombre moyen d'allèles par *locus* variait de 4.4 pour le MG à 14.9 pour la PC (Table 2.2). Ce niveau de polymorphisme semblait satisfaisant pour les analyses de génétique des populations et du paysage envisagées, bien qu'il soit plus restreint pour le MG que pour les quatre autres espèces.

Espèces	Nombre moyen d'allèles	Nombre minimum et maximum d'allèles observés pour un <i>locus</i>
Grive à Pieds Jaunes	8,7	4 - 18
Moqueur Grivotte	4,4	2 - 7
Saltator Gros-Bec	7,1	2 - 17
Sporophile Rouge-Gorge	9,9	3 - 18
Paruline Caféiette	14,9	5 - 25

**TABLE 2.2 :** Résultats concernant le niveau de variabilité des *loci* génotypés pour les cinq espèces considérées. Nombre moyen d'allèles : moyenne de nombre d'allèles observés à chaque *locus*.

### 2.6.3. Equilibre de Hardy-Weinberg

Afin de détecter d'éventuels problèmes d'amplification, nous avons testé la déviation à l'équilibre d'Hardy-Weinberg pour chaque *locus*, dans chaque population et pour chaque espèce avec le logiciel GENEPOP (Rousset 2008). Les résultats, synthétisés dans la Table 2.3, ont montré qu'une grande majorité des *loci* est à l'équilibre d'Hardy-Weinberg. Sept tests restaient significatifs après correction de Bonferroni, et tous correspondaient à un déficit d'hétérozygotes. Néanmoins, ces tests concernaient un *locus* en déséquilibre dans une unique population. Dans le cas d'une problématique d'allèles nuls, on s'attendrait à observer un

déséquilibre du même *locus* dans plusieurs populations, ce qui n'était pas le cas ici. L'ensemble des *loci* séquencés ont donc été conservés pour chaque espèce pour la suite des analyses.

Espèce	Nombre de <i>loci</i> en déséquilibre	Sens du déséquilibre éventuel
Grive à Pieds Jaunes	0 / 220 tests	-
Moqueur Grivotte	1 / 170 tests	Déficit d'hétérozygotes
Saltator Gros Bec	0 / 336 tests	-
Sporophile Rouge Gorge	4 / 290 tests	Déficit d'hétérozygotes
Paruline Caféïette	2 / 320 tests	Déficit d'hétérozygotes

**TABLE 2.3** : Résultats des tests de déviation par rapport à l'équilibre de Hardy-Weinberg, après correction de Bonferroni, pour chaque *locus*, dans chaque population, et pour les cinq espèces considérées

#### 2.6.4. Déséquilibre de liaison

Enfin, nous avons vérifié l'indépendance des différents *locus* en testant les déséquilibres de liaison entre toutes les paires de *locus* dans chaque population et pour chaque espèce, en utilisant le logiciel GENEPOP (Rousset 2008).

Espèce	Nombre de paires de <i>loci</i> en déséquilibre
Grive à Pieds Jaunes	0 / 990 tests
Moqueur Grivotte	0 / 765 tests
Saltator Gros-Bec	0 / 2184 tests
Sporophile Rouge-Gorge	1 / 1305 tests
Paruline Caféïette	4 / 1782 tests

**TABLE 2.4** : Résultats des tests de déséquilibre de liaison entre toutes les paires de *loci* dans chaque population, après correction de Bonferroni.

Les résultats (Table 2.4) ont confirmé l'indépendance statistique des *loci* pour la quasi-totalité des tests réalisés. Une paire de *loci* dans une population de SR présente un déséquilibre, mais ce résultat n'est pas inquiétant car elle ne ressort comme étant significativement liée dans aucun des autres tests réalisés. Pour la PC en revanche, trois déséquilibres sur quatre surviennent dans une même population (F080). La lecture des génotypes pour cette population

a été contrôlée une seconde fois, sans que nous ne puissions repérer d'anomalie. Cette population devra donc faire l'objet d'une attention particulière lors de l'interprétation des analyses de structure génétique pour cette espèce.

## 2.7. Données spatiales d'occupation du sol

Dans la suite du manuscrit, ces données génétiques sont mises en relation avec des données spatiales utilisées pour modéliser la structure des paysages. Les données spatiales utilisées dans ce manuscrit pour les analyses empiriques de génétique du paysage sont présentées brièvement ici. Le paysage a été modélisé par l'occupation des sols en s'appuyant sur plusieurs sources de données pour la Guadeloupe :

- *Le CORINE Land Cover data (2012)* : Il s'agit d'une base de données européenne d'occupation des sols issue de l'interprétation visuelle d'images satellitaires et dotée d'une résolution spatiale de 20 à 25m. Les différentes catégories d'occupation du sol qui y sont recensées ont été agrégées dans nos analyses pour atteindre entre 9 et 10 types d'occupation, en fonction des besoins des analyses.
- *La BD TOPO (2012)* de l'Institut National de l'Information Géographique et Forestière (IGN) : cette base de données propose une représentation vectorielle des infrastructures du territoire. Nous en avons extrait les zones bâties et les axes routiers principaux.
- *Le diagnostic de la végétation forestière de Guadeloupe* de 2010 et de 1950, réalisé par le Conseil Départemental de la Guadeloupe, l'IGN, et la Direction Régionale de l'Office National des Forêts (ONF). Cette source de données nous a permis de bénéficier d'une résolution thématique très fine de l'habitat forestier.

Ces trois couches de données spatiales ont été superposées et rastérisées à différentes résolutions en fonction des analyses, de façon à optimiser le ratio entre précision et temps de calcul (résolution de 20m pour la construction des graphes paysager mais de 100m pour l'optimisation des surfaces de résistance par exemple). Les cartes d'occupation du sol ainsi obtenues ont servi de support à différentes analyses de connectivité, telles que la construction des graphes paysagers, au calcul des métriques de connectivité structurelle, ainsi qu'à l'optimisation des surfaces de résistance. Ces analyses sont détaillées par la suite, au fil des chapitres.





# CHAPITRE 3

## VALIDITÉ DES MODÈLES DE L'EFFET DE LA CONNECTIVITÉ SUR LA VARIABILITÉ GÉNÉTIQUE

---

<b>3. VALIDITÉ DES MODÈLES DE L'EFFET DE LA CONNECTIVITÉ SUR LA VARIABILITÉ GÉNÉTIQUE.....</b>	<b>53</b>
<b>3.1. Validation par simulation d'une méthode d'optimisation de distances-coût à partir de données génétiques.....</b>	<b>55</b>
3.1.1.Introduction .....	55
3.1.2.Matériels et méthodes .....	55
3.1.3.Résultats principaux .....	56
3.1.4.Article : What can optimized cost distances based on genetic distances offer?.....	57
<b>3.2. Validation empirique des modèles de graphes paysagers reflétant l'effet de la connectivité des habitats sur le flux génétique.....</b>	<b>87</b>
3.2.1.Introduction .....	87
3.2.2.Matériels et méthodes général.....	87
3.2.3.Résultats principaux .....	88
3.2.4.Article : Validating graph-based connectivity models with independent presence-absence and genetic data sets.....	89
<b>3.3. Conclusions.....</b>	<b>105</b>

Dans ce chapitre, nous nous attacherons à évaluer les performances, la validité et les conditions d'application de deux outils couramment exploités en génétique du paysage, à savoir l'algorithme d'optimisation ResistanceGA (Peterman, 2018) ainsi que les graphes paysagers. Le premier outil s'utilise pour inférer la relation entre les données génétiques et spatiales, tandis que le second modélise les contraintes paysagères liées à la dispersion à partir de données spatiales. Ces deux outils reposent sur des approches et concepts très différents. ResistanceGA est un algorithme d'optimisation piloté par la donnée (approche inductive, ou '*data-driven*'), à visée inférentielle et utilisé pour caractériser des patrons génétiques d'isolement par la résistance du paysage. Les graphes paysagers reposent quant à eux sur la théorie des graphes pour modéliser de façon exhaustive et peu coûteuse (i.e., sans nécessairement de données empiriques) les taches d'habitat et leur implication dans les processus de dispersion. Ces deux outils s'utilisent donc tous deux dans l'objectif général de modéliser l'effet de l'hétérogénéité du paysage sur les flux de gènes. En effet, ResistanceGA infère des valeurs de coûts représentant la résistance des éléments paysagers aux flux de gènes, tandis que les graphes paysagers fournissent des métriques de connectivité paysagère supposées refléter le rôle de la connectivité sur la structure génétique. Ces graphes paysagers sont construits à partir de paramètres de coûts similaires à ceux inférés par l'algorithme ResistanceGA, à la différence que l'approche par graphe nécessite une connaissance *a priori* de ces paramètres tandis que RGA les infère depuis des données observées. Ces deux approches sont très utilisées en génétique du paysage, mais rarement de façon conjointe et sans que leur validité écologique ait préalablement été évaluée. Dans ce chapitre, nous mobiliserons donc des données empiriques et simulées pour évaluer si i) les valeurs de coûts inférées par ResistanceGA fournissent des proxys performant de la structuration génétique des populations, et si ii) les graphes paysagers reflètent l'effet de la connectivité de l'habitat sur la structure génétique des populations.

### 3.1. Validation par simulation d'une méthode d'optimisation de distances-coût à partir de données génétiques

#### 3.1.1. Introduction

Modéliser la connectivité des habitats est un enjeu méthodologique central pour la conservation de la biodiversité. Cette modélisation s'appuie souvent sur des surfaces de résistance, qui modélisent spatialement la propension au déplacement, le coût physiologique et le risque de mortalité encourus par les individus se déplaçant à travers des environnements hétérogènes. Une surface de résistance est ainsi paramétrée par un scénario de coût, associant une valeur de coût à chaque élément paysager cartographié. Le choix de ces paramètres de coût est une étape clef conditionnant la validité écologique de ces surfaces. Bien que beaucoup d'études fondent ce paramétrage sur des connaissances naturalistes et des dires d'experts, plusieurs méthodes ont récemment été développées pour inférer de façon plus objective des valeurs de coût reflétant les contraintes au processus de dispersion. ResistanceGA (RGA) est un algorithme d'optimisation couramment utilisé pour paramétrer ces surfaces de résistance (Peterman, 2018). Il s'agit d'un algorithme génétique (voir (Mitchell, 1995) qui explore aléatoirement des valeurs de ces paramètres de coûts, jusqu'à obtenir celles maximisant l'ajustement statistique entre les distances génétiques et les distances-coût calculées à partir des valeurs de ces paramètres. En dépit de son utilisation généralisée dans le domaine de la génétique du paysage (voir par exemple Antunes *et al.*, 2023; Mulvaney *et al.*, 2021), la pertinence écologique et les performances de cet algorithme ont rarement été étudiées. Ainsi, dans cette étude, nous avons évalué les capacités et les conditions optimales d'application de RGA, à la fois à des fins prédictives et d'interprétation des propriétés traits paysagers vis-à-vis des flux de gènes.



### 3.1.2. Matériels et méthodes

Nous avons pour cela effectué des simulations démo-génétiques dans des paysages réels, pour des espèces virtuelles aux profils écologiques distincts (c'est-à-dire caractérisées par des capacités de dispersion et des niveaux de spécialisation contrastés). A chaque profil écologique correspondait un scénario de coût de déplacement de référence. Nous avons ainsi simulé les distances génétiques résultant de plusieurs générations durant lesquelles les individus se déplaçaient en fonction de ces scénarios de coût et des capacités de dispersion propres à leur espèce virtuelle. Nous avons ensuite utilisé ces distances génétiques simulées pour paramétrer les surfaces de résistance en utilisant RGA. Cela revenait à vérifier la capacité de cet algorithme à retrouver les valeurs de coût des scénarios utilisés. Les performances d'optimisation de l'algorithme ont été évaluées selon différents critères. Tout d'abord, nous avons évalué la capacité de RGA à distinguer correctement les principales hypothèses couramment invoquées pour caractériser le moteur des patrons génétiques simulés. Il s'agissait de déterminer si la solution optimale de l'algorithme suggérait que les patrons génétiques résultaient d'un phénomène d'isolement par la résistance du paysage (IBR), par la distance (IBD), ou n'étaient pas principalement liés aux distances écologiques ou euclidiennes entre populations (hypothèse nulle). Lorsque l'hypothèse d'IBR était retenue (hypothèse correcte compte-tenu des paramètres de nos simulations), nous avons évalué le pouvoir prédictif de ces modèles basés sur l'IBR, ainsi que les paramètres optimisés par RGA. En d'autres termes, nous avons évalué la capacité des valeurs de coût inférées à prédire les distances génétiques simulées. Enfin, nous avons évalué la capacité de RGA à retrouver le scénario de coût de référence à partir des données génétiques. Cela a consisté à évaluer la similarité des scénarios de coût optimisés et du scénario de référence.

### 3.1.3. Résultats principaux

Nous avons mis en évidence la sensibilité de RGA à la nature du phénomène qui structure spatialement les patrons de différenciation génétique. Les inférences permettaient en effet de distinguer correctement les patrons d'IBR, des patrons d'IBD ou nul. Deuxièmement, nous avons démontré l'excellent pouvoir prédictif des modèles optimisés par RGA. En effet, les scénarios de coûts optimisés permettaient de prédire les distances génétiques avec acuité,

parfois même paradoxalement avec davantage de précision que les scénarios de référence. Cette performance dépendait toutefois de l'interaction entre les capacités de dispersion et le niveau de spécialisation de l'espèce focale. De plus, la considération de l'échelle spatiale à laquelle les processus de dispersion influencent les patrons génétiques (par élagage des matrices de distance) permettait d'améliorer les performances prédictives de l'algorithme. Enfin, nous avons démontré que RGA ne parvenait que rarement à retrouver exactement le scénario de coût de référence. En effet, les valeurs des coûts optimisées s'écartaient souvent des valeurs de référence, reflétant davantage les conséquences d'un sur-ajustement statistique (« *overfitting* ») que l'influence des éléments paysagers sur les processus écologiques. L'interprétation des valeurs de coût inférées par RGA en termes de perméabilité aux flux de gènes, ainsi que leur extrapolation à d'autres paysages, sont donc potentiellement fallacieuses. Nos résultats démontrent donc le grand intérêt de cet algorithme d'optimisation des surfaces de résistance pour la modélisation à visée prédictive, mais appellent à une interprétation prudente de la valeur absolue des paramètres optimisés. Nous proposons également dans cette étude des pistes et perspectives pour une utilisation optimale de cet outil.

#### 3.1.4. Article

### **What can optimized cost distances based on genetic distances offer? A simulation study on the use and misuse of ResistanceGA**

Daniel, Alexandrine ; Savary, Paul ; Foltête, Jean-Christophe ; Vuidel, Gilles ; Faivre, Bruno ; Garnier, Stéphane ; Khimoun, Aurélie

Cet article a été accepté sous réserve de modifications mineures dans la revue *Molecular Ecology Resources* et est en cours de révision.

## What can optimized cost distances based on genetic distances offer?

### A simulation study on the use and misuse of ResistanceGA

Daniel, Alexandrine<sup>1\*</sup>; Savary, Paul<sup>2</sup>; Foltête, Jean-Christophe<sup>3</sup>; Vuidel, Gilles<sup>3</sup>; Faivre, Bruno<sup>1</sup>; Garnier, Stéphane<sup>1,4</sup>; Khimoun, Aurélie<sup>1,4</sup>

1: Biogéosciences, UMR 6282 CNRS, Université Bourgogne-Franche-Comté, 6 Boulevard Gabriel 21000 Dijon, France

2: Department of Biology, Concordia University, 7141 Sherbrooke Street West, Montreal, QC, H4B 1R6, Canada

3: ThéMA, UMR 6049 CNRS, Université Bourgogne-Franche-Comté, 32 Rue Mégevand, 25030 Besançon, Cedex, France

4 : These authors contributed equally to this work.

\* corresponding author

Keywords: landscape genetics; resistance optimization; ResistanceGA; gene flow; genetic simulation

#### 3.1.4.1. Abstract

Modeling population connectivity is a central stake in biodiversity conservation and often relies on resistance surfaces reflecting individual movements. ResistanceGA (RGA) is an optimization framework commonly used to parametrize these surfaces by maximizing the statistical fit between genetic distances and cost distances. As the reliability of this framework has rarely been studied, we investigated the conditions that maximize its accuracy, for both predictive purposes and the interpretation of landscape features' permeability. We ran demographic simulations in real landscapes, for distinct ecological profiles (i.e., contrasted dispersal capacities, specialization levels), and corresponding reference scenarios of movement costs. We then parametrized resistance surfaces from the simulated genetic distances using RGA. We first evaluated whether RGA correctly distinguished the main drivers of the simulated genetic patterns (i.e., distinguished Isolation-by-Resistance (IBR) patterns from Isolation-by-Distance or spatial patterns unrelated to ecological nor Euclidean distances). We then assessed RGA predictive performance for IBR patterns and its ability to recover the reference cost scenarios. IBR patterns were well detected and genetic distances were predicted with great accuracy. This performance depended on the interaction between dispersal capacities and habitat specialization, and was further enhanced by pruning distance matrices for matching the spatial scale of dispersal influence on genetic patterns. However, the absolute optimized cost values often departed from the reference values, making their interpretation and extrapolation to other landscapes potentially dubious. We therefore demonstrate the great value of this optimization framework for predictive modeling, but call for caution and provide additional guidance for its optimal use.

### 3.1.4.2. Introduction

Defined as the movement of an individual from its native population to another location for breeding, dispersal is a key ecological process that directly affects evolutionary dynamics by moderating gene flow. The resulting movement of individuals and genetic variations shapes both species and genetic diversity, and influences biotic interactions (Richardson *et al.* 2016; Spielman, Brook, and Frankham 2004; Frankham 2015; Schlägel *et al.* 2020). Considering dispersal is therefore critical in the current context of habitat fragmentation and species range shifts triggered by human activities and global changes (Crispo *et al.* 2011; Crooks *et al.* 2017; Manel and Holderegger 2013). Accordingly, spatial models of connectivity have been of great help for identifying population connectivity drivers, and for implementing sound biodiversity conservation policies (Correa Ayram *et al.* 2016; Newmark *et al.* 2023; Rudnick *et al.* 2012).

Connectivity modeling approaches often represent the landscape as a resistance surface embedding the movement propensity, physiological cost, and mortality risk incurred by individuals dispersing across heterogeneous environments (Zeller, McGarigal, and Whiteley 2012; Diniz *et al.* 2020). As such, resistance values are commonly associated with discrete landscape features and the corresponding resistance surfaces are spatialized as raster maps (Spear *et al.* 2010; Spear, Cushman, and McRae 2015). Modeling connectivity then consists in translating these spatially-explicit resistance assumptions and our mechanistic understanding of dispersal movements into connectivity estimates, for detecting landscape barriers, mapping corridors, or computing connectivity metrics (Dutta *et al.* 2022, Foltête *et al.* 2014). From the above, assigning a cost value to each landscape feature is a critical modeling decision, with far-reaching consequences for the reliability of connectivity analyses.

Several methods have been developed over the past 20 years to bring cost values closer to the reality of dispersal processes (Peterman *et al.* 2019; Zeller, McGarigal, and Whiteley 2012; Dutta *et al.* 2022; Vanhove and Launey 2023; Peterman and Pope 2021). Although many connectivity studies rely upon expert opinion to assign cost values (Spear *et al.* 2010), Beier *et al.* (2008) argued that inferring landscape resistance from genetic or animal movement data improves the ecological relevance of connectivity models. In this context, the ResistanceGA (RGA) framework has received a great deal of interest. This recent resistance surface optimization method is based on genetic data and assumes that genetic distances reflect landscape functional connectivity by providing insights into gene flow (Zeller *et al.* 2017). Developed and released in the form of an R package by Peterman (2018), this approach has

been increasingly used since 2018 (Zeller *et al.* 2023; Antunes *et al.* 2022; McCluskey *et al.* 2022; Atzeni *et al.* 2023). In essence, this framework infers dispersal costs from the statistical fit between genetic and cost distances. Its power stems from the use of a genetic optimization algorithm, which explores the cost parameter space, without requiring any *a priori* restriction of the cost value range.

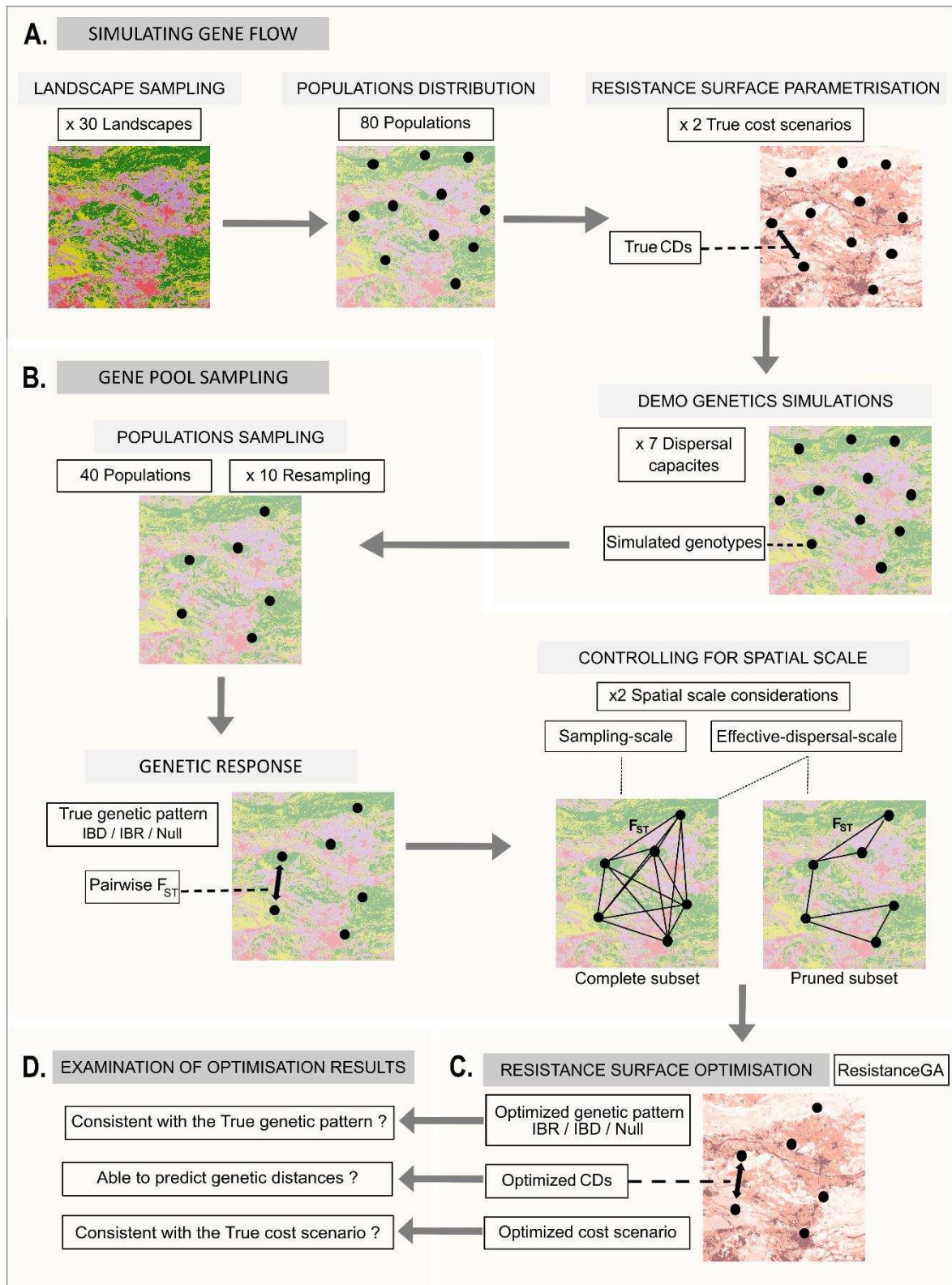
Despite the common use and flexibility of this optimization tool, very few studies have examined the validity of its inferences (but see Peterman *et al.* 2019; Beninde, Wittische, and Frantz 2023; Winiarski, Peterman, and McGarigal 2020). Although Peterman *et al.* (2019) and Winiarski *et al.* (2020) obtained encouraging results regarding RGA performance, assessing high correlations between true and RGA optimized resistance surfaces, they called for further studies relying on direct simulations of genetic processes across a large range of resistance scenarios. According to these authors, several population genetics mechanisms or landscape structure properties could potentially affect landscape genetic inference and remain unexplored. Their call for future work on the sensitivity of RGA to landscape composition or sampling design was followed by the study of Beninde *et al.* (2023). The latter study examined RGA performance across numerous genetic distance metrics and a wide range of resistance scenarios, which were assumed to reflect variation in landscape fragmentation, the number of features limiting gene flow, and species dispersal capacity. In this study, RGA better recovered the true resistance surface when few landscape features restricted gene flow. Yet, these works called for more explicit examination of the influence of the ecological profile and degree of habitat specialization of the species under study. For instance, we can reasonably assume that the interaction between dispersal capacity and the level of habitat specialization determines the spatial scale at which populations are linked by dispersal events. Assessing and considering this spatial scale in the inference may allow discarding pairwise genetic distances which are mainly driven by genetic drift, and as such generate stochastic noise hindering the inference of landscape structure effect on gene flow.

More generally, Beninde *et al.* (2023), Winiarski *et al.* (2020) and Peterman *et al.* (2019) called for a more critical look at the resistance surfaces optimized with RGA from empirical data, especially when the results guide the design of conservation measures. Of particular concern is the overfitting effect inherent to the optimization process, which may impede the accurate assessment of the relative resistance values of each landscape feature. Indeed, on the one hand, it is commonly known that unsupervised data-driven approaches are prone to overfitting when several parameter sets can result in the same pattern. As such, further research

testing whether bootstrapping or leave-one-out cross-validation could be used to limit overfitting is necessary. On the other hand, while not without limits for the understanding of the processes at play, the data-driven nature of the ResistanceGA stochastic search algorithm (Scrucca 2013) would make it perfectly suited for predictive modeling (Shmueli 2010). Indeed, only a handful of studies have explored the predictive capacity of cost-distance inferences, despite a call for landscape genetics research to contribute to the prediction of genetic responses to landscape changes (Maarten J. van Strien *et al.* 2014; Storfer *et al.* 2007; Balkenhol *et al.* 2009).

To provide the landscape genetics community with evidence-based guidelines regarding the use and misuse of resistance surface optimization frameworks, we evaluated the performance of RGA across a wide range of realistic landscape structures, ecological profiles, sampling designs, and spatial scales. For that purpose, we simulated gene flow using categorical resistance surfaces reflecting the level of habitat specialization and dispersal capacities of distinct virtual species. We then optimized cost distances (i.e., the sum of cost values along the least cost path between two populations) from these surfaces, based on the simulated genetic distances using the ResistanceGA algorithm. Our simulation framework aimed at assessing:

- i) the sensitivity and the specificity of the optimization framework for distinguishing the drivers of genetic differentiation (i.e., distinguishing Isolation by Resistance (IBR) patterns over Isolation by Distance (IBD) or spatially structured patterns unrelated to ecological nor Euclidean distances),
- ii) the accuracy of the optimized cost distances to predict genetic differentiation
- iii) the congruence between the optimized cost scenario and the true cost scenario shaping the simulated genetic patterns



**FIGURE 3.1.1:** Overall methodology for assessing the ability of optimized cost distances to reliably reflect landscape effect on genetic structure. A) Workflow for simulating gene flow among virtual populations on real landscapes, depending on 2 scenarios of movement costs and 7 dispersal capacities. B) Preparation of genetic inputs for the optimization process by sampling 10 times 40 populations and evaluating their related pairwise genetic distance. Two genetic distance datasets were considered depending on spatial scale. C) Optimization of cost distances according to these genetic inputs using the RGA workflow. D) Evaluation of RGA inferences according to the sensitivity of the algorithm to genetic pattern, the predictive power of optimized cost distances and the ability to recover the true cost scenario. CDs: Cost distances, IBR: Isolation-by-Resistance, IBD: Isolation-by-Distance, Null: No spatial structure.

### 3.1.4.3. Materials & Methods

#### *Overall Approach*

To assess whether optimized cost distances were reliable for detecting and predicting landscape effect on genetic structure, we performed demo-genetic simulations in real landscapes for virtual species, varying in their dispersal costs and capacities (Figure 3.1.1). We then used RGA and assessed the performances of this optimization framework in predicting the resulting genetic differentiation patterns and recovering the “true” dispersal costs used for the simulations.

#### 3.1.4.3.1. Real landscape sampling

We selected 30 real landscapes (40\*40 km, 100m cell resolution) in metropolitan France that maximized variations in the amount and spatial configuration of the four land cover types considered (forest, grassland, agricultural, and urban areas), using the 2012 CORINE Land Cover data (“Union Européenne – SOeS, CORINE Land Cover” 2012). See Annexe 2-S3.1.1 for more details about landscape sampling.

#### 3.1.4.3.2. True resistance surfaces parametrization and cost-distances calculation

We designed two cost scenarios characterizing the resistance of each land cover type to individual dispersal. We based the construction of these scenarios on the negative exponential function introduced by Keeley, Beier, and Gagnon (2016) (See Annexe 2-S3.1.2 for details). The two resulting cost scenarios represented the following contrasted ecological profiles: i) a virtual forest-specialist species experiencing high dispersal resistance across all land cover types but forest: {Cost<sub>forest</sub>=1; Cost<sub>grassland</sub>= 700; Cost<sub>agricultural</sub> =900; Cost<sub>urban</sub>=1000} ; ii) a more generalist virtual species with a smoother range of resistance values, experiencing strong resistance across urban areas only: {Cost<sub>forest</sub>=1; Cost<sub>grassland</sub>= 50; Cost<sub>agricultural</sub> =200; Cost<sub>urban</sub>=1000}. Hereafter, these true cost scenarios will be referred to as a proxy for species specialization levels.

We located the virtual populations by randomly sampling 80 forest pixels in each of the 30 landscapes. We ensured that the sampled pixels were more than 500 m apart, at least 2 km from the edge of the raster to avoid edge effects, and located in forest patches of more than 25 ha.



Then, we computed the least-cost paths and corresponding cost distances among populations on each true resistance surface, using the `graph4lg` package (Savary *et al.* 2021b).

#### 3.1.4.3.3. Demo-genetics simulations and gene pool sampling

We used the `PopGenReport` package (Adamack and Gruber 2014) to simulate gene flow and resulting individual genotypes. The simulated dispersal was conditioned by the level of specialization (according to the two “true” cost scenarios) and dispersal abilities of the virtual species. We considered 7 different maximum dispersal distances. Local population sizes remained constant and equal to 25 individuals throughout simulations. Further details about simulation parameters are given in Annexe 2-S3.1.3. We ran a total of 420 simulations (30 landscapes x 2 levels of specialization x 7 dispersal capacities) over 200 generations each, to obtain a steady pattern of genetic differentiation. Then, we used individual genotypes to calculate pairwise genetic distances between populations, using pairwise  $F_{ST}$  (Weir and Cockerham 1984).

To test for the effect of sampling design on RGA performance, we randomly sampled 40 of the 80 populations in each landscape (10 sampling iterations) and extracted the corresponding  $F_{ST}$  matrices between the sampled populations. Every combination of a landscape, a specialization level, a dispersal capacity level, and a sampling design (4,200 combinations) will be referred to as a ‘statistical individual’ in our analytical workflow.

#### 3.1.4.3.4. Spatial scale of landscape influence on genetic structure

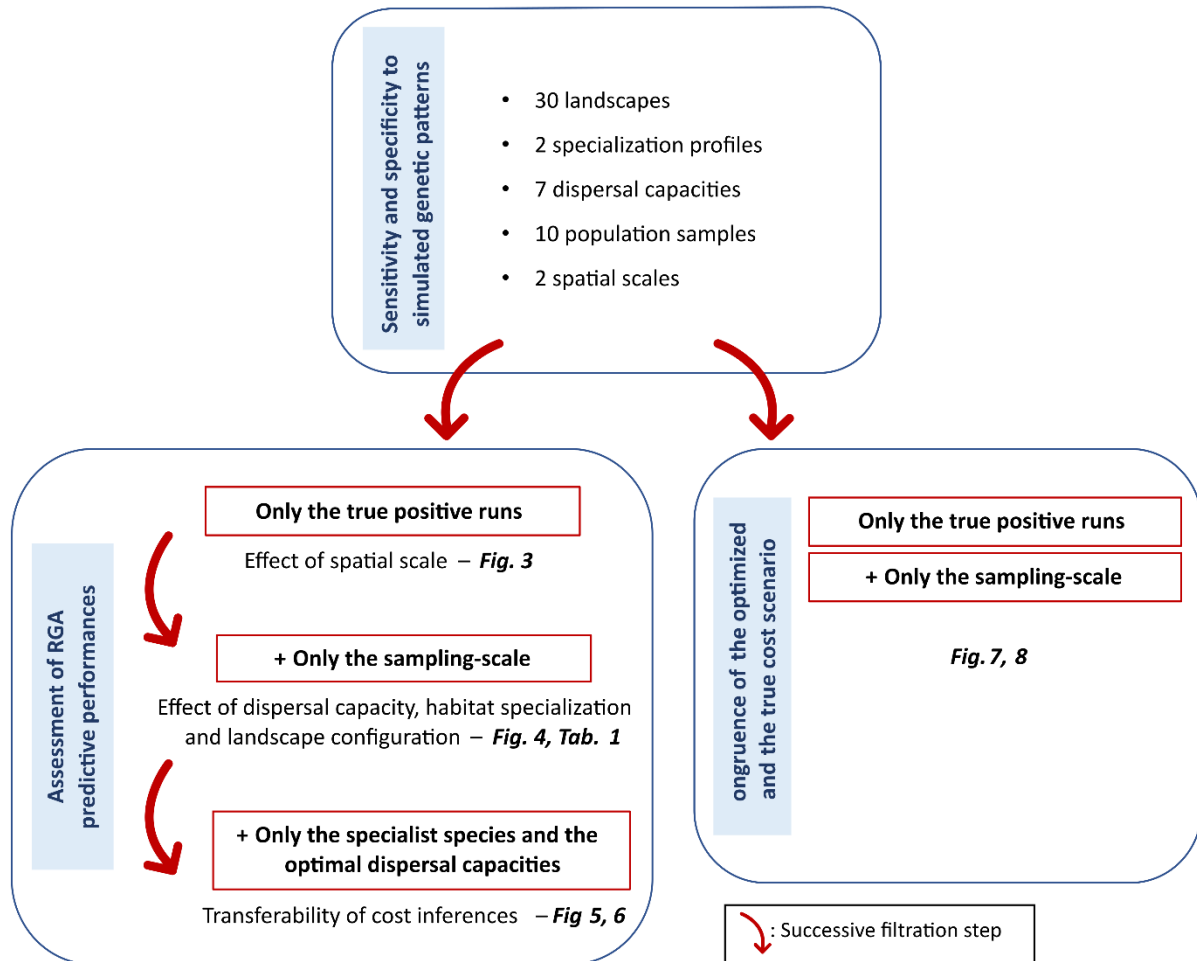
To test whether considering the spatial scale of dispersal would enhance RGA performance, we built two types of pairwise matrices for genetic and cost distances, differing in the spatial scale of the connections considered. The first “sampling-scale” dataset considered the sampling spatial scale and thus included all population pairs. Consequently, for a species with low dispersal capacity, some pairwise genetic differentiation values may mostly result from genetic drift rather than dispersal. For the second “effective-dispersal-scale” dataset, we considered a variety of spatial scales. For each pairwise matrix (statistical individual in our analysis), we excluded drift-driven pairwise genetic differentiation values and kept only those

corresponding to reachable spatial distances, more likely to be predominantly driven by dispersal. To prune these matrices, we relied on a graph-theoretical method aiming to select links between populations directly exchanging migrants, based on the conditional independence principle (Dyer and Nason 2004)(See Annexe 2-S3.1.4 for more details).

#### 3.1.4.3.5. RGA workflow and new implementations

The RGA algorithm optimizes resistance surfaces from genetic distances using a genetic algorithm, which efficiently explores the parameter space until maximizing the statistical fit between pairwise cost distances and pairwise genetic distances. Each of the 60 resistance surfaces (30 landscapes x 2 levels of specialization) was optimized by RGA based on its 140 related  $F_{ST}$  matrices (7 dispersal capacities x 10 sampling designs x 2 spatial scales). Each optimization was replicated 5 times to assess the stability of the algorithm inferences (42,000 runs in total).

For each run, the optimization seeks to maximize a quantitative objective criterion describing the fit of maximum likelihood population effect (MLPE) models, modeling pairwise genetic distances as a function of pairwise cost distances. These MLPE models accounts for the non-independence inherent to pairwise data (See Annexe 2-S3.1.5 for more details on RGA workflow). To prevent model optimization from overfitting the data, we implemented a new objective criterion within the ResistanceGA workflow. We adapted the Leave One Out Cross Validation (LOOCV) method and computed a validation  $R^2$  to quantify the prediction error. When fitting the models, we iteratively removed one of the 40 populations, before predicting the genetic distances involving this population using the calibrated model. The mean of the predicted “out-of-sample” genetic distances was compared with the corresponding observed genetic distances to assess the predictive quality (see Daniel *et al.* (2023) for more details). Besides, we adapted the RGA algorithm to pruned distance matrices for assessing spatial scale influence on inference accuracy (Code available online: <https://gitlab.com/psavary3/rga>).



**FIGURE 3.1.2:** Diagram summarizing the successive stages of data filtering throughout the analyses. The filtration rules are shown in the red boxes. The arrows indicate the successive filtration rules.

### 3.1.4.3.6. Reliability of cost inferences from RGA optimization

The reliability of cost inferences was evaluated from i) the sensitivity and the specificity of the optimization algorithm to the simulated genetic pattern (i.e. Isolation by Resistance (IBR), Isolation by Distance (IBD) or Null pattern assuming that the pairwise genetic distances are related neither to cost distance nor Euclidian distances), ii) the accuracy of optimized cost-distances to predict population genetic structure, iii) the congruence between the optimized cost scenario and the true cost scenario shaping genetic patterns. The following performance analysis workflow involved successive steps of data filtration (see Figure 3.1.2).

*RGA optimization sensitivity to the pattern of spatial genetic structure*

To identify the spatial genetic pattern resulting from RGA optimization, we compared the AIC values deriving from the MLPE models explaining genetic distances with either true or optimized cost distances (“Cost distances model”), Euclidean distances (“Euclidean model”) or a constant (“Null model”). We distinguished three possible patterns:

- i) Isolation-by-Resistance (IBR) pattern when:  

$$AIC_{\text{Euclidean model}} - AIC_{\text{Cost distances model}} > 2$$
- ii) Isolation-by-Distance (IBD) pattern when:  

$$AIC_{\text{Euclidean model}} - AIC_{\text{Cost distances model}} < 2 \text{ and } AIC_{\text{null model}} - AIC_{\text{Euclidean model}} > 2$$
- iii) Null pattern when:  

$$AIC_{\text{Cost distances model}} - AIC_{\text{null model}} > -2 \text{ and } AIC_{\text{Euclidean model}} - AIC_{\text{null model}} > -2$$

Consequently, we classified the optimization runs into four categories: true positives (IBR detected from both true and optimized cost distances), true negatives (IBR never detected), false positives (IBR detected from optimized but not from true cost distances) and false negatives (IBR detected from true but not from optimized cost distances) (See Annexe 2-S3.1.9.a). We then computed the sensitivity (True positives / (True positives + False negatives)) and the specificity (True negatives / (True negatives + False positives)) indices for each spatial scale.

These indices allowed us to assess whether the spatial scale of the pairwise connections considered significantly affected the inference (See Annexe 2-S3.1.9.b). Besides, we also evaluated the influence of landscape structure and the virtual species ecological profile. To this aim, we modeled the sensitivity or specificity indices as a function of the level of habitat specialization, dispersal capacity and 9 FRAGSTAT metrics describing landscape structure. We used the glmmTMB package (Brooks *et al.* 2017) to fit generalized linear mixed models with beta distribution (beta GLMM). The focal landscape was set as a random effect (30 levels, intercept-only) to control for the non-independence of the data. We checked for variable collinearity and performed a stepwise model selection based on AIC criteria, using the R package Buildmer (Voeten 2020).

In the following analysis, we only focused on the True Positive runs (Figure 3.1.2) and all beta GLMMs were run similarly, with focal landscape as an intercept-only random effect.

*Predictive power of optimized cost distances*

First, we tested the effect of restricting pairwise connections to the effective-dispersal-scale on the predictive ability of optimized cost distances, as reflected by the validation  $R^2$ . For each run, as we compared the predictive power of the pairwise cost distances optimized from both the sampling and the effective-dispersal-scale datasets, we focused on the population pairs shared by both datasets for comparison purposes. We ran beta GLMM explaining validation  $R^2$  by a binary categorical variable distinguishing the spatial scales (sampling or effective-dispersal one).

Second, we compared the predictive accuracy (validation  $R^2$ ) of optimized cost distances across dispersal capacities, specialization levels, and landscape structure variations, using a beta GLMM. We checked for variable collinearity and performed a model selection based on AIC criteria. The effective-dispersal-scale dataset included statistical individuals with varying numbers of populations. Therefore, for the sake of reliable comparison, we only considered for this analysis and the following the True Positives runs for the sampling-scale dataset, as they shared the same number of population pairs (Figure 3.1.2).

Then for the two subsequent analyses on predictive power, we focused on specialist species and considered only the dispersal distance that maximized the fit between genetic and cost distances (Figure 3.1.2). This allowed us to reduce the number of parameters to be considered while reflecting the conditions optimizing RGA performance. We assessed the transferability of cost inferences from one landscape to another. To do so, we optimized a cost scenario on a landscape and we evaluated its predictive power for genetic distances between the 40 remaining out-of-sample populations, assessing its interpolation ability. Then, we assessed its extrapolation ability, related to its predictive power for 40 out-of-sample populations from another randomly selected landscape, different from the calibration one. We then fitted beta GLMM explaining validation  $R^2$  with a binary variable representing the origin of the inferred costs (interpolation: calibration and prediction on the same landscape, extrapolation: calibration on landscape  $i$  and prediction on landscape  $j$ ).

Finally, we assessed how much the predictive performance of the optimized cost distances differed from the true cost distances. For each landscape, we fitted a beta GLMM explaining validation  $R^2$  with a categorical variable accounting for the type of scenario used for calibration, either the true or the optimized one.

*Congruence between optimized and true cost scenarios*

We tested whether the cost values of optimized cost scenarios recovered the true ones, and whether they could be interpreted from a biological point of view to identify landscape features acting as barriers or facilitators of gene flow.

These analyses were carried out with both the specialist and the generalist cost scenarios (cf. 3.1.4.3.2.), and on two variants reversing the resistance ranking for urban, agricultural and grassland areas in the specialist scenario: 1<sup>st</sup> variant {Cost<sub>forest</sub>=1; Cost<sub>grassland</sub>= 900; Cost<sub>agricultural</sub>=700; Cost<sub>urban</sub>=1000} and 2<sup>nd</sup> variant {Cost<sub>forest</sub>=1; Cost<sub>grassland</sub>= 1000; Cost<sub>agricultural</sub>=900; Cost<sub>urban</sub>=700}. We ensured that these variants responded in the same way as the specialist and the generalist scenarios to the previous sensitivity and performance analyses. Their consideration at this point allows us to introduce resistance-order variations, relevant to the evaluation of the recovery indices described below.

We assessed the proportion of runs assigning each land cover to each resistance rank, over the 30 landscapes x 10 sampling designs x 4 cost scenarios considered at the optimal dispersal capacities (i.e., maximizing the fit between genetic and cost distances) in the sampling scale dataset. Considering the true scenarios, we would ideally expect 100% of the runs assigning rank 1 to forest, rank 2 to grassland, rank 3 to agricultural and rank 4 to urban areas for the specialist and the generalist scenarios.

Then, we quantified the similarity between the true and optimized scenarios by calculating the Spearman correlations of land cover ranks between these two scenarios (see Annexe 2-S3.1.10 for the correspondence table of the land cover ranking recovery index and the associated land cover type ranking for the four true cost scenarios). We also assessed the contrast between the maximum and the minimum costs for each optimized scenario by computing the log<sub>10</sub> ratio  $\log_{10} \left( \frac{\text{opt. costmax}}{\text{opt. costmin}} \right)$ , and comparing it to that of the true scenarios (equals to  $\log_{10} \left( \frac{1000}{1} \right) = 3$ ) by the difference  $3 - \log_{10} \left( \frac{\text{opt. costmax}}{\text{opt. costmin}} \right)$ . The closer this index was to zero, the better the recovery of the true contrast. The correlation of land cover ranking and the contrast ratio were referred to as recovery indices.

We ran beta GLMM and Linear Mixed Models (LMM) to test for the influence of specialization level, dispersal capacities, and landscape structure (with the 9 FRAGSTAT metrics) on these recovery indices. We set landscapes as a random effect, checked for variable collinearity and performed a stepwise model selection based on AIC criteria. When modeling the contrast, we used the lmer and MuMin R packages for the LMM and model selection, respectively (Bartoń 2013). Finally, we tested whether the recovery indices affected the predictive quality (validation  $R^2$ ) of the optimized cost scenario, running a beta GLMM.

### 3.1.4.4. RESULTS

#### 3.1.4.4.1. Landscape sampling and demo-genetic simulations

124 landscapes out of the 1,000 landscapes initially sampled were selected based on land cover proportion criteria. We then restricted the dataset to 30 landscapes, sampled along the two first PCA axes (accounting for 61% of the variance). Within them, urban areas had a lower representation (5% of the area on average) than forest, agricultural, and grassland areas (See S6 for more information).

After 200 generations of genetic simulation, the observed genetic pattern depended on the interaction between species dispersal capacity, their specialization profile and the spatial scale of the analysis (Annexe 2-S3.1.7). For the generalist species, the genetic structure detected was similar when analyzing both the sampling-scale and the effective-dispersal-scale datasets. For most simulations (83% when considering the sampling-scale dataset and 77 % for the effective-dispersal-scale dataset), we detected an IBR pattern for a dispersal capacity of 1,000 cost units (cu), whereas above this threshold, the majority of simulations (82% for both datasets) resulted in a null pattern. For the specialist species and the sampling-scale dataset, 47 % of the simulations resulted in a pattern of IBD at a small dispersal capacity (1,000 cu). Simulations mainly resulted in an IBR pattern up to an optimal scale of 20,000 cu, before gradually shifting towards a null pattern beyond that scale (Annexe 2-S3.1.7). In certain regions of the parameter space (specialist species with low dispersal capacities in particular), a bimodal distribution of  $F_{ST}$  values resulted in the identification of spurious IBD patterns (see Annexe 2-S3.1.8 for more details). These spurious patterns tended to be replaced by a null pattern in the effective-dispersal-scale dataset (only 7% of IBD patterns at 1000 cu; Annexe 2-S3.1.7.C).

#### 3.1.4.4.2. Optimization sensitivity and specificity to simulated genetic patterns

The sensitivity (i.e., true positives / (true positives + false negatives)) of the RGA algorithm to the simulated genetic pattern (IBR, IBD or null pattern) was 92% for the effective-dispersal-scale dataset and 90% for the sampling-scale dataset (Annexe 2-S3.1.9.a). A chi-squared test (See Annexe 2-S3.1.9.b) showed that it comes from a slightly higher number of false negatives when using the sampling-scale dataset and indicates that pruning the distance matrices improved the identification of the main driver of genetic patterns ( $X^2 = 5.3$ ,  $df = 1$ ,  $p=0.02$ ).

The specificity index (i.e. true negatives / (true negatives + false positives)) was 88% and 92% for the effective-dispersal-scale and the sampling-scale datasets respectively (see Annexe 2-S3.1.9.b). This significantly better performance of the sampling scale dataset was associated with a higher than expected number of false positives with the effective-dispersal-scale dataset ( $X^2 = 16.11$ ,  $df = 1$ ,  $p=5.1e^{-5}$ ).

We did not detect any significant effects of landscape, specialization level and dispersal capacity on the sensitivity and specificity. Then, the following results only concern the analysis of the true positive runs (Figure 3.1.2).

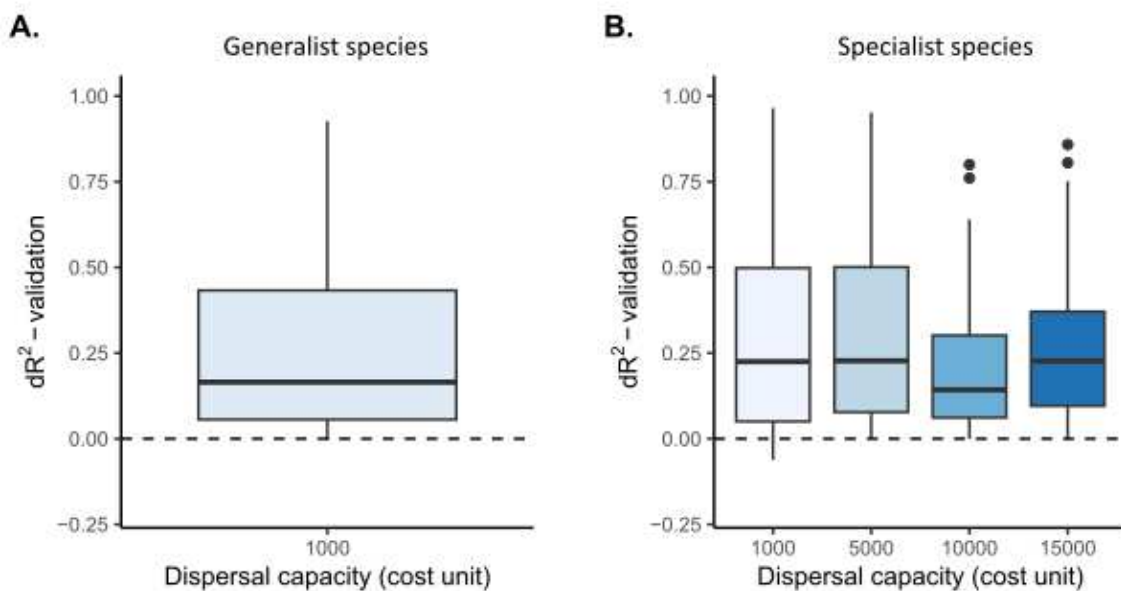
#### 3.1.4.4.3. Assessment of RGA predictive performances

##### *Effect of spatial scale on predictive performance*

In the effective-dispersal-scale dataset, 49% of statistical individuals associated with a generalist scenario and a dispersal capacity of 1000 cu showed drift-driven genetic differentiation and were therefore concerned by pruning. Those with a specialist scenario were pruned in decreasing proportion until 15,000 cu (85% at 1000 cu, 65% at 5000 cu, 50% at 10,000 cu and 28% at 15,000 cu).



The accuracy of  $F_{ST}$  predictions for the links remaining after pruning and shared between the effective-dispersal and the sampling-scale datasets was significantly better when the model was calibrated on the effective-dispersal-scale dataset rather than with the sampling-scale dataset (estimate for sampling-scale  $\pm$  SE=  $-1.5 \pm 0.1$ , pvalue  $< 2e^{-16}$ ). Indeed, for both the generalist and the specialist virtual species and for all values of dispersal capacity considered, the validation  $R^2$  was higher when cost distances were optimized based on the effective-dispersal-scale dataset (Figure 3.1.3A-B). However, for the sake of reliable comparison, we focused on the runs optimized from the sampling-scale dataset in subsequent analyzes (Figure 3.1.2).



**FIGURE 3.1.3:** Comparison of predictive performance of RGA models based either on the sampling-scale or the effective-dispersal-scale datasets, as a function of dispersal capacities and specialization levels. The  $dR^2$ -validation axis represents differences in validation  $R^2$  between models based on cost distance matrices sharing the same pairwise connections for both the effective-dispersal-scale and the sampling-scale datasets for generalist (A) and specialist species (B), respectively. The dashed line represents equal predictive power of RGA models between the sampling-scale and the effective-dispersal-scale datasets. A positive  $dR^2$  values indicates a higher performance of RGA models based on the effective-dispersal-scale dataset. Each box describes the delta  $R^2$  for the 30 landscapes and their 10 related samples populations. The effective-dispersal-scale dataset differs from the sampling-scale dataset only for the distance matrices corresponding to simulations of virtual species with dispersal capacities of 1000 cost units for generalist species and below 20,000 cost units for specialist species. The comparison between the two datasets is therefore only made for these dispersal capacities.

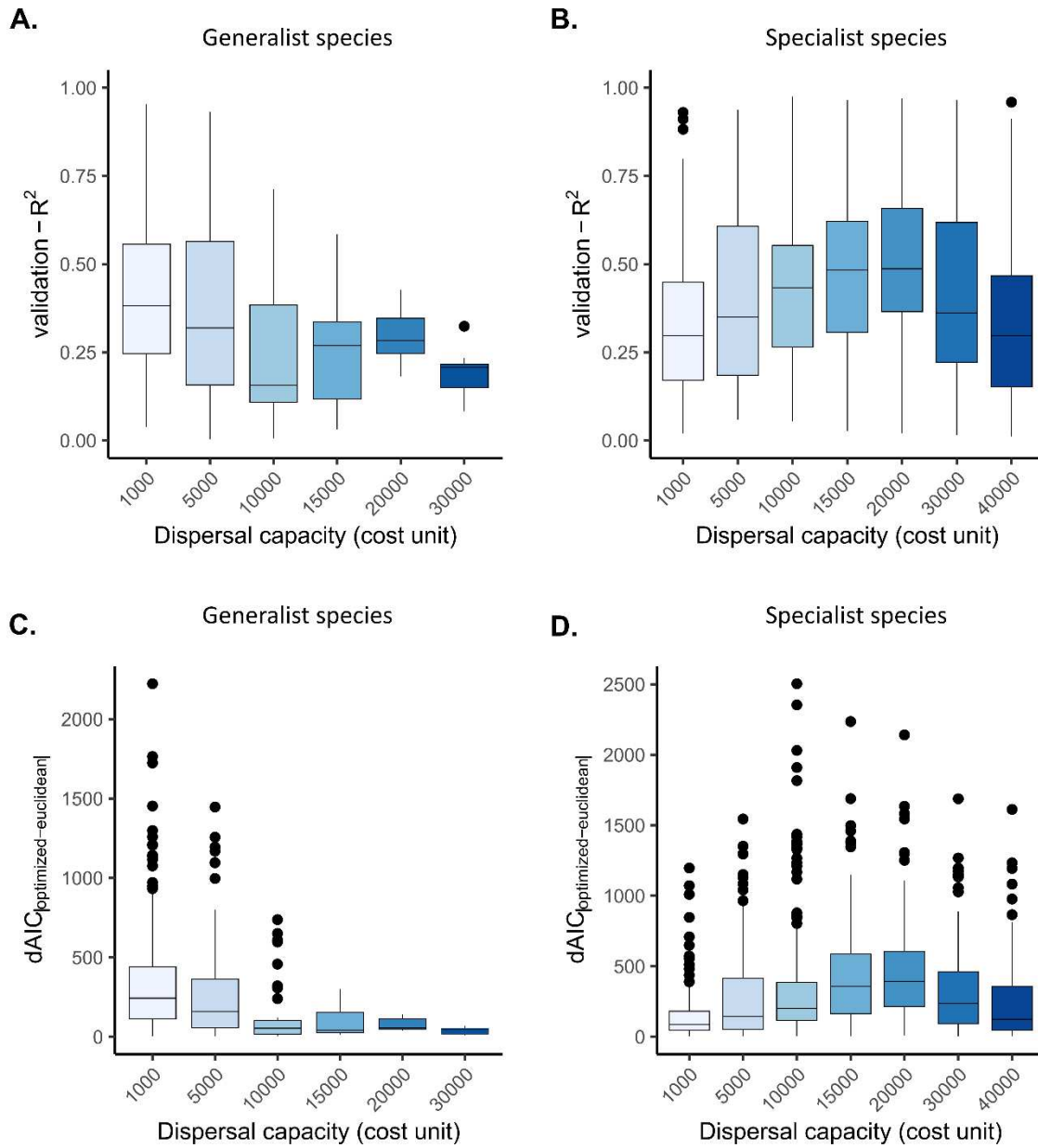
*Effect of dispersal capacity, habitat specialization and landscape configuration on predictive performance*

The predictive performance of the IBR models (assessed by the validation  $R^2$ ) depended on dispersal capacities, as evidenced by a quadratic effect of dispersal distances in cost units, although mainly dependent on its interaction with specialization levels (Table 3.1.1). Genetic differentiation was slightly better predicted for specialist species than for generalist, with optimum median values of validation  $R^2$  reaching 0.48 at 15,000 and 20,000 cu for the specialist species, and 0.38 at 1,000 cu for the generalist species (Table 3.1.1, Figure 3.1.4 – A, B). For both species, forest aggregation improved the predictive performance (estimate= 4.7,  $p = 1.4e^{-4}$ ), whereas grassland aggregation decreased the predictive performance (estimate= -2.8,  $p = 0.01$ ) (Table 3.1.1).

Overall, the strength of the IBR pattern, assessed by AIC differences between Euclidean distance versus cost distance models, reached an optimum at 20,000 cu for the specialist species, and at 1,000 cu for the generalist species (Figure 3.1.4 - C,D). This AIC difference was affected by dispersal capacities in the same way as the validation  $R^2$  (quadratic relationship, Figure 3.1.4 – B, C). Hence, both the predictive power and the IBR strength reached an optimum at greater dispersal distances for the specialist than for the generalist species. In the following sections, we thus focus on the results obtained at the optimal dispersal capacity maximizing the intensity of IBR signal, for the specialist virtual species (Figure 3.1.2).

Validation $R^2$ Random effect : 1   landscape	
Fixed Effect	Estimate (SE)
Generalist scenario	-0.94 (0.09) ***
Dispersal capacity	6.4 (1.4) ***
Dispersal capacity <sup>2</sup>	-8.4 (1.0) ***
Forest aggregation	4.7 (1.2) ***
Grassland aggregation	-2.8 (1.1) *
Generalist sc. * Disp.	-39.4 (3.4) ***
Generalist sc. * Disp. <sup>2</sup>	<i>ns</i>

**TABLE 3.1.1:** Results of statistical model explaining the predictive performance of optimized IBR models from the sampling-scale dataset (assessed by the validation  $R^2$  index) as a function of the level of specialization, the dispersal capacity and two landscape structure metrics. Significance : \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ ; *ns* : non-significant ; Disp. : Dispersal capacity

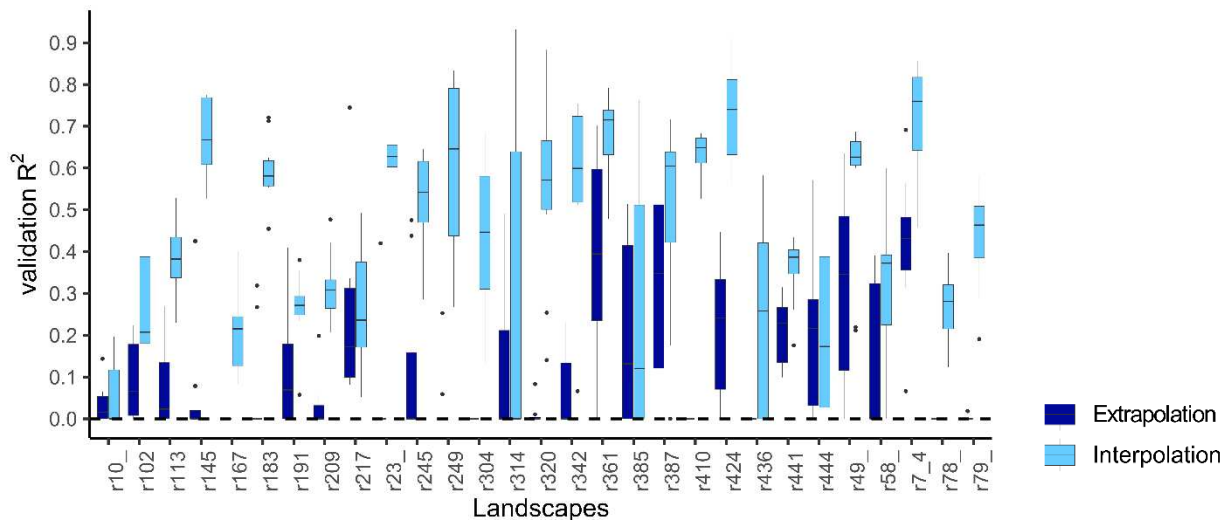


**FIGURE 3.1.4:** Performance criteria as a function of dispersal capacities and specialization levels. Only sampling-scale datasets coming from true positive runs have been considered here. Each box corresponds to a dispersal capacity level and includes the model performance criteria for 30 landscapes and their 10 corresponding population samples. A – B). Predictive accuracy as assessed by the validation  $R^2$  of the IBR models explaining the genetic distances by the optimized cost-distances, for the generalist (A) and specialist species (B). The higher the validation  $R^2$ , the better the predictive power of IBR models. C – D). Strength of the IBR patterns as assessed by the delta AIC between the Euclidean distance models and the IBR models for the generalist (C) and specialist species (D). A positive value indicates that the IBR model performs better than the Euclidean model. The higher the  $dAIC$ , the stronger the IBR signal. As the data were filtered to select only the true positive runs, the IBR models always have more support than the Euclidean models (i.e., only positive  $dAIC$  are shown here).

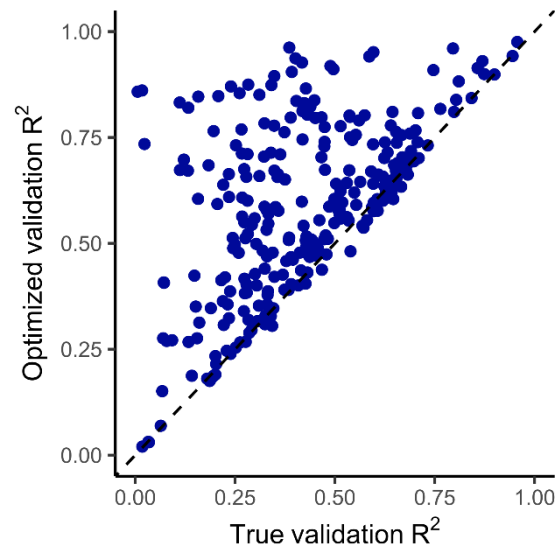
### Transferability of cost inferences

Transferring cost inference to population pairs from an independent landscape resulted in poorer  $F_{ST}$  predictions, as compared with interpolated predictions among populations from the landscape considered for cost inference (extrapolation estimate  $\pm$  SE =  $-2.27 \pm 0.12$ ,  $p = 2e^{-16}$ ). Overall, interpolations resulted in a mean validation  $R^2$  of 0.44, whereas the predictive capacity of extrapolations was characterized by a lower mean validation  $R^2$  of 0.12 (Figure 3.1.5). In addition, large variations in predictive performances both within and among landscapes (median validation  $R^2$  ranging from 0 to 0.7 across landscapes) suggested that certain landscape structures improved RGA performances (Figure 3.1.5). Similarly, the variability of  $R^2$  validation across the 10 population samples within the same landscape (i.e., within-box variations in Figure 3.1.5) also depended on the landscape. This variability may be related to differences in population sampling design, as all other variables remained equal. This suggests that RGA performance might be more sensitive to sampling design in certain landscapes.

Finally, it is worth mentioning that the predictive capacity of optimized cost scenarios was much better than that of true cost scenarios (true cost scenario estimate  $\pm$  SE =  $-0.78 \pm 0.06$ ,  $p < 2e^{-16}$ , Figure 3.1. 6), which indicates that the optimization consistently resulted in overfitting.



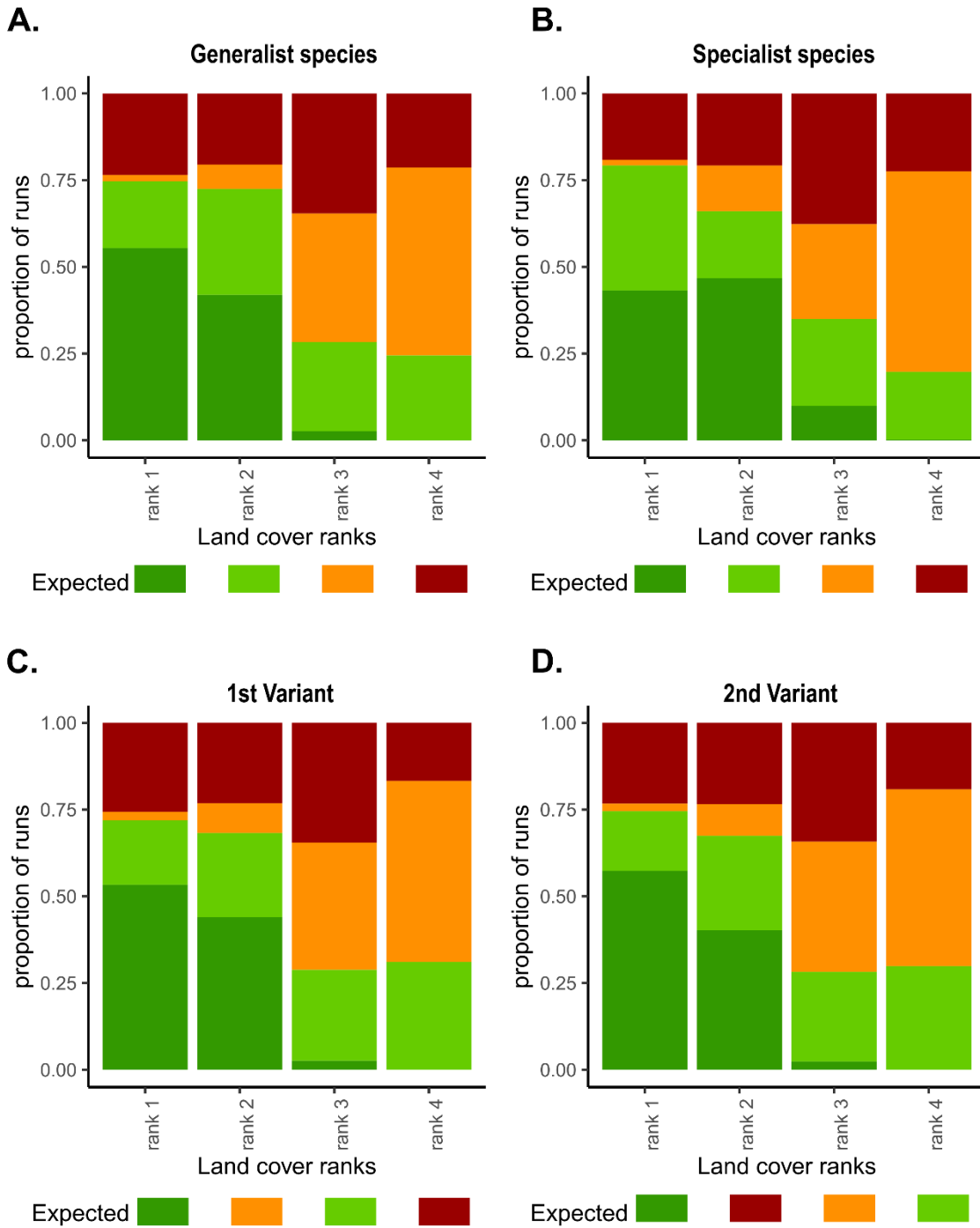
**FIGURE 3.1. 5:** Evaluation of the interpolation and the extrapolation capacities of the optimized IBR models for a specialist species. Only the true positive runs and dispersal capacity maximizing the intensity of IBR signal at sampling-scale are displayed here. The Figure shows the validation  $R^2$  of IBR models as a function of the model calibration landscapes. Each pair of light and dark blue boxes represents runs calibrated on the same landscape (displayed in the x axis), for their 10 related samples of populations. Light blue bars correspond to interpolation results, showing validation  $R^2$  for data located in the calibration area, whereas dark blue bars correspond to extrapolation results, showing validation  $R^2$  for data located on another landscape, away from the calibration area. Only 25 of the 30 landscapes were represented. Indeed, 5 landscapes had less than 6 population samples at the selected dispersal distance that resulted in true positive runs. They were discarded to ensure enough statistical individuals per box (here  $6 \leq n \leq 10$ ).



**FIGURE 3.1.6:** Comparison of predictive performances (assessed by the validation  $R^2$  of IBR models) between the true and the optimized cost scenarios for a specialist species. Only the true positive runs with a dispersal capacity maximizing the intensity of IBR signal at the sampling-scale are shown. Each point corresponds to a combination of landscape x population sample. The points above the  $y=x$  line indicate a better predictive power of the models using optimized cost distances as compared with the models using the true cost distances used for simulating the data (i.e., the true scenarios).

#### 3.1.4.4.4. The congruence between optimized and true cost scenarios

The ranking of optimized costs for land cover types remained almost identical for the 4 ecological profiles. Yet, this ranking differed from that expected under the true cost scenarios for 90% of the optimized runs overall (Figure 3.1.7). Forest areas were correctly assigned the lowest resistance (i.e., rank 1) in about 50% of the cases, and the second lowest otherwise. Urban and grassland areas were assigned to each of the four ranks in equal proportions, while agricultural areas were mostly associated with the highest resistance rank (rank 4 in 53% of the cases). The median contrast between the lowest and highest optimized cost values was consistent with the true scenarios (i.e.,  $\log_{10}(1000) = 3$ ) (Figure 3.1.8A). Moreover, we found a correlation of 0.51 between the land cover recovery index (i.e., the Spearman correlation between true and optimized scenarios) and the contrast recovery index (i.e.,  $3 - \log_{10}(\text{optimized contrast})$ ). This suggested that the closer the optimised classification of land cover types is to the true ranking, the closer the contrast in optimised values is to the true contrast, as the contrast was mostly overestimated.

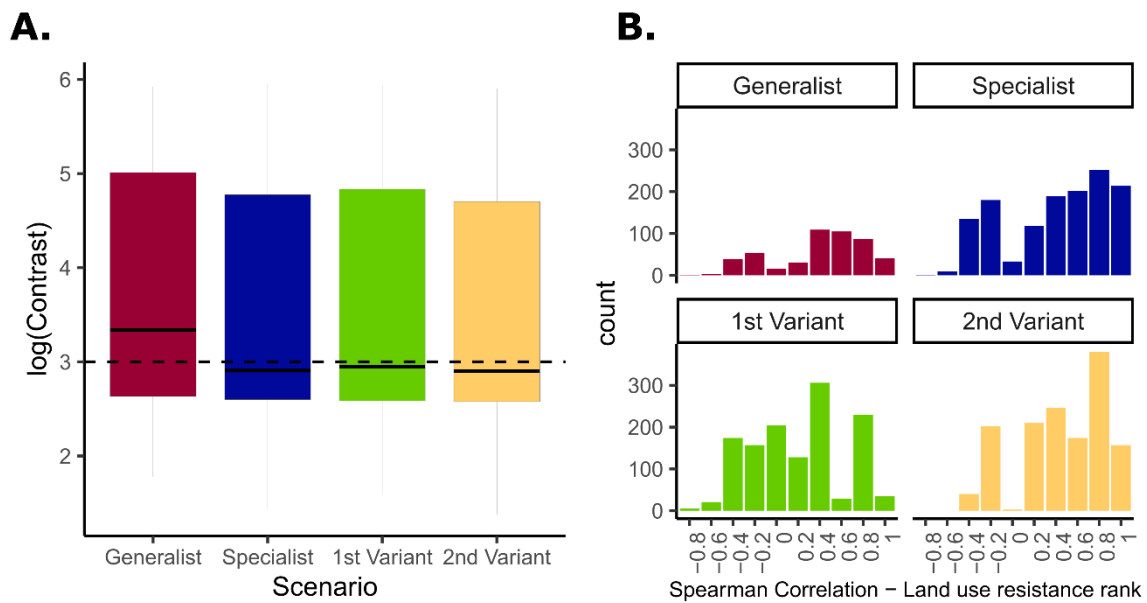


**E.**

Land cover	Specialist species	Generalist species	1 <sup>st</sup> variant	2 <sup>nd</sup> variant
Forest	1	1	1	1
Grassland	700	50	900	1000
Agricultural	900	200	700	900
Urban	1000	1000	1000	700

■ Urban areas  
■ Agricultural areas  
■ Grassland areas  
■ Forest areas

**FIGURE 3.1.7:** Assignment of each land cover type to the 4 ranks of relative resistance in the optimized cost scenarios for A) a generalist species, B) a specialist species, C & D) virtual species with ecological profiles derived from variations in the specialist profile (the order of land use resistance has been reversed as compared with the specialist scenario). Each bar represents the proportion of each land cover type assigned to every resistance rank, across the 30 landscapes and their 10 related population samples from the sampling-scale dataset. The “Expected” line below each bar plot corresponds to the land cover type expected under the true cost scenario, represented by its color. The recovery of the land cover ranking is accurate if the expected colors match the dominant colors in the corresponding bar plot. E) Table of the cost values assigned to every land cover type in the four true cost scenarios considered. Each column of the Table (i.e., each true cost scenario) has to be compared with the bar plot describing the related optimized cost scenario.



**FIGURE 3.1.8:** A) Characterization of the optimized cost scenarios according to their contrast as a function of the ecological profile, B) Distribution of the ranking consistency with true cost scenarios of optimized runs (i.e., Spearman correlation between the ranking of land covers in the optimized cost scenario and in the true cost scenario) according to the ecological profile

The models explaining the two recovery indices by landscape and ecological variables only evidenced an effect of the specialization profile. The recovery of land cover ranks showed significantly poorer performance with variant 1 ( $p < 0.001$ ), and the contrast was poorly assessed with the generalist profile ( $p < 0.001$ ). Finally, the ranking of the land cover types and the contrast values did not affect the validation  $R^2$  of optimized cost scenarios.

### 3.1.4.5. Discussion

We demonstrated that RGA cost inferences are very sensitive to the effect of landscape on gene flow. Moreover, our results underlined the strong predictive performance of these inferences. This predictive power depends on several factors shaping the structure of the dispersal network, such as species dispersal capacity and specialization level, or sampling design. Moreover, we found that pruning part of the pairwise genetic distances enhances RGA predictive power, allowing a better fit between the spatial scales of dispersal and that of the analyzes. Finally, we draw attention to the fact that optimized costs assigned to land cover types do not always properly reflect their permeability to gene flow.

#### 3.1.4.5.1. The spatial scale of landscape influence on genetic structure depends on the topology of the effective dispersal network

Our results demonstrated that detecting an IBR pattern depended on the interaction between species specialization level and dispersal capacity. These two parameters affect the scale at which the genetic pattern emerges, and thereby the scale to consider for properly detecting it. For instance, in a species experiencing moderate movement costs across the landscape (generalist species in our study), IBR patterns can only emerge and be detected if this species covers short distances overall. In other words, the dispersal limitation responsible for IBR patterns in heterogeneous landscapes (Orsini *et al.* 2013) is caused by the interplay of the movement costs experienced and the dispersal distances. This dual cause of effective dispersal limitation also explains why, in our results, IBR patterns were detected for larger dispersal distances in the specialist species profile. This result is consistent with other studies that have demonstrated the impact of species specialization on effective dispersal and its consequences for population genetic differentiation through dispersal limitation (Khimoun *et al.* 2016; Harris and Reed 2002). In the context of resistance surface optimization, an essential pre-requisite for proper inferences is the existence of a spatially structured genetic pattern, as subsequent analyzes assume its presence. The corresponding spatial scale of landscape influence on genetic structure (here referred to as the effective-dispersal scale) depends on the interplay of specialization level and dispersal ability. This interplay defines the topology of the effective dispersal network among populations (Savary *et al.* 2021a).



#### 3.1.4.5.2. Pruning can enhance RGA predictive performance under strong dispersal limitation

We demonstrated that RGA predictive performances were maximized when considering the subset of population pairwise connections matching the effective-dispersal-scale. In other words, better results are surprisingly obtained from a reduced dataset. It is expected from theory that low gene flow will lead to isolated populations whose genetic differentiation is mainly driven by stochastic genetic drift (Hutchison & Templeton, 1999), and consequently difficult to predict from their location on the landscape. Conversely, if gene flow is too strong, landscape constraints to dispersal may no longer be a limiting factor, thus making their effects difficult to detect. Accordingly, the effective dispersal scale to consider for selecting pairwise population connections allowing for the detection and assessment of landscape effect on gene flow depends on the degree to which gene flow is restricted (Savary *et al.*, 2021b).

Considering the above, we introduced a new approach to infer cost values, involving the conditional pruning method (Dyer and Nason 2004) that restricts the pairwise datasets to match the effective-dispersal spatial scale. We showed that pruning the distance matrices improved RGA performances as it ensured that genetic differentiation was only modeled between populations supposed to be linked by significant disperser exchanges and corresponding gene flow (Dyer and Nason 2004; Keller, Holderegger, and Van Strien 2013; Murphy *et al.* 2010; Van Strien, Holderegger, and Van Heck 2015). Daniel *et al.* (2023) had already shown a similar positive effect of reducing the genetic dataset to capture better the spatial scale of effective dispersal using empirical data. These results demonstrate the usefulness of considering the topology of dispersal networks in the context of resistance surface optimization, thereby extending similar conclusions regarding common landscape genetic relationships assessment methods (Savary *et al.* (2021b) and Van Strien (2017)).

### 3.1.4.5.3. Sampling design and landscape structure affect optimization performance

We found that in a given landscape, the predictive performances of the optimized cost values were likely to vary greatly across the 10 samples of populations, all other things being equal. This suggests a significant influence of the population sampling design on the inference outcome, potentially related with the capacity of this sampling to properly capture the dispersal network. Indeed, McRae (2006) and Van Strien (2017) already demonstrated that the spatial distribution of the whole set of populations determines the topology of the dispersal network shaping genetic patterns. In this context, the absence of some populations in the sampled dataset could prevent landscape genetic analysis from reliably capturing the spatial drivers of genetic structure (Naujokaitis-Lewis *et al.* 2013; Van. Strien 2017). In other words, when the sampling design fails to adequately approach the real topology of the whole set of natural populations, landscape effects on gene flow cannot be estimated accurately because part of the spatial signal is missed. These results were further supported by our finding of a significant relationship between the validation  $R^2$  and the aggregation of forest areas, a factor that influences the spatial distribution of the sampled populations. Forest aggregation was highly positively correlated with the aggregation of agricultural areas and the whole landscape. Interestingly, the results showed that in coarse-grained and aggregated landscapes, RGA performances may be enhanced. One explanation could be that the topology of the dispersal network may be easier to capture in an aggregated landscape, reducing the risk of missing a high-centrality population (i.e., a population with high topological importance in the dispersal network) during sampling. M J van Strien, Holderegger, and Van Heck (2015) showed that population topology was tightly linked to landscape structure and habitat distribution and our results suggest that some landscape configurations and compositions are more resilient to partial sampling. It should be noted that our land cover map is at a low spatial and thematic resolution. For future studies, it would be interesting to test whether the sensitivity of the predictive power of RGA to the aggregation level of forest areas could be stronger when using higher resolution maps.

#### 3.1.4.5.4. Optimization is not causation: good predictive abilities at the expense of accurate causal interpretation

Many studies have used the results of RGA cost inferences to rank landscape features according to their resistance to gene flow (Mapelli *et al.* 2020; Mulvaney, Matthee, and Cherry 2021; Reyne *et al.* 2023; Martin *et al.* 2023; Antunes *et al.* 2022; Khimoun *et al.* 2017). These interpretations often served as the basis for a supposedly mechanistic understanding of landscape effect on dispersal and gene flow. However, our results evidenced a frequent mismatch between the optimized and true cost rankings across land cover types. Only 10% of the optimized runs showed a land cover cost ranking in line with the simulated biological reality. Surprisingly, these overall trends in land cover cost assignment remained fairly identical regardless of the type of true cost scenario shaping genetic structure. Moreover, this cost ranking often seemed to match landscape composition rather than its actual permeability to gene flow. For instance, we found that the land cover types covering the lowest landscape proportions, i.e., grassland and urban areas, were equally likely to be assigned a cost value at any one of the 4 resistance ranks. They probably served as adjustment variables in the optimization process, mainly driven by the final goodness-of-fit.

We also found that the more the optimized ranking differed from the true ranking, the more the contrast of the optimized cost scenario was overestimated relative to the true contrast.

Therefore, our results call for great caution when interpreting the resistance of landscape features and suggest a strong dependence of cost allocation on landscape structure. This is consistent with former results obtained by Peterman *et al.* (2019) and Beninde, Wittische, and Frantz (2023). In these two studies, without discussing the issues in detail, the authors pointed out possible difficulties in deriving reliable cost values with the RGA framework. The overfitting effect inherent to the optimization process (Yates *et al.* 2018; Winiarski, Peterman, and McGarigal 2020; Peterman *et al.* 2019) was mentioned as a possible consequence. Given our result on the better predictive performance for optimized cost scenarios than for true cost scenarios, we could reasonably assume that overfit lets a strong imprint of landscape structure on the calibration of cost values.

Based on our results, we can argue that the dependence of the inference on the topology of the dispersal network exacerbates the adverse effect of overfitting on inference accuracy. Indeed, the good predictive accuracy of RGA models obtained in the calibration area, even for population pairs excluded from the calibration data set, dropped when the models were

extrapolated to population pairs located in another landscape. By making cost inference highly dependent on landscape composition and population topology, a cost scenario optimization subject to overfitting reflects the structure of a landscape rather than its permeability to gene flow. It follows that its extrapolation in another landscape is unlikely to provide information about neither the genetic structure of populations nor the landscape effect on dispersal in that other landscape.

Despite the above-mentioned limitations of the RGA framework, our results emphasize the good predictive performance of its cost inferences, thereby reflecting the common duality between explanatory and predictive modeling (Shmueli 2010; Yates *et al.* 2018). Based on a data-driven process and an optimization algorithm, RGA provides accurate predictions for genetic differentiation, which could then lead to a better understanding of the effect of population topology or landscape configuration on genetic structure. However, establishing a causal link between the cost distances driving dispersal and gene flow from the modeling results and interpreting it from an ecological point of view is potentially dubious. This echoes the fact that predictive models, such as RGA, rarely provide insight into the underlying causal mechanism, and “*an optimal model for prediction purposes may be different from the one obtained by estimating the true model*” (Shmueli 2010).

Some very promising landscape genetics approaches, e.g., using deep and machine learning, have recently been implemented to predict genetic connectivity. Kittlein *et al.* (2022) showed that convolutional neural networks could provide highly accurate predictions for small-scale genetic differentiation and diversity, while Pless *et al.* (2021) conducted a least-cost transect analysis to predict gene flow in the context of *Aedes aegypti* vector regulation. As soon as the common limitations of data-driven approaches are acknowledged, predictive models could be of great interest to capture complex patterns and relationships, otherwise difficult to predict using theory-based models. They may allow to identify new potential causal mechanisms, generate new hypotheses and improve already existing explanatory models (Murphy *et al.* 2010; Vanhove and Launey 2023; Lucas 2020; Shmueli 2010).

#### 3.1.4.5.5. The use of RGA in landscape genetics: conditions and prospects

We outline below a set of guidelines regarding the use of RGA and list them in the order in which they should be considered when designing a landscape genetic study.

First, because the level of genetic structuring increases the qualitative relevance of cost inferences, we call for preliminary assessments of spatial genetic structuring before the use of optimization approaches (e.g., through genetic clustering analyses or the study of IBD patterns). When the genetic structuring is weak, great care should be taken when interpreting the optimization results.

Second, one needs to capture the spatial scale of the observed genetic signal properly (Savary *et al.*, 2021b). To this end, we recommend performing inferences at multiple spatial scales, either using Moran's Eigenvector Maps (Dray *et al.*, 2006; Galpern *et al.*, 2014) or through iterative analyses with multiple pruning thresholds (van Strien *et al.* 2015), and selecting the one that leads to the best predictive performance of the optimised IBR model.

Third, population sampling design plays a key role in the resistance optimization workflow. Our results suggest that the reliability of RGA inferences may be improved by favoring exhaustive sampling designs to correctly model topological effects and limit their confounding effect (Maarten J. Strien 2017). If this is too costly, preliminary sensitivity analyzes, e.g., based on genetic simulations, could determine the ideal set of populations to sample in the focal landscape. (Naujokaitis-Lewis *et al.* 2013). After empirical data have been sampled and used for optimization, the stability of the optimized resistance surface when refitting it with a subset of the populations can also be assessed with a new bootstrap procedure implemented in RGA (Peterman *et al.* 2019).

Fourth, attention must be paid to the structure of the studied landscape and the thematic resolution, particularly if the aim is to interpret the cost values optimized by RGA. When some land cover types are poorly represented, or for some land cover configurations, RGA inferences may be more sensitive to the landscape itself than they are to the ecological process of interest. A first indicator for this concern is the stability of the land use ranking: highly unstable cost values point towards unreliable inferences. Here too, simulating gene flow in the focal landscape, and assessing the algorithm's ability to capture the correct cost scenario might prevent drawing spurious conclusions.

Fifth, it is strongly recommended to control the overfitting of the optimized model and ensure that it does not fit the noise in the data rather than the targeted signal (Winiarski, Peterman, and McGarigal 2020; Peterman *et al.* 2019; Shmueli 2010; Lucas 2020). To do this, out-of-sample performance could be assessed by withholding some of the data and testing the predictive quality of the model on this independent dataset, as performed in the present study (for other examples using this cross-validation method, see Maarten J. van Strien *et al.* 2014; Daniel *et al.* 2023). This also provides an assessment of the transferability of the inferences made.

Finally, although we have not examined it here, Beninde, Wittische, and Frantz (2023) found that the choice of genetic distance metric had a strong impact on landscape genetic inferences. They call for the adoption of eigenvector-based quantifications of pairwise genetic distances, which showed superior model support and fit, and very stable inferences regarding the effect of landscape features on gene flow.

Although these guidelines might improve the reliability of optimization results, they cannot lead to any improvements in the presence of significant noise in the data or when a key spatial process or covariate is missed (Lucas 2020). Note also that the sensitivity analyses recommended above can help assessing the importance of population topology and habitat distribution for spatial genetic patterns. If these variables are of great importance, including them in explanatory models is likely to improve their performance (Keller, Holderegger, and Van Strien 2013; Maarten J. van Strien *et al.* 2014; Maarten J. Strien 2017). Moreover, comparing levels of predictive performance across landscapes can provide information on the landscape types or features that limit dispersal.

In conclusion, the strength of the RGA workflow is its excellent ability to predict genetic distances, although its lack of transferability limits the prospective use of its inference to predict the impact of landscape change on gene flow. This characteristic can still be very useful for operational studies in conservation and population genetics (Maarten J. van Strien *et al.* 2014). It may even be interesting to use the genetic distances predicted by RGA to inform operational models. For example, we could imagine using RGA's cost inference to weight the links of habitat network models and derive relevant functional connectivity metrics, design dispersal corridors, or identify restoration areas (Foltête, Girardet, and Clauzel 2014).

#### 3.1.4.6. Acknowledgements

We thank Julien Pergaud for his help with the computer programming aspect. Simulations and optimization analyses were carried out on the computing cluster of the University of Bourgogne, with Henri Gaulard as our cluster correspondent.

## 3.2. Validation empirique des modèles de graphes paysagers reflétant l'effet de la connectivité des habitats sur le flux génétique

### 3.2.1. Introduction

La connectivité des habitats peut être modélisée par des graphes paysagers. Ces graphes sont constitués de taches d'habitat (leurs 'nœuds') reliées par des chemins de dispersion potentiels constituant leurs 'liens'. Ils permettent de modéliser la contribution de chaque tache d'habitat à la connectivité du réseau à partir de données spatiales peu coûteuses à acquérir et de peu de paramètres. Ils présentent donc un fort intérêt opérationnel pour la mise en œuvre des politiques de conservation de la biodiversité. Si cette approche semble prometteuse, sa validité écologique reste toutefois à démontrer. En effet, ces graphes paysagers sont souvent construits à partir de dires d'experts ou de modèles de distribution d'espèces (SDM) et manquent donc d'une validation empirique à partir de données reflétant plus étroitement la connectivité fonctionnelle. Dans cette étude, nous avons donc évalué si les métriques de connectivité issues des graphes paysagers sont une représentation fiable des effets de la connectivité fonctionnelle sur les patrons de biodiversité, et plus précisément sur les flux génétiques qui sont un des principaux processus éco-évolutifs.

### 3.2.2. Matériels et méthodes

Dans cet objectif, nous avons modélisé la connectivité de l'habitat d'une espèce d'oiseau forestière, la Paruline caféïette (*Setophaga plumbea*), en Guadeloupe. Trois graphes paysagers ont été construits selon : i) l'opinion d'experts, ii) un indice de spécialisation à l'habitat basé sur des données de présences, iii) un modèle de distribution de cette espèce basé sur des données de présences et des données climatiques (SDM). En parallèle, des données génétiques (microsatellites) ont été recueillies sur 712 individus de Paruline caféïette issus de 27 populations. Nous avons calculé des indices de diversité génétique locale et des distances génétiques par paires de populations. Ce jeu de données génétiques, censé refléter notamment la conséquence d'épisodes de dispersion contraints par l'hétérogénéité du paysage, a été



utilisé comme outil de validation empirique des trois graphes paysagers. Nous avons pour cela mis en œuvre deux approches évaluant les relations entre : i) les distances-coûts sur les graphes et les distances génétiques, ii) les métriques de connectivité issues des graphes et les indices de diversité génétique, en utilisant des modèles de distance MLPE<sup>1</sup> et des corrélations de Spearman, respectivement.

### 3.2.3. Résultats principaux

Nous avons montré que, dans l'ensemble, les graphes paysagers reflètent de manière fiable l'influence de la connectivité fonctionnelle sur la structure génétique des populations, avec des  $R^2$  de validation associés aux modèles de distances allant jusqu'à 0,30 et des coefficients de corrélation entre métriques locales allant jusqu'à 0,71. Cependant, la relation entre la validité écologique des graphes et leurs exigences en matière de données ou la complexité de leurs méthodes de construction n'était pas linéaire. En effet, le graphe basé sur la méthode de construction la plus complexe (SDM) et sur des données empiriques de présence-absence conduisait parfois à des métriques de connectivité plus faiblement liées aux indices génétiques que les autres graphes. En comparaison, la mobilisation de opinions d'experts pour construire les modèles s'avérait donc très satisfaisante dans ce contexte. Nous avons donc confirmé la pertinence des graphes paysagers pour modéliser les connectivités paysagères, mais nous recommandons de prêter attention au rapport coût-bénéfice de leurs méthodes de construction. Ces résultats devront être confortés par des études similaires, explorant également la validité de ces modèles *a posteriori*, c'est-à-dire à partir de données biologiques indépendantes de leur construction, et cela pour une plus grande diversité d'espèces et de paysages.

---

<sup>1</sup> Maximum-Likelihood Population Effect models, voir introduction générale.

### 3.2.4. Article

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

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

Cet article a été publié au sein de la revue *Conservation Biology* : DOI: 10.1111/cobi.14047

# 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

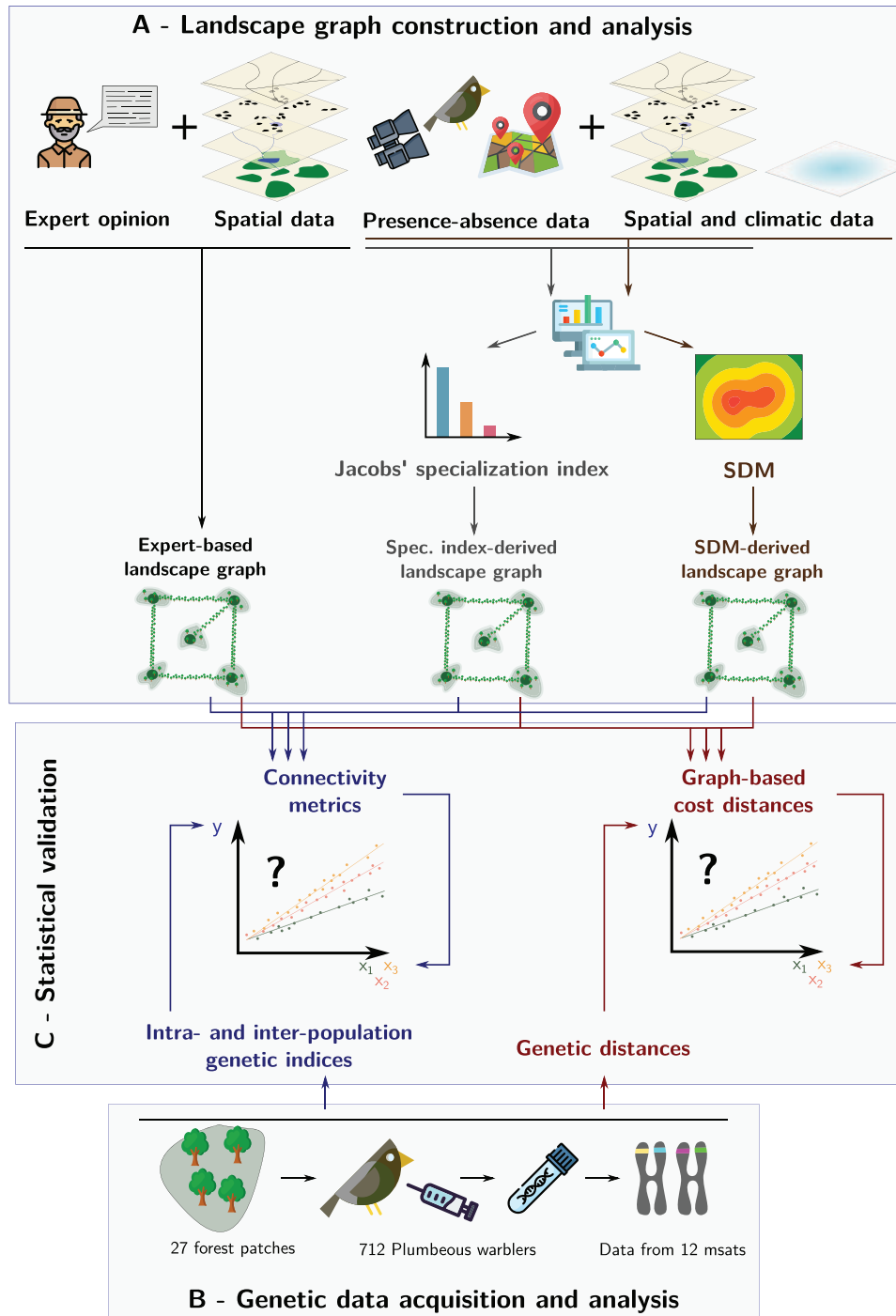
To build landscape graphs, we had to define 2 main parameters: habitat patches (graph nodes) and dispersal cost distances among pairs of nodes (weight of graph links) calculated from movement resistance surfaces relying on cost scenarios. We considered 3 methods for defining nodes and cost scenarios 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

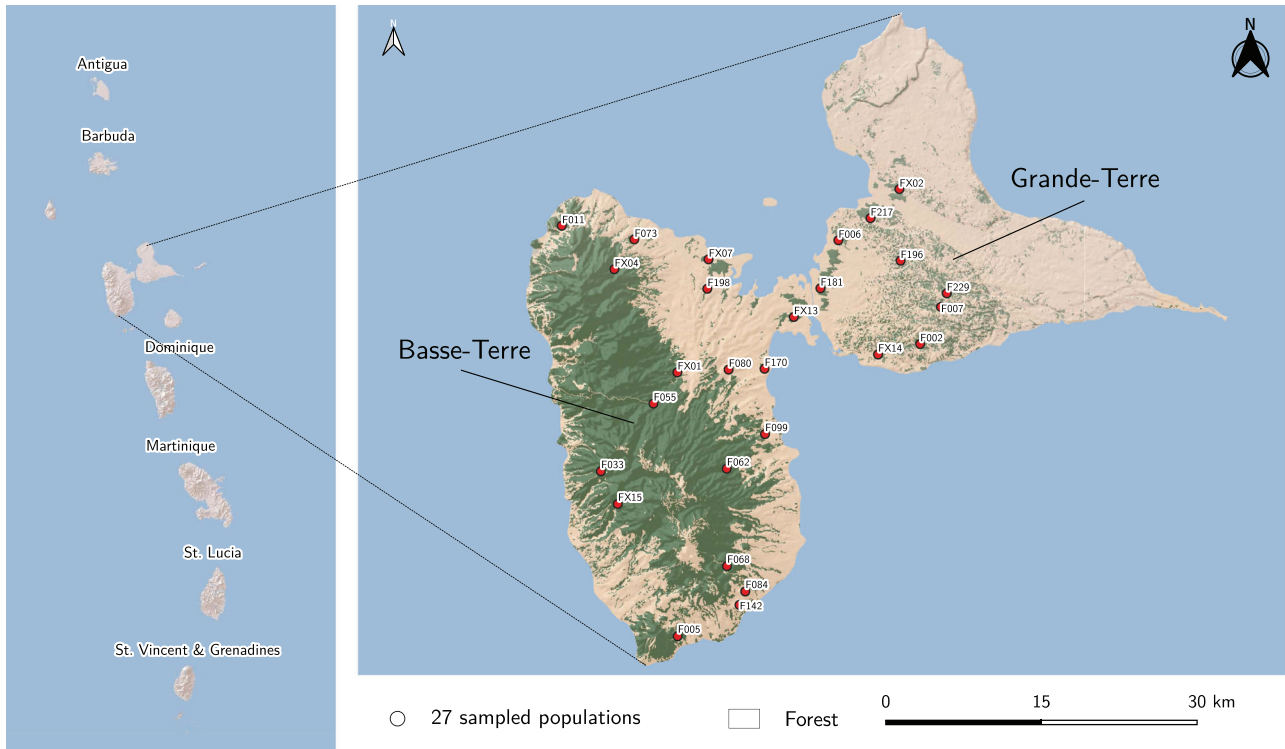


**FIGURE 1** Overall methodology in validating graph-based connectivity models with genetic data sets: (A) landscape graph construction and analysis from spatial data and expert opinion-only and from presence-absence, spatial, and climatic data. Graph-based cost distances between sampled patches and 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:

$$cost_i = e^{\frac{-\ln(cost_{max}) \times \ln(Jacobs_i + 1)}{Jacobs_{habitat} + 1}} \times cost_{max},$$

where  $cost_i$  and  $Jacobs_i$  are, respectively, the dispersal cost value and Jacobs’ index associated with land-cover type  $i$ ;  $cost_{max}$  is the maximum cost value; and  $Jacobs_{habitat}$  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 multiloaded 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–

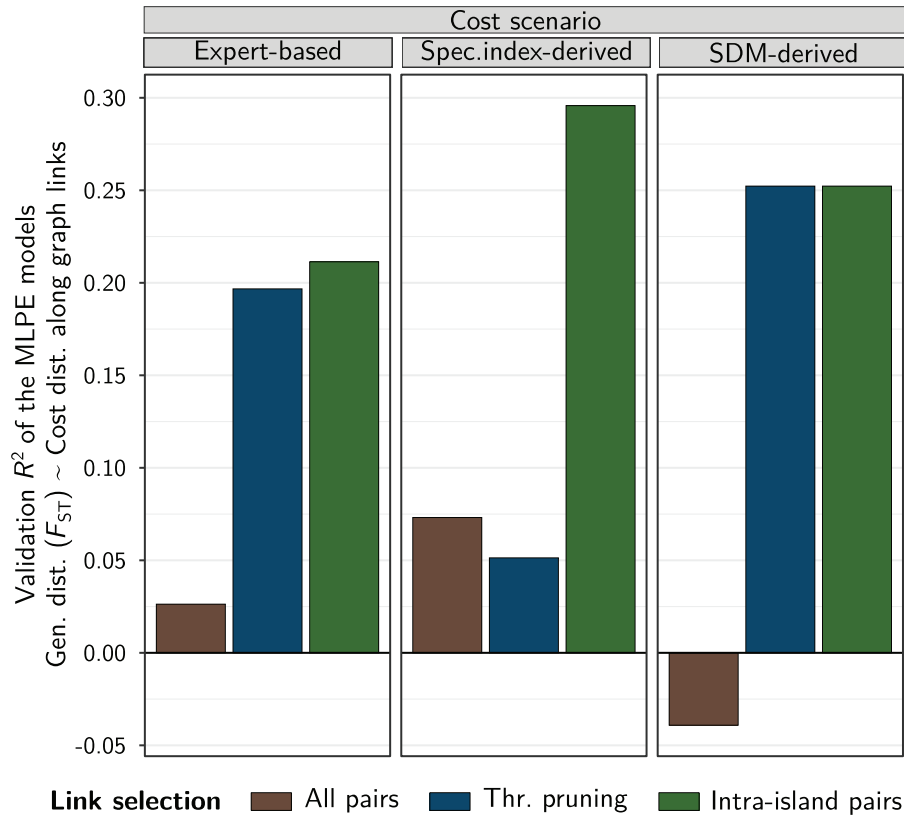
GT) were larger than the cost distances between population pairs from BT or GT (BT–BT or GT–GT). However, because of the high cost values assigned to both agricultural and artificial areas when building the graph from the SDM, the difference between the cost distances computed across the isthmus and 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).

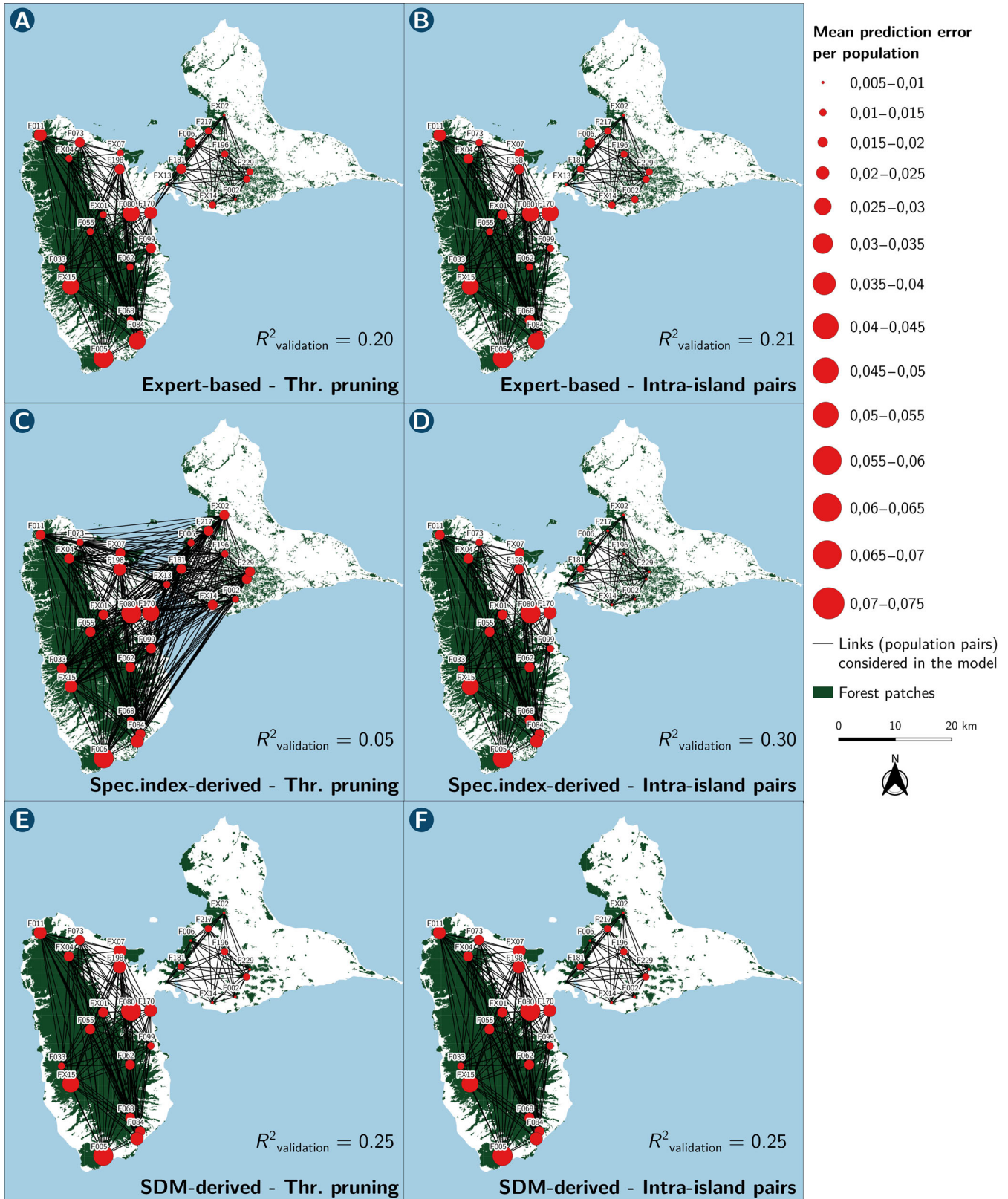
The set of links to consider for maximizing the predictive performance of the model was not the same depending on the cost distances used (Figures 3 & 4). For the specialization index-derived method, the best MLPE model relied on a set of intra-island links (BT–BT and GT–GT links; BT–GT links discarded) (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., Dufloy, 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>
- 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>

### 3.3. Conclusions

Dans ce chapitre, nous avons évalué les conditions d'application et la validité écologique de deux modèles couramment employés en génétique du paysage. Le premier vise à inférer à partir de distances écologiques et génétiques l'influence de l'hétérogénéité du paysage sur la structure génétique (algorithme ResistanceGA). Le second modélise sous forme de graphes paysagers la façon dont la connectivité des habitats est censée façonner cette structure génétique dans des paysages hétérogènes. Nous avons montré que l'algorithme d'optimisation ResistanceGA était un outil très performant pour identifier un patron d'IBR et prédire des distances génétiques entre des populations situées au sein du paysage échantillonné. De plus, nous avons démontré la capacité des graphes paysagers à refléter l'effet de la connectivité sur les flux génétiques, tels que reflétés par des distances génétiques mesurées empiriquement. Dans l'étude consacrée aux graphes paysagers, nous avons comparé l'utilisation de trois types de données et de méthodes de construction des graphes. Il serait très intéressant de reconduire cette analyse en ajoutant un quatrième type de données traduisant plus directement les processus de dispersion, telle que les données génétiques. On pourrait en effet imaginer paramétrer un graphe paysager avec un scénario de coût optimisé à partir de données génétiques, et valider les métriques de connectivité issues de ce graphe avec un second jeu de données génétiques indépendant de l'étape de paramétrage. D'après nos résultats sur les performances de ResistanceGA, le scénario de coût optimisé permettrait de pondérer les liens du graphe par des distances coûts prédisant avec précision les distances génétiques entre les populations occupant les taches d'habitat du graphe. On peut donc faire l'hypothèse qu'une telle utilisation combinée des graphes paysagers et de ResistanceGA pourrait améliorer la pertinence écologique des modèles de connectivité issus des graphes paysager. Le coût de construction d'un tel graphe étant élevé, il serait utile d'évaluer l'éventuel gain apporté par rapport à des méthodes moins onéreuses mais performantes, telles que des opinions d'experts ou un indice de spécialisation.

Plusieurs facteurs encouragent l'utilisation conjointe de ResistanceGA et des graphes paysagers. Tout d'abord, tant l'approche RGA que les graphes paysagers proposent une modélisation à l'échelle des liens entre populations. De plus, les performances de ces deux outils présentent une même dépendance à l'échelle spatiale. Nous avons en effet montré qu'élaguer les matrices de distances pour restreindre l'échelle spatiale à celle à laquelle ont

lieu les flux de dispersion permettait d'améliorer les inférences de ces deux outils. Le nouveau module implémenté dans ResistanceGA pour contrôler l'échelle spatiale repose ainsi sur la matrice d'adjacence d'un graphe génétique. Par ailleurs, la question de la méthode de paramétrage du scénario de coût est centrale pour la construction d'un graphe paysager. Or, RGA permet d'inférer un scénario de coût représentant étroitement la perméabilité du paysage au flux de gènes. De plus, alors que les graphes sont peu coûteux en données (les données spatiales étant les seules indispensables) et fournissent une représentation exhaustive du réseau d'habitat, RGA est une approche très onéreuse en données et en temps de calcul et ne considère qu'un échantillon de populations. Enfin, alors que les inférences de coût de RGA ne peuvent pas être interprétées spatialement en termes de chemin de dispersion, les graphes paysagers permettent une représentation spatiale des ensembles paysagers contribuant activement à la connectivité et des corridors de dispersion. Ainsi, par tous ces aspects, ces deux outils ont un potentiel de complémentarité qu'il serait intéressant d'explorer. Un exemple d'utilisation conjointe de ResistanceGA et des graphes paysager sera ainsi illustré dans le chapitre suivant.





# CHAPITRE 4

## DOMAINE DE VALIDITÉ MULTI- SPÉCIFIQUE D'UN MODÈLE DE CONNECTIVITÉ EN GÉNÉTIQUE DU PAYSAGE



<b>4.</b>	<b>DOMAINE DE VALIDITÉ MULTI-SPÉCIFIQUE D'UN MODÈLE DE CONNECTIVITÉ EN GÉNÉTIQUE DU PAYSAGE.....</b>	<b>109</b>
4.1.	Introduction .....	110
4.2.	Matériels et méthodes .....	111
4.3.	Résultats principaux .....	112
4.4.	Article : Is landscape effect on gene flow shared across species? .....	113
4.5.	Conclusions .....	141

## 4.1. Introduction

Maintenir la connectivité des populations et de leurs habitats est un enjeu majeur de conservation de la biodiversité. Les objectifs généraux de la conservation actuelle visent à préserver les fonctionnalités des écosystèmes en tablant notamment sur le lien positif qui existe entre fonctionnement des écosystèmes et diversité biologique. En suivant ce principe, pour insérer la connectivité dans une logique de préservation de la diversité fonctionnelle à l'échelle des écosystèmes, cet enjeu doit être abordé dans une perspective multi-espèces. Cependant, l'approche exhaustive demeurant hors de portée pour diverses raisons (*e.g.*, connaissances encore parcellaires, coût des analyses), une solution opérationnelle peut être de se focaliser sur un cortège optimal d'espèces, qui donnerait l'opportunité, selon les principes des espèces parapluies, de préserver la diversité fonctionnelle et taxonomique de l'ensemble de l'écosystème.

L'identification de ce cortège optimal d'espèces reste toutefois une problématique délicate. À notre connaissance, aucune étude n'a aujourd'hui validé la capacité d'une espèce à prédire la réponse d'une, ou plusieurs, autres espèces à la connectivité sur la base de données biologiques reflétant la dispersion des individus, telles que les données génétiques. Nous avons donc mené une étude de validation croisée interspécifique portant sur cinq espèces d'oiseaux tropicaux couvrant un gradient de spécialisation pour l'habitat forestier, dans un contexte de fragmentation de cet habitat. Nous avons évalué si le modèle de résistance du paysage pour une espèce focale pouvait prédire de manière fiable la réponse génétique d'une autre espèce au même paysage, et ceci d'autant plus que leurs niveaux de spécialisation étaient proches. Notre objectif était ainsi d'identifier une ou plusieurs espèces pouvant servir d'ombrelles aux autres, et sur lesquelles concentrer les mesures de préservation des connectivités paysagères pour potentiellement produire un effet global sur toute la communauté.

## **4.2. Matériels et méthodes**

Dans cette perspective, nous avons tout d'abord optimisé une surface de résistance pour chacune des 5 espèces sur la base des données génétiques les concernant. Nous avons pour cela utilisé l'algorithme d'optimisation ResistanceGA (Peterman, 2018), dont l'évaluation a été présentée au chapitre précédent. Nous avons ensuite adopté une démarche de validation croisée, en prédisant la structure génétique d'une espèce à partir d'un modèle de résistance calibré pour une autre espèce (i.e., un modèle de résistance correspond à la surface de résistance optimisée), en considérant toutes les paires d'espèces possibles. Une bonne capacité de prédiction inter-espèce suggère que l'espèce à la base de la prédiction serait une bonne espèce parapluie ou du moins indicatrice des effets de la connectivité pour l'espèce prédite. Nous avons ensuite évalué dans quelle mesure les occupations du sol les plus résistantes au mouvement pour une espèce indicatrice étaient identiques à celles contraignant le mouvement pour des espèces dont la structure génétique était prédite par les contraintes paysagères de cette espèce indicatrice. Nous avons ainsi répondu à la question : la capacité des modèles calibrés pour des espèces indicatrices à prédire la structure génétique d'autres espèces trouve-t-elle son origine dans l'homogénéité de l'influence des éléments paysagers sur la dispersion de ces espèces ?

Dans le chapitre 3.1, nous avons montré que les scénarios de coûts optimisés par RGA ne pouvaient pas être interprétés de façon brute, car ils reflétaient souvent davantage le sur-ajustement des modèles que des effets écologiques extrapolables au-delà du contexte de leur optimisation. Ainsi, si nous avons précédemment évalué la capacité d'extrapolation spatiale des inférences de coût, nous cherchons maintenant à valider empiriquement leurs capacités d'extrapolation 'spécifique'.

Par ailleurs, nous proposons dans ce chapitre une utilisation conjointe des surfaces de résistance optimisées et de la modélisation des graphes paysagers pour inférer le profil de sensibilité au paysage d'une espèce. Pour chaque espèce, nous avons donc construit un graphe paysager pondéré par des distances-coûts optimisées et calculé la connectivité de chaque tache d'habitat avec la métrique de Flux. Enfin nous avons réalisé une Analyse en Composante Principale (ACP) de ces métriques de connectivité (5 métriques car 5 espèces) afin d'identifier les espèces présentant une réponse génétique similaire à la connectivité de l'habitat et la traduction spatiale de ces similarités. Notre hypothèse était que les espèces



ayant un même niveau de spécialisation pour l'habitat forestier présenteraient des réponses similaires à la résistance exercée par le paysage, et seraient donc de bonnes indicatrices les unes des autres.

### **4.3. Résultats principaux**

Nos résultats ont mis en évidence le fait que le modèle de connectivité d'une espèce spécialiste de l'habitat forestier pouvait être extrapolé pour inférer la connectivité d'une autre espèce aussi spécialiste de cet habitat, mais pas le patron de connectivité d'une espèce plus moins exigeante dans ses préférences écologique. A l'inverse, les deux espèces les plus généralistes présentaient de bonnes capacités de prédiction l'une envers l'autre. Elles étaient en revanche de mauvaises prédictrices du patron de connectivité affectant les espèces spécialistes. En outre, nous avons démontré que les taches d'habitat étaient classées de manière similaire en termes de connectivité pour les espèces ayant des niveaux de spécialisation similaires. Cela suggère que le modèle de résistance du paysage d'une espèce focale peut prédire la structure génétique d'une autre espèce à condition que ces espèces partagent le même modèle de structuration génétique et niveau de spécialisation pour l'habitat.

Par conséquent, une mesure de conservation ciblant une espèce sera potentiellement bénéfique à d'autres espèces aux comportements de dispersion et préférences d'habitat similaires. On peut donc faire l'hypothèse que le cortège optimal d'espèces à considérer pour préserver la diversité fonctionnelle d'un écosystème doit être constitué d'espèces présentant des profils écologiques variés. Pour vérifier cette hypothèse, il serait intéressant d'élargir cette étude avec un échantillon d'espèces présentant des profils de dispersion et de spécialisation plus contrastés, ou occupant d'autres types d'habitats.

#### **4.4. Article**

**Is landscape effect on gene flow shared across species? A multispecies cross-validation landscape genetic study.**

Daniel, Alexandrine ; Savary, Paul ; Foltête, Jean-Christophe ; Vuidel, Gilles ; Faivre, Bruno ; Garnier, Stéphane ; Khimoun, Aurélie.

Article en cours de préparation.

## **Is landscape effect on gene flow shared across species? A multispecies cross-validation landscape genetic study.**

Daniel, Alexandrine ; Savary, Paul ; Foltête, Jean-Christophe ; Vuidel, Gilles ; Faivre, Bruno ; Garnier, Stéphane ; Khimoun, Aurélie.

### **4.4.1. Abstract**

Maintaining population connectivity is an important stake for biodiversity conservation, and must be achieved with a multi-specific perspective to preserve ecosystem functioning. Yet, identifying the optimal suite of species to consider for modeling connectivity is still a major hurdle, as corridors identified for one species do not necessarily benefit other species. To our knowledge, no study has validated the ability of one species to predict the response of another species to habitat connectivity, based on biological data reflecting individual movements, such as genetic data. Therefore, we conducted an interspecific cross-validation study focusing on five tropical bird species spanning a specialization gradient for forest habitat. We assessed whether the pattern of landscape resistance of a focal species can reliably predict the genetic response of another species to the same landscape. First, we optimized a resistance surface for each species based on their genetic populations structure. Then, adopting a cross-validation approach, we assessed the ability of the ecological distances derived from resistance surfaces optimized for a species to reliably predict the genetic differentiation pattern of the four other species. Finally, for each species, after converting the habitat maps and resistance surfaces into habitat graphs, we calculated connectivity metrics. Multivariate analyses of these metrics determined whether species with similar levels of specialization shared habitat connectivity patterns and genetic responses to landscape structure. We found that the best cross-predictions of genetic structure according to landscape resistance were made between the most specialist species on the one hand, and the most generalist species on the other. Moreover, we demonstrated that habitat patches were similarly ranked in terms of connectivity for species having similar habitat preferences. This suggested that for the landscape resistance pattern of a focal species to predict the genetic structure of another, these species must share similar habitat preferences. Consequently, effective conservation policies for corridor design and habitat patch prioritization and restoration need to be aligned when they target suites of species with similar dispersal behaviors and habitat requirements.

#### 4.4.2. Introduction

Maintaining species functional connectivity is necessary to enable them to cope with current climate change perturbations (Correa Ayram *et al.*, 2016; Heller and Zavaleta, 2009; Kebaïli *et al.*, 2023). To preserve ecosystem functioning more broadly, this connectivity issue must be addressed from a multi-specific perspective (Beier *et al.*, 2008; Cushman and Landguth, 2012; Meurant *et al.*, 2018). The challenge therefore lies on how we can simultaneously satisfy the requirements of multiple species in terms of habitat amount, distribution, and connectivity given their varying dispersal abilities.

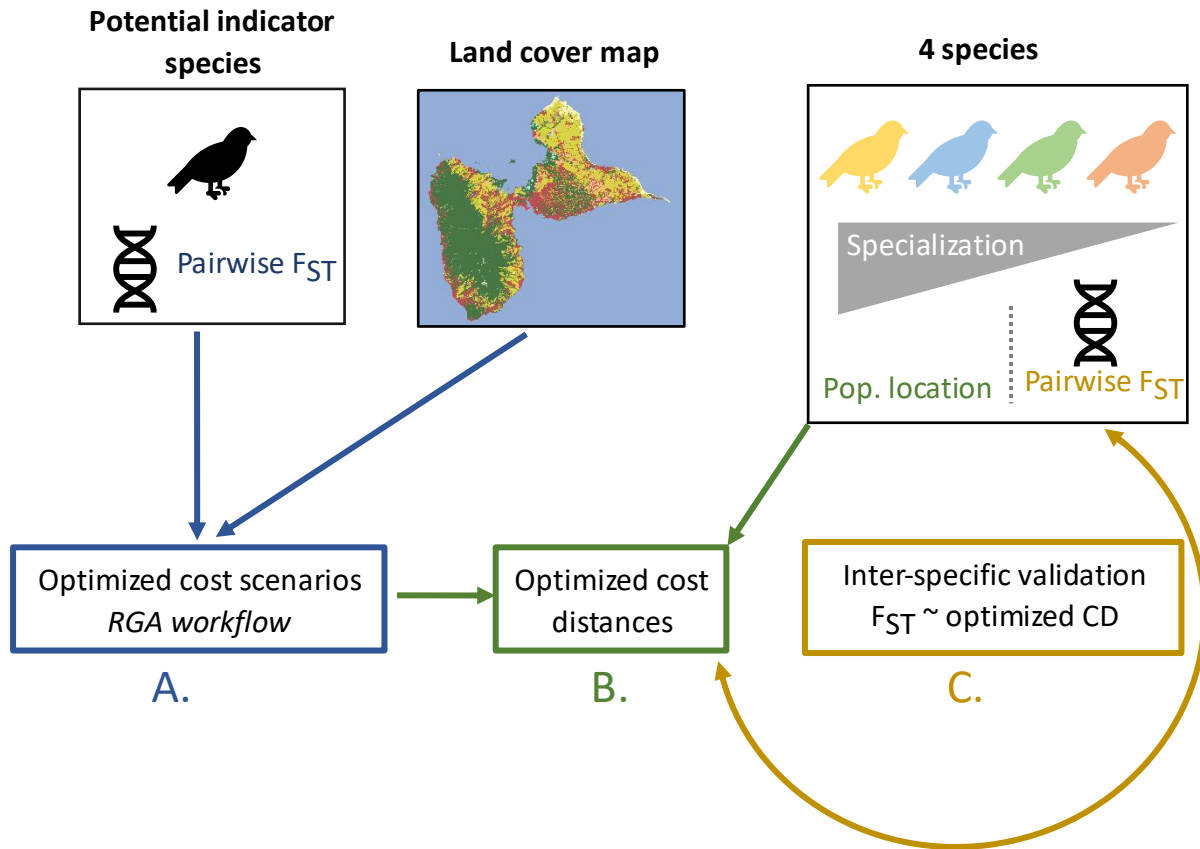
Conservation biologists have been addressing this challenge by developing a growing number of approaches to model connectivity and guide efforts to maintain or restore it for a wide range of species. For example, linkage planning (Beier *et al.* 2008) relied on the identification of a suite of species whose characteristics would ensure multiple species movement and the propagation of ecosystem processes between two connected sets of habitat patches. Choosing the appropriate surrogate species is key for ensuring the ecological effectiveness of such conservation measures (Meurant *et al.*, 2018), as corridors identified for one species do not necessarily benefit other species (Beier *et al.*, 2009).

Considering umbrella species, i.e., species with a wide home range that encompasses that of many other species (Roberge and Angelstam 2004), is a common strategy for the design of conservation measures. Yet, this may be less effective for linkage planning than considering a suite of species representing a diversity of habitat and movement requirements (Meurant *et al.* 2018). The latter approach nonetheless requires evaluation and Cushman and Landguth (2012) were the first to explicitly assess the potential of a species to act as a connectivity indicator for others, depending on its dispersal ability. They found that habitat specialist species with limited dispersal abilities were weak indicators of connectivity for other species. The choice of the species is here crucial and could explain these disappointing performances. We can indeed hypothesize that a species will correctly estimate the connectivity of another species only if they share similar life history traits and responses to disturbance (Brodie *et al.* 2015).

Although identifying the optimal suite of species for connectivity modeling is still a major hurdle, to our knowledge, no study has validated the ability of one species to predict the response of another species to connectivity based on biological data reflecting the movement of individuals. Genetic data would be very informative in this context as they provide direct

insights into demographic processes such as gene flow, and thereby reflect accurately habitat connectivity variations. Several comparative studies in the field of landscape genetics (Manel *et al.*, 2003; Storfer *et al.*, 2007) have tackled the multispecies issue. They explored the differences in species genetic response to connectivity as a function of life-history strategies and spatial scale, among others (Burkhart *et al.*, 2017; Emel *et al.*, 2019; Engler *et al.*, 2014; Gutiérrez-Rodríguez *et al.*, 2017; Mims *et al.*, 2015). Some of these studies demonstrated that landscape effects on gene flow strongly depend on species specialization degree, specialist species being more sensitive to dispersal limitations and underlying connectivity patterns than generalist species (Engler *et al.* 2014, Khimoun *et al.* 2016). Furthermore, the spatial scale considered seems to influence the results of these interspecific comparisons. For example, Radford *et al.* (2021) found no evidence of reduced gene flow at the regional scale, but detected a signal of dispersal limitation at the local scale, depending on the 10 woodlands bird species studied.

Here, we harnessed landscape genetics approaches to address both fundamental knowledge about species movements and habitat preferences, and corresponding conservation needs. For that purpose, we assessed whether a pattern of landscape resistance inferred from genetic data for a focal species can reliably predict the genetic response of another species to the same landscape. We conducted an interspecific cross-validation study focusing on five tropical bird species spanning a specialization gradient for forest habitat (Figure 4.1). First, we optimized a resistance surface for each species based on their genetic data. Then, we performed a cross-validation to assess the ability of each optimised resistance surface to predict the genetic structure of the four other species. Finally, for each species, after converting the habitat maps and resistance surfaces into habitat graphs, we calculated connectivity metrics to determine whether species with similar levels of specialization shared habitat connectivity patterns and genetic responses to landscape structure. The results would allow us to refine our understanding of the similarity of genetic responses to landscape heterogeneity across multiple species. We predicted that bird species with similar levels of specialisation for the forest habitat would exhibit some similarity in their genetic response to landscape structure and in their habitat connectivity pattern. In this case, genetic data from a single species could then be used as a surrogate dataset for identifying dispersal corridors for several equally specialised species when data about those species are missing.



**FIGURE 4.1:** General methodology for the multi-species cross-validation approach. A.) Optimization of a cost-scenario with ResistanceGA (RGA) from the genetic data of a potential indicator species and a landcover map. B.) Based on this optimized cost-scenario, computation of pairwise cost distances among sampled populations for each of the 4 other species not considered in the optimization process. C.) Inter-specific validation by evaluation of the performance of MLPE models explaining the pairwise genetic distances of the independent species with the pairwise cost-distances deriving from the focal species optimization workflow. *Pop. location: spatial location of the sampled population.*

Species	PC	GJ	MG	SR	SB
SSI	0.784	0.665	0.242	0.065	0.060

**TABLE 4.1:** Species specialization index (SSI) based on occurrence data across forest and non-forest patches. This index varies from -1 (strict avoidance of the forest habitat) to 1 (strict preference for forest habitat).

### 4.4.3. Materials & Methods

#### 4.4.3.1. Species and study area

The Guadeloupe archipelago is part of the Lesser Antilles and consists of two main islands linked by a narrow and urbanised isthmus: Basse-Terre (BT) to the west and Grande-Terre (GT) to the east. While BT exhibits a steep topography, an extensive forest cover, and is home to the Guadeloupe National Park, GT landscapes have been much more modified by human settlements and agriculture since the 16th century, which has substantially reduced and fragmented its forest cover (Figure 4.2).

In a previous study, Khimoun *et al.* (2016) computed a species specialization index (SSI) based on occurrence data across forest and non-forest patches to quantify the relative specialization level of eight bird species co-occurring in Guadeloupe. To test our prediction, we selected five of these eight bird species with contrasting ecological profile and SSI values: the Lesser Antillean bullfinch (*Loxigilla noctis*, thereafter referred to as SR), the Lesser Antillean Saltator (*Saltator albicollis*, SB), the Forest Thrush (*Turdus Iherminieri*, GJ), the Scaly-breasted Thrasher (*Allenia fusca*, MG) and the Plumbeous Warbler (*Setophaga plumbea*, PC). PC and GJ are the most specialist species and only occur in forests (Table 4.1). MG is also a forest species but has less stringent requirements as it can occur in forest edges and semi-open woodlands. Finally, SR and SB are the most ubiquitous species and can be observed in woodlands as well as in urban green areas. Preliminary results suggest the existence of a population genetic structure at the Guadeloupe scale for these species (Khimoun *et al.*, 2016). Besides, their genetic response to forest fragmentation is linked to their level of specialisation, with specialist species showing higher levels of genetic differentiation among populations than generalists (Khimoun *et al.* (2016)).

#### 4.4.3.2. Sampling design and molecular analysis

Birds were mist-netted in 45 forest patches, following the protocol outlined by Khimoun *et al.* (2017). As each species was not observed at every site, the number of populations ranged from 22 to 29 across species (Annexe 3-S4.1). Among the 45 sites, 38 were sampled in 2012-2013 and the other seven in 2020. Four sites were sampled twice and temporal genetic

differentiation between 2015 and 2020 was tested for each species using genepop (Raymond and Rousset, 1995). This temporal genetic differentiation was not significant, except for one PC population and one GJ population (Annexe 3-S4.1). Because the relative level of genetic differentiation between these populations that had slightly differentiated over time and the other ones remained stable from 2015 to 2020, we pooled the data.

The total DNA extraction of the 2015 dataset was performed following a standard phenol-chloroform protocol (optimized from Hillis *et al.* 1996) from 240  $\mu$ L of blood in QLB, after the first step of proteinase K digestion (overnight incubation at 55 °C). The DNA from 2020 samples was extracted using Blood Genomic DNA Mini-Preps Kits (BIO BASIC INC., Markham, Canada). Depending on species, individuals were genotyped at 10-14 microsatellite loci (Annexe 3-S4.2) following PCR conditions reported in Khimoun *et al.* (2016a) and in Annexe 3-S4.1. Loci were amplified in simplex in a Dyad thermal cycler (Bio-Rad, Hercules, CA, USA). PCR products were multi-loaded for analyses in an automated sequencer (ABI3730), and allele scoring was performed using GENEIOUS R.8 (Kearse *et al.*, 2012).

#### 4.4.3.3. Population genetics analyses

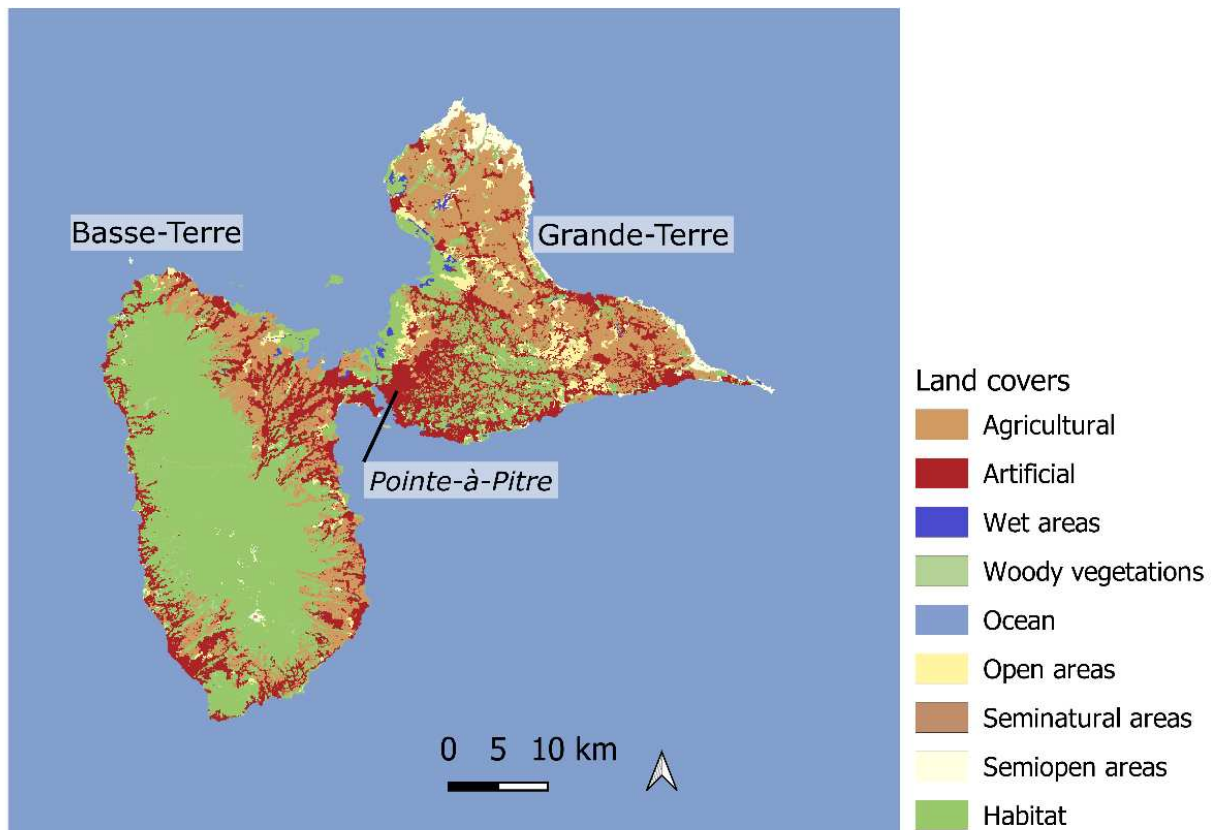
For each species, we checked for Hardy-Weinberg Equilibrium for each locus in each population and for the absence of linkage disequilibrium between pairs of loci using genepop (Raymond and Rousset, 1995).

We estimated genetic differentiation between populations with pairwise  $F_{ST}$  values (Weir and Cockerham 1984) and tested for genotypic differentiation using a log-likelihood based exact G-test. Significance levels were adjusted using Bonferroni corrections for multiple tests. We evaluated the Isolation-By-Distance (IBD) signal for each species with maximum-likelihood population-effect (MLPE) models (Clarke, Rothery, and Raybould 2002) modeling the linearized pairwise  $F_{ST}$  values ( $F_{ST}/(1-F_{ST})$ ) as a linear function of the logarithm of Euclidean distances between populations. MLPE models account for the non-independence of pairwise data by including a population-level random effect. To test for the presence of IBD patterns at the whole scale of Guadeloupe and at the local scale of the two islands BT and GT, we performed MLPE models for two sets of population pairs: one including all population pairs and the other including only the pairs of populations located on the same island (either BT or



GT). All these calculations were performed with the R software (R Core team, 2022) and the `graph4lg` package (Savary *et al.*, 2021).

Finally, we described the spatial patterns of genetic structure using a Bayesian model-based clustering algorithm implemented in STRUCTURE 2.3.4 (Pritchard, Stephens, and Donnelly 2000). Within each species, we tested for the presence of  $K$  genetic clusters,  $K$  ranging from 1 to 10 clusters. We used an admixture model with correlated allele frequencies and the LOCPRIOR option. We performed 10 independent runs of 100,000 iterations following a burn-in period of 20,000 iterations. We inferred the number of genetic clusters following the approach proposed by Garnier *et al.* (2004), focusing on successive increases of the posterior probability of the data for increasing values of  $K$ .



**FIGURE 4.2:** Map of the 9 landcover types considered in this study, with a 100m resolution.

#### 4.4.3.4. Spatial data and landcover map

We wanted to test whether resistance surfaces inferred for a focal species could be used to predict genetic differentiation in other species. For modeling resistance surfaces, we first needed to map landscape features hypothesised to influence gene flow. Alike Daniel *et al.* (2023), we used the following land cover databases for that purpose: CORINE Land Cover data (2012), Institut National de l'Information Géographique et Forestière (IGN) BD TOPO (2012), and the 2010 map of woody vegetation formations created by the Conseil Départemental of Guadeloupe, the IGN, and the Direction Régionale de l'Office National des Forêts (ONF). The initial land cover map included 10 land cover types and had a resolution of 20m (see Daniel *et al.* (2023) for more details), before we gathered the inland water surfaces and wetlands, ending up with 9 land cover types. Finally, we aggregated the map at a 100 m spatial resolution to make it compatible with the very time-consuming optimisation approach (Figure 4.2).

#### 4.4.3.5. Optimization of cost scenarios with ResistanceGA and computation of cost distances

From this land cover map, we built for each species the resistance surface that best fits and predicts its genetic differentiation pattern, using an optimization algorithm for an efficient exploration of the cost parameter space. A resistance surface aims at making explicit how landscape features influence gene flow by moderating the effect of geographic distance on genetic differentiation. On this surface, each pixel is assigned a cost depending on the landscape feature it represents. This cost value then serves as a coefficient weighting every distance unit covered during species movements. It thus represents the physiological costs and risks incurred by individuals dispersing across an heterogeneous landscape. The combination of costs for all landscape features is referred to as a cost-scenario, and our goal was to identify the cost scenario providing the best fit in models predicting pairwise genetic differentiation for each species.

To parametrize resistance surfaces deriving from the landcover map, we used a genetic algorithm called ResistanceGA (RGA), commonly used in landscape genetics (Antunes *et al.*, 2023; Mulvaney *et al.*, 2021). The RGA algorithm provides an efficient way to explore the multidimensional cost parameter space and to optimize cost scenarios. The optimisation stops when the goodness-of-fit of the relationship between pairwise cost distances (i.e., the

accumulation of pixel costs along a least cost path between two populations) and pairwise genetic distances (here pairwise  $F_{ST}$  values) reaches a maximum. Based on genetic data simulated for virtual species with contrasted ecological profiles, Daniel *et al.* (submitted) showed that this optimisation framework led to cost scenarios that were very good predictors of genetic distances. Here we used the modified version of the RGA workflow described in Daniel *et al.* (submitted). Accordingly, the goodness-of-fit objective criterion was the validation  $R^2$  introduced in this study, in order to limit the risk of overfitting during optimization (see Daniel *et al.* (2023) for more details on its calculation). Other parameters of the genetic algorithm were kept at default values.

Daniel *et al.* (submitted) demonstrated that making flexible the spatial scale and specific set of population pairs included in the optimisation could significantly improve the reliability of the results. Following their methodological guidelines, we implemented a pruning option that makes it possible to only assess and optimize the relationship among distances for a specific subset of population pairs. The rationale behind this methodological choice is here the potentially strong dispersal limitation between Basse-Terre and Grande-Terre, as reflected by previous analyses (Arnoux *et al.*, 2014; Daniel *et al.*, 2023). We therefore considered two sets of population pairs reflecting two hypotheses about the shape of the dispersal network:

- A complete set including all population pairs – ***Whole Guadeloupe scale***: it reflects the assumption that all the populations of the dispersal network may be connected by dispersal events.
- A pruned set restricted to pairs of populations located within the same island (BT-BT or GT-GT pairs) – ***Within-island scale***: this dissociates two spatially coherent entities, assuming there is little dispersal between them.

For each of the five considered species, we conducted the optimization workflow under each of the two dispersal scale assumptions (and corresponding pruning methods), ending up with two optimized cost scenarios. This resulted in 10 optimized resistance surfaces for the same landscape. Each optimized model was characterized by its support against the Euclidean model (based on AIC comparisons), its predictive quality of genetic differentiation (based on the validation  $R^2$ ) and the contrast of its related cost scenarios (i.e., the 10-logarithm ratio between the maximum and minimum cost values).

We then used these resistance surfaces to compute cost distances among populations. We wanted to know whether these cost-distances optimized for a species could be a good predictor of the genetic differentiation of other co-occurring forest bird species. For that purpose, we implemented an inter-specific cross-validation approach, considering species A as the focal species and species B as the extrapolation species. We examined the relationship between the pairwise genetic distances among populations of species B using MLPE models, and their related cost distances on the resistance surface optimized for species A, at both the whole Guadeloupe scale and the within-island scale (Figure 4.1). We adapted a leave-one-out cross-validation method to assess the predictive power of this model. This method iteratively removes a population and its associated pairwise distances for the calibration step, then predicts the genetic distances in which this population is involved and compares them with the observed genetic distances. A validation  $R^2$  statistic is computed to summarise the differences between observed and predicted genetic distances among all populations (more details on this method can be found in Daniel *et al.* 2023). We examined the model performances for all possible combinations of focal and extrapolation species.

#### 4.4.3.6. Habitat connectivity models from landscape graphs and optimized cost scenarios

We tested whether a species identified as a good indicator of the genetic connectivity of another species also had a habitat connectivity pattern congruent with that species. For that purpose, we modelled habitat connectivity for the five focal species using landscape graphs. A landscape graph consisted of nodes corresponding to habitat patches, connected by links representing potential dispersal paths. Here, the five landscape graphs shared the same nodes, corresponding to the forest patches of the landcover map. The links were weighted by pairwise cost distances according to the optimized cost scenarios, on a species-specific basis. The landscape graph topology was planar, as it efficiently approximates the complete graph topology in this context. Then, for each graph, we computed a local connectivity metric assessing the contribution of each habitat patch to emigration and immigration flows.

This metric, inspired by the incidence function of Hanski and Gilpin (1991), is called the flux index (F) and was computed as follows:

$$F_i = \sum_{j=1; j \neq i}^n A_j \times e^{-\alpha \times d_{ij}}$$

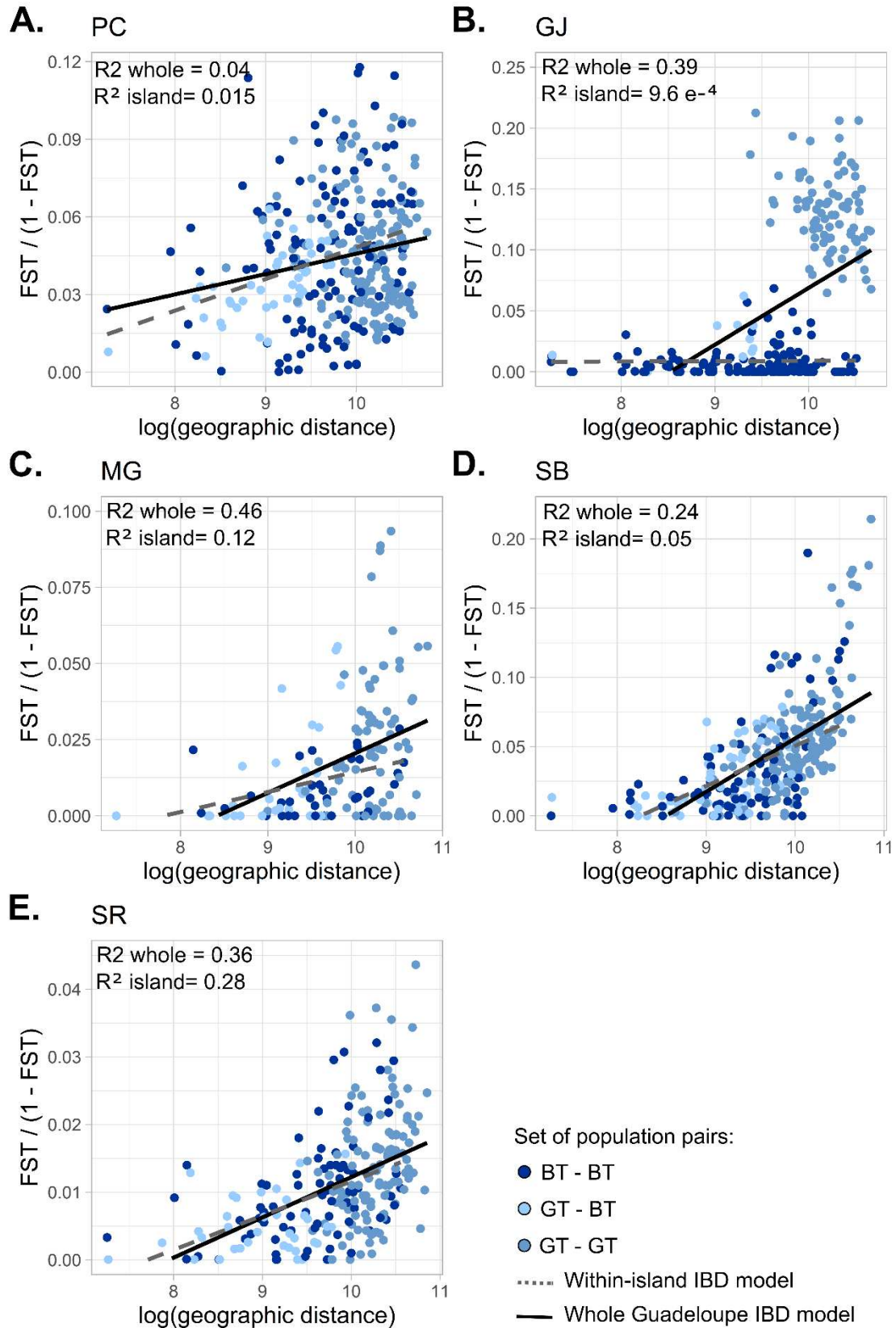
With  $i$  the index of the focal patch,  $j$  the index of all the other  $n$  patches,  $A$  the patch area and  $d_{ij}$  the cost-distance between patches  $i$  and  $j$ .  $\alpha$  sets the spatial scale at which pairwise connections were the most important for the computation of this index. Indeed,  $\alpha$  determines the shape of the exponential decay of dispersal probability with the distance between patches  $i$  and  $j$  ( $p = e^{-\alpha \times d_{ij}} = 0.05$ ). We performed a preliminary analysis and chose  $\alpha$  such that patches located more than 1,000 m away were considered to be outside the immediate neighbourhood and reached with a probability below 5%.

Finally, we performed a Principal Component Analysis (PCA) to project all habitat patches in the space of the Flux indices computed for the five species. We predicted that the importance of patches in dispersal flows would depend on the degree of specialization of the species for forest habitats.

#### 4.4.4. Results

##### 4.4.4.1. Population genetic results

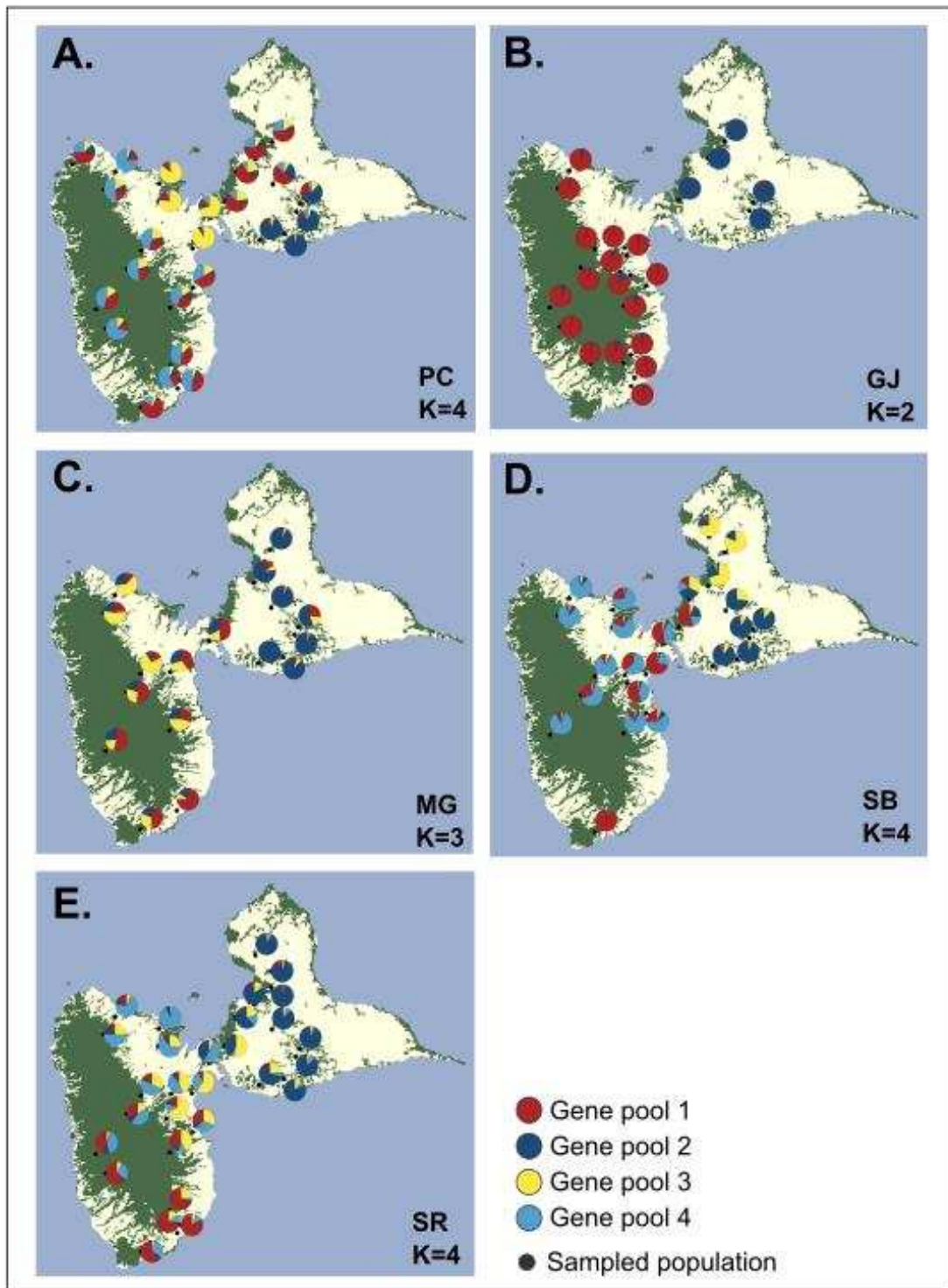
The vast majority of loci were in Hardy-Weinberg equilibrium and independent (Annexe 3-S4.2). Both the detection of genetic clusters with STRUCTURE and the analysis of pairwise genetic differentiation point towards the presence of two main allelic pools separating BT populations from GT populations for all species (Figures 4.3, 4.4 and Table 4.2). The intensity of this overall structuring and the existence of finer levels of differentiation vary among species.



**FIGURE 4.3:** Graphical visualization of Isolation-by-Distance (IBD) models explaining linearized  $F_{ST}$  by the logarithm of geographical distances. Two IBD models were performed for each species: one with all pairwise data (whole Guadeloupe scale model with the full regression line) and one accounting for populations pairs located in the same island (i.e., only BT-BT and GT-GT pairwise data; within-island model with dashed regression line). The calibration  $R^2$  of these IBD models are displayed on the top-left corner of each graph. Pairwise data are plotted under the regression lines according to their spatial location. BT-BT: pairwise data from populations located within BT; GT-BT: pairwise data from populations located on either side of the isthmus; GT-GT: pairwise data from populations located within GT.

Pop. set	Species	Mean $F_{ST} \pm SD$ (min - max)	Proportion of pop. pairs significantly differentiated
BT-BT	Plumbeus Warbler (PC)	0,043 $\pm$ 0,026 (0 – 0,105)	79 %
GT-BT	Plumbeus Warbler (PC)	0,044 $\pm$ 0,019 (0,011 – 0,090)	98 %
GT-GT	Plumbeus Warbler (PC)	0,031 $\pm$ 0,013 (0,006 – 0,059)	92 %
Total	Plumbeus Warbler (PC)	0,042 $\pm$ 0,022 (0 – 0,105)	90 %
BT-BT	Forest Thrush (GJ)	0,006 $\pm$ 0,013 (0 – 0,064)	0,7 %
GT-BT	Forest Thrush (GJ)	0,110 $\pm$ 0,025 (0,060 – 0,180)	94 %
GT-GT	Forest Thrush (GJ)	0,025 $\pm$ 0,018 (0 – 0,059)	20 %
Total	Forest Thrush (GJ)	0,047 $\pm$ 0,056 (0 – 0,180)	35 %
BT-BT	Scaly-breasted Thrasher (MG)	0,008 $\pm$ 0,008 (0 – 0,028)	0
GT-BT	Scaly-breasted Thrasher (MG)	0,024 $\pm$ 0,021 (0 – 0,086)	26 %
GT-GT	M Scaly-breasted Thrasher (MG)	0,012 $\pm$ 0,017 (0 – 0,053)	0
Total	MG Scaly-breasted Thrasher (MG)	0,018 $\pm$ 0,019 (0 – 0,086)	14 %
BT-BT	Lesser Antillean bullfinch (SR)	0,010 $\pm$ 0,008 (0 - 0,030)	21 %
GT-BT	Lesser Antillean bullfinch (SR)	0,014 $\pm$ 0,008 (0 – 0,042)	74 %
GT-GT	Lesser Antillean bullfinch (SR)	0,005 $\pm$ 0,004 (0 – 0,015)	14 %
Total	Lesser Antillean bullfinch (SR)	0,011 $\pm$ 0,008 (0 – 0,042)	46 %
BT-BT	Lesser Antillean Saltator (SB)	0,031 $\pm$ 0,034 (0 – 0,160)	29 %
GT-BT	Lesser Antillean Saltator (SB)	0,051 $\pm$ 0,033 (0 – 0,180)	85 %
GT-GT	Lesser Antillean Saltator (SB)	0,024 $\pm$ 0,02 (0 – 0,073)	53 %
Total	Lesser Antillean Saltator (SB)	0,040 $\pm$ 0,033 (0 – 0,180)	63 %

**TABLE 4.2:** Graphical visualization of Isolation-by-Distance (IBD) models explaining linearized  $F_{ST}$  by the logarithm of geographical distances. Two IBD models were performed for each species: one with all pairwise data (whole Guadeloupe scale model with the full regression line) and one accounting for populations pairs located in the same island (i.e., only BT-BT and GT-GT pairwise data; within-island model with dashed regression line). The calibration  $R^2$  of these IBD models are displayed on the top-left corner of each graph. Pairwise data are plotted under the regression lines according to their spatial location. BT-BT: pairwise data from populations located within BT; GT-BT: pairwise data from populations located on either side of the isthmus; GT-GT: pairwise data from populations located within GT.



**FIGURE 4.4:** Spatialization of the results from the clustering analyses with STRUCTURE for each of the five species. Only the higher levels of structuring are displayed here. See Annexe 3-S4.4 for coarser levels. Each pie chart illustrates the genetic composition of a sampled population, according to the number of clusters considered for each species. K represents the optimal number of clusters for the finest genetic structuring (See Annexe 3-S4.4 for information on its determination).



The PC exhibited the highest level of genetic structure, 90 % of its population pairs being significantly differentiated, including 98% of inter-island population pairs (Table 4.2). Four genetic clusters could be identified for this species, respectively separating the populations from BT, the isthmus, and the north and south of GT (Figure 4.4A). The extremely weak IBD pattern ( $R^2 = 0.04$ ) (Figure 4.3A) suggested that straight line geographical distances poorly reflect the constraints to dispersal shaping genetic differentiation patterns.

The GJ only exhibited two clusters distinguishing BT and GT populations (Figure 4.4B), as supported by low genetic differentiation values within these islands, particularly in BT (Table 4.2). The inter-island genetic distances were notably higher than their intra-island counterparts, resulting in an artificially strong fit of the IBD model at the whole Guadeloupe scale ( $R^2=0.39$ ) (Figure 4.3B). Its strength sharply decreased when considering only the pairs of populations located on the same island ( $R^2= 9.6e^{-4}$ ), suggesting an isolation by barrier pattern, precisely the isthmus in that case.

The MG was the least genetically structured species, without any differentiated population pairs within neither BT nor GT (Table 4.2). Nevertheless, it showed an IBD pattern at the whole Guadeloupe scale ( $R^2=0.24$ , Figure 4.3C), as reflected by the genetic clusters detected by STRUCTURE along a north-south gradient on BT. (Figure 4.4C).

The SB and the SR showed similar levels of genetic differentiation, with highest structure among inter-island than intra-island population pairs (Table 4.2). They showed a clear IBD signal, both when considering all population pairs ( $R^2=0.46$  and  $0.39$  respectively, Figure 4.3D,E) or only those located on the same island ( $R^2=0.28$  and  $0.12$  respectively). Four genetic clusters could be identified for both species, that may reflect the IBD pattern (Figure 4. 4D, E). The SR showed three clusters forming a north-south gradient on BT and isolating the homogeneous populations of GT. For the SB, one cluster isolated a population in the far south of BT from the northern ones, and two other clusters structured the GT populations.

#### 4.4.4.2. Optimized cost scenarios from the genetic data of the five studied species

For all species, the IBR models derived from optimized cost scenarios had better support than the IBD model based on Euclidean distances. However, the comparison of the optimized cost scenarios obtained at both spatial scales revealed some discrepancies between species. At the whole Guadeloupe scale, the contrast of cost scenarios decreased with specialization level. Moreover, whereas the forest habitat was the least resistant land cover at both spatial scale for the PC, GJ and MG (Table 4.3 and Annexe 3-S4.3), SB and SR showed greater variability in the distribution of their land cover resistance values. At within island scale, the genetic differentiation patterns of PC, GJ, and MG were poorly supported by a model involving only Euclidean distances. On the contrary, the cost-distance models of SB and SR only had a slightly better fit than the Euclidean distance models at this scale. It should be noted that, according to the SSI calculated by Khimoun *et al.* (2016), these differences followed the degree of specialisation of these species for the forest habitat, with GJ and PC being more specialist than MG, which in turn is more specialist than SB and SR.

Notably, the PC showed the highest contrast between its lowest and highest cost values at the whole Guadeloupe scale, with a ratio of  $10^{3.99}$  (Table 4.3). Contrary to the other species, the support and predictive quality of the models for the PC improved when considering only population pairs located within the same island (whole Guadeloupe scale:  $\Delta AIC = -55$ ,  $vR^2=0.2$ , within-island scale:  $\Delta AIC=-76$ ,  $vR^2=0.49$ ). The GJ resistance model received the best support and predictive quality at the whole Guadeloupe scale, strikingly better than for the other species ( $\Delta AIC = -370$ ,  $vR^2=0.9$ ). However, these performances dropped sharply at the within-island scale ( $\Delta AIC=-28$ ,  $vR^2=0.24$ ). The  $vR^2$  and model support for the MG remained stable around respective values of  $vR^2 = 0.5$  and  $\Delta AIC = -40$  at both spatial scales. The SB showed the lowest cost value contrast and the weakest support for the cost-distance model, with a relatively uniform distribution of cost values (from 1 to 58), thus translating into a cost-distance distribution not very different from that of Euclidean distances. Its predictive accuracy was relatively high, although it decreased from  $vR^2 = 0.56$  to  $vR^2 = 0.33$  when restricting the spatial scale of pairwise connections to the within-island scale. The predictive quality of the SR model remained stable around a  $vR^2$  of 0.35 at both scales. Unlike others, its cost value contrast jumped from  $10^{2.9}$  to  $10^4$  when moving from the whole Guadeloupe to the within-island scale.

<i>Spatial scale</i>	<i>Whole Guadeloupe scale</i>					<i>Within-island scale</i>				
<b>Species</b>	<b>PC</b>	<b>GJ</b>	<b>MG</b>	<b>SR</b>	<b>SB</b>	<b>PC</b>	<b>GJ</b>	<b>MG</b>	<b>SR</b>	<b>SB</b>
<b>Contrast</b>	3,99	3,60	3,20	2,90	1,80	2.92	3.6	3.3	4.5	1.9
<b><math>\Delta</math>AIC</b>	-65	-370	-42	-42	-23	-72	-28	-30	-8	-7
<b>vR<sup>2</sup></b>	0,20	0,90	0,50	0,36	0,56	0,49	0,24	0,52	0,35	0,33

**TABLE 4.3:** Summary table of the optimization results at the two spatial scales considered and for the five species. Contrast: the logarithm to the base 10 of the ratio between the highest and the lowest cost value of the optimized cost scenario (e.g. for the {1, 10, 100, 1000} scenario, the contrast is  $\log_{10}(1000/1) = 3$ ).  $\Delta$ AIC: AIC difference between the optimal isolation by landscape resistance (IBR) model and the Euclidean model. A negative value of  $\Delta$ AIC means that the IBR model received more support than the Euclidean model. vR<sup>2</sup>: validation R<sup>2</sup> criterion, standing for the predictive quality of the IBR model.

#### 4.4.4.3. Interspecies cross-validation for genetic distance modelling

When evaluating to what extent the cost scenario optimized for a species translates into cost-distances that are good predictors of the genetic differentiation pattern of another species, we first notice that interspecific predictive performances were broadest but often weaker at the whole Guadeloupe scale than at the intra-island scale (Figure 4.5). At both spatial scales, we can distinguish two species groups (represented by two modules on the network-like representation of Figure 4.5) associated with great “cross-prediction” capacities: the most specialist species (GJ, PC, MG) on the one hand, and the most generalist species (SB and SR), on the other hand. This means that the optimized cost values supposed to reflect the dispersal behavior of a species from one of these two groups were poorly informative regarding the dispersal-driven genetic pattern of a species from the other group.

Note that the figure 4.5 can be read on two levels. If there is no connection between a predictive and a predicted species, then the vR<sup>2</sup> associated with this cross prediction is zero. If there is a connection, then the cost-distance model associated with the cross-prediction may be i) less supported than the model based on Euclidean distances (suggesting an IBD pattern), or ii) more supported than the Euclidean distances model (suggesting an IBR pattern).

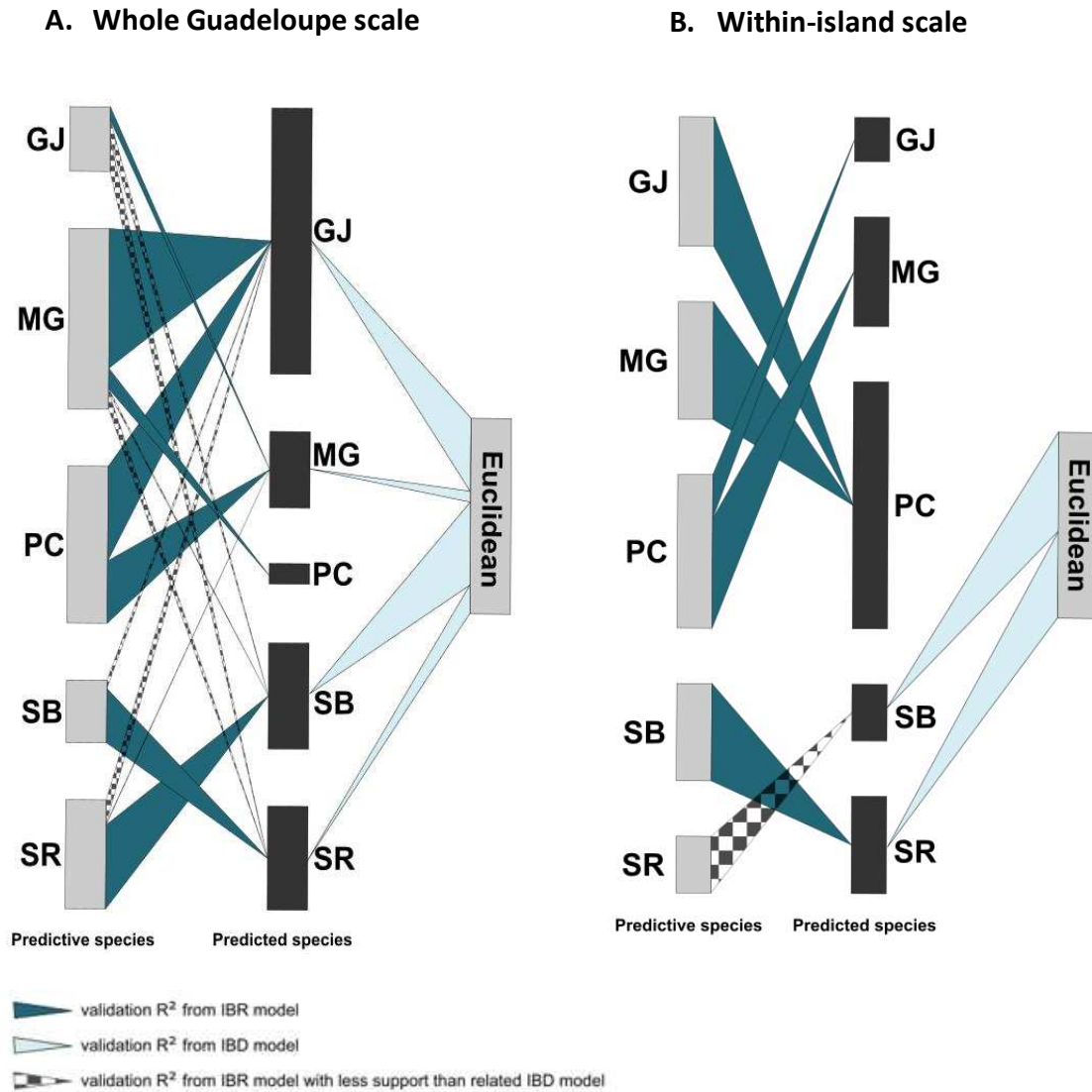
#### 4.4.4.3.1. Whole Guadeloupe scale

The cost scenario optimized from the PC data could reliably predict the MG and the GJ genetic distance ( $vR^2 = 0.32$  and  $0.48$  respectively) (Figure 4.5A, Annexe 3-S4.5). The MG best cost-scenario was a good predictor of genetic differentiation in GJ ( $vR^2=0.72$ ), but also slightly predicted PC. The GJ appeared to be a poor predictor of other species genetic patterns. The SB and SR cost scenarios were only good at predicting each other's genetic structure ( $vR^2 = 0.43$  and  $0.28$  respectively). Overall, the GJ, SB, and SR were the best predicted species, and the MG and the PC were the best predictive species (Figure 5A, Annexe 3-S4.5). One should notice that GJ, SB, and SR were predicted very accurately by the Euclidean model. For GJ, only the cross predictions based on cost-distances deriving from the MG and PC cost-scenarios were better supported than Euclidean models. For SB and SR, only the cross predictions derived from each other model received better support than the Euclidean model. These results suggested that the numerous weak cross-validation  $R^2$ s characteristic of that spatial scale may result from an IBD pattern common across species. However, the specialist species (PC, GJ, MG) on the one hand and the generalist species (SR and MG) on the other hand shared a common structuring signal that differed from IBD.

#### 4.4.4.3.2. Within-island scale

The PC, MG, and GJ only predicted each other relatively well, and their cost-scenarios were poorly informative of the genetic patterns of the other two generalist species (Figure 4.5B, Annexe 3-S4.5). The PC reliably predicted both the MG ( $vR^2 = 0.25$ ) and the GJ genetic differentiation ( $vR^2 = 0.10$ ), but the GJ and the MG only predicted the PC genetic distances with good accuracy ( $vR^2 = 0.30$  and  $0.27$  respectively). These four cost distance models of cross-species validation received better support than Euclidean model, showing that the overlap in the structuring signal between these species was not due to a common IBD signal. In turn, SB and SR were always very good predictors of each other ( $vR^2$  of  $0.13$  and  $0.22$  respectively) but did not reveal good at predicting any specialist species (Figure 6). However, the cost-distance model cross-predicting SB genetic structure from SR one received less support than Euclidean model ( $\Delta AIC_{\text{cost-distance} - \text{Euclidean distance}} = 16.2$ ). In turn, the cost-distance model, which cross-predicted the genetic structure of the SR from the SB one, received quite the same support

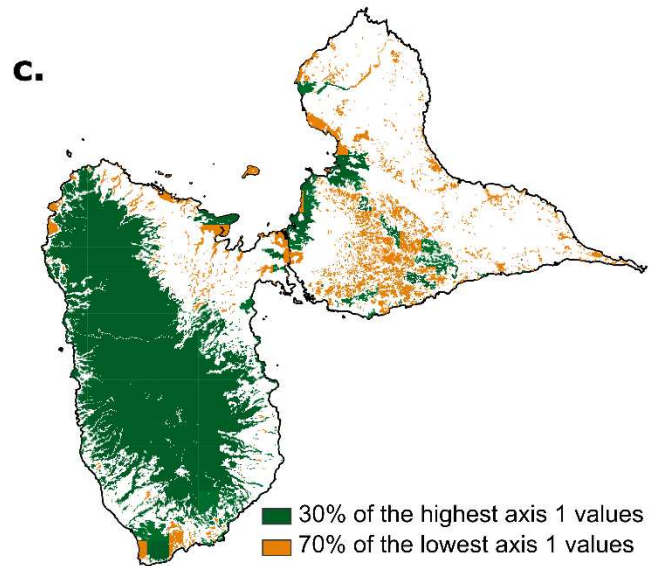
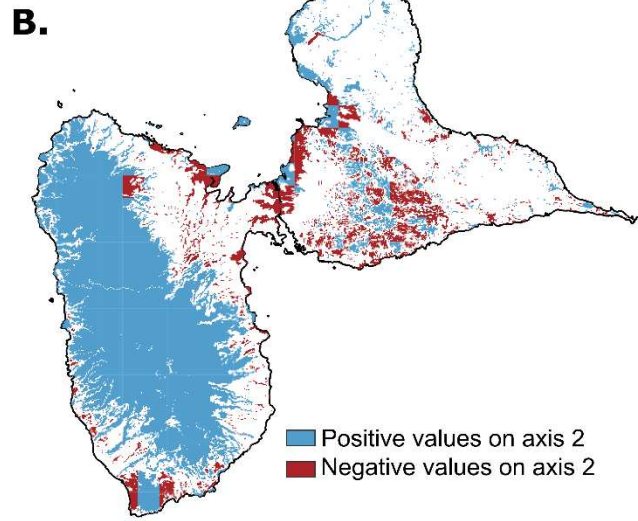
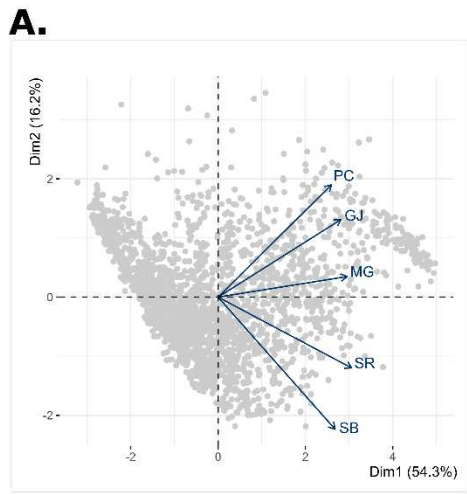
as the Euclidean model ( $\Delta AIC_{\text{cost-distance} - \text{Euclidean distance}} = -1.9$ ). This suggested that the good predictive power of these cost-distance models mainly rely on the IBD signal common to SB and SR, rather than on a shared pattern of genetic differentiation shaped by landscape resistance to dispersal.



**FIGURE 4.5:** Graphical representation of the results of the multi-species cross-validation models. A.) Models extrapolated for other species at the whole Guadeloupe scale, i.e., considering all pairwise data. B.) Models extrapolated at the within-island scale, i.e., considering data from population pairs located either on BT or GT, but not across these two islands. The grey boxes on the left represent the “predictor” species (from which the optimized cost scenarios were derived), and the dark boxes represent the “predicted” species (whose genetic pattern is predicted). The grey box on the right corresponds to an Euclidean scenario. The box length represents the quality of the species as a predictor or predicted species. The greater, the higher the validation  $R^2$ s ( $vR^2$ ). Each arrow, from a predictor to a predicted species, corresponds to the  $vR^2$  of a model explaining the pairwise genetic data of the predicted species with the pairwise cost distances derived from the predictor’s cost-scenario. The larger the arrow, the higher the predictive power ( $vR^2$ ). A missing arrow between two species means that their related  $vR^2$  is null. A checkerboard arrow corresponds to a strictly positive  $vR^2$  derived from an IBR model receiving less support than the related IBD model. The predictive quality of the IBD model for each species is represented by the light blue arrows, starting from the Euclidean scenario on the right.).

#### 4.4.4.4. Landscape graph properties and PCA analysis on Flux values

The five landscape graphs shared the same 2,306 nodes (mean forest patch area = 28.3 ha). The average cost-distance between these nodes was highest for PC (mean=2,991 cost units (cu)), followed by GJ (mean=985 cu), MG (mean=558 cu), SR (mean=95), and finally SB (mean=33 cu). The first two axes of the PCA, which discriminated the 2,306 habitat patches in the space of the five series of Flux indices, accounted for 54.3 and 16.2% of the variance respectively (Figure 4.6A). Patches with high scores on the Axis I were associated with high potential dispersal flows in the five species, thereby being important for multispecies habitat connectivity. They were mainly located within the dense forest cover of BT (Figures 4.6B, S4.6). On the contrary, many of the small habitat patches in the south of GT were associated with low connectivity levels. The Axis II distinguished the patches according to their importance for the connectivity of the five species independently (Figure 4.6A). The discrimination of patches along this axis matched the specialisation gradient. Thus, a patch with a high value on Axes I and II is a patch to be conserved for specialist species (PC, GJ and MG) connectivity. A patch with a high value on Axis I and a negative value on Axis II benefits the habitat connectivity of generalist species (SB and SR). Representing simultaneously the information carried by both axes (Figures 4.6B, S4.6) shows that small patches at the edge of the dense BT forest and across the isthmus were well connected in the dispersal network of generalist species, but appeared to be disconnected for species more specialist of the forest habitat.



#### 4.4.5. Discussion

##### 4.4.5.1. Interspecific cross-validation of genetic structure predictability

The interspecific predictive network at the whole Guadeloupe scale revealed good predictive accuracies within both specialist (PC, GJ, MG) and generalist (SB, SR) species, and numerous lower predictive powers among species from either group. We found that the models associated with these weak cross-predictions received less support than the Euclidean model. This suggested that the weakly overlapping signals of the generalist and specialist species may actually stem from a shared IBD pattern. The strong divergence between BT and GT populations, detected in all species, may also explain to a large extent this result. Indeed, the Guadeloupe archipelago has a peculiar particular topography and land cover, with a highly urbanised land bottleneck surrounding the isthmus separating the two islands and corresponding to Pointe-à-Pitre, the most populated city and the economic capital. This area seems to reduce inter-island gene flow based on the genetic differentiation patterns observed in all five species. Although it could also result from a secondary contact between divergent GT and BT gene pools, due to colonisation events, this signal represents a significant fraction of the genetic structure of the species.

The strongest predictive powers at the large spatial scale involved the same species as at the smaller intra-island scale: specialist on the one hand, and generalist on the other. This suggests two conditions for the landscape resistance pattern of a focal species to predict the genetic structure of another: i) that focal species must be structured by the distance or the landscape at the scale of interest, and ii) the species to be predicted must share the same structuring process. For the latter condition to be met, they should share common ecological requirements, in our case they were similar in terms of habitat specialisation. These conditions are even more evident at the intra-island level, with a complete absence of cross-predictive power between the specialist and the generalist species. We found that the specialist cross-predictions were better supported than those from the Euclidean model, suggesting that specialist species shared a congruent structuring signal different from IBD. In contrast, the specialist species SB and SR, which were equally effective in cross-predicting each other genetic structure, showed a dominant IBD pattern. Their ability to cross-predict is therefore not primarily based on a species-specific structuring signal, but on a common null model. It would



be interesting to develop this type of study further with two groups of species structured by IBR but sensitive to different landscape features. This would strengthen the assessment of RGA's ability to capture structuring signals common to multiple species, more finely than general pattern such as IBR *versus* IBD.

Overall, our results are consistent with others demonstrating that a specialist species will be more sensitive to landscape heterogeneity and therefore more likely to show an IBR pattern than the generalist species (Emel *et al.*, 2019; Engler *et al.*, 2014). Kierepka *et al.* (2016) found that the genetic structure of specialist species of agricultural rodents was consistently related to habitat configuration, whereas the generalist species were unaffected by landscape features. Therefore, a specialist species may not be a good surrogate for a generalist species and vice versa. Cushman and Ladnguth (2012) found a strong dependence of a species' surrogate status on its dispersal ability. They showed that habitat specialists with low dispersal ability rarely share connectivity patterns with others, unlike large dispersers. Penjor *et al.* (2023) argued that common species with low trophic levels, large home ranges, and long-term persistence are well suited to serve as umbrella species. Our study considered a less diverse pool of species as the five bird species share relatively similar life-history traits. Rather than looking for an umbrella species, we wondered if one species in this cluster might be indicative of the connectivity patterns of others and act as a focal species, as defined by Penjor (2023). We found that two indicator species were needed to accurately reflect the connectivity of these five species, representing specialist and generalist requirements respectively. Our results recall Meurant *et al.*'s (2018) which showed that a species can be a good indicator of other species with similar habitat and movement needs. They demonstrated that sampling multiple indicators representing a diversity of species requirements was a powerful method for multi-species connectivity planning.

#### 4.4.5.2. Population genetic structure across species spanning a forest specialization gradient

The contribution of landscape resistance to the explanation of genetic differentiation patterns appeared to be related to the degree of species specialization for the forest habitat. Working on a very close set of species in Guadeloupe, Khimoun *et al.* (2016) already showed that specialist species are more affected by forest fragmentation than generalist ones. Here, we

found that the more specialist a species is, the greater the contrast between the resistance values and the greater the support for the resistance model as compared to the Euclidean model. However, the specialist species did not necessarily have the finest genetic structure. Indeed, the GJ and MG had the second and third highest SSI values but showed only two and three genetic clusters respectively. Only the PC displayed the expected relationship between genetic structure and specialization, as it had the highest level of specialisation, the finest clustering structure and the highest level of genetic divergence both within the two islands and in Guadeloupe as a whole. Studying landscape genetics relationships for five forest bird species, Mulvaney *et al.* (2021) found that specialist species had a stronger level of genetic spatial structuring and less within-population genetic diversity than generalist species. They related these variations in genetic patterns to the contractions in glacial refuges during Pleistocene climatic variations. The trade-offs between the degree of specialisation, dispersal ability, genetic diversity and differentiation are therefore not obvious and may depend on multiple processes operating at different temporal and spatial scales.

The predictive power of the optimised landscape resistance models was not related to their relative support compared to the Euclidean model, nor to the degree of specialisation of the species. At the entire Guadeloupe scale, the SR, the SB and the MG showed up to twice the predictive power of the PC, despite having the lowest contrast and model support compared to the Euclidean model. This lack of correlation between model predictive quality and support is based on the feature of RGA to overfit cost scenarios with low contrast for species with a significant IBD pattern (Daniel *et al.* in prep). Our study showed that species sharing similar life-history traits and living in the same landscape can have different genetic patterns, and thereby, different responses to the same landscape. Similarly, Radford *et al.* (2021) and Mulvaney *et al.* (2021) evidenced some species-specific relationships with landscape features. Both studies found that the same landscape could either promote or limit gene flow depending on the focal species, according to its dispersal capacity or behaviour. According to Delord *et al.* (2023), the genetic pattern may vary as a function of demographic, phenological or evolutionary history parameters, independently of the level of specialization. By co-sampling 18 fish species in the Maroni river (Guiana), they found that several drivers such as the heterogeneity in population effective size acted differently on genetic patterns among species, and that the dispersal capacity and reproductive guilds were not related to the patterns of population structure.

Finally, considering two spatial scales allowed us to highlight variations in the effect of landscape resistance across scales. While at the larger scale, the landscape appears to limit gene flow for the 5 species, at the finer intra-island scale it only affects three species. It also refines our understanding of species clustering for cross-prediction. Indeed, the stability of species groups (specialist versus generalist) at both spatial scales supports this finding. On the contrary, the slight differences between the two spatial scales, especially regarding the overlapping signal between specialist and generalist species, provide insight into the drivers of genetic structuring. Palm *et al.* (2023) explored the influence of landscape features on the genetic structure of grizzly bears across a gradient of spatial scale. They found that, while geographic distance was the best driver of genetic differentiation at all spatial scales, the effect of landscape resistance varied among scales, canopy cover being only important at the broadest ones. Radford *et al.* (2021) also evidenced some discrepancies between local-scale and regional-scale genetic connectivity patterns, focusing on 10 bird species in Australia. Both studies explained these variations according to the process at play at the different scales, such as foraging and mating patterns at small scales versus stochastic juvenile long-distance dispersal at larger scales. Short Bull *et al.* (2011) also stressed that these variations could result from modelling issues, and in particular, the varying statistical power provided by analyses depending on the scale considered. Importantly, a landscape feature has to be highly variable to be supported in a landscape genetics model, which depends on the composition of the landscape at a given scale.

#### 4.4.5.3. The conservation ecology value of optimising resistance zones using genetic data

Promising cross-validation approaches are emerging in landscape genetic studies (Daniel *et al.*, 2023; Palm *et al.*, 2023; Savary *et al.*, 2022b; van Strien *et al.*, 2014). However, to our knowledge, this is the first study addressing a multispecies connectivity issue by performing cross-validations of landscape genetics models across species. Due to the cost of collecting genetic data from many species, studies often rely on habitat suitability models to perform cluster and spatial intersection analysis rather than model cross-validation (Cushman and Landguth, 2012; Penjor *et al.*, 2024). However, Savary *et al.* (2022) showed that the spatial overlap of least-cost path corridors may depend more on resistance model parameters than on actual ecological processes if the cost-scenarios are optimized on the basis of their resulting cost-distances only, without ways to assess whether the spatial trajectory of the least-cost paths

fits the data as well. Therefore, the use of an objective criterion independent of spatial projection (such as a  $vR^2$ ) may be more reliable than overlap analysis. However, the different number or locations of sampled populations across species may be a limitation of our cross-validation approach, as differences in statistical power may act as a confounding factor when comparing predictive quality between species. Here we had 12 fewer populations for the MG than for the SR. Despite this, the MG showed good levels of cross prediction, sometimes outperforming the SR, which seems to rule out this problem.

We shed light on a new application of the RGA optimisation workflow in a multi-specific context. Indeed, some previous studies have used RGA to compare IBR patterns among species, focusing on the comparison of cost values (Gutiérrez-Rodríguez *et al.* 2017; Mulvaney *et al.* 2021). This allowed them to assess the stability of landscape feature resistance to gene flow across a range of species. However, Daniel *et al.* (in prep) showed that the interpretation of such cost values is uncertain and often biased due to overfitting issues. Here we proposed instead to compare the contrasts of values in the optimised cost scenarios, the cost-distance model relative support as compared to the Euclidean model, and the model predictive power. The cost contrasts and model supports provided information about the intensity of the IBR pattern, while the predictive power reflected the quality of the model and its extrapolation performance. A strength of the RGA approach is its ability to provide optimised cost scenarios with good predictive power for species with a weak IBR signal, such as generalist species. Yet, users might lose the more mechanistic understanding that would provide a more reliable identification of landscape elements resistant to gene flow. However, they gain the ability to predict genetic differentiation, even across several species. This can be particularly useful to reduce the number of species to sample when sampling is expensive and complicated. One species could then be used as a surrogate for another. Finally, integrating the optimised cost scenario into a graph-based landscape connectivity model can help identify dispersal corridors and facilitate multi-species linkages planning.

#### 4.4.5.4. Landscape graph derived from optimized cost scenarios: implications for conservation

Landscape graphs provide information on the centrality of habitat patches in dispersal networks, highlighting conservation priorities. Weighting the links of these graphs with optimised cost distances may improve their biological relevance, as we show here that these distances can be reliable predictors of genetic distances. It could thus be interesting to test this hypothesis with two independent genetic datasets, one for cost optimisation and subsequent graph construction, and the to assess the predictive performance of the latter optimised cost distances. This is particularly promising given that landscape graphs are emerging as a common approach for multi-species connectivity models. For example, Santini *et al.* (2016) proposed a method for combining multiple single-species landscape graphs into a unique graph. The latter pinpoints conservation and restoration priorities for a species assemblage. In the same vein, Clauzel and Godet (2020) combined single-species graph results with PCA analyses to identify patches associated with high connectivity values for multiple amphibian species. They succeeded in highlighting restoration areas that were simultaneously strategic for 5 pond-dwelling species showing different ecological requirements and dispersal abilities. Here, we followed their approach and found that the pattern of well-connected habitat patches supporting the dispersal network was highly congruent across the fivebird species studied. Focusing conservation efforts on the dense forest cover of BT, the hook-shaped forest corridor across the isthmus, the swamp forests north of GT, and the small fragmented patches to the south may benefit all species. Some connectivity patterns were also species-specific and reflected the level of specialisation for forest habitats. We found that patches nearby yet not connected to the largest forest central patches in BT had high connectivity values for generalist species, but not for specialist species. This probably reflects the ability of generalist species to disperse more easily across the landscape matrix than specialist species, and thus to reach more distant patches. Restoration measures on these disconnected patches in BT must therefore focus on a specialist ecological profile. These results show that preserving species that are good indicators of genetic connectivity for other species translated into the same prioritisation of habitat patches as for preserving those other species. Therefore, the use of optimised cost scenarios from genetic data provides insight into the dispersal and habitat patterns of species with similar specialisation profiles.

## **4.5. Conclusions**

Améliorer notre compréhension de la réponse d'une communauté d'espèces aux changements paysagers est primordial à l'heure où les enjeux de conservation se placent à l'échelle des écosystèmes. Pour maximiser l'efficacité des mesures de restauration des connectivités en l'étendant à un spectre d'espèces le plus large possible, un gestionnaire doit savoir quelles données collecter en priorité. On peut supposer que les espèces les plus spécialistes sont les plus sensibles à l'altération de leur habitat et que, par conséquent, les mesures de conservation les concernant bénéficient aux autres espèces de même spécialisation, mais aussi à celles présentant des besoins moins étroits. Cependant, cette hypothèse ne semble pas entièrement soutenue par nos résultats. Nous avons en effet montré que plus les espèces ont un patron spatial de déplacement et des habitats similaires, plus leur perception du paysage et l'effet de son hétérogénéité sur le flux génétique seront également similaires. Cela pourrait suggérer que les mesures de conservation pour une espèce spécialiste seraient susceptibles de bénéficier principalement aux espèces présentant les mêmes préférences d'habitat. Nous pourrions également aller plus loin en considérant un panel d'espèces bien plus large et diversifié que celui considéré ici. Dans une zone géographique hétérogène du point de vue des habitats qui la caractérise, ces différentes espèces pourraient présenter une spécialisation relative à divers habitats, complexifiant la comparaison des zones paysagères à enjeux pour l'ensemble de ces espèces. Nous pourrions alors nous inspirer des travaux de Bellil (2023) basés sur les graphes paysagers, en interpolant à l'ensemble du paysage des métriques de connectivité associées aux chemins de dispersion (e.g. le flux du lien) propre à chaque espèce. La superposition de ces cartes modélisant les connectivités locales pour chaque espèce identifierait les zones associées à un enjeu de connectivité pour un maximum d'espèce. Les graphes paysagers constituent en effet un appui analytique puissant pour la mise en place de mesures de conservation à l'échelle des communautés ou/et des écosystèmes.

L'intérêt de la combinaison des graphes paysagers avec des données génétiques, que ce soit en amont pour leur construction ou a posteriori pour leur validation ou l'amélioration des connaissances sur l'effet des connectivités reste encore à approfondir. Comme évoqué au chapitre précédent, il serait intéressant d'évaluer le gain apporté par un scénario de coût issu de données génétiques pour la pondération des liens d'un graphe. On pourrait également

évaluer la capacité des graphes paysagers à refléter l'effet de la connectivité sur les flux génétiques pour un cortège d'espèces différant par leur niveau de spécialisation. Enfin, il serait intéressant d'étudier si les surfaces de connectivité, par exemple issues d'un graphe paramétré avec des données génétiques, permettraient de prédire l'abondance d'une espèce. Cette étude pourrait se décliner sur un volet multi-espèces, en évaluant si des surfaces de connectivité combinées seraient susceptibles de prédire la diversité spécifique d'un site.

Ces modèles évaluant l'effet de l'hétérogénéité du paysage sur la structure génétique reposent sur plusieurs hypothèses couramment admises. Pourtant, comme discuté dans l'article associé à ce chapitre à propos de la résistance exercée par les zones urbaines de l'isthme ce chapitre à propos de la résistance exercée par les zones urbaines de l'isthme, une hypothèse non vérifiée pourrait rendre caduque l'interprétation de ces modèles de génétique du paysage. Nous nous attarderons donc dans le prochain chapitre à l'exploration de ces postulats implicites aux modèles de génétique du paysage.







# CHAPITRE 5

## DOMAINE DE VALIDITÉ TEMPORELLE D'UN MODÈLE DE CONNECTIVITÉ EN GÉNÉTIQUE DU PAYSAGE

---

<b>5. DOMAINE DE VALIDITÉ TEMPORELLE D'UN MODÈLE DE CONNECTIVITÉ EN GÉNÉTIQUE DU PAYSAGE.....</b>	<b>145</b>
5.1. Influence des effets phylogéographiques sur la distribution contemporaine de la variabilité génétique .....	147
5.1.1. Origine historique et contemporaine de la structure génétique.....	147
5.1.2. Génétique du paysage et phylogéographie.....	148
5.1.3. Histoire de colonisation des Petites Antilles par les passereaux.....	149
5.1.4. Les outils de la phylogéographie : modèles de coalescence et approche 'ABC'.....	152
5.1.5. Inférer l'histoire évolutive des populations de Paruline caféïette en Guadeloupe.....	158
5.1.5.1. Paramètres populationnels.....	158
5.1.5.2. Echantillonnage, extraction d'ADN et séquençage RAD-seq.....	158
5.1.5.3. Travail en cours.....	160
5.2. Équilibre temporel de l'interaction entre variabilité génétique et variabilité paysagère .....	162
5.2.1. Introduction.....	162
5.2.2. Matériels et méthodes .....	163
5.2.3. Résultats principaux.....	163
5.2.4. Article: Disentangling past landscape legacies from the contemporary genetic structure of a forest bird using a historical map.....	164
5.3. Conclusions .....	185

Les modèles de génétique du paysage reposent sur l'hypothèse que la structure génétique échantillonnée est majoritairement déterminée par des processus micro-évolutifs (i.e., agissant à faible échelle spatiale et temporelle), tels que la dérive et le flux génétique lorsqu'on s'intéresse à la diversité neutre (Manel *et al.*, 2003). Or d'autres processus, notamment à l'échelle macro-évolutive, sont susceptibles d'influencer la structure génétique. Une divergence génétique entre deux groupes de populations pourrait aussi bien résulter d'une restriction récente des flux génétiques que d'une différence d'histoire évolutive mobilisant des horizons temporels beaucoup plus. Identifier la source principale de la structuration génétique permettrait d'évaluer si le modèle se positionne dans le cadre de la génétique du paysage, ou au contraire s'il relève davantage de la phylogénétique par exemple. De plus, au sein même du cadre spatio-temporel propre aux processus micro-évolutifs, une autre hypothèse sous-tend communément les modèles de génétique du paysage. En effet, il est classique d'étudier les facteurs de la structure génétique contemporaine à l'aide de variables mesurées dans le paysage contemporain. Cela fait le postulat implicite d'une réponse immédiate de la variabilité génétique à une modification du paysage (Bolliger *et al.*, 2014; Manel and Holderegger, 2013) ou d'une absence modification majeure du paysage dans un passé récent. Cependant, il a été théoriquement démontré qu'un intervalle de temps est nécessaire à la variabilité génétique pour atteindre un nouvel état d'équilibre suite à une perturbation (Varvio *et al.*, 1986). Les distances génétiques sont ainsi parfois susceptibles de refléter des patrons historiques de flux génétiques. Prendre en compte ce délai de réponse dans les modèles de génétique du paysage pourrait permettre d'aboutir à des inférences plus fiables, et de se soustraire à l'hypothèse simplificatrice de la détectabilité immédiate d'un changement paysager.

Dans ce chapitre, nous explorerons ces problématiques d'échelle spatio-temporelle des études de génétique du paysage, en nous basant sur le modèle de la Paruline caféïette en Guadeloupe. Dans un premier temps, nous présenterons le travail préliminaire et actuellement en cours visant à inférer l'histoire évolutive de ce passereau. L'objectif est d'évaluer si la structure génétique contemporaine de la Paruline caféïette résulte davantage d'un effet du paysage contemporain ou de l'existence de deux lignées évolutives distinctes. Dans un second temps, nous évaluerons si cette structure génétique contemporaine est encore marquée par l'influence du paysage historique de la Guadeloupe, tel que caractérisé en 1950, ou si elle porte déjà la signature des changements paysagers récents.

## 5.1. Influence des effets phylogéographiques sur la distribution contemporaine de la variabilité génétique

### 5.1.1. Origine historique et contemporaine de la structure génétique

La divergence génétique neutre contemporaine entre populations résulte de contraintes contemporaines et historiques exercées sur les processus évolutifs de la dérive et du flux de gènes. L’empreinte des contraintes historiques est susceptible de devenir moins perceptible avec le temps, s’effaçant derrière celles des contraintes contemporaines. Parfois au contraire, l’empreinte de ces contraintes historiques reste dominante par rapport à celles contemporaines, rendant ainsi la structure génétique contemporaine héritière d’évènements passés. Le patron de divergence génétique entre les populations de Basse-Terre (BT) et de Grande-Terre (GT) observé de façon commune pour 5 espèces de passereaux en Guadeloupe (chapitre IV) pourrait s’interpréter de deux façons. Tout d’abord, cette divergence pourrait résulter d’une restriction contemporaine des flux de gènes entre BT et GT suite à l’urbanisation croissante de l’isthme les connectant. Elle témoignerait alors d’un processus contemporain de divergence *in situ*, lié à un isolement des populations par résistance du paysage à une échelle spatiale réduite (soit l’île de la Guadeloupe, et plus particulièrement son isthme). D’autre part, cette divergence génétique pourrait également être le vestige d’un isolement passé de deux lignées ayant chacune colonisé BT ou GT de façon indépendante. Cette divergence témoignerait alors d’un contact secondaire entre deux lignées évolutives, et résulterait d’un processus plus large, ayant eu lieu à l’échelle de l’archipel des Petites Antilles il y a plusieurs millions d’années. Latta and Mitton (1997) ont été confrontés à la dualité historique *versus* contemporaine de l’origine d’un patron génétique, lors de leur étude des populations de pins flexibles le long d’un gradient d’altitude dans les montagnes Rocheuses du Colorado. Ils ont démontré que la divergence génétique entre les populations résultait davantage d’un contact secondaire entre deux lignées distinctes, plutôt que d’un effet du gradient paysager désynchronisant la phénologie des populations. Toutefois, si les facteurs historiques n’avaient pas été explorés, la coïncidence spatiale de ce gradient avec la zone de contact secondaire aurait facilement pu conduire à une corrélation fallacieuse entre structure génétique et paysage contemporain. Aussi, pour éviter ces écueils et améliorer la compréhension des patrons de variation génétique,

Dyer *et al.* (2010) ont appelé à contrôler les relations entre variations génétiques et paysages contemporains par des facteurs phylogéographiques. Ils ont en effet montré que l'utilisation d'une variable réponse représentant la différenciation génétique conditionnée par l'appartenance des populations à des lignées évolutives distinctes permettait d'améliorer la précision et la fiabilité des modèles de résistance du paysage aux flux de gènes.

Ainsi, inférer l'histoire évolutive des populations de passereaux en Guadeloupe nous permettrait d'améliorer notre compréhension de l'effet du paysage contemporain sur les flux génétiques, en considérant comme modèle nul la part de variance génétique expliquée par des facteurs historiques. Dans cette partie, nous nous attacherons à évaluer la vraisemblance des hypothèses de divergence *in situ* ou de contact secondaire entre les populations de Paruline caféïette de BT et GT sur l'île de la Guadeloupe.

### 5.1.2. Génétique du paysage et phylogéographie

L'étude des sources contemporaines et historiques de variation génétique fait appel à deux champs disciplinaires distincts mais complémentaires : la génétique du paysage et la phylogéographie. Ces deux disciplines partagent l'objectif commun de l'étude de la distribution spatiale de la variation génétique, mais se positionnent chacune à des échelles spatio-temporelles distinctes. En effet, la phylogéographie évalue les processus historiques à l'origine de la distribution des lignées évolutives à large échelle spatiale, tandis que la génétique du paysage évalue les processus contemporains affectant les patrons de variation génétique à fine échelle spatiale. La phylogéographie se positionne donc à l'échelle du million d'années là où la génétique du paysage considère l'échelle du siècle (Manel *et al.*, 2003; Rissler, 2016; Wang, 2010).

Ces objectifs et cadres différents se traduisent également par l'utilisation de marqueurs génétiques et de méthodes analytiques différentes. Les marqueurs microsatellites permettent de capturer des processus micro-évolutifs très récents et sont principalement utilisés en génétique du paysage, tandis que les marqueurs mitochondriaux ou chloroplastiques sont particulièrement adaptés à la détection de signaux historiques. La différence de profondeur temporelle entre ces marqueurs résulte du taux de mutation  $10^4$  fois plus important de l'ADN microsatellite par rapport à l'ADN des organites. Enfin, alors que la génétique de paysage

s'appuie sur des méthodes d'étude des fréquences alléliques tel que les tests d'assignation, la phylogéographie fait appel aux modèles d'arbres phylogénétiques ou de coalescence spatiale (Rissler, 2016; Wang, 2010).

Ces deux disciplines sont toutefois complémentaires. Comme rapporté par Manel *et al.* (2003), le biogéographe De Candolle (1778–1841) avait attribué les variations spatiales des patrons de diversité à des causes physiques présentement en action, ainsi qu'à des forces n'étant plus en présence. Cette hypothèse précurseur de la dualité historique et contemporaine des sources de variation génétique illustre le potentiel d'une utilisation conjointe de la phylogéographie et de la génétique du paysage. Ainsi, la mobilisation simultanée d'approches de partitionnement génétique, de modèles de niches, d'arbres phylogénétiques et de modèles de coalescence a permis à Pease *et al.* (2009) de démontrer que l'isolement génétique de populations de cerfs en Californie était un phénomène récent induit par les modifications du paysage contemporain, et non liée à des barrières historiques entre populations à l'époque du dernier maximum glaciaire. Ce résultat était inattendu, compte tenu du caractère généraliste et de la grande capacité de dispersion de l'espèce ciblée. Notre objectif est donc d'approfondir nos connaissances sur l'histoire de colonisation de la Paruline caféïette en Guadeloupe, pour améliorer la robustesse des inférences sur les moteurs de sa structuration génétique contemporaine.

### 5.1.3. Histoire de colonisation des Petites Antilles par les passereaux

Robert E. Ricklefs et Eldredge Bermingham ont fortement contribué à la compréhension des processus de colonisation, de spéciation et d'extinction à l'origine de la diversité actuelle de la faune des Antilles, considérant cet archipel comme « un laboratoire de biogéographie et d'évolution » (Ricklefs and Bermingham, 2008). En particulier, leur étude sur la colonisation des Petites Antilles par la faune aviaire (Ricklefs, 2010) pose un cadre solide et les fondements nécessaires à l'évaluation plus fine de l'histoire démographique de la Paruline caféïette.

L'archipel des Petites Antilles s'est formé il y a 20 millions d'années, le long de l'arc volcanique induit par la subduction des plaques Nord et Sud Atlantique sous la plaque Caribéenne. Cet archipel n'a jamais été relié au continent par voie terrestre, même lors des forts abaissements du niveau des océans lors des épisodes glaciaires du Pléistocène (Ricklefs

and Bermingham, 2008). Contrairement aux archipels des Galápagos ou de Hawaï caractérisés par un isolement extrême, les Petites Antilles sont suffisamment éloignées du continent pour qu'y émergent des endémismes, mais toutefois assez proches pour maintenir avec celui-ci une interaction dynamique. Ces échanges réguliers avec le continent ont par ailleurs limité l'émergence de radiations adaptatives dans l'avifaune (Ricklefs and Bermingham, 2007).

La colonisation des Petites Antilles par la faune aviaire a débuté il y a 10 millions d'années. Les fortes contraintes de dispersion, liées à la traversée de vastes étendues d'eau, ont résulté en une richesse spécifique aviaire moins importante que celle du continent (38 espèces de passereaux recensées en 2010) (Ricklefs, 2010). Deux voies de colonisation de l'archipel ont été identifiées : par le Sud depuis le Venezuela et Trinité-et-Tobago, et par le Nord depuis les Grandes Antilles (Porto Rico en particulier). Les événements de colonisation issus du Nord se sont étendus à l'ensemble de l'arc et jusqu'au Venezuela par le biais d'îles et îlets jouant le rôle de *stepping-stones*. À l'inverse, la voie de colonisation par le sud s'est arrêtée assez tôt dans sa remontée vers les Grandes Antilles. Cette asymétrie dans le succès de colonisation n'est pas expliquée mais plusieurs hypothèses prévalent, comme par exemple l'influence des vents dominants sur les contraintes à la dispersion (Ricklefs and Bermingham, 2008).

Les taux de colonisation et d'extinction n'ont pas été homogènes au cours des dix derniers millions d'années (Ricklefs and Bermingham, 2001). En particulier, une rupture de pente dans la courbe d'accumulation d'espèces en fonction du temps indique une perturbation de l'équilibre extinction-colonisation il y a environ un million d'années, soit durant les épisodes de glaciation majeurs du Pléistocène. Deux hypothèses pourraient expliquer cette perturbation. D'une part, les modifications d'habitats et la baisse du niveau des océans liés au climat froid et sec de cette période seraient susceptibles d'avoir induit une augmentation d'un facteur 10 du taux de colonisation. D'autre part, un événement d'extinction de masse impactant 90% des lignées et lié aux fortes variations climatiques caractéristiques de la période aurait pu conduire à un fort déclin du nombre d'espèce (Ricklefs, 2010; Ricklefs and Bermingham, 2001).

La colonisation de l'archipel a donc été un processus transitoire plutôt que continu, résultant des variations temporelles des capacités de dispersion des individus (Ricklefs, 2010). Ricklefs et Cox (1972) ont introduit la notion de *cycle du taxon* pour décrire les quatre phases successives d'expansion et de contraction des aires de distribution des espèces. La première

phase correspond à l'expansion de l'espèce, alors largement distribuée en populations homogènes sur plusieurs îles. Il s'en suit une phase de différenciation, au cours de laquelle les populations se différencient fortement génétiquement. La troisième phase correspond à la fragmentation de l'aire de distribution, qui s'accroît jusqu'à atteindre l'endémisme, dernière phase où l'espèce n'est présente que sur une seule île. Cette séquence temporelle est souvent accompagnée d'une modification de la niche écologique de l'espèce, avec notamment le passage d'un habitat sec et ouvert lors des premiers stades, vers un habitat humide et forestier dans les derniers stades (R. Ricklefs and Bermingham 2008; R. Ricklefs and Bermingham 2004). Enfin, une espèce est susceptible de retourner à la première phase d'expansion à tout moment du cycle. L'élément déclencheur de la dynamique d'expansion n'est pas clairement identifié. Il ne semble pas lié aux phénomènes climatiques, car des espèces présentant des écologies similaires (i.e., susceptibles de présenter une réponse similaire à une perturbation environnementale) s'observent à différentes phases du cycle. La phase d'expansion pourrait en revanche être liée à des dynamiques hôtes-parasites. En effet, l'apparition d'une résistance de l'hôte par pression de sélection, ou la baisse de virulence du parasite, pourrait entraîner une augmentation de l'effectif des populations et déclencher une phase d'expansion de l'espèce (R. Ricklefs and Cox 1972; R. Ricklefs and Bermingham 2008). Trois groupes d'espèces peuvent ainsi être distingués : les « jeunes » espèces ayant récemment colonisé l'archipel, les espèces endémiques anciennement établies et en phase de contraction de leur aire de distribution, et enfin les espèces endémiques anciennement établies ayant connu ou connaissant une seconde phase d'expansion (R. Ricklefs and Bermingham 2004).

Les espèces endémiques en 4<sup>e</sup> phase du cycle sont supposées présenter des populations âgées, de faible effectif et fragmentées, rendant ces espèces vulnérables à l'extinction. Cependant, aucune extinction de fond (i.e., extinction locale d'une espèce pour des raisons qui lui sont particulières et qui n'influencent pas la dynamique des autres espèces) n'a pu être mise en évidence au cours des 10 millions d'années d'évolution de l'avifaune sur l'archipel des Petites Antilles (R. Ricklefs and Bermingham 2001). Le temps de persistance moyen d'une espèce sur une île est ainsi de 2 millions d'années, et peut durer plus de 5 millions d'années pour les espèces endémiques des grandes îles centrales de la Guadeloupe. Cette particularité pourrait s'expliquer par la nature transitoire du processus de colonisation, permettant aux espèces endémiques de basculer ponctuellement en phase d'expansion et de recoloniser les îles voisines. De plus, aucune île de l'archipel ne présente aujourd'hui une saturation de sa richesse spécifique aviaire. Or le modèle de diversité insulaire de MacArthur



et Wilson stipule une augmentation du nombre d'espèces jusqu'à un seuil d'équilibre des taux d'extinction et de colonisation (MacArthur and Wilson, 1967). L'absence d'extinction de fond, l'accumulation d'espèces âgées et la perturbation remontant à 1 million d'années montrent que les taux de colonisation et d'extinction n'ont pas été homogènes. La diversité de l'avifaune des Petites Antilles n'est pas à l'équilibre et se situe probablement encore sous le seuil décrit par MacArthur et Wilson (R. Ricklefs and Bermingham 2001).

#### 5.1.4. Les outils de la phylogéographie : modèles de coalescence et approche 'ABC'

Suite à cette mise en contexte, nous allons maintenant modéliser de façon plus fine l'histoire de colonisation de la Guadeloupe par la Paruline caféïette. En particulier, nous souhaiterions déterminer si la population actuelle dérive de deux lignées évolutives distinctes, ou d'une seule lignée dont les populations se seraient fragmentées suite à des pressions contemporaines. Pour cela, nous comparerons la vraisemblance de ces deux scénarios évolutifs, modélisés par coalescence.

Le coalescent est un modèle mathématique décrit par Kingman en 1982 permettant d'inférer l'histoire évolutive d'une population à partir de sa structure génétique contemporaine (Kingman, 1982). La particularité de cette approche est qu'elle est rétrospective : elle vise en effet à reconstruire l'histoire généalogique des gènes en remontant dans le temps jusqu'à leur ancêtre commun le plus récent (Encadré 5.1). Le processus par lequel les généalogies de deux gènes se rejoignent à rebours en un ancêtre commun est précisément ce qu'on appelle la coalescence. La théorie de la coalescence considère donc uniquement la généalogie des lignées aboutissant à des descendants représentés par une génération actuelle pour un échantillon de gènes. Elle ne s'intéresse donc pas à la généalogie exhaustive de l'ensemble des populations car elle fait abstraction des lignées perdues (par effet de la dérive par exemple) au cours du temps.

### Encadré 5.1 : Principe de la coalescence

Considérons un échantillon issu d'une population haploïde de taille  $2N$ . Le processus de coalescence décrit la marche aléatoire à rebrousse temps modélisant les passages successifs de  $n$  à  $n-1$  lignées, de  $n-1$  à  $n-2$  lignées, et ainsi jusqu'à la coalescence finale des 2 dernières lignées en un ancêtre commun. Considérons maintenant que cet échantillon présente  $j$  lignages à la génération actuelle. On note  $P(j)$  la probabilité d'un évènement de coalescence entre ces  $j$  lignées à la génération précédente. Sachant que la probabilité que deux gènes partagent un ancêtre commun à la génération précédente vaut  $\frac{1}{2N}$ , et qu'il est possible de former  $\frac{j(j-1)}{2}$  paires de lignées parmi les  $j$  en présence à la génération  $t$ , on a donc :

$$P(j) = \frac{1}{2N} \times \frac{j(j-1)}{2} = \frac{j(j-1)}{4N}$$

On considère alors maintenant le temps de coalescence  $T_j$ , correspondant au nombre de génération écoulées avant qu'il n'y ait un évènement de coalescence induisant le passage de  $j$  à  $j-1$  lignées. La probabilité qu'il se soit écoulé  $t$  générations sans coalescence parmi les  $j$  lignées correspond donc à la probabilité qu'il n'y ait pas eu d'évènements de coalescence pendant  $t-1$  générations et qu'un évènement de coalescence survienne à la génération  $t$ , soit :

$$P(T_j = t) = [1 - P(j)]^{t-1} \times P(j)$$

Le temps de coalescence  $T_j$  suit donc une loi géométrique de paramètre  $P(j)$ .

L'espérance et la variance du temps de coalescence valent alors :

$$E(T_j) = \frac{4N}{j(j-1)}$$

$$V(T_j) = \frac{16N^2}{j^2(j-1)^2}$$

Les temps moyens de coalescence augmentent de façon exponentielle avec la marche dans le passé. Il faut compter en moyenne  $2N$  générations pour le dernier évènement de coalescence permettant de passer des deux dernières lignées à l'ancêtre commun de l'échantillon. Cependant, la variance autour de cette valeur moyenne est très forte. Deux gènes indépendants seront donc susceptibles d'avoir des temps de coalescence très différents.

Il est également possible de prendre en compte les processus mutationnels dans les modèles de coalescence, de façon à améliorer la compréhension des phénomènes évolutifs ayant façonné le polymorphisme génétique. La modélisation d'un processus de coalescence avec mutation consiste à considérer que des mutations se produisent aléatoirement le long de branches de la généalogie, selon un taux défini au préalable par un modèle mutationnel.

Cette théorie permet de modéliser l'évolution de la variabilité génétique selon différents scénarios démo-génétiques. Dans une approche de coalescence inférentielle, un échantillon de gènes d'une génération contemporaine correspond à la réalisation d'un processus stochastique caractérisé par un ensemble  $P$  de paramètres démographiques et génétiques. Inférer le modèle démo-génétique sous-jacent à la structure génétique contemporaine revient à estimer cet ensemble de paramètres. Cette estimation repose sur le théorème de Bayes. Il s'agit en effet de calculer la probabilité *a posteriori* d'un scénario démo-génétique (i.e., de ses paramètres  $P$ ) sachant les données génétiques observées (Figure Encadré 5.2, Figure 5.11). Cette probabilité peut être approximée par la vraisemblance de ce scénario démo-génétique (i.e., probabilité d'observer de telles données génétiques sachant ce scénario) multipliée par la probabilité *a priori* de celui-ci. La probabilité *a priori* du scénario s'appuie sur des connaissances préalables des paramètres démo-génétiques et est donc connue, tandis que la vraisemblance du scénario doit être estimée. Lorsque le scénario évolutif est simple et que le jeu de données génétique est réduit, cette estimation peut être réalisée par de multiples itérations, en appliquant la méthode de Monte Carlo par chaîne de Markov (MCMC). En revanche, un scénario évolutif complexe associé à un large jeu de données génétiques rend impossible l'estimation exacte de la vraisemblance par l'approche MCMC.

La méthode de l'«*Approximate Bayesian Computation*» ('ABC') permet alors de contourner cet écueil (Encadré 5.2, Figure 5.1.1, 5.1.2). Cette méthode approxime la vraisemblance d'un scénario évolutif par une mesure de similarité entre les données génétiques observées et des données simulées par un modèle de coalescence (Beaumont *et al.* 2002). Les travaux de Tavaré *et al.* (1997), Fu and Li (1997), Weiss and Von Haeseler (1998), Pritchard *et al.* (1999) furent les pionniers de cette approche, qui a été par la suite formalisée et généralisée par Beaumont *et al.* (2002). Le développement de l'approche 'ABC' a permis de s'affranchir des contraintes liées à la formalisation de la fonction de vraisemblance, ouvrant ainsi la voie à la modélisation de scénarios évolutifs complexes plus réalistes et

s'appuyant sur de larges jeux de données génétiques. Depuis sa formalisation en 2002, la méthode 'ABC' a fait la preuve de performances plus satisfaisantes que celles des approches basées sur le calcul exact de la vraisemblance. Elle a notamment été employée pour réaliser des inférences complexes en biologie des populations et d'épidémiologie (Bertorelle *et al.*, 2010; Lopes and Beaumont, 2010).

Cornuet *et al.* (2008) ont rendu cette approche accessible aux biologistes non-spécialistes des statistiques bayésiennes en développant le programme informatique DIYABC. Ce programme permet d'inférer l'histoire évolutive d'une population en estimant les distributions postérieures de paramètres démo-génétiques, en comparant les probabilités postérieures de scénarios évolutifs alternatifs, et en évaluant quantitativement la crédibilité de ces résultats. De nombreuses études ont ainsi utilisé cette méthode pour explorer l'impact d'événements historiques sur la variabilité génétique contemporaine, comme par exemple l'impact des épisodes de glaciation du Quaternaire (Frantz *et al.*, 2014). Bien qu'une majorité de ces études de phylogéographie portent sur des zones tempérées (Europe et Etats-Unis), Garg *et al.* (2020) ont utilisé les modèles de coalescence mis en œuvre par DIYABC pour évaluer l'impact du dernier maximum glaciaire sur les populations de deux espèces d'oiseaux tropicaux en Papouasie-Nouvelle-Guinée. Ils ont démontré que cette période concordait avec des phases d'expansion des populations, et non de contraction comme en milieu tempéré, et soulignent le manque d'études sur les effets des cycles de glaciation en milieu tropical. Par ailleurs, Barbosa *et al.* (2017) ont appelé à considérer l'histoire phylogéographique des espèces endémiques, défendant qu'elle était aussi riche que celles des espèces largement répandues et plus communément étudiées. Enfin, on peut noter que beaucoup d'études infèrent des scénarios démographiques à très large échelle spatiale, comme celle du continent Européen (Frantz *et al.*, 2014; Senczuk *et al.*, 2021) ou de l'océan indien (Takeuchi *et al.*, 2020), et que peu se positionnent à l'échelle de la colonisation d'un territoire restreint comme une île (mais voir Cornuet *et al.* 2008).

Le programme DIYABC étant couramment utilisé pour inférer des scénarios évolutifs, la validité de ses inférences a fait l'objet d'évaluations. Cabrera et Palsbøll (2017) ont étudié la capacité du programme à retrouver le scénario démo-génétique à l'origine d'une structuration génétique parmi plusieurs scénarios alternatifs, à partir de 520 jeux de données génétiques simulés. Ils ont montré que le taux d'identification du scénario correct s'élevait à 49%. Ce taux de réussite s'améliorait pour l'inférence de scénarios simples (60%) confrontés à peu d'alternatives, mais diminuait pour des scénarios complexes (i.e., comptant plus de 3

variations démographiques, taux de réussite = 30%) ou confrontés à un grand nombre de scénarios alternatifs. Ces auteurs recommandent donc de considérer moins de 5 scénarios candidats, très dissimilaires entre eux et ne modélisant que les évènements démographiques majeurs. Pelletier and Carstens (2014) ont souligné l'importance que revêt l'étape du choix de ces scénarios candidats, montrant qu'un choix subjectif ne conduit pas toujours à l'identification du scénario le plus probable, et encouragent donc une exploration plus systématique des scénarios candidats.

### **Encadré 5.2 : Principes de l'ABC**

La méthode du calcul bayésien approché (ABC) consiste à remplacer le calcul de la vraisemblance d'un scénario démo-génétique par le calcul d'une distance entre les données observées et des données simulées à partir de ce scénario. Cette méthode permet de rendre le calcul de la probabilité postérieure d'un scénario démo-génétique indépendant de la formalisation de la fonction de vraisemblance, et donc réalisable pour des scénarios très complexes.

Cette méthode ABC repose sur plusieurs étapes. Tout d'abord, des données génétiques sont modélisées par coalescence selon le scénario démo-génétique dont on souhaite évaluer la probabilité postérieure. Les jeux de données génétiques observés et simulés sont ensuite résumés par des statistiques (*e.g.*, nombre moyen d'allèles, taux d'admixture, diversité génétique moyenne). Puis, la similarité entre ces jeux de données est évaluée par une mesure de distances dans l'espace déterminé par ces statistiques résumées. Il s'agit alors de remplacer l'estimation de la vraisemblance du scénario démo-génétique par le calcul de la probabilité que les données simulées et observées soient suffisamment proches selon ce scénario :

$$P(\text{datasimulées} - \text{dataobservées} \leq \epsilon / \text{Scenario})$$

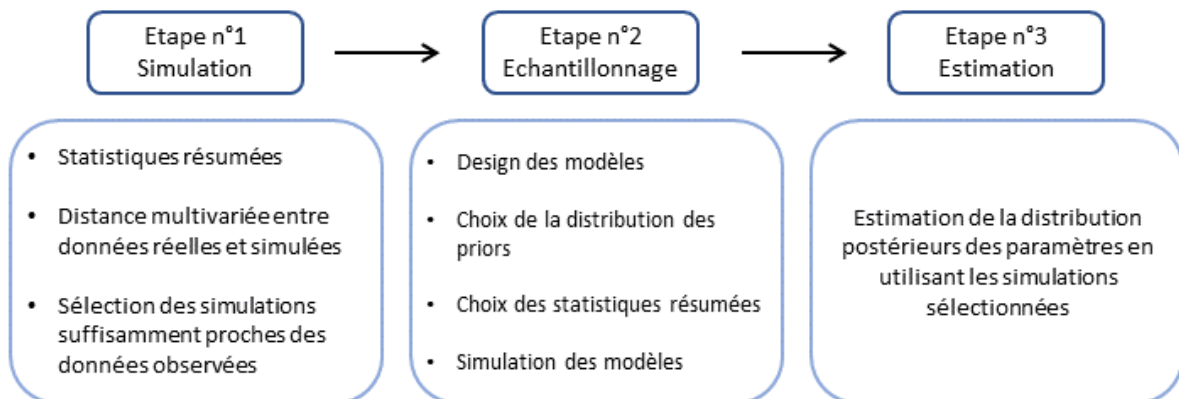
Cette approximation permet enfin de calculer la distribution de la probabilité postérieure du scénario démo-génétique, et d'estimer à partir de cette distribution les paramètres démographiques et génétiques de ce scénario (Figure 5.1.1).

$$P(\text{Scenario}/\text{data}) = \frac{P(\text{Scenario}) \times P(\text{data}/\text{Scenario})}{P(\text{data})} \sim P(\text{Scenario}) \times P(\text{data}/\text{Scenario})$$

**Prior**  
Connu
**Vraisemblance**  
• A estimer : MCMC  
• A approximer : **ABC**

**Posterior**  
Recherché
Constante

**FIGURE 5.1.1:** Formule du théorème de Bayes pour le calcul de la probabilité postérieure d'un scénario évolutif (Scenario) conditionné aux données génétiques observées (data). La probabilité *a priori* d'observer de telles données génétiques ( $P(\text{data})$ ) est considéré comme une constante de normalisation impossible à calculer et qu'il est possible de négliger pour obtenir une approximation de la probabilité postérieure du scénario évolutif. MCMC : Monte Carlo par Chaîne de Markov. ABC : Approximate Bayesian Computation.



**FIGURE 5.1.2 :** Principes de l'ABC résumé en 3 étapes, d'après Lopes and Beaumont (2010) et Bertorelle, Benazzo, and Mona (2010).

### 5.1.5. Inférer l'histoire évolutive des populations de Paruline caféïette en Guadeloupe

Notre objectif est d'évaluer la vraisemblance des hypothèses de divergence *in situ* ou de contact secondaire entre les populations de BT et GT de la Paruline caféïette (PC). Nous modéliserons pour cela ces deux scénarios démo-génétiques par coalescence à partir de la structure génétique contemporaine caractérisée par des marqueurs SNP. Nous utiliserons l'approche ABC mis en œuvre sur DIYABC (Cornuet *et al.*, 2008) pour comparer ces modèles de coalescence, et inférer le scénario évolutif le plus probable.

#### 5.1.5.1. Paramètres populationnels

La PC a colonisé les Petites Antilles il y a 4,5 millions d'années (Ricklefs and Bermingham, 2001) et appartient au groupe des anciennes espèces endémiques de l'archipel. Son aire de distribution est restreinte à deux îles adjacentes au centre de l'arc volcanique : la Guadeloupe et la Dominique. Khimoun *et al.* (2016) ont caractérisé le patron de structuration génétique de 8 espèces de passereaux dont la PC, entre la Guadeloupe et la Dominique. Ils ont mis en évidence une forte différenciation génétique des populations de PC entre la BT et la GT, ainsi qu'entre ces deux sous-ensembles de la Guadeloupe et Dominique ( $F_{ST}$  global de 0.063). La différenciation génétique des populations de BT et GT était toutefois plus faible entre elles qu'avec celles de Dominique ( $F_{ST}$  BT-GT = 0.026 ;  $F_{ST}$  BT-Dominique = 0.095 ;  $F_{ST}$  GT-Dominique = 0.107). La taille actuelle de la population de Paruline caféïette en Guadeloupe s'élèverait autour de 100 000 individus, dont 25% dans le Parc National sur BT (Eraud *et al.*, 2012).

#### 5.1.5.2. Echantillonnage, extraction d'ADN et séquençage RAD-seq

Les inférences phylogéographiques reposent souvent sur l'utilisation de marqueurs nucléaires tels que les microsatellites. Cependant, le faible nombre de marqueurs disponibles pour une espèce limite la puissance statistique pour la résolution des problématiques phylogéographiques complexes. Etant bi-alléliques, les SNP présentent un polymorphisme

plus restreint que les microsatellites et portent donc moins d'information. Cependant, la moindre puissance statistique conférée par le manque de résolution de chaque marqueur est largement compensée par le nombre de marqueurs disponibles pour chaque individu. En effet, le séquençage RAD-seq permet l'identification d'un millier de marqueurs SNP par organisme, distribués de manière plus dense et uniforme dans le génome que les marqueurs microsatellites. De plus, les taux d'erreur lors du génotypage sont plus faibles pour les SNP que pour les microsatellites, les SNP étant moins sensibles au risque d'homoplasie. Ainsi, Jeffries *et al.* (2016) ont montré que le patron d'IBD était plus fort, les clusters génétiques plus distincts et les scénarios évolutifs mieux identifiés en utilisant des SNP plutôt que des microsatellites. Cette méthode RAD-seq étant toutefois coûteuse, un compromis est à trouver entre le nombre de populations et d'individus à génotyper. Nous avons donc choisi de caractériser la structure génétique contemporaine de la Paruline caféïette en génotypant les marqueurs SNP de 51 individus issus de 21 populations de Guadeloupe et Martinique.

712 individus de Paruline caféïette ont été échantillonnés dans 27 taches d'habitat forestier en Guadeloupe et 11 individus ont été échantillonnés en Dominique. Des prélèvements sanguins ont été réalisés (cf. Ch. 2.4,5). Parmi les 712 individus de Guadeloupe, 40 ont été sélectionnés dans 20 populations (2 individus par population, 10 populations sur GT et BT respectivement) représentatives des différents clusters génétiques identifiés dans de précédentes études. Ces 40 individus de Guadeloupe associés aux 11 individus de Dominique constituent le jeu de données de cette analyse (n=51).

L'ADN de ces 51 échantillons sanguins a été extrait en utilisant les kits commerciaux Blood Genomic (BIO BASIC Inc., Markham, Canada) puis dosé par Qubit HS-dsDNA. La construction des banques de RAD-seq a été réalisée selon le protocole de Baird *et al.* (2008), en utilisant l'enzyme de restriction PstI et en réalisant la sélection de taille sur Pippin HT. La construction des banques s'est achevée par 12 cycles d'amplification PCR. La qualité des librairies a été validée par quantification d'ADN sur Fragment Analyser et par qPCR. Le séquençage a été réalisé sur un NovaSeq 6000 d'Illumina grâce à la technique SBS (Sequence By Synthesis) à l'aide du kit NovaSeq Reagent Kits. Les pools ont été séquencés sur une piste d'une flow cell S4 en paired-end 150 nt.

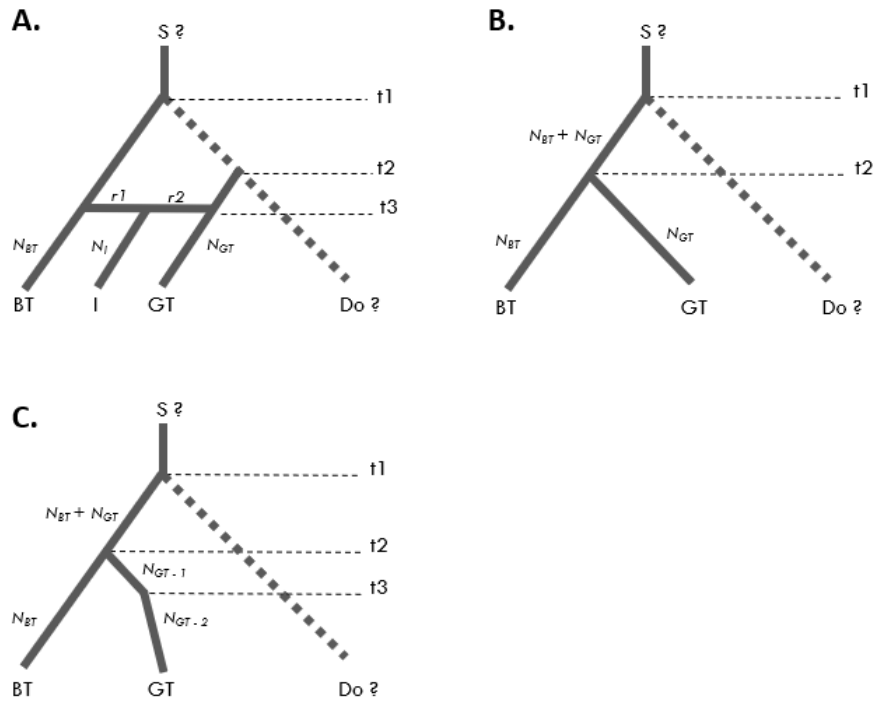


### 5.1.5.3. Travail en cours

Nous disposons actuellement des données brutes de séquençage. Ces données doivent encore être filtrées et démultiplexés en utilisant le pipeline STACKS.

Nous souhaitons ensuite mettre en oeuvre l'approche ABC à l'aide du logiciel DIYABC pour inférer l'histoire évolutive de la Paruline caféïette en Guadeloupe à partir de la structure génétique décrite par les marqueurs SNP. Plusieurs scénarios démographiques seront comparés. Il s'agirait en premier lieu de déterminer si la colonisation de la Guadeloupe viendrait de la Dominique ou d'une source aujourd'hui éteinte au Nord de l'île. Selon ce scénario, la Dominique aurait alors été colonisée par les populations de Guadeloupe, selon un schéma d'expansion en *stepping stone*. Ce contexte posé, nous modéliserons plus finement l'histoire de colonisation à l'échelle de la Guadeloupe, pour déterminer si les populations de BT et GT résultent de deux événements de colonisation distincts ou d'une divergence *in situ* après une unique phase d'expansion. Nous explorerons donc les scénarios suivants (Figure 2) :

- Deux événements de colonisation indépendants sur BT et GT (depuis la Dominique ou une autre source au Nord), suivis d'un contact secondaire entre les deux populations.
- Un unique événement de colonisation (depuis la Dominique ou une autre source), suivi d'une phase d'expansion à l'échelle de l'île, conduisant à une divergence *in situ* sous l'effet d'un isolement par la distance.
- Un unique événement de colonisation (depuis la Dominique ou une autre source) suivi d'une phase d'expansion, puis d'une phase de contraction sur GT lié à une rupture des flux de gènes au niveau de l'isthme.



**FIGURE 5.3 :** Scénarios évolutifs pressentis pour l’analyse phylogéographique. La probabilité postérieure de ces scénarios sera calculée par méthode ABC avec le logiciel DIYABC. Ces scénarios représentent A) deux évènements de colonisation suivis d’un contact secondaire, B) un unique évènement de colonisation suivi d’une expansion et d’un isolement par la distance entre les populations de BT et GT, C) un unique évènement de colonisation suivi d’une contraction démographique sur GT.  $S ?$  : population source inconnue ;  $t_1$ - $t_2$ - $t_3$  : paramètres temporels datant les évènements démographiques ;  $N_{BT}$ ,  $N_{GT}$ ,  $N_I$  : paramètres démographiques correspondant aux effectifs sur BT, GT et l’isthme respectivement.  $r_1$ ,  $r_2$  : taux d’admixture. BT : Basse-Terre ; GT : Grande-Terre ; I : Isthme ; Do ? : Dominique, potentiellement colonisée par les populations de Guadeloupe dans l’hypothèse d’une colonisation depuis le Nord vers le Sud.

## 5.2. Équilibre temporel de l'interaction entre variabilité génétique et variabilité paysagère

### 5.2.1. Introduction

L'époque actuelle connaît des modifications rapides et intenses des paysages, sous l'effet des pressions anthropiques et de leurs conséquences climatiques. Ces changements sont susceptibles de mettre en péril la biodiversité, par la destruction des habitats, des ressources et des connectivités. Pour pallier les risques d'extinction, une meilleure compréhension des héritages paysagers des patrons de diversité est nécessaire. En particulier, il est admis que les patrons de diversité génétique répondent avec un temps de latence, à une perturbation paysagère, telle que la destruction et fragmentation des habitats. La prédiction des effets du paysage sur la variabilité génétique est un enjeu majeur pour la persistance des populations et doit considérer ce temps de latence, en intégrant l'héritage des paysages historiques sur la diversité génétique contemporaine. La persistance et la détectabilité de ces héritages paysagers dépendent de différents facteurs tels que la taille de la population, le temps de génération de l'espèce ciblée, sa spécialisation pour un habitat ou ses capacités de dispersion. Cependant, peu d'études ont examiné conjointement l'effet de ces facteurs sur le délai de réponse de la structure génétique. Cette lacune découle en partie de la rareté des données paysagères historiques, contraignant leur intégration dans les modèles de génétique du paysage. Dans cette étude, nous avons donc évalué la part d'influence de l'héritage du paysage historique dans la structure génétique contemporaine d'un oiseau forestier de la Guadeloupe, la Paruline caféïette.

La Guadeloupe est demeurée densément boisée jusqu'à l'arrivée des colons européens au XVI<sup>e</sup> siècle. Ces derniers ont ensuite largement déforesté le territoire à des fins agricoles, convertissant en particulier les forêts en cultures de canne à sucre. Plus récemment, à la fin du XX<sup>e</sup> siècle, le paysage guadeloupéen a connu une importante expansion urbaine liée à une forte croissance démographique (passage de 225000 habitants en 1954 à 450000 en 2007 d'après l'INSEE). Nos objectifs étaient donc de déterminer si i) les empreintes des anciennes modifications du paysage, notamment de la fragmentation historique de l'habitat forestier, pouvaient être encore détectées dans la structure génétique contemporaine, ii) si les barrières

paysagères récentes liées à l'explosion urbaine avaient déjà affecté le patron de variabilité génétique.

### 5.2.2. Matériels et méthodes

Pour cela, nous avons cartographié l'occupation du sol de la Guadeloupe en 1950 (image des paysages en place depuis la déforestation à but agricole initiée au XVI<sup>e</sup> siècle) et en 2008 (image des paysages remaniés par le développement urbain de la seconde moitié du XX<sup>e</sup>). Nous avons ensuite paramétré des surfaces de résistance au déplacement des individus pour ces deux périodes temporelles, selon 114 scénarios de coûts<sup>1</sup> construits systématiquement. Nous avons enfin modélisé la différenciation génétique contemporaine entre 27 populations de Paruline caféïette par les distances-coûts issues des surfaces de résistance de 1950 et de 2008 entre ces mêmes populations. A partir de ces modèles, nous avons pu déterminer si la structure génétique contemporaine était impactée par le paysage historique de 1950, par le paysage contemporain de 2008, ou par les deux, et évaluer la magnitude de ces effets.

### 5.2.3. Résultats principaux

Nous avons ainsi montré que la structure génétique de la Paruline caféïette était significativement liée à la structure des paysages de 1950 et de 2008, mais que cette relation était plus forte pour le paysage de 2008. La différenciation génétique contemporaine semble donc être un héritage de la fragmentation et de la destruction passée de l'habitat forestier, fortement exacerbé par l'urbanisation récente du territoire. Ces résultats mettent à la fois en évidence une persistance et une impression rapide des effets de résistance paysagère sur un patron de variabilité génétique. La structure génétique de la Paruline caféïette porte ainsi l'empreinte de modifications du paysage survenues il y a moins de 50 ans. Ces résultats sont alarmants du point de vue de la conservation, puisque le temps de réaction pour pallier une potentielle chute de diversité est donc faible. Toutefois, cette réponse rapide de la structure

---

<sup>1</sup> Un scénario de coût correspond à l'association d'une valeur de coût à un trait paysager, ici une occupation du sol. Ce coût représente le coût physiologique, le frein comportemental et le risque de mortalité encouru par un individu lors de sa traversée de cette occupation du sol.

génétique donne également l'espoir d'une réponse tout aussi rapide à des mesures de restauration de la connectivité de l'habitat.

#### 5.2.4. Article

**Disentangling past landscape legacies from the contemporary genetic structure of  
a forest bird using a historical map**

Daniel, Alexandrine ; Savary, Paul ; Foltête, Jean-Christophe ; Roy, Damien ; Vuidel, Gilles ;  
Faivre, Bruno ; Garnier, Stéphane ; Khimoun, Aurélie.

Article en cours de préparation.

## **Disentangling past landscape legacies from the contemporary genetic structure of a forest bird using a historical map**

Daniel, Alexandrine ; Savary, Paul ; Foltête, Jean-Christophe ; Roy, Damien ; Vuidel, Gilles ; Faivre, Bruno ; Garnier, Stéphane ; Khimoun, Aurélie.

### 5.2.4.1. Abstract

At a time when rapid landscape changes are putting biodiversity at risk, our ability to prevent further extinction relies on a better understanding of landscape legacies. Theoretically, genetic patterns are expected to respond to landscape changes, such as habitat destruction, after a time lag. Given the crucial role of genetic diversity in population persistence, these legacy effects must be considered for predicting accurately landscape effects on genetic variability. Past landscape changes leave an imprint on current genetic patterns whose persistence and detectability depend on several species traits such as population size, generation times, habitat specialization, or dispersal abilities. However, few studies have investigated the combined effect of these potential factors. This research gap stems from the scarcity of historical landscape data, which prevents modeling past landscape constraints on genetic drift and gene flow. We therefore disentangled past landscape legacies from the effects of current landscape drivers on the contemporary genetic structure of a forest bird in Guadeloupe. After major deforestation episodes in the 16th century, the Guadeloupe landscape has recently undergone important urban expansion. Our objectives were to determine whether i) the imprints of past landscape changes could still be detected, ii) the recently-created landscape barriers have already affected the current genetic structure. We show that overall, the genetic structure of the forest bird studied was significantly related to landscape structure, as modeled at both dates (1954 and 2020). Current genetic differentiation patterns therefore seem to be a legacy of past forest destruction and subdivision, while having been exacerbated by recent urbanization. These results are alarming from a conservation perspective as the genetic pattern already carries the imprint of landscape changes that occurred less than 50 years ago. However, this rapid response of genetic structure may also give hope regarding the potential efficiency of habitat connectivity restoration.

#### 5.2.4.2. Introduction

Landscapes are inherently heterogeneous and dynamic, and their variations shape biodiversity patterns at all levels. Over the past millenia, the climatic and landscape variations of the Quaternary have produced waves of expansion and contraction of species ranges, resulting in sequences of connexion and isolation between species and populations (Hewitt, 2004). At shorter time scales, the last century has witnessed a sustained surge in human activities, causing profound climatic and environmental upheaval. At a time when rapid landscape changes are putting biodiversity at risk (Fischer and Lindenmayer, 2007) our ability to prevent further extinction and foster resilience relies on a better understanding of landscape legacies (Foster *et al.*, 2003; Tappeiner *et al.*, 2021).

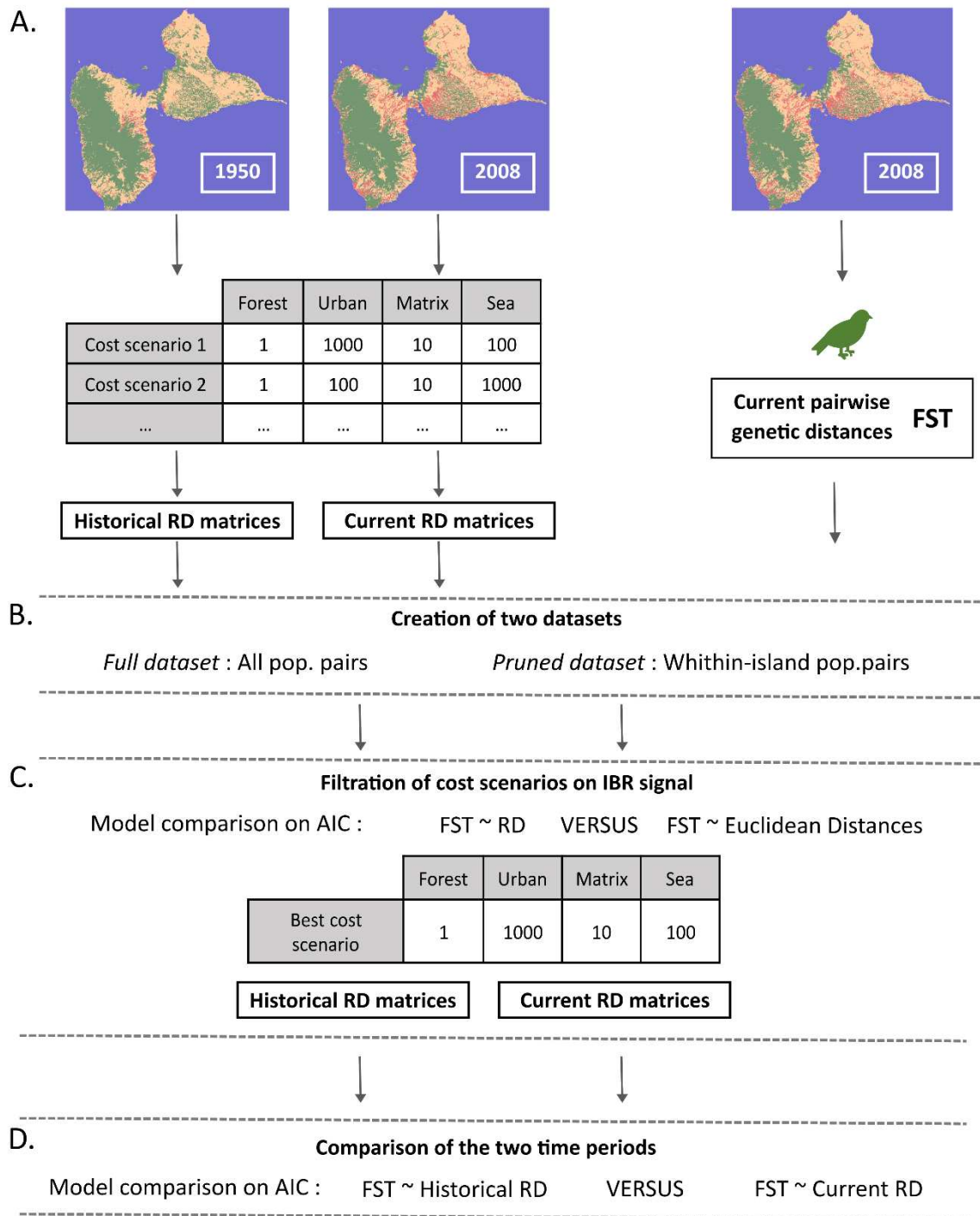
Numerous studies have highlighted the genetic responses to habitat destruction and subdivision, raising alarm about the potentially dramatic consequences for population persistence and global biodiversity (Haddad *et al.*, 2015; Keyghobadi, 2007; Templeton *et al.*, 1990). Conserving genetic diversity is therefore a major challenge, which depends on our ability to understand and predict its response to an increasingly changing environment. The field of landscape genetics aims to study genetic variation in relation to landscape composition and configuration (Manel *et al.* 2003; Storfer *et al.* 2007). Unlike phylogeography, this discipline focuses on processes at fine spatial and temporal scales (Manel *et al.* 2003), and has first relied on 'snapshots' of data assessing population and landscape structure in their contemporary state (Draheim *et al.* 2018; but see Keyghobadi *et al.* 2005; Holzhauer *et al.* 2006; Honnay *et al.* 2006 for precursor studies). It therefore assumes a temporal correspondence between the genetic pattern and the structuring process. However, theoretical studies show delayed responses of genetic patterns to landscape changes, as a time lag is usually observed following a perturbation before genetic diversity adjusts to a new equilibrium (Crow and Aoki, 1984; Latter, 1973; Varvio *et al.*, 1986). This equilibrium may even never be reached if the landscape changes faster than it takes for genetic diversity to reach it (Epps and Keyghobadi, 2015; Holzhauer *et al.*, 2006). Therefore, these time lag effects must be considered when predicting landscape effects on genetic variability and a growing number of reviews have recently called for greater attention to be paid to this issue (Anderson *et al.*, 2010; Bolliger *et al.*, 2014; Epps and Keyghobadi, 2015; Manel and Holderegger, 2013). Incorporating this temporal dimension into landscape genetics studies would provide a deeper understanding of evolutionary mechanisms, as well as guidelines to support conservation policies. Disentangling the effects

of past and present landscapes on current genetic structure could help to determine whether a landscape feature is isolating (Landguth *et al.*, 2010), and whether its effect is not yet detectable. It is also useful to determine where connectivity existed previously, and thus to predict whether restoration is likely to be effective (Epps and Keyghobadi, 2015; Miller *et al.*, 2013; Pavlacky Jr *et al.*, 2009).

Studies that have investigated the time lag in the genetic response to landscape disturbance present very heterogeneous results concerning the duration of this response lag and its potential explanatory factors. Indeed, they showed that past landscapes could leave a persistent imprint on current genetic patterns, with time scales ranging from a decade to a thousand years, depending on several factors such as population size, generation times, habitat specialisation or dispersal ability. (Epps and Keyghobadi, 2015). Moreover, these studies have tackled the problem of time lag in a variety of ways, including studying the correlation between genetic structure and the present or past landscape. (Draheim *et al.*, 2018; Miller *et al.*, 2013; Nowakowski *et al.*, 2015; Pavlacky Jr *et al.*, 2009), the comparison of markers or genetic metrics reflecting genetic processes with different temporalities (Blair *et al.*, 2015; Epps *et al.*, 2013), or the study of the relationship between genetic and landscape distance controlling for phylogeographic history (Dyer *et al.*, 2010; Vandergast *et al.*, 2007). Given this diversity of methods and results, and the lack of consensus, there is a need for more empirical contributions. In particular, the scarcity of historical landscape data prevents testing of whether the current pattern is best explained by past landscape constraints on genetic drift and gene flow (Epps and Keyghobadi, 2015).

We therefore conducted an empirical study to disentangle past landscape legacies from the effects of current landscape features on the contemporary genetic structure of a forest bird (the plumbeous warbler [*Setophaga plumbea*]) in the Guadeloupe archipelago, in which landscape has changed over the last centuries (Figure 5.2.1). Indeed, the island was almost entirely forested until the arrival of European settlers in the 16<sup>th</sup> century. Until the 18<sup>th</sup> century, these settlers intensively deforested the landscape, to the benefit of the agricultural activity, in particular the banana and sugar production (Ricklefs and Bermingham, 2004). More recently, at the end of the 20<sup>th</sup> century, population growth and the development of tourism led to major urban expansion (Eraud 2009). Our objectives were to determine whether i) the imprint of forest cover fragmentation can still be detected in the current genetic variation of a forest bird known to be sensitive to fragmentation (Khimoun *et al.*, 2017, 2016), ii) the recent urbanization has already affected the current genetic structure.



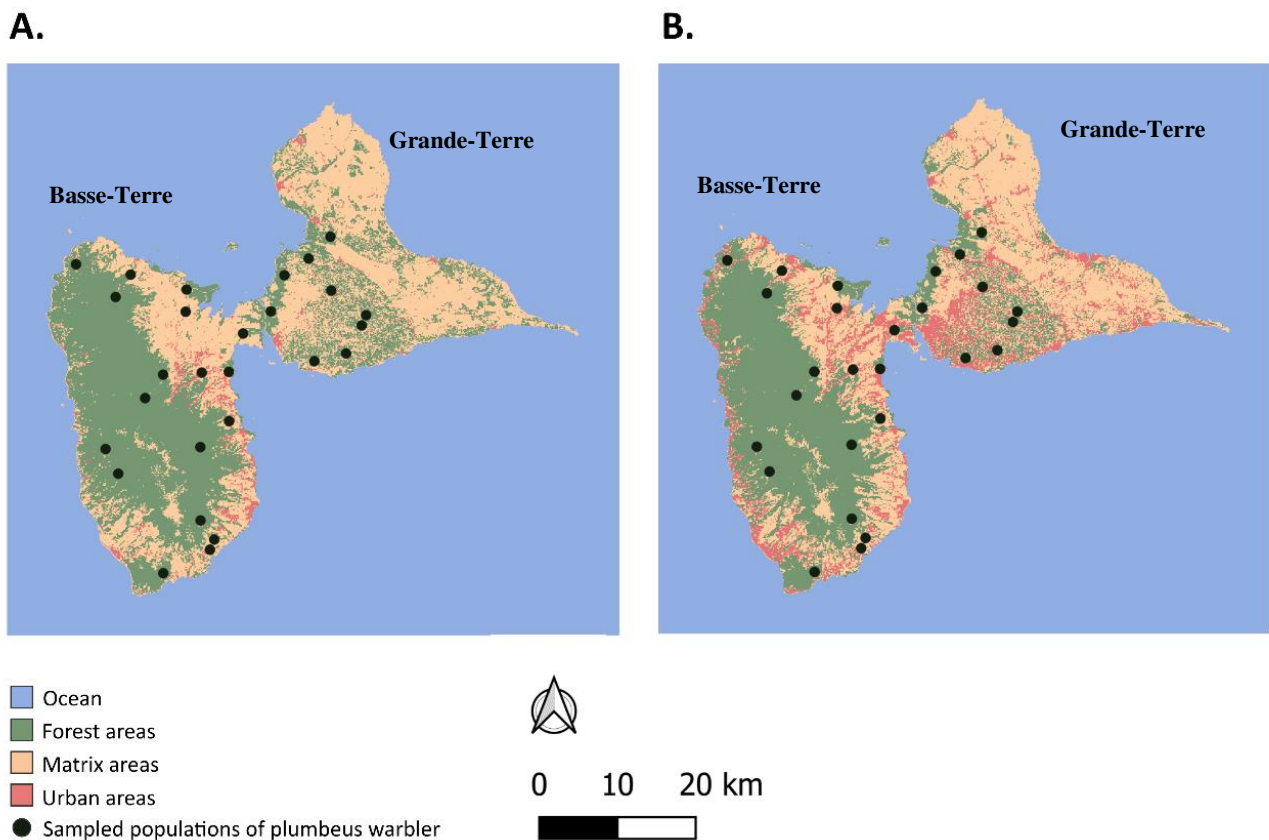


**FIGURE 5.2.1:** General methodology for the diachronic landscape genetics approach. A) Computation of resistance distances on both the current and the historical landscape based on 114 cost scenarios, and calculation of current genetic differentiation between populations. B) Consideration of two spatial scales: the whole Guadeloupe scale and the within-island scale. C) Selection of cost scenarios to retain only those leading to an IBR pattern. D) Comparison of current genetic data with current and historical resistance distances at the two spatial scales.

### 5.2.4.3. Materials and Methods

#### 5.2.4.3.1. Study area

Located in the Lesser Antilles, the Guadeloupe archipelago is composed of two main islands: the Grande-Terre (GT) to the east and Basse-Terre (BT) to the west. It had remained densely forested until the 16th century, when the arrival of European settlers led to extensive deforestation for agricultural purposes (Ricklefs and Bermingham, 2004). BT is still largely forested and is home to the Guadeloupe National Park, while GT is more anthropized with a sparser forest cover. These two main islands are connected by a narrow and urbanised isthmus (Figure 5.2.2).



**FIGURE 5.2.2:** Land cover map of Guadeloupe in A) 1950 and B) 2008. The dark dots correspond to the localisations the 27 populations of plumbeus warbler that were sampled in 2012-2013 and 2020.

#### 5.2.4.3.2. Study model

We focused on the plumbeous warbler (*Setophaga plumbea*), a forest specialist species (Lovette *et al.*, 1998) endemic to Guadeloupe and Dominica. Previous studies have already shown that this species is genetically structured both at the whole Guadeloupe scale and within the BT and GT scales, partly due to landscape resistance (Khimoun *et al.*, 2017, 2016). This makes it a reliable model for a more detailed study of the time lag of genetic response to landscape change.

#### 5.2.4.3.3. Sampling, molecular analysis and genetic structure index

We relied on the dataset previously described in Daniel *et al.* (2023), consisting of 712 birds mist-nested in 27 forest patches (Figure 5.2.2). Twenty sites were sampled in 2012-2013 (7 and 13 from GT and BT, respectively) and seven more in 2020 (2 and 5 in GT and BT respectively). Before pooling the data, we ensured genetic homogeneity between these two sampling periods. 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.*, 2017, 2016). Loci were amplified in simplex in a Dyad thermal cycler (Bio-Rad, Hercules, CA, USA), PCR products were multiloading for analyses in an automated sequencer (ABI3730), and allele scoring was performed using GENEIOUS R.8 (Kearse *et al.*, 2012).

We evaluated genetic differentiation between populations with pairwise  $F_{ST}$  (Weir and Cockerham, 1984) and tested for genotypic differentiation using a log-likelihood-based exact G-test. Significance was adjusted using Bonferroni correction when multiple tests were performed (initial threshold of 0.05). Finally, we described the patterns of genetic structure using a Bayesian model-based clustering algorithm implemented in STRUCTURE 2.3.4 (Pritchard *et al.*, 2000). We tested for the presence of  $K$  genetic clusters, with  $1 < K < 10$ , and used an admixture model with correlated allele frequencies and LOCPRIOR. We performed 10 independent runs of 100,000 iterations following a burn-in period of 20,000 iterations. We inferred the number of genetic clusters following the approach proposed by Garnier *et al.*

(2004), focusing on successive increases of posterior probability of the data for increasing values of  $K$ .

#### 5.2.4.3.4. Construction of the historical and current land cover maps

To map the past and the current forest covers, we used the 2010 and 1950 maps of 17 woody vegetation formations created by the Conseil Départemental of Guadeloupe, the Institut National de l'Information Géographique et Forestière (IGN) and the Direction Régionale de l'Office National des Forêts (ONF). 11 of the 17 woody vegetation types were classified as potential forest habitat (Annexe 4-S5.2.1) and constituted the forest vector maps. The mapping of the current urban areas was based on the building layer from the IGN BD TOPO database of 2008. These data were used to produce a building layer of 1950, following a protocol based on aerial ortho-photographs provided by the IGN BD ORTHO Historique. This protocol consisted in identifying by photo-interpretation the buildings of 2008 already present in the photograph of 1950. From these two building layers, we applied a dilatation of 50 m followed by an erosion of 40 m to each set of buildings in order to define the patches of artificial areas, i.e., the areas including buildings and their close surrounding. In addition, we used the Geographic Information Portal for Guadeloupe (Karugeo) to recover the coastlines from 1950 and 2008.

Finally, we created a categorical raster surface at a spatial resolution of 100 m by overlaying and rasterizing all these layers for both periods. As a result, we obtained two land cover maps including four land cover classes: forests, artificial areas, sea and matrix. The sea area corresponded to the pixels in a 5 km buffer zone from the coastline. All pixels within the coastline that were not assigned to the forest or artificial areas were classed as matrix pixels, corresponding mainly to agricultural areas.

For the sake of simplicity, and despite the fact that the composite data sources differ between 2008 and 2020, the contemporary map will be referred to as the 2012 map. The past landscape map is referred to as the 1950 map.

#### 5.2.4.3.5. Landscape description

We know that Guadeloupe has undergone severe urban expansion since the mid-20<sup>th</sup> (Eraud 2009). To characterise the associated landscape changes, we compared the composition and the configuration of the current and the historical landscapes using 7 FRAGSTAT metrics. From the two land cover maps, we computed four class metrics related to each of the three terrestrial land covers: the number of patches, the area, and three aggregation metrics that were the cohesion, the clumpy and the aggregation indexes. We also considered the proportion of forest areas and urban areas that were common between 1950 and 2012.

#### 5.2.4.3.6. Modelling landscape resistance: systematic screening of cost scenarios and resistance distances

We modelled the landscape influence on genetic structure by focusing on the resistance exerted by the land cover on individual dispersal. This modelling relied on the parametrisation of a resistance surface which associates cost values with discrete landscape features (this association will hereafter be referred to as a cost scenario). A resistance surface makes spatially explicit the resistance assumptions on dispersal movement. The assignment of cost values to landscape features (i.e., the design of cost scenarios) is a keystone in resistance modelling. Here, we based the construction of these cost scenarios on the negative exponential function introduced by Keeley *et al.* (Keeley *et al.*, 2016) for converting habitat suitability scores into cost values:

$$R = (\text{Cost max}) - (\text{Cost max} - 1) * ((1 - \exp(c*H))/(1-\exp(-c)))$$

Where Cost max is the maximum cost values, R is the resistance of the landscape feature, H is the suitability of the landscape features for the species, and c determines the shape of the curves. We arbitrarily considered four suitability scores for land cover types (0%, 33%, 66% and 100%) and set Cost max at 1000. We considered several shapes of relationship, with 10 values of c ranging from -8 to 8. This resulted in 10 sets of cost values with contrasts (i.e., the ratio between the maximum and minimum cost) ranging from 1000:2 to 1000:78.

To explore cost scenarios with more varied contrasts, we derived 8 more from Shirk *et al.*'s (2010) approach:

$$R_i = (\text{Rank}_i / \text{Rank}_{\max})^x \times C_{\max}$$

Where  $R_i$  represents the resistance of the land cover  $i$ ,  $\text{Rank}_{\max}$  the number of land covers considered,  $\text{Rank}_i$  the rank of the land cover  $i$  over the 1 to  $\text{Rank}_{\max}$  resistance ranks,  $C_{\max}$  the maximum cost values, and  $x$  the shape of the relationship. We considered here four resistance ranks, set the maximum cost at 1000 and evaluated 8 values of  $x$  (0.1, 0.25, 0.5, 1, 2, 4, 8, 16). This resulted in 8 sets of cost values with contrasts ranging from 1000:1 to 1000:871.

We pooled the cost scenarios derived from both methods (Annexe 4-S5.2.2). For each scenario, we assigned the forest to the lowest cost value and made all possible permutations of the three remaining cost values, randomly assigning them to artificial areas, matrix or sea. It resulted in 114 cost scenarios. Then we evaluated each cost scenario on both the historical and the current land cover map, computing resistance distances between population pairs using CircuitScape (Shah and McRae, 2008) (Figure 5.2.1A). Based on circuit theory, resistance distances reflect a random exploration of the landscape and lead to the identification of multiple movement pathways. This approach was therefore well suited to discriminate minor landscape changes through distance matrices. We finally got 114 matrices of resistance distances derived from the current landscape, and 114 others from the historical one.

#### 5.2.4.3.7. Statistical analysis for current and historical model comparison

As the shape of the Guadeloupe archipelago is very peculiar, with densely urbanised areas concentrated in the vicinity of the isthmus between the two major islands, we chose to study the effect of the landscape at two different spatial scales (Figure 5.2.1B). The first was the whole Guadeloupe archipelago scale, which included all pairs of BT and GT populations (hereafter referred to as the *whole-scale dataset*). The second scale corresponded to the logical units of BT and GT perceived as independent. It only considered genetic differentiation between pairs of populations within the same island. The related dataset (hereafter referred to as the *intra-island-scale dataset*) removed pairwise connections between BT and GT populations, thus reducing the spatial scale to an 'intra-island' scale and discarding the modelling of divergence between inter-island populations.

We repeated the following analysis workflow for both the whole-scale and the intra-island-scale datasets. For each cost scenario, we employed maximum-likelihood population-effect (MLPE) models (Clarke *et al.*, 2002) explaining current genetic distances by either current and historical resistance distances. MLPE models account for non-independence among the pairwise data by including a population random effect. The 228 resulting models (114 cost scenarios \* 2 time periods) were characterized by their Akaike Information Criterion (AIC) value. We also evaluated an Euclidean distance model, explaining genetic distances by Euclidean distances between population pairs (hereafter referred to as the null model). We computed the delta AIC (dAIC) between each resistance model and the null model. We first extracted cost scenarios leading to an Isolation by Resistance (IBR) pattern (i.e., showing a dAIC greater than two compared to the null model) with both the historical and the current resistance model (Figure 5.2.1C). Among these cost scenarios, we retained only those leading to the strongest IBR signal within a 7-point dAIC range. Then, for each of these selected scenarios, we compared the statistical quality between the historical model and the current model based on the AIC (Figure 5.2.1D). If the difference in AIC between the historical model and the current model was less than -2, then the best model was the historical one. If the difference was greater than 2, the best model was the current one. If the difference was between -2 and 2, the two models were considered equivalent. We therefore evaluated which of the current or the historical models appeared to best explain the current genetic data, for each best-performing cost scenario.

#### 5.2.4.4. Results

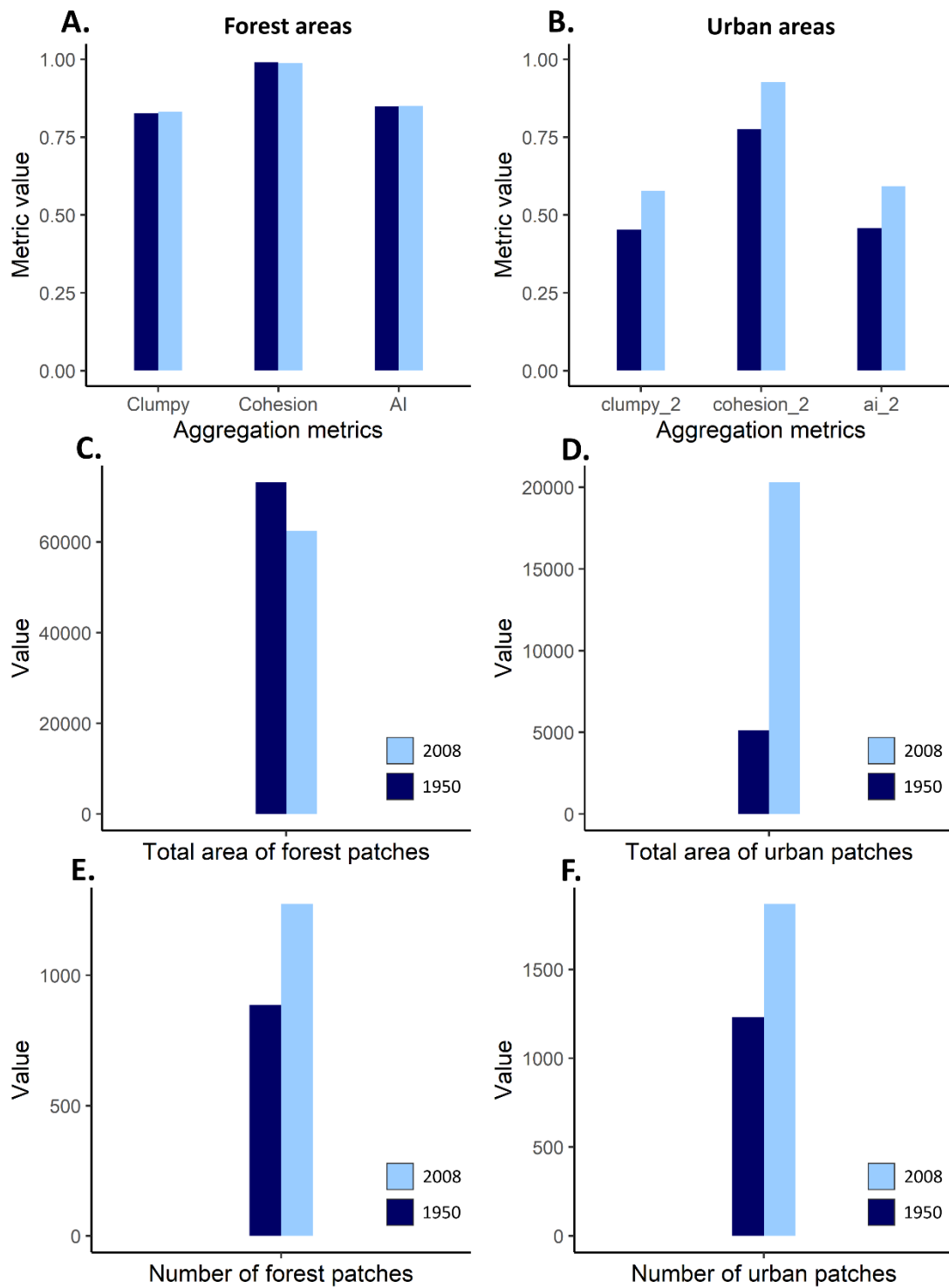
##### 5.2.4.4.1. Changes in the landscape between 1950 and 2012

The land cover maps of the two periods are shown in Figure 5.2. 2. The most significant change in land cover occurred in urban areas (Figure 5.2.3). The urban area in 1950 was only 25% of its current size: 15,206 ha have been added (Figure 5.2.3D) and the number of patches has increased by 52% from 1950 to 2008 (Figure 5.2.3F). The urban cover appeared to be currently more aggregated than in 1950, but this may reflect the concentration of urban areas around a few places (Figure 5.2.3B). The change in forest cover over the period 1950-2008 was smaller, with 84 % of the historical forest cover and 94 % of the current forest cover remaining between 1950 and 2008. However, there was a loss of 10,709 ha and a 44% increase in the number of patches, reflecting the fragmentation resulting from this forest loss (Figure 5.2.3C, E). The aggregation metrics did not allow for the identification of any clear change in the configuration of the forest cover between 1950 and 2008 (Figure 3A). Thus, major landscape changes have mainly affected matrix and urban areas rather than forest habitat.

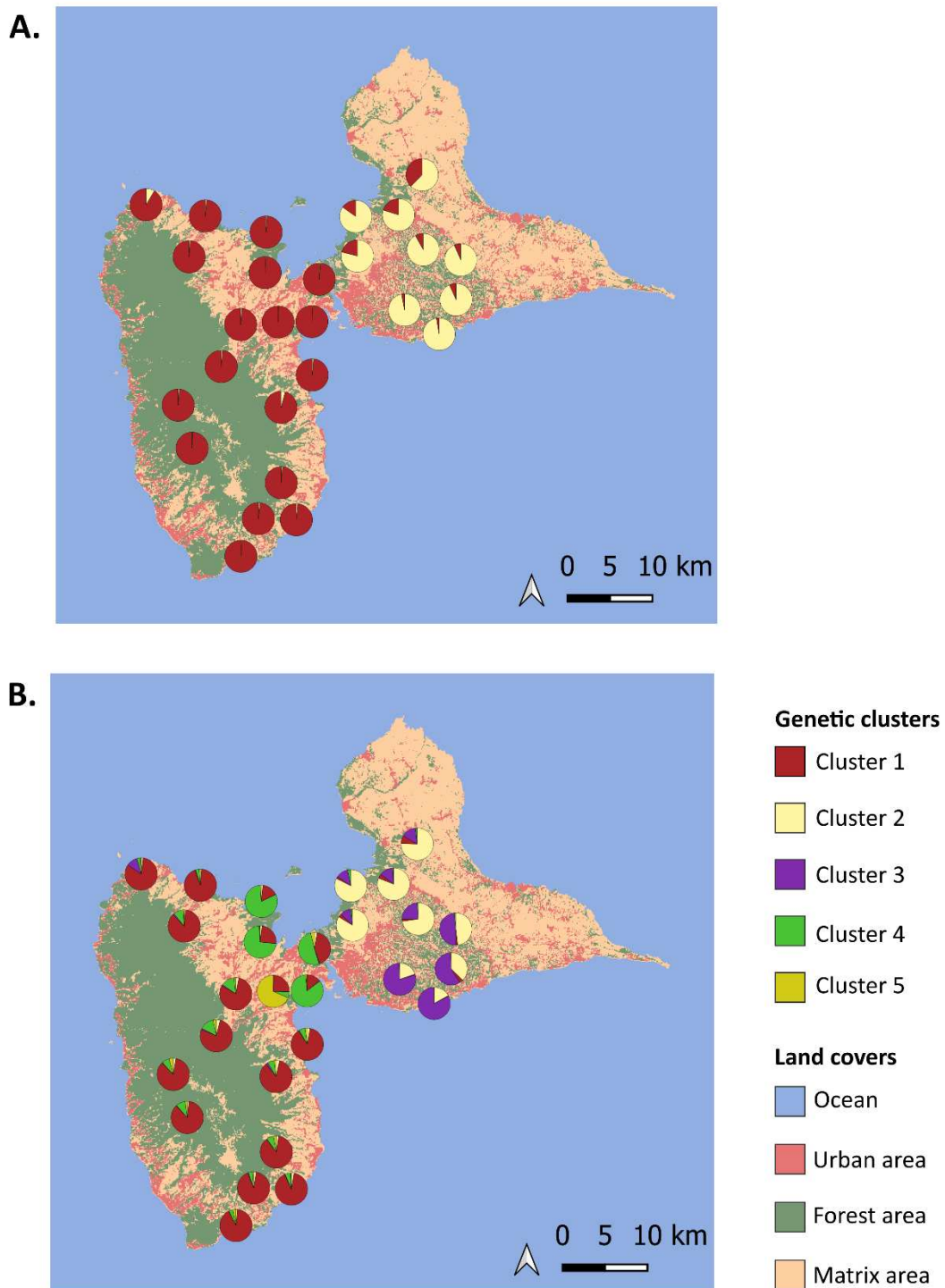
##### 5.2.4.4.2. Population genetic structure

The plumbeus warbler showed high levels of genetic structuring, with 90 % of population pairs significantly differentiated (Table 5.2.1). The highest values of genetic differentiation were related to population pairs located on either side of the isthmus, and to populations disconnected from the forest mega-patch within BT. These results were consistent with the genetic clusters inferred with STRUCTURE (Figure 5.2.4). The plumbeus warbler exhibited two main distinct allelic pools, separating the populations from BT and GT (Figure 5.2.4A). A finer investigation allowed to identify five genetics clusters (Figure 5.2.4B). Two of these separated the GT populations and a third corresponded to the isthmus population on the BT side. The fourth and fifth isolated a population from the others in BT.





**FIGURE 5.2.3:** Values of FRAGSTAT landscape metrics characterizing both the current and the historical landscape. A – B) Aggregation metrics for the forest and the urban classes respectively, C – E) Area and number of patches of forest class, D – F) Area and number of patches for the urban class. The light blue bars correspond to metrics values for the current landscape, while the dark ones characterize the historical one.



**FIGURE 5.2.4:** Spatial visualization of genetic clusters inferred by STRUCTURE. A) Partition in two clusters, B) Finer partitioning into 5 clusters. Each pie chart represents the genetic composition of a sampled population.

Genetic Index	Pop. pairs	Mean	Sd	Min	Max	Pop. pairs significantly differentiated
<b>FST</b>	Tot	0,044	0,023	0,0002	0,106	90%
<b>FST</b>	BT_BT	0,045	0,027	0,0002	0,106	82%
<b>FST</b>	BT_GT	0,046	0,021	0,011	0,101	92%
<b>FST</b>	GT_GT	0,033	0,012	0,006	0,059	98%

**TABLE 5.2.1:** Statistical description of genetic differentiation values between sampled populations of plumbeus warbler. The column “Pop. pairs significantly differentiated” refers to the results of the log-likelihood based exact G-test, showing the proportion of population pairs with significant genetic differentiation. Several groups of population pairs were studied: all population pairs (Tot), those within Basse-Terre (BT\_BT), within Grande-Terre (GT\_GT), and between Basse-Terre and Grande-Terre (BT\_GT). Sd: standard deviation, Min: minimum, Max: maximum, Pop. pairs: population pairs.

#### 5.2.4.4.3. Identification of best cost scenarios for landscape resistance modelling with the whole-scale dataset

111 out of 114 cost scenarios led to resistance models with a better fit than the Euclidean model, providing a median AIC improvement of -40 points against it.

We considered the four best models within a range of 7 dAIC points. These four best models were all calibrated on the current landcover map. They derived from two pairs of similar cost scenarios (Table 5.2.2) exhibiting a high contrast (1000:4 and 1000:2) and assigning either the matrix or the urban areas to the highest resistance values. The best model calibrated on the historical map ranked 8<sup>th</sup> and was associated with the same cost scenario as the first best model calibrated on the current map: {Forest, Urban, Matrix, Sea} = {4,317,1000,63}. Considering the best cost scenarios, there was an AIC difference of 7 to 12 points in favour of the current resistance models compared to the historical ones (Table 5.2.2).

Three scenarios ({Forest, Urban, Matrix, Sea} = {1, 1, 1000, 1}, {1,101,1000,4} and {1,11,1000,1}) resulted in an equal or worse fit than the Euclidean model with the historical map and were discarded for the rest of the analysis.

Scenario Forest - Urban - Matrix - Sea	Model	dAIC RD - Euclidean	dAIC Current – Historical	Best support period
4 – 1000 – 317 - 63	FST ~ Current RD	-76,31	-7,43	Current
2 - 1000 - 255 – 55	FST ~ Current RD	-74,58	-8,23	Current
4 – 317 - 1000 – 63	FST ~ Current RD	-73,90	-9,18	Current
2 – 255 – 1000 - 55	FST ~ Current RD	-72,25	-12,69	Current

**TABLE 5.2.2:** Table of cost scenarios resulting in the best resistance models according to the AIC criterion for the *WHOLE-SCALE MODELS*. dAIC RD - Euclidean: Difference in AIC between the resistance distance model and the Euclidean distance model. dAIC Current – Historical: Difference in AIC between the resistance distance model derived from the current map and that derived from the historical map.

Scenario Forest - Urban - Matrix - Sea	Model	dAIC RD - Euclidean	dAIC Current – Historical	Best support period
7 - 1000 - 556 - 237	FST ~ Current RD	-62,64	-3,81	Current
7 - 556 - 1000 - 237	FST ~ Current RD	-62,53	-4,79	Current
10 - 643 - 1000 - 314	FST ~ Current RD	-62,05	-3,26	Current
5 - 1000 - 442 - 154	FST ~ Current RD	-62,00	-5,24	Current
10 - 1000 - 643 - 314	FST ~ Current RD	-61,92	-3,02	Current
7 - 237 - 1000 - 556	FST ~ Current RD	-61,40	-2,67	Current
5 - 442 - 1000 - 154	FST ~ Current RD	-61,38	-7,46	Current
5 - 154 - 1000 - 442	FST ~ Current RD	-61,17	-4,90	Current
13 - 698 - 1000 - 370	FST ~ Current RD	-60,88	-2,57	Current
7 - 237 - 556 - 1000	FST ~ Historical RD	-60,67	0,56	Identical

**TABLE 5.2.3:** Table of cost scenarios resulting in the best resistance models according to the AIC criterion for the *WITHIN-ISLAND MODELS*. Only the top 10 models are presented here, the other 58 are in appendices S4. dAIC RD - Euclidean: Difference in AIC between the resistance distance model and the Euclidean distance model. dAIC Current – Historical: Difference in AIC between the resistance distance model derived from the current map and that derived from the historical map.

#### 5.2.4.4. Identification of best cost scenarios for landscape resistance modelling with the intra-island-scale dataset

All cost scenarios led to resistance models more informative than the Euclidean model, with a median AIC improvement of -43 points against it. We selected the 68 best models within a range of 7 dAIC points (Table 5.2. 3 and Annexe 4-S5.2.4). Among them, 10 corresponded to cost scenarios best supported by the current map, and 58 to cost scenarios showing identical support for both time periods (Table 5.2.3 and Annexe 4-S5.2.4). These scenarios were derived from 8 similar source scenarios with high contrasts (from 1000:2 to 1000:24) and no trend in the assignment of a resistance value to a land use type. The AIC difference between current and historical models for each of these best cost scenarios ranged from -13 to 2, with a median value of -3.25 for models best supported by the current map and -0.3 for those with equal support. Finally, the large number of models leading to a similar improvement compared with the null model, the large majority of models associated with identical support between the 1950 and 2012 maps, and finally the low average AIC deviation of models with greater support for the contemporary map suggest that there is no major effect of recent urbanisation on an intra-island scale.

#### 5.2.4.5. Discussion

We demonstrated that the genetic structure of the plumbeus warbler was significantly related to landscape structure, mainly of the current time, but also of the past. Current genetic differentiation would therefore be a legacy of past forest fragmentation, exacerbated by recent urbanisation. However, the fast response of the genetic structure to these landscape changes with specific configuration questions the origin of the genetic differentiation observed between the two main islands, as well as our ability to detect landscape legacies.

The expansion of urban areas, particularly intense on the isthmus, was the main landscape change between 1950 and 2000. This expansion occurred at the expense of the agricultural and open matrix. The forest cover was already fragmented by this matrix in the 1950s and had changed very little by the year 2012. As we found better support towards the current than the historical landscape at the whole Guadeloupe scale, it is likely that the emergence of urban areas has already affected the current genetic structure of the plumbeus

warbler, making dispersal between habitat patches even less permeable than before, especially between both islands. This was further supported by the within-island results, showing equivalent explanatory power for both the 1950 and 2012 maps. The difference between the whole-Guadeloupe and the within-island datasets lies in the pairwise connections between BT and GT, which were discarded in the within-island dataset. Thus, if the current map is more accurate than the historical map at the whole Guadeloupe scale, but the two maps are equivalent at the within-island scale, this means that the urban expansion on the isthmus area provides additional information about the genetic divergence between BT and GT. Therefore, the current genetic structure of the plumbeous warbler may be the result of two consecutive steps of habitat fragmentation. First, it would be inherited from the past fragmentation of the forest cover, as supported by historical models. Then, it would result from the barrier effect of the resistant urban matrix on the isthmus. Therefore, the plumbeous warbler's genetic structure would carry the imprint of landscape changes that occurred less than 50 years ago, involving a very fast genetic time lag.

We could consider three hypotheses to explain why landscape changes would rapidly translate into patterns of neutral genetic divergence in the plumbeous warbler. First, despite the lack of naturalistic knowledge on this species, we could explain this fast genetic response regarding life history traits and population dynamics. Indeed, estimates suggest there are around 100,000 individuals of plumbeous warbler in Guadeloupe, which is lower than the demographic threshold of the most common species (Eraud 2012). Moreover, this species is known to be highly specialised in forest habitats and sensitive to fragmentation (Khimoun *et al.* 2017, Eraud 2012). This makes its spatial distribution heterogeneous and dependent on dense and contiguous patches of rainforest (Eraud 2012). It has been shown that the smaller the population size, the faster a new equilibrium is reached after a disturbance due to a stronger drift effect (Varvio *et al.*, 1986). Furthermore, for a constant number of individuals, several small populations induce a faster time lag than a large population (Epps and Keyghobadi, 2015). It is therefore likely that the highly specialized character of the plumbeous warbler has led to the isolation of populations within forest fragments, making them vulnerable to drift-induced divergence and resulting in fast genetic time lag. This finding echoes that of Nowakowski *et al.* (2015), who showed that a tropical generalist amphibian species, with long generation times and spatially autocorrelated population sizes, was less sensitive to current landscape structure than a more specialist species. Münzbergová *et al.* (2013) also argued that some life history traits could impact the pace of genetic response to landscape changes. Indeed, they explained a 150-year time lag in the genetic

structure of a perennial grass by its long lifespan, overlapping generations, persistent bank seed and infrequent seedling reproduction. Finally, Holzhauer *et al.* (2006) summarised the issue by suggesting that the intensity of a time lag depends on the interplay between several temporal and spatial scale parameters, such as generation time, dispersal ability, landscape structure and rate of change. In the same vein, Zellmer and Knowles (2009) evidenced a very fast response of an amphibian genetic structure (less than 50 generations) to landscape change by metapopulation dynamics. They argued that an intense colonization-extinction dynamic had enhanced drift-induced divergence, thus accelerating the achievement of a state of equilibrium with the constraints of the contemporary landscape. Therefore, further research on the population dynamic of the plumbeus warbler may provide insights into the drivers of genetic differentiation.

Second, because genetic structure in natural populations results from multiple interacting evolutionary processes, phylogeographic histories rather than landscape effects, may account for the current genetic divergence observed between BT and GT. As shown by the STRUCTURE results, the two main allelic pools correspond to these geographical entities and could be inherited from two independent waves of colonisation. Thus, assuming that the genetic structure results from secondary contact between two historically distinct populations, we could explain the better support associated with the current map as a modelling artefact. The emergence of a 4th urban land use, mainly on the isthmus, provides an additional feature for modelling the separation between BT and GT by associating it with a very high resistance value. Distinguishing between phylogeographic and landscape processes is challenging, but several demographic modelling approaches, such as Approximate Bayesian Computation methods (Cabrera and Palsbøll, 2017), have been developed to infer the impact of anthropogenic change on contemporary species structure. Identifying the evolutionary process at work could therefore be seen as a primary step in studies seeking to disentangle the effects of current or historical landscape structure on genetic divergence.

Third, we could question our method's ability to detect landscape legacies accurately. Because of the difficulty to access historical data, we based our analysis on a landcover map with a low thematic resolution. This led to a coarse description of landscape changes. Moreover, we attempted to capture them using resistance distances. Although resistance modelling seemed to be more appropriate than cost-distance modelling because it captures information on all potential pathways rather than on a single optimal path prone to ignore significant parts of the landscape, it is still based on pairwise distance modelling. We could wonder if a pairwise

distance can detect landscape changes concentrated in a particular area. While Miller *et al.* (2013) also used distance metrics, they proposed a refined version of resistance distances by considering the residuals of the regression of resistance distance on geographic distance. This measure of connectivity redundancy represents the availability of suitable habitat for a given species between two locations, independently from geographic distances. They found a clear link between the genetic structure of a tropical bird and the connectivity redundancy of a historical landscape. Because this metric showed the 35-year time lag more clearly than classic resistance distances, they suggested using it when gene flow is severely restricted and the effects of drift blur the IBD signal. As the genetic structure of the plumbeous warbler showed a strong restriction of gene flow, it would be interesting to look further in this direction. Finally, we did not have information about the genetic structure of 1950. We assumed that genetic differentiation increased between 1950 and 2012, particularly between the BT and GT populations, but we cannot assure it nor assess the magnitude of this genetic change. A promising approach to disentangle past legacies from contemporary genetic structure would rely on a joint study of gene and landscape time series. For example, Draheim *et al.* (2018) provide the first landscape genetics analyses based on time series across multiple generations of the same population of black bears in Michigan (USA). Over a 9-year monitoring period, they evidenced an association between the magnitude of landscape change and the change in the black bear's genetic structure. Some landscape features that were not informative at one point in time have become so as the landscape has evolved, demonstrating the spatial and temporal heterogeneity of landscape structure and its consequences for genetic processes. This type of study could therefore be of great interest to further investigate the imprint of contemporary and past landscapes on genetic patterns, and to improve predictions of how future landscape changes will affect genetic differentiation. However, time series studies appear to be costly in terms of time and data collection, and as the lack of knowledge of the time lag makes their necessary duration uncertain, the use of simulations could be very promising.

The results obtained are alarming from a conservation perspective. The small temporal and spatial scales at which the effects of landscape change impact the genetic structure shed light on the importance of connectivity for the focal species. Indeed, our results show that increasing urbanisation could severely restrict gene flow and we could expect future declines in the genetic diversity of the isolated populations as well, because of a slower approach to equilibrium of within-patch genetic diversity than genetic differentiation (Keyghobadi *et al.*, 2005). These two elements threaten the persistence of populations, for example by limiting



rescue or adaptation to climate change (Ingvarsson, 2001). However, this rapid response in genetic structure may give us hope for an even faster response in the opposite direction, towards the restoration of dispersal corridors (Epps and Keyghobadi, 2015; Landguth *et al.*, 2010).

#### 5.2.4.6. Perspectives

This manuscript is a work in progress. Several avenues remain to be explored before its completion. In particular, we plan to repeat these analyses using a genetic distance index that is likely to respond more quickly to landscape change by transcribing the most recent genetic flows, such as the  $D_{PS}$  (Murphy *et al.*, 2010). We could expect an even weaker adjustment of genetic structure to the landscape structure of 1950 than with the  $F_{ST}$ . Then we got some doubt with the high permeability assigned to the sea, which might be an artefact of the resistance distance algorithm. We are therefore considering some additional analyses, such as comparing the 'unrealistic' scenarios with a 'land IBR' scenario, assigning high resistance to the sea and low and homogeneous cost to the land area. We could also, as in Khimoun *et al.* (2017), divide the sea occupation into two features, with a coastal strip of one kilometre width and the open sea elsewhere. Finally, it would be interesting to relate this issue of time lag in genetic response to the process of extinction debt that has been documented in community ecology (Halley *et al.*, 2016). To investigate the parallel with the time lag at the level of species richness, we could examine the time lag of genetic diversity, in addition to the time lag of genetic differentiation observed here.

### 5.3. Conclusions

Tenir compte de l'histoire évolutive des populations semble être un préalable indispensable aux études de la variabilité génétique contemporaine pour que les inférences de génétique du paysage soient fiables et précises. Dans notre cas d'étude, dans l'hypothèse où la BT et la GT auraient été colonisées par deux lignées distinctes de Paruline caféïette, interpréter la divergence génétique entre les populations de BT et de GT comme résultant de la résistance du paysage urbain les séparant serait fallacieux. Inférer le scénario de colonisation de la Guadeloupe nous permettra donc de construire un modèle nul de la variabilité génétique, auquel pourra être comparé les modèles de résistance du paysage. Toutefois, compte tenu du caractère d'ancienne espèce endémique de la Paruline caféïette et de sa présence sur l'île depuis 4,5 millions d'années sans événement évident de recolonisation, compte tenu également du fort patron d'isolement par la résistance du paysage observé malgré l'utilisation de marqueurs microsatellites présentant une profondeur temporelle plus faible que les marqueurs d'ADN mitochondriaux et conférant une puissance statistique moindre que celle fournie par les SNP, nous pouvons faire l'hypothèse qu'une part non négligeable de la variabilité génétique entre les populations résulte de facteurs paysagers.

Il est ainsi probable que la différenciation génétique soit en partie expliquée par le paysage contemporain, et en particulier par l'urbanisation majeure de l'isthme entre BT et GT. Cette modification du paysage aurait creusé la différenciation génétique entre BT et GT en moins de 50 ans. On peut noter la brièveté de ce temps de latence, bien que la considération d'un plus grand nombre de pas temporels dans les données génétiques et paysagères serait nécessaire pour l'estimer avec plus de précision. Une mesure plus précise du taux et de la vitesse de la variation génétique en fonction de l'intensité des changements paysagers serait particulièrement utile pour prédire les conséquences de futurs changements paysagers. Contrôler les modèles de génétique du paysage par l'histoire évolutive des populations et reconduire ces modèles à différents pas temporels pour évaluer le délai de réponse de la structure génétique n'est pas toujours réalisable. Aussi, il faudrait davantage d'études évaluant les facteurs susceptibles d'allonger ou de raccourcir ce délai de réponse. Enfin, on peut imaginer encourager l'étude de l'histoire évolutive dans les cas où des connaissances *a priori* suggèrent un effet de facteurs historiques, comme par exemple des connaissances sur la localisation des refuges glaciaires ou sur les routes de colonisation.



# CHAPITRE 6

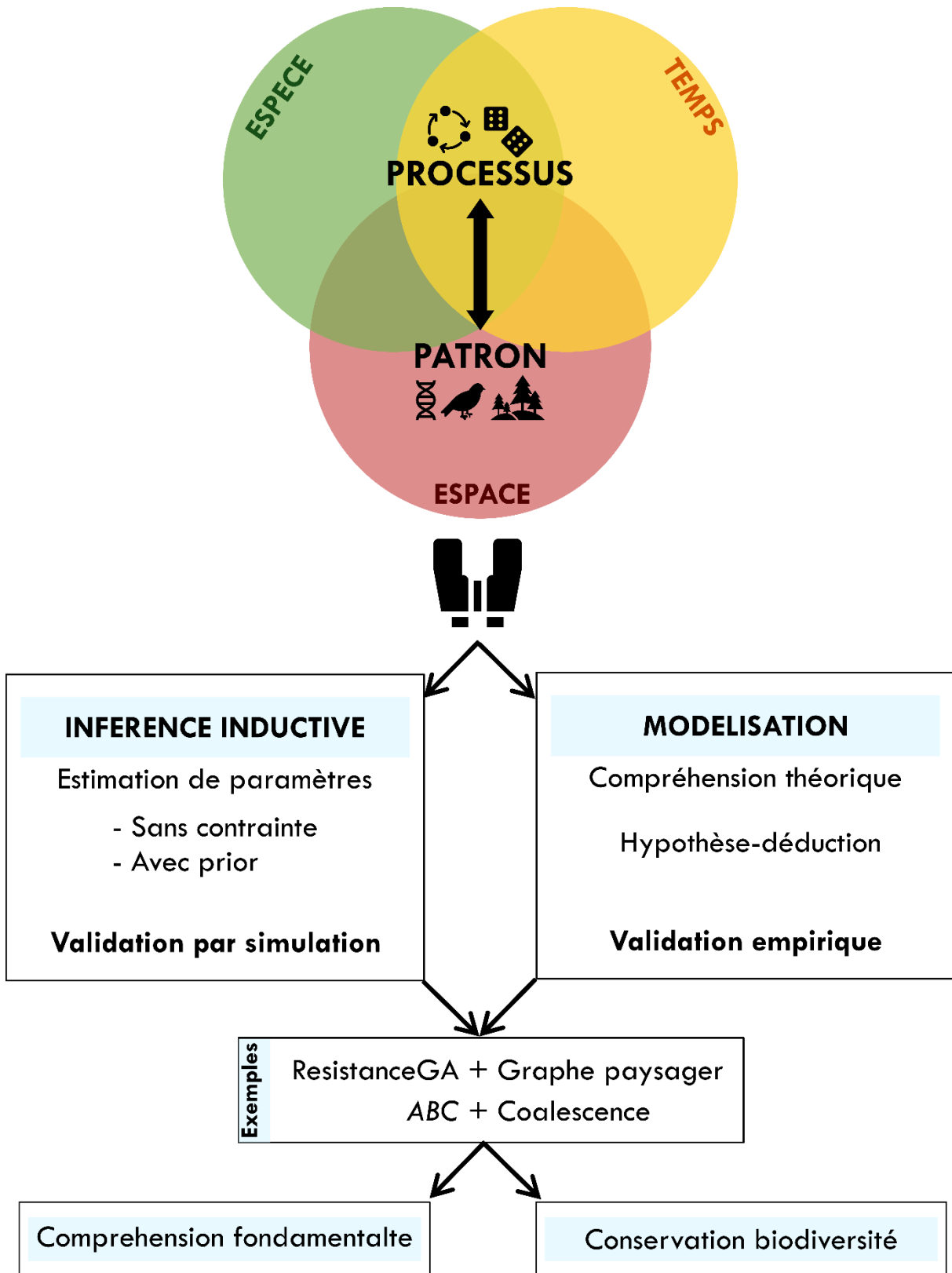
## DISCUSSION GÉNÉRALE

---

<b>6. DISCUSSION GÉNÉRALE</b> .....	187
6.1. Valider des modèles de connectivités, comment et pourquoi ?.....	190
6.1.1. Validation de ResistanceGA à partir de données simulées .....	192
6.1.2. Validation empirique des modèles de connectivité .....	202
6.2. Domaine de validité spatial, temporel et multi-spécifique des modèles de connectivité .....	210
6.2.1. Considération de l'échelle spatiale dans les modèles de connectivité.....	212
6.2.2. Considération de l'échelle temporelle dans les modèles de connectivité .....	218
6.2.3. Modèles de connectivité multi-spécifiques .....	221
6.3. Perspectives pour des modèles de connectivité valides.....	223
6.3.1. Vers une vision dynamique des modèles de connectivité.....	225
6.3.2. Vers une diversification des variables explicatives de la connectivité.....	226
6.3.3. Modéliser l'influence du paysage sur la diversité adaptative.....	227
6.3.4. Traduction des inférences en mesures de conservation .....	230

Les résultats présentés dans ce manuscrit de thèse gravitent autour de la notion de validité des modèles de connectivité. Un modèle de connectivité est le reflet de notre compréhension de l'interaction entre des processus éco-évolutifs non observables (tels que le flux génétique, la migration ou la dérive) et des patrons observables paysagers, génétiques ou spécifiques (Figure 6.1). Un modèle est jugé valide si les processus considérés dans la modélisation conduisent à des prédictions conformes à la réalité du patron observé. La validité d'un modèle repose donc à la fois sur la validité de notre compréhension, mais également sur la validité intrinsèque des méthodes d'inférences permettant de paramétrer ce modèle (Figure 6.1). Enfin, un modèle est souvent associé à des postulats implicites ou à des conditions d'application déterminant son domaine de validité théorique. On postule ainsi généralement que les processus modélisés sont à l'équilibre, que l'échelle spatio-temporelle correspond à celle à laquelle se joue l'interaction modélisée, ou encore que l'écologie de l'espèce étudiée est conforme à certains prérequis de modélisation.

Nous discuterons donc ici de la validité des modèles de connectivité en génétique du paysage en i) présentant les méthodes de validation par simulation et empiriques, et ii) en évaluant les limites spatiales, temporelles et multi-spécifiques du domaine de validité de ces modèles. Nous terminerons en évoquant les perspectives pour améliorer la validité de ces modèles, et pour leur application au domaine de la conservation de la biodiversité.



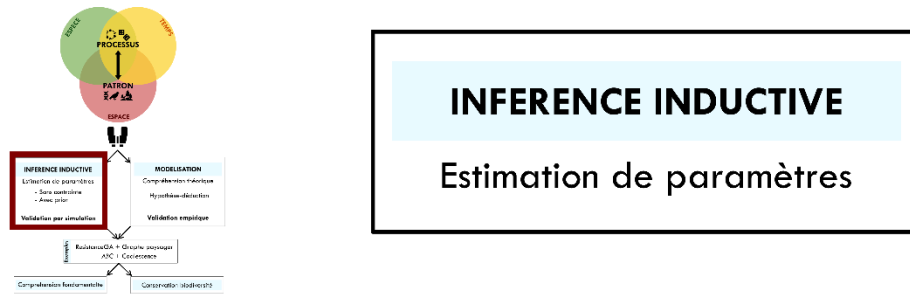
**FIGURE 6.1** : Schéma général synthétisant les deux approches développées dans cette thèse, concernant l'exploration de l'interaction entre patron et processus en génétique du paysage.

## 6.1. Valider des modèles de connectivités, comment et pourquoi ?

La préservation de la connectivité des populations étant un enjeu majeur de conservation, de nombreuses méthodes de modélisation des connectivités ont vu le jour depuis le début du siècle. L'avancée de l'effort interdisciplinaire ayant abouti à ces modèles de connectivité repose maintenant sur un nouveau défi : la validation et la comparaison de ces modèles (Encadré 6.1). Outre l'avancée des connaissances scientifiques sur le processus de dispersion, explorer la validité des modèles de connectivité est nécessaire pour assurer la pertinence et l'efficacité des mesures de conservation qui en dérivent. Cela permettrait aux gestionnaires et décideurs de conceptualiser et appliquer des mesures adaptées, et apporterait le poids et la crédibilité nécessaire pour le financement et la reconduction de ces mesures. Malgré certaines études au caractère précurseur (Driezen *et al.*, 2007), l'évaluation de la validité écologique des modèles de connectivité ne s'est réellement développée qu'au cours des dix dernières années. Nous axerons cette première partie sur deux approches de validation, en nous intéressant à leur construction et leur portée. Dans un premier temps, nous verrons comment la validité d'une méthode d'inférence des paramètres d'une surface de résistance (ResistanceGA) peut être évaluée. Enfin, dans un second temps, nous traiterons de la validation empirique des modèles théoriques de la connectivité, et notamment des graphes paysagers (Figure 6.1).

### **Encadré 6.1 : Approche de validation en écologie**

Plusieurs objets peuvent faire l'objet de validation en écologie. Il est ainsi possible de distinguer la validation des modèles conceptuels de celle des méthodes d'inférences (Figure 6.1). Dans le premier cas, la validation intervient comme l'étape finale d'un processus hypothético-déductif et elle est généralement conduite à partir de données empiriques. Des hypothèses sont proposées pour expliquer le processus conduisant au patron observé, puis transcrites sous forme de modèle, et enfin confrontées avec des données empiriques reflétant les conséquences du processus modélisé. Une validation empirique consiste donc à valider notre compréhension de l'interaction entre un processus et un patron, ainsi que notre capacité à la traduire dans un modèle. Dans le second cas, il s'agit d'évaluer la capacité d'une méthode à estimer les paramètres d'interaction entre un processus et un patron. Les approches de simulations sont généralement adaptées aux validations de ces méthodes d'inférences, car elles permettent de contrôler ces paramètres, et de vérifier si la méthode est en capacité de les estimer correctement à partir du patron observé.

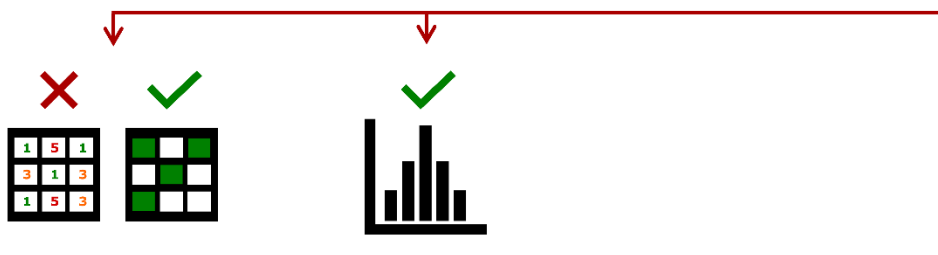


## VALIDATION PAR SIMULATION : paramètres correctement estimés ?



Paramètres → Simulation → Méthode → Prédiction des paramètres

### Exemple : RGA



Coûts + Paysage → Distances génétiques → **RGA** → Coûts optimisés

### Apprentissage

- des distances génétiques
- de la topologie des populations
- mais pas de la distribution des coûts

### Conclusion :

Bonne capacité de prédiction des distances génétiques en contexte d'interpolation.

**FIGURE 6.2** : Focus depuis la figure 6.1 sur l'approche d'inférence inductive. En génétique du paysage, ce type d'approche permet d'estimer des paramètres intervenant dans l'interaction patron-processus étudiée. Par exemple, ResistanceGA permet d'estimer les paramètres de résistance des éléments paysagers au flux génétique. La validité d'une méthode d'inférence est classiquement étudiée *via* le recours à la simulation.



### 6.1.1. Validation de ResistanceGA à partir de données simulées

Dans le chapitre 3.1, nous nous sommes intéressés aux surfaces de résistances paramétrées par un processus d’optimisation à partir de données génétiques. Une surface de résistance étant une retranscription spatiale de notre compréhension du processus de dispersion en milieu hétérogène, nous aurions pu tester la validité empirique de cette retranscription. Toutefois, l’étape d’optimisation rend cette démarche bancal. En effet, une surface de résistance optimisée sans contrainte ne repose pas sur une théorie, et ne représente pas notre compréhension du processus de dispersion. Elle est le résultat d’un apprentissage automatique à partir d’un grand nombre de données : on parle de modélisation ‘*data-driven*’, par opposition aux modèles ‘*theory-driven*’ (Yates *et al.*, 2018). Dans le chapitre 3.1, nous avons donc évalué, à partir de simulations démo-génétiques, la validité de ResistanceGA (RGA), un outil permettant de produire des inférences inductives sur la résistance du paysage au flux génétique. Dans cette partie, nous présenterons les intérêts de notre approche de validation, puis questionnerons les qualités de l’apprentissage de RGA, avant de mentionner les limites de notre étude (Figure 6.2).

#### 6.1.1.1. Apports méthodologiques de notre approche de validation

Les études de Beninde *et al.* (2023), Peterman *et al.* (2019), et Winiarski *et al.* (2020) ont évalué la concordance entre des surfaces de résistance simulées et celles optimisées par RGA, en fonction de différents facteurs, tels que la nature des variables génétiques réponses et les profils de résistance (Beninde *et al.*, 2023), l’effort d’échantillonnage et la variabilité du patron génétique (Winiarski *et al.*, 2020), la nature contrainte ou non contrainte de l’algorithme d’optimisation, ou enfin la nature catégorielle ou continue de la surface à optimiser (Peterman *et al.*, 2019). Les conclusions de ces études sont sujettes à de nombreuses controverses. En effet, tandis que Winiarski *et al.* (2020) soutiennent la capacité de RGA à optimiser avec précision les surfaces de résistance, Beninde *et al.* (2023) émettent plus de réserves et restreignent le champ de validité des inférences de RGA aux paysages peu complexes et peu résistants. Face à ce constat, notre étude apporte des éléments méthodologiques et conceptuels nouveaux pour discuter de la validité des inférences de RGA et de leur usage.

*L'utilisation de paysages réels*

Tout d'abord, nous avons travaillé sur des paysages réels et non simulés, ce qui nous rapproche des conditions réelles d'utilisation de l'algorithme. Le nombre important de paysages considérés permettait d'explorer l'effet des variations paysagères sur les performances de ResistanceGA.

*Des distances génétiques issues de simulation démo-génétiques*

Par ailleurs, nous avons utilisé des simulations démo-génétiques pour produire les variables génétiques réponses, ce qui nous a fourni une distribution de la variation génétique *a priori* plus conforme à la réalité que le proxy de distance paysagères bruitées utilisé par (Winiarski *et al.*, 2020) (Figure 6.2).

*Différentes espèces virtuelles aux profils écologiques distincts*

Nous avons également comparé les performances de RGA pour des espèces virtuelles ayant plusieurs profils écologiques, définis par le niveau de préférence pour l'habitat forestier et la capacité de dispersion. Cette prise en compte fait écho aux travaux de Beninde *et al.*, (2023), qui ont évalué les effets de différentes perceptions du paysage par l'espèce focale, en termes de fragmentation, de résistance de la matrice et d'intensité du signal d'IBD.

*Une approche de validation indépendante fondée sur trois critères*

De plus, nous avons réalisé une validation croisée indépendante offrant des résultats robustes sur les performances prédictives de RGA (Figure 6.2). Nous avons également dans cette étude évalué un nouveau critère d'optimisation : le  $R^2$  de validation. Cet indice mesure l'adéquation entre les valeurs de différenciation génétiques observées et prédites par le modèle. Utiliser cet indice comme critère objectif d'optimisation permet de maximiser les capacités prédictives du scénario de coût inféré, et donc ainsi de contrebalancer les effets du sur-ajustement propres aux processus d'optimisation. Afin d'évaluer le potentiel de cet indice, il serait toutefois nécessaire de reconduire les analyses réalisées en utilisant un critère plus classique tel que l'AIC, et comparer les performances des deux optimisations. Enfin, un apport majeur de notre étude concerne les métriques de validation considérées. Contrairement aux travaux précédents nous n'avons pas évalué la corrélation entre la surface 'vraie' et la surface optimisée, mais avons considéré trois métriques complémentaires basées sur la sensibilité de l'algorithme au type de patron génétique simulé, ses capacités prédictives et la concordance des scénarios de coût 'vrais' et optimisés. Cette dernière métrique est celle se rapprochant le plus de la métrique de corrélation spatiale entre surfaces, couramment utilisée (Beninde *et al.*, 2023; Peterman *et al.*, 2019; Winiarski *et al.*, 2020). Les deux autres

métriques apportent en revanche un nouvel éclairage sur les potentialités de RGA, en particulier sur son utilisation à des fins prédictives. Elles éclairent également sur l'utilisation des approches 'data-driven' à des fins d'inférence, faisant l'objet de nombreuses questions à l'heure de la démocratisation de l'apprentissage automatique dans le domaine de l'écologie.

#### 6.1.1.2. Les critères de validité des inférences de RGA mis en perspective avec sa qualité d'apprentissage

RGA repose sur un algorithme génétique, qui est une méthode d'apprentissage automatique inspirée du processus d'évolution naturelle (Mitchell, 1995). De façon caricaturale, cet algorithme suit une forme d'apprentissage dans lequel il confronte son exploration de l'espace des paramètres de résistance au déplacement avec des données de distances génétiques. Cet apprentissage le conduit à prédire la distribution des distances génétiques dans un paysage hétérogène. Dans cette partie, nous nous intéressons à la validité des inférences de coût de RGA au regard de cette forme d'apprentissage. Nous verrons notamment que i) l'algorithme n'apprend pas les paramètres que l'on cherche à estimer, ii) mais qu'il capture au contraire des patrons moins évidents, et peut-être moins pertinents par rapport notre problématique et enfin iii) que cet apprentissage conduit à des capacités de prédiction très prometteuses.

##### *L'apprentissage des paramètres de coût ne fonctionne pas*

RGA est principalement utilisé pour inférer les effets de la matrice paysagère sur les flux de gènes, notamment en interprétant les valeurs de coût optimisées et liées au déplacement à travers chaque occupation du sol (Draheim *et al.*, 2021; Khimoun *et al.*, 2017; Martin *et al.*, 2023; Mulvaney *et al.*, 2021). Nous émettons donc l'hypothèse que RGA est capable de retrouver ces valeurs de coût à partir de données de distances génétiques. Pour évaluer cette capacité, nous avons confronté de 'vrais' scénarios de coût simulés avec ceux optimisés par l'algorithme. Nos résultats n'ont pas été concluants. En effet, seulement 10% des optimisations ont conduit à un classement correct des occupations du sol considérées dans l'ordre des valeurs de résistance simulées. De plus, le contraste de coût entre l'habitat favorable et l'occupation du sol la plus imperméable était souvent surestimé. Cette faible performance est cohérente avec les résultats issus d'études antérieures. En effet, Peterman *et al.* (2019) ont obtenu une corrélation modérée ( $r = 0.63$ ) entre des surfaces catégorielles 'vraies' et optimisées, qu'ils ont expliqué par le faible taux de recouvrement de l'ordre de

résistance réel des occupations du sol (3 optimisations sur 9 ont conduit à un arrangement conforme). Beninde *et al.* (2023) ont montré que les optimisations réalisées à partir de différents indices génétiques résumant le même jeu de données conduisaient à des surfaces de résistance différentes, en particulier pour des espèces faiblement structurées génétiquement. De plus, ces auteurs ont mis en évidence un effet du profil de résistance sur la corrélation entre les surfaces ‘vraies’ et optimisées. Cette corrélation chutait à une valeur médiane inférieure ou égale à 0.5 dans les cas de forte résistance du paysage, c’est-à-dire lorsque l’ensemble des éléments paysagers, hormis un unique type d’occupation du sol favorable, restreignait le flux génétique. Ces résultats appellent donc à la prudence lors de l’interprétation des valeurs de coût optimisées, en particulier à des fins de conservation, et questionne l’apprentissage de l’algorithme. Quel patron cet algorithme apprend-il lors du processus d’optimisation, si ce n’est pas celui de la distribution des valeurs de coût ?

#### *Les patrons spatiaux capturés par RGA*

Tout d’abord, les performances duales de l’algorithme observées au chapitre 3.1 (entre faible taux de recouvrement du scénario de coût et forte capacité de prédiction des distances génétiques interpolées), suggèrent un sur-ajustement de la distribution spatiale des valeurs de coût à la distribution des données génétique. Plus que le caractère perméable d’un élément paysager, une valeur de coût optimisée semble refléter la proportion et la configuration spatiale de cet élément. Les caractéristiques structurelles de l’élément paysager conditionnent des processus de compensation entre les paramètres à optimiser, rendant plusieurs combinaisons de valeurs de coût toutes aussi performantes pour décrire une distribution spatiale de la variation génétique. De plus, si la structuration génétique est faible, les effets structuraux seront également plus faibles, rendant l’optimisation encore moins contrainte et les potentielles compensations plus fréquentes. Enfin, si la distribution des valeurs de différenciation génétique change, en raison de l’utilisation d’une statistique résumée différente par exemple, la distribution des valeurs de coût optimisées va s’ajuster à ce changement en fonction des contraintes spatiales. Ne pas observer de convergence des surfaces de résistance optimisées, particulièrement pour des espèces faiblement structurées, n’est donc pas forcément surprenant. Il faut également noter que si l’indice de distance génétique influence fortement le paramétrage de la surface de résistance, il n’impacte que très peu la capacité de RGA à retrouver le ‘vrai’ scénario de coût (Beninde *et al.*, 2023). Autrement dit, aucune métrique génétique ne conduit à une amélioration significative de la pertinence écologique de l’optimisation. Cela semble cohérent avec notre hypothèse que seule

la distribution spatiale des éléments paysagers et les contraintes exercées par le niveau de structuration génétique influencent le succès de recouvrement d'un scénario de coût.

Par ailleurs, nous avons démontré la capacité de RGA à prédire des distances génétiques indépendantes, pour des paires de populations situées sur le paysage de calibration du modèle. En revanche, ce pouvoir prédictif diminuait drastiquement lorsqu'il s'agissait de prédire des distances génétiques sur un paysage indépendant des données de calibration (chapitre III-A). Cette mauvaise capacité d'extrapolation est cohérente avec l'hypothèse du sur-ajustement des valeurs de coût au signal de structure spatiale des éléments paysagers. En effet, nos résultats suggèrent une forte dépendance de l'assignation d'une valeur de coût à la proportion et la configuration de l'élément paysager, ainsi que de forts effets de compensation entre éléments considérés. Il n'est donc pas surprenant qu'un scénario de coût optimisé sur un paysage ne soit pas pertinent sur un autre paysage présentant une structure différente des éléments paysagers. La performance décevante de RGA en contexte d'extrapolation a plusieurs conséquences. Tout d'abord, elle empêche la généralisation des inférences de résistance (réalisées avec RGA pour une espèce sur un territoire donné) sur d'autres territoires. Un modèle de résistance est donc entièrement dépendant du paysage d'étude, et si des inférences sont requises pour mettre en place des mesures conservatoires sur une autre zone de l'aire de distribution de l'espèce, un nouvel échantillonnage génétique devra être réalisé.

Nous pourrions également faire l'hypothèse que ces mauvaises performances de recouvrement et de prédiction en contexte d'extrapolation résultent moins d'un effet de sur-ajustement que de l'apprentissage d'un patron spatial capturé par les données de distances mais négligé de notre cadre conceptuel. Il est possible que les données de distances-coût capturent des informations sur la configuration du paysage, qui sont apprises par RGA. En effet, la distribution des valeurs de distances-coût dépend à la fois de la résistance exercée par la matrice (et donc des valeurs de coût), mais également de la distribution des populations (et donc des taches d'habitat) dans le paysage. Un même scénario de coûts peut aboutir à différentes distributions de distances-coût pour deux distributions spatiales de populations différentes. Dans le cas de la Guadeloupe, la forme de l'île impose une topologie des populations particulière, susceptible d'expliquer à elle seule une part significative d'un patron de différenciation génétique. Cette influence de l'hétérogénéité spatiale du réseau de dispersion sur les patrons de diversité, probablement capturée par RGA, est discutée par Savary *et al.* (2023) et Van Strien, (2017). Nous avons projeté, dans le cadre de cette thèse,

d'évaluer cet effet de la forme de l'île, et sa conséquence pour la topologie des réseaux de populations, à partir de simulations démo-génétiques. Nous aurions ainsi pu simuler des données génétiques en considérant le paysage terrestre de la Guadeloupe comme homogène, et en attribuant une très forte résistance à la mer. Nous aurions alors évalué le patron de différenciation génétique obtenu à partir de ces données, et vérifié s'il concordait avec celui d'une divergence BT/GT, comme observé empiriquement pour nos 5 espèces d'oiseaux. Nous avons toutefois manqué de temps pour mener à bien ces analyses.

Il est donc probable que l'approche d'optimisation de RGA repose à la fois sur le patron de résistance de la matrice paysagère et sur celui de la distribution spatiale des populations, c'est-à-dire des taches d'habitat. Cette hypothèse soutiendrait le résultat d'une moindre performance de l'algorithme en contexte d'extrapolation, puisque le scénario optimisé porterait l'information de la configuration de l'habitat propre à un paysage. Nous avons observé une meilleure performance prédictive médiane des distances-coût optimisées par rapport aux distances-coût 'vraies'. Or, un modèle prédisant les distances génétiques à partir des distances-coût 'vraies' et d'une variable de topologie de l'habitat (*e.g.*, indice de 'Betweenness Centrality', 'Mean of Inverse Weight' euclidien) présentait une capacité prédictive médiane équivalente à celle des distances-coût optimisées seules (résultats préliminaires non exposés dans ce manuscrit de thèse). Ces résultats soutiendraient l'hypothèse d'un processus d'optimisation et d'apprentissage fortement guidé par la configuration du paysage de calibration (Figure 6.2).

#### *Qualité de l'apprentissage de RGA à des fins prédictives*

Notre étude invite à une considération nouvelle de l'outil RGA. L'apprentissage automatique de RGA conduit à des prédictions précises des distances génétiques en contexte d'interpolation (Figure 6.2). Ce pouvoir prédictif est une caractéristique des outils d'inférences basées sur les données (Shmueli, 2010; Yates *et al.*, 2018), mais ne semble pas avoir été évalué dans les études passées. RGA est donc un outil intéressant pour prédire la structure génétique d'une population à un instant  $t$  et dans un paysage donné. Cette capacité peut trouver plusieurs applications. Tout d'abord, cela permettrait de limiter l'effort d'échantillonnage coûteux de données génétiques, et ainsi d'augmenter la puissance statistique des inférences en génétique des populations. Une étude de sensibilité permettrait d'identifier le nombre de populations et le protocole d'échantillonnage nécessaires pour inférer avec précision les distances génétiques entre toutes les populations potentiellement abritées dans les habitats cartographiés. RGA serait donc un outil performant pour passer d'un

échantillon de données à des inférences exhaustives. Cela trouve toute son utilité dans les modèles de connectivité requérant des métriques d'isolement entre toutes les taches d'habitats. En effet, paramétrer les liens d'un graphe paysager à partir de distances génétiques inférées par RGA (comme réalisé au chapitre 4) pourrait permettre d'améliorer la pertinence écologique des métriques de connectivité qui en sont dérivées. Cette hypothèse pourrait être évaluée en conduisant une analyse de validation empirique telle que présentée au chapitre 3.2). Cela apporterait un éclairage sur l'usage de RGA à des fins descriptives, afin d'améliorer la compréhension des effets du paysage sur le flux génétique et proposer des mesures de conservation adaptées.

Enfin, notre étude met en avant la capacité de RGA à identifier les moteurs de la structuration spatiale de la variabilité génétique. En effet, lorsqu'un patron d'IBR est détecté à partir du 'vrai' scénario de coût, le scénario de coût optimisé par RGA conduit également dans 90% des cas à la détection d'un patron d'IBR (Ch. 3.1). Si l'utilisation de RGA pour l'interprétation du niveau de perméabilité des éléments paysagers doit rester prudente, cet algorithme peut donc être utilisé pour détecter l'existence d'une résistance du paysage. Cela permettrait d'évaluer si les flux de gènes pour une espèce dans un milieu donné sont contraints par la structure du paysage, et d'orienter en conséquence des études complémentaires de compréhension plus fine de ce patron.

*Optimisation contrainte : une alternative à l'apprentissage automatique ?*

Les résultats de Beninde *et al.* (2023), attestant d'une meilleure performance de recouvrement des coûts lorsque peu d'éléments paysagers influencent le flux génétique, posent une autre question. RGA serait-il performant pour identifier correctement la résistance des éléments paysagers pour des surfaces binaires, c'est-à-dire représentant un type d'habitat perméable et un type d'habitat imperméable aux flux de gènes ? Cela reste à tester plus directement. Les résultats de Peterman *et al.* (2019) et ceux de notre étude (Ch. 3.1) tendent effectivement à confirmer la capacité de ResistanceGA à identifier correctement l'habitat le plus perméable au flux génétique. Cependant, ce résultat n'est pas valide pour les occupations du sol plus résistantes. En effet, nous avons observé que RGA attribuait de façon équiprobable un rang de résistance allant du plus faible au plus fort à l'occupation du sol la plus résistante dans nos simulations. La question de la pertinence de l'utilisation de RGA pour le paramétrage de surfaces binaires se pose toutefois, car cela reviendrait à déployer une méthode coûteuse pour l'optimisation de deux uniques paramètres. Une alternative à l'utilisation de ResistanceGA, notamment pour des profils binaires 'simples', pourrait être

d'adopter une méthode d'optimisation contrainte, consistant à tester l'ajustement aux données génétiques de plusieurs scénarios de coûts construits systématiquement. Peterman *et al.* (2019) ont montré que l'optimisation contrainte de modèles MLPE conduisait à des performances de recouvrement du 'vrai' scénario similaires à supérieures à celles de ResistanceGA pour des surfaces catégorielles. De plus, dans le cadre d'analyses préliminaires liées au chapitre 4 (non présentées dans ce manuscrit de thèse), nous avons observé des performances prédictives équivalentes pour les modèles optimisés par RGA et des modèles issus d'une optimisation contrainte dans un espace de 100 scénarios de coûts. Enfin, nous pouvons comparer les résultats des modèles issus de l'optimisation par RGA (Ch. 4), et ceux issus d'une optimisation contrainte basées sur 114 scénarios (Ch. 4.2) pour la Paruline caféïette. À l'échelle de la Guadeloupe, les  $R^2$  de validation valaient 0.20 avec l'optimisation de RGA et 0.37 pour l'optimisation contrainte. À l'échelle réduite intra-Basse-Terre et intra-Grande-Terre, les  $R^2$  de validation valaient 0.49 et 0.51 avec RGA et l'optimisation contrainte respectivement. En dépit des différences méthodologiques (résolutions et types de distances écologiques différentes), et du fait que l'on traite ici de pouvoir prédictif et non de capacité de recouvrement d'un scénario de coût, nous pouvons souligner la performance équivalente, si ce n'est meilleure, de l'optimisation contrainte par rapport à celle de ResistanceGA. Cela pourrait être lié à la convergence de RGA vers un optimum à travers l'espace des paramètres de coûts. Si l'exploration est insuffisante, l'algorithme pourrait rester bloqué vers un optimum local, moins pertinent que celui de la réalité écologique. Dans ces analyses, nous n'avons pas exploré les paramètres par défaut de l'algorithme déterminant le comportement d'exploration et de convergence, tel que le taux de mutations. Cela pourrait être une piste à tester, de même que l'apport coût-bénéfice de l'optimisation objective et non contrainte de ResistanceGA par rapport à une optimisation réduite à une gamme de scénarios systématiques.



### 6.1.1.3. Les limites de notre étude

#### *Le type de distance écologique*

Nous avons utilisé dans cette étude des distances-coût (Adriaensen *et al.*, 2003) et non des distances de résistance dérivées de la théorie des circuits (McRae, 2006). Nos simulations démo-génétiques et modèles d'optimisation font donc l'hypothèse d'une espèce focale spécialiste de l'habitat forestier ayant une bonne connaissance *a priori* du territoire et se déplaçant le long des chemins de moindre coût. Nous avons fait ce choix pour des raisons de simplicité d'implémentation du module de calcul de distances-coûts dans le programme informatique de ResistanceGA, ainsi qu'en raison de la rapidité des temps de calcul. Cependant, comme mentionné précédemment, la pertinence écologique des distances-coût ou des distances de résistance semble varier en fonction des analyses et des modèles d'études (McClure *et al.*, 2016; Zeller *et al.*, 2018). Les deux types de distances écologiques diffèrent non seulement par leur hypothèse sous-jacente sur la perception du paysage par l'espèce focale, mais aussi par leur sensibilité à la résolution spatiale de l'étude, à la distribution des populations, ainsi qu'à l'agrégation du paysage (Marrotte et Bowman, 2017). Nous pouvons toutefois faire l'hypothèse que cela ne modifierait pas fondamentalement l'orientation de nos conclusions, car ces dernières sont en partie soutenues par les précédentes études de Beninde *et al.*, (2023) et de Peterman *et al.*, (2019) ayant toutes deux utilisé des distances 'commute-time' analogues aux distances de résistance issues de la théorie des circuits.

#### *Sensibilité à la résolution thématique*

Dans le cadre d'études préliminaires aux analyses du chapitre 4 (non présentées dans ce manuscrit), nous avons constaté une sensibilité des capacités de prédiction de RGA à la résolution thématique : le passage d'une résolution de cinq à neuf types d'occupation du sol augmentait fortement les capacités de prédiction des modèles, proches de zéro avec la résolution la plus grossière. Cette sensibilité questionne les résultats obtenus dans le chapitre 3.1, où nous avons considéré une résolution thématique de quatre occupations du sol pour évaluer la validité de RGA. L'effet de la résolution thématique n'a pas été testé ouvertement dans les précédentes études d'évaluation de l'algorithme, et son exploration serait donc intéressante afin de poursuivre l'identification des conditions optimales d'utilisation de cet outil dans le cadre de futures études empiriques.

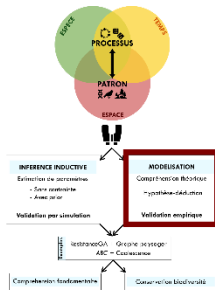
*Effet de l'indice de différenciation génétique*

Par ailleurs, si Beninde *et al.* (2023) ont évalué l'effet de la métrique de distance génétique sur les performances de recouvrement du 'vrai' scénario de coût lors de l'optimisation d'une surface de résistance, l'effet de cette métrique sur les performances prédictives des modèles optimisés reste à évaluer. Nos résultats avaient montré une absence de dépendance des capacités prédictives d'un modèle de résistance à la congruence entre scénario de coût 'vrai' et optimisé. Nous pouvons donc faire l'hypothèse que l'absence de congruence des surfaces de résistance pour différentes métriques génétiques, mise en évidence empiriquement par Beninde *et al.* (2023), ne signifie pas forcément qu'une métrique conduira à de meilleures performances prédictives, de la même façon qu'aucune ne conduit à un meilleur recouvrement du 'vrai' scénario.

Dans cette partie, nous nous sommes attardés sur la façon de valider une méthode d'inférence de paramètres de connectivité et sur de l'intérêt de cette validation. Nous allons maintenant nous intéresser à la validation, non pas d'une méthode en elle-même, mais du cadre théorique dans lequel cette méthode est susceptible d'être utilisée. En particulier, nous verrons comment valider empiriquement le modèle des graphes paysagers, traduisant notre compréhension de l'effet de l'hétérogénéité du paysage sur le flux génétique.

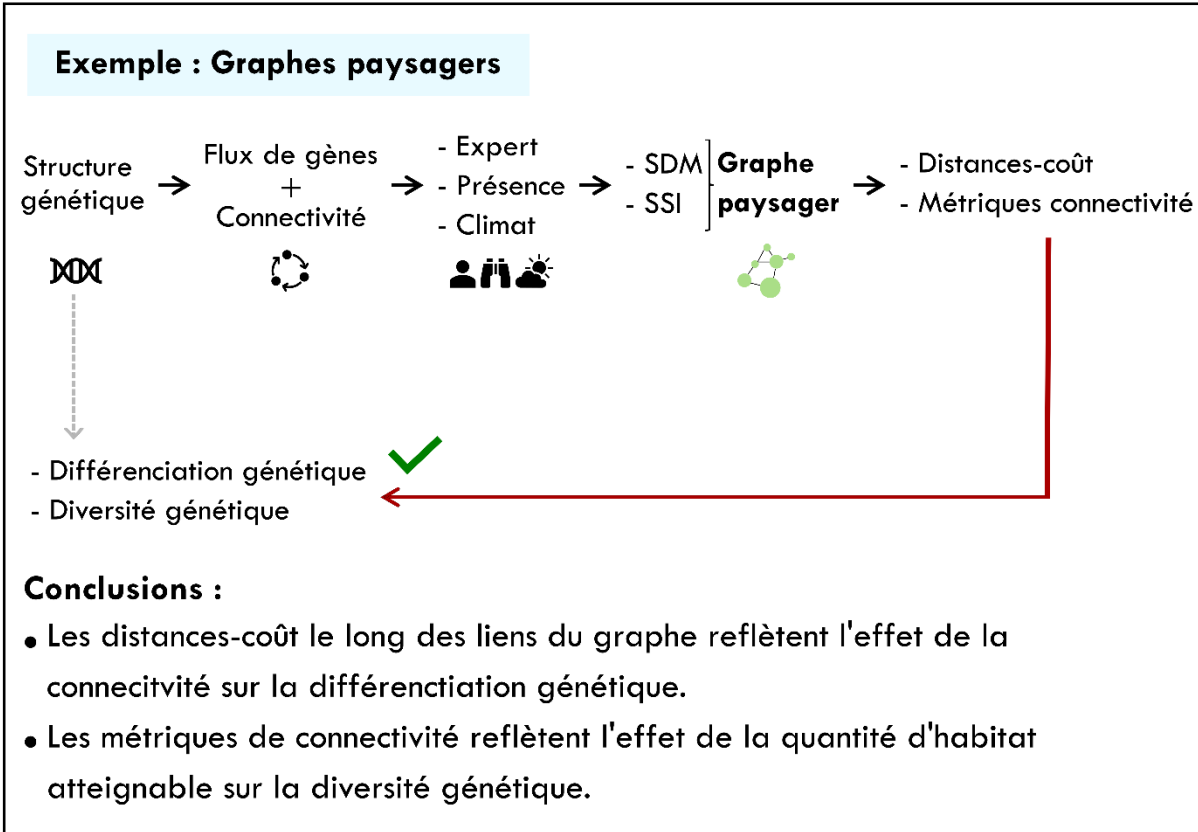
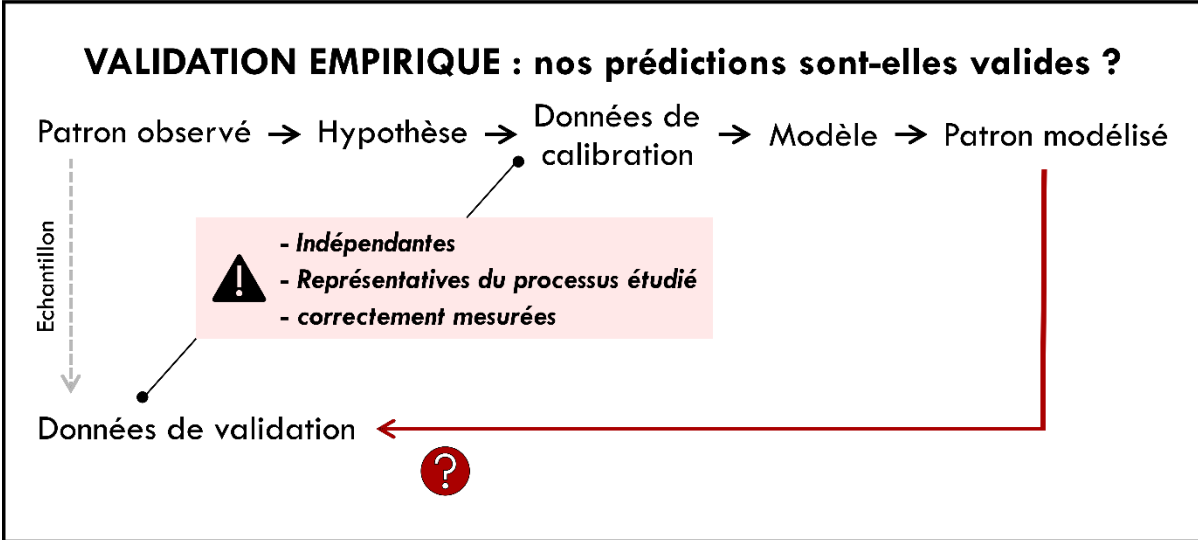
### 6.1.2. Validation empirique des modèles de connectivité

Une approche de validation empirique permet de valider un modèle théorique d'un processus écologique. Dans cette thèse, nous avons exploré la validité de notre compréhension de l'impact de la connectivité des habitats sur la dispersion des individus (Ch. 3.2). Pour cela, nous avons modélisé un réseau de dispersion potentiel contraint par l'hétérogénéité du paysage, en utilisant le cadre théorique des graphes paysagers. Notre hypothèse était que la quantité d'habitat atteignable au voisinage d'une population déterminait la fréquence des événements de dispersion, depuis et à destination de cette population. Pour valider cette prédiction, nous avons confronté les métriques de connectivités issues de notre modèle conceptuel de dispersion à des données empiriques supposées refléter les conséquences biologiques de cette dispersion, à savoir des données génétiques traduisant le processus de flux génétique. Les bons niveaux de corrélations entre les métriques de connectivité et les données génétiques tendaient à valider notre hypothèse relative à l'influence de la connectivité sur le processus dispersion, ainsi que le modèle théorique des graphes paysagers. La robustesse d'une approche de validation repose sur deux points : la qualité du modèle conceptuel d'une part, et la qualité des données de validation d'autre part. Dans cette partie, nous détaillerons donc ces deux points, avant de conclure sur la validité des graphes paysagers comme modèle de connectivité des habitats (Figure 6.3).



## MODELISATION

### Compréhension théorique



**FIGURE 6.3 :** Focus depuis la figure 6.1 sur l'approche de modélisation. En génétique du paysage, un modèle est la transcription de notre compréhension de la relation entre patrons génétiques et paysagers et processus éco-évolutifs. Évaluer empiriquement un modèle de connectivité revient à tester les hypothèses émises sur cette relation. Par exemple, le modèle des graphes paysagers traduit l'hypothèse que la quantité d'habitat atteignable au voisinage d'une population influence les processus de dispersion et détermine son niveau de variabilité génétique.

### 6.1.2.1. Calibration du modèle : comment modéliser le processus de dispersion en interaction avec le paysage ?

#### *Le type de données*

La première étape est donc de calibrer un modèle traduisant une théorie, à savoir ici que la quantité d'habitat disponible autour d'une population conditionne les événements de dispersion (Figure 6.3). Le cadre conceptuel des graphes paysagers offre la possibilité de calibrer un tel modèle sans nécessairement utiliser de données biologiques. Toutefois, plusieurs études suggèrent que la pertinence écologique des graphes paysagers pourrait être influencée par leur couplage avec des données empiriques (Clauzel et Godet, 2020; Foltête *et al.*, 2020). De façon générale, le niveau de validité des modèles de connectivité semble dépendre du type de données de calibration (Zeller *et al.*, 2018). Un grand nombre des modèles évalués dans la littérature reposent sur des surfaces de connectivité paramétrées par des fonctions de sélection de ressources utilisant des données de présences ou de déplacement (Aiello *et al.*, 2023; Bourdouxhe *et al.*, 2020; Clauzel et Godet, 2020; Laliberté et St-Laurent, 2020). Selon Beier *et al.* (2008), les données génétiques ou de déplacement seraient les plus performantes pour modéliser la résistance du paysage au déplacement. Cette hypothèse est soutenue empiriquement par les résultats de Zeller *et al.* (2018). Dans notre étude (Ch. 3.2), les modèles calibrés avec des données de présences présentaient une performance équivalente au modèle calibré à dire d'expert et indépendant de données biologique. Bien que perçu comme arbitraire, le paramétrage selon les savoirs d'experts peut donc conduire à une modélisation valide de l'effet des connectivités (Koen *et al.*, 2014). Le couplage de cette méthode avec des données biologiques présente de plus un potentiel pour optimiser le compromis entre qualité et complexité des modélisations (Laliberté et St-Laurent, 2020).

#### *Comparaison des méthodes de calibration*

Nos résultats ont montré que les méthodes de calibration des modèles, davantage que les types de données, pouvaient conduire à des performances inégales. En effet, le graphe calibré par un indice de spécialisation était associé à une capacité prédictive plus élevée ou plus faible, en fonction de l'échelle spatiale, que celui calibré par un modèle de distribution d'espèce. Pourtant, ces deux modèles reposaient sur les mêmes données de présence. L'étude de l'influence des méthodes de calibration sur la validité des modèles n'est pas si fréquente dans la littérature (mais voir Bond *et al.*, 2017; Zeller *et al.*, 2018). Les études comparatives les plus répétées concernent les algorithmes de modélisation des corridors selon les distances-circuit ou les distances de moindres coûts (Bond *et al.*, 2017; Laliberté et St-Laurent, 2020;

McClure *et al.*, 2016). Dans certains cas, les distances-circuit apparaissent comme plus pertinentes que les distances-coût (Laliberté et St-Laurent, 2020), tandis que dans d'autres la tendance s'inverse (Zeller *et al.*, 2018). McClure *et al.* (2016) ont suggéré une dépendance de la performance de ces deux algorithmes à l'écologie du modèle d'étude. Dans leurs travaux, les distances-coût conduisaient à une meilleure inférence des routes de dispersion des élans en raison de leur fidélité aux itinéraires ancestraux de migration, tandis que les distances circuits étaient plus pertinentes pour modéliser les déplacements des loups, dont la dispersion juvénile s'effectue en terrain inconnu. Dans notre étude, nous avons uniquement évalué des graphes paysagers pondérés par des distances-coût. Il est possible de calculer des distances de résistance le long des liens d'un graphe, en l'assimilant à un circuit électrique dont chaque lien représente une résistance, et il serait intéressant d'évaluer la robustesse de nos résultats avec ce type de distances-circuit. On peut cependant noter que ces deux méthodes couramment utilisées peuvent être considérées comme aussi performantes l'une que l'autre (Simpkins *et al.*, 2018), et qu'intégrer ainsi l'effet de la matrice paysagère dans les modèles de connectivité n'est pas systématiquement pertinent. Ainsi, Poli *et al.* (2020) ont démontré que la métrique de connectivité  $dPC_{flux}$  expliquait aussi fortement les taux de migration d'un insecte, qu'elle soit dérivée d'un graphe pondéré par des distances euclidiennes, donc indépendantes de la résistance exercée par la matrice, que d'un graphe pondéré par des distances écologiques (coûts et circuits).

#### *Modélisation de la dispersion par une métrique de distance*

Une majorité des études de génétique du paysage modélisent le processus de dispersion par des distances, en se positionnant à l'échelle des liens entre populations (Storfer *et al.*, 2010). Toutefois, la capacité d'une métrique de distance à résumer des variations spatiales peut être questionnée. En effet, Savary *et al.* (2022a) ont montré que deux scénarios de coût différents, aboutissant à des chemins de moindre coût spatialement distincts, peuvent engendrer des matrices de distances-coût fortement corrélées. De plus, nous avons montré (III-A.) que des distances-coût optimisées, très ajustées à des données génétiques simulées, pouvaient dériver de scénarios de coûts différents de ceux utilisés pour la simulation de ces données génétiques. Ces résultats suggèrent donc une redondance et un manque de spécificité des métriques de connectivité basées uniquement sur des distances. D'après le cadre conceptuel défini par Wagner et Fortin (2013) et DiLeo et Wagner (2016), les modélisations au niveau des nœuds seraient très pertinentes pour évaluer l'effet de l'hétérogénéité du

paysage sur le processus de dispersion et le patron variabilité génétique en dérivant. Les graphes paysagers sont particulièrement adaptés à ce niveau d'analyse. Nous avons ainsi montré (III-B) que trois métriques de connectivité locale, décrivant la taille de la tache d'habitat, la quantité d'habitat atteignable à son voisinage et son importance pour la traversabilité du réseau, reflétaient la richesse allélique et le niveau relatif de différenciation génétique des populations de Paruline caféïette. D'autres études empiriques ont mis en évidence des relations entre diversité et différenciation génétique locale et quantité et fragmentation de l'habitat (Barr *et al.*, 2015; Flavenot *et al.*, 2015; Peterman *et al.*, 2015; Savary *et al.*, 2022c). Un intérêt de ces métriques de connectivité locales réside dans leur prise en compte des effets conjoints de la configuration du paysage, de la résistance de la matrice et de la quantité d'habitat atteignable. La quantité d'habitat influence directement la taille efficace des populations, qui conditionne leur sensibilité aux effets de la dérive génétique, tandis que la configuration de l'habitat et la résistance de la matrice influencent les processus de dispersion, et donc le flux génétique. Ainsi, la modélisation au niveau des nœuds considère l'effet des deux principales forces évolutives de la diversité neutre (la dérive et le flux de gène), et permet de décrire les deux composantes de la structure génétique (la diversité et la différenciation génétique). En dépit de résultats prometteurs, une faible proportion des études de génétique du paysage évaluent l'effet du voisinage des taches, et beaucoup négligent donc la modélisation des effets de la quantité d'habitat au profit de ceux de la configuration du paysage et de la résistance de la matrice (DiLeo et Wagner, 2016). Pourtant, Prunier *et al.* (2017) ont montré à partir de données simulées et empiriques que la prise en compte d'un proxy de la taille efficace des populations ( $N_e$ ), telle que la quantité d'habitat, permettait d'expliquer une part importante de la différenciation génétique entre deux populations, en particulier dans des conditions de faible taux de migration et de forte hétérogénéité de taille des populations.

### 6.1.2.2. Qualité des données de validation : les données empiriques, gage d'une validation indépendante reflétant la réalité du terrain

Une fois le modèle traduisant notre compréhension du processus de dispersion en milieu hétérogène calibré, la seconde étape consiste à le valider. La validation empirique permet de confronter notre prédiction à la réalité du terrain, en comparant des métriques issues d'un modèle avec des données biologiques reflétant les conséquences du processus modélisé. Dans notre cas, nous avons modélisé le processus de connectivité lié à la dispersion des individus, et avons validé les métriques de connectivités issues de ces modèles par des données génétiques reflétant le processus de dispersion. Bien que Zeller *et al.* (2018) aient démontré que les données génétiques reflétaient correctement l'effet des connectivités paysagères sur ce processus de dispersion, notre étude figure parmi les rares à avoir utilisé ce type de données pour valider des modèles de connectivité (voir aussi Savary *et al.*, 2022). Les données biologiques de validation correspondent plus couramment à des données de déplacement relevées par télémétrie, ou à des données de présence issues de pièges, d'inventaires protocolés ou, à défaut, de relevés naturalistes plus opportunistes (Bond *et al.*, 2017; Carroll *et al.*, 2020; Koen *et al.*, 2014; Laliberté et St-Laurent, 2020; McClure *et al.*, 2016). L'interprétation du niveau de validité des modèles de connectivité dépend cependant de la capacité de ces données de validation à refléter la dispersion effective des individus. Cette capacité dépend de trois facteurs principaux (Figure 6.3). Tout d'abord, les données de validation doivent refléter un patron principalement influencé par la dispersion. Ainsi, se baser sur des distances génétiques peut être pertinent si le patron de différenciation génétique est majoritairement influencé par la dispersion, et non par la dérive ou l'histoire phylogéographique des populations. En revanche, l'utilisation de données d'occurrence peut sembler moins évidente car de nombreux facteurs indépendant de la dispersion, tels que la disponibilité des ressources, les conditions climatiques ou la qualité des habitats, sont susceptibles d'avoir un effet prépondérant sur la distribution des individus par rapport au processus de dispersion (Keeley *et al.*, 2017). Le deuxième facteur déterminant la qualité des données de validation concerne la fiabilité de leur protocole d'échantillonnage. Ces données doivent en effet être des proxys fiables du patron qu'elle représente. Par exemple, des taux de détection trop faibles par pièges photographiques ne constituent pas un proxy fiable des événements de dispersion et seront ainsi peu liés aux métriques de connectivité (Laliberté et St-Laurent, 2020). Dans le cas de nos données génétiques, nous nous sommes assurés de disposer d'un nombre suffisant d'individus et de populations échantillonnées, et avons



contrôlé le niveau de polymorphisme et les éventuels déséquilibres de liaison des *locus* génotypés. Enfin, le dernier point d'importance concerne l'indépendance des données de validation par rapport aux données de calibration. Cette indépendance garantit la robustesse du modèle de connectivité et la fiabilité de ses prédictions. Dans notre étude (Ch. 3.2), les données génétiques de validation étaient donc indépendantes des données de présences utilisées pour la calibration des modèles. Peu d'études ont effectué des validations indépendantes des modèles de connectivité (voir Bond *et al.*, 2017; Koen *et al.*, 2014; Laliberté et St-Laurent, 2020; Savary *et al.*, 2022b). Outre la question de la pertinence et de la qualité des données de validation, l'approche de validation dépend de la méthode utilisée pour mettre en relation les données du modèle et les données empiriques. Il existe une diversité de méthodes et de métriques de validation. McClure *et al.* (2016) ont par exemple formalisé une méthode de validation des surfaces de résistance reposant sur trois métriques complémentaires comparant les valeurs empiriques de connectivité (e.g. taux de d'occurrence) avec les valeurs de connectivité inférées au niveau des points de d'occurrence et de points aléatoires sur le paysage. Les surfaces de résistance n'étant qu'une base de construction pour les graphes paysagers pondérés, nous n'avons pas adopté la même démarche de validation. Notre approche a consisté à évaluer i) la corrélation entre les métriques de diversité génétique et les métriques de connectivité au niveau des nœuds du graphe, et ii) la dépendance des distances génétiques aux distances-coût le long des liens du graphe.

### 6.1.2.3. Conclusions et perspectives quant à la validité des graphes paysagers

La question de la validité écologique des modèles de connectivité occupe une place grandissante dans la littérature. Notre travail propose une approche rigoureuse permettant d'évaluer la capacité de métriques de connectivité dérivées des graphes paysagers à refléter l'influence de l'hétérogénéité du paysage sur le flux génétique entre les populations d'un passereau forestier (Figure 6.3). Quel que soit les méthodes et les données de calibration utilisées, les trois modèles de connectivité issus des graphes paysagers présentaient de capacités de prédiction correctes des distances génétiques ( $R^2$  de validation  $\sim 0.25$ ), ainsi que de bons niveaux de corrélation avec les indices de diversité génétique ( $r_{\text{Spearman}} > 0.5$ ), une fois l'échelle optimale identifiée (Figure 6.3). Ces niveaux de performance rejoignent ceux d'autres modèles. Laliberté et St-Laurent (2020) ont ainsi obtenu une corrélation de 0.61 entre la densité de collision et le niveau de connectivité du paysage pour des populations d'élan, à

partir d'une surface de densité de courant paramétrée par des données d'expert et de télémétrie. Poli *et al.* (2020) ont obtenu un pseudo  $R^2$  de 0.44 pour le modèle expliquant des taux d'immigration par la métrique  $dPC_{flux}$  issue des graphes paysagers.

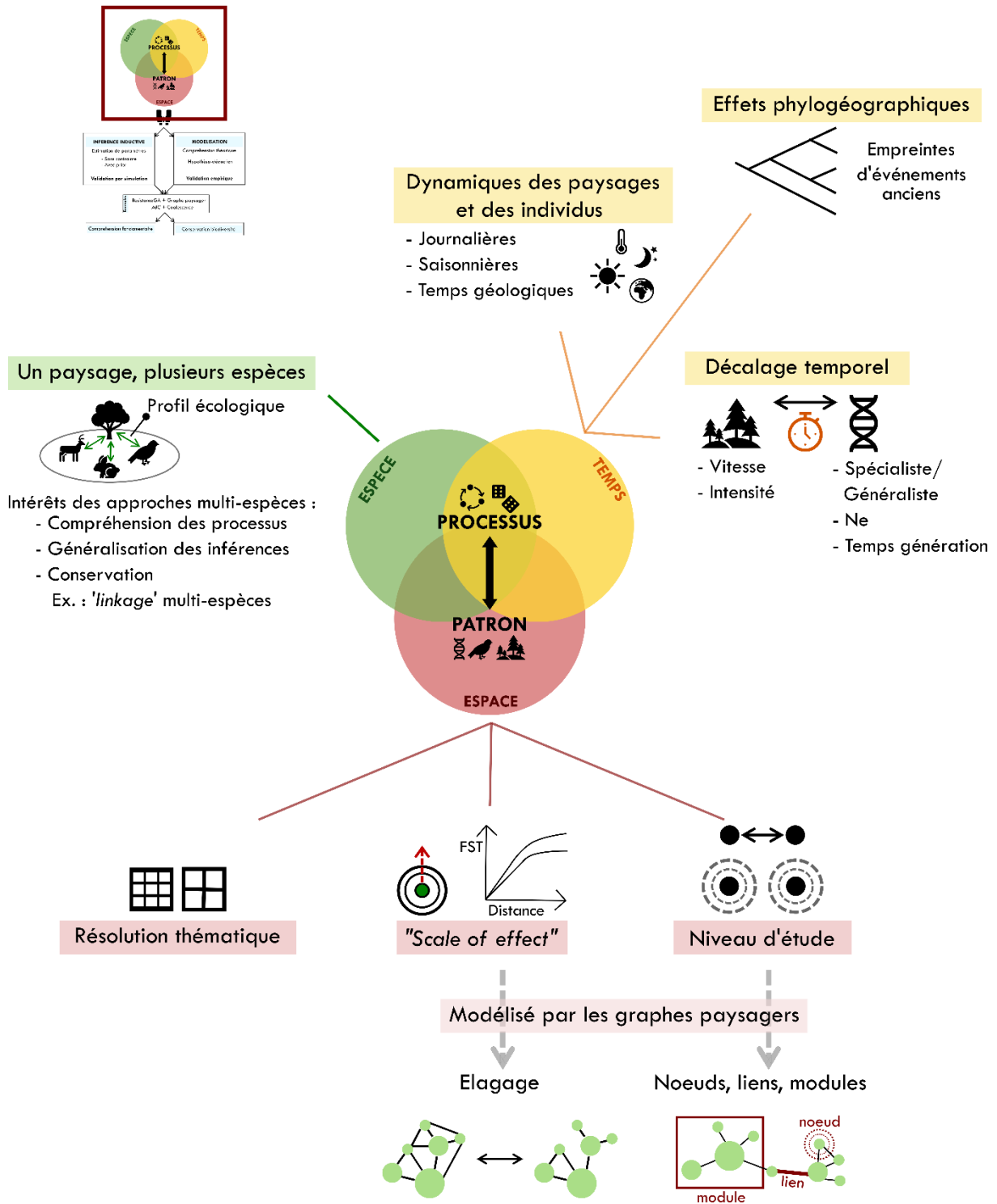
Ces résultats prometteurs concernent une espèce très spécialiste de l'habitat forestier. L'identification des taches d'habitat pour la construction d'un graphe paysager est donc dans ce cas facilitée. Le caractère insulaire et restreint de la zone d'étude permet de plus une modélisation exhaustive du domaine vital de cette espèce, s'affranchissant d'éventuels effets de bords. Enfin, la stabilité du couvert forestier depuis plusieurs dizaines à centaines d'années, en particulier sur la Basse-Terre, favorise l'ancrage de la signature paysagère dans la structure génétique de la Paruline caféïette. La sensibilité du flux de gènes à l'hétérogénéité du paysage avait déjà été démontrée pour cette espèce (Khimoun *et al.*, 2017), et sa structure génétique contemporaine porte déjà les traces des modifications du paysage des dernières décennies (Ch. 5). Toutes ces conditions sont très favorables à la modélisation de l'effet des connectivités sur les flux de gènes. Elles nous assuraient également un certain niveau de confiance dans les prédictions testées par nos modélisations. Pour généraliser la pertinence écologique des graphes paysagers, il serait donc nécessaire de réaliser de nouvelles études de validation basées sur des jeux de données empiriques et indépendants, et couvrant une diversité d'espèces et de paysages. Ainsi, Savary *et al.* (2022) ont observé des corrélations entre des métriques de connectivité issues de graphes et des indices de diversité génétique pour des populations de criquet ensanglanté (*Stethophyma grossum*) plus faibles que celles observées dans le chapitre 3.2 pour la Paruline caféïette. Cet écart de performance pourrait provenir de la différence d'écologie entre les deux espèces ainsi que des caractéristiques du paysage étudié. L'influence du profil écologique de l'espèce cible est soulignée par Driezen *et al.* (2007), qui interprètent la faible validité écologique de leur modèle de connectivité du hérisson commun (*Erinaceus europaeus*) comme résultant du caractère très généraliste de cette espèce.

La validité d'un modèle de connectivité est ainsi généralement évaluée autour d'un couple unique d'une espèce et d'un paysage. Mais ce modèle conserve-t-il sa validité lorsqu'il est appliqué à une autre espèce, un autre paysage ou sur une autre période temporelle ? Dans la partie suivante, nous nous intéresserons au domaine de validité des modèles de connectivité, selon les espèces, l'espace et le temps.

## 6.2. Domaine de validité spatial, temporel et multi-spécifique des modèles de connectivité

La distribution spatiale de la variabilité génétique résulte de l'interaction entre différents patrons et processus écologiques et évolutifs, dynamiques dans le temps et dans l'espace (Wagner et Fortin, 2013) (Figure 6.4). Les processus écologiques, tels que les interactions biotiques et les déplacements d'individus, et les processus évolutifs, comme la dérive et le flux génétique étudiés ici, sont influencés par le caractère spatialement et temporellement hétérogène des paysages dans lesquels ils se déroulent. L'hétérogénéité des paysages dépend de la distribution spatiale et de la qualité des taches d'habitat, ainsi que de la structure et de la résistance au déplacement exercée par la matrice paysagère. Cette hétérogénéité spatiale varie également dans le temps, au gré des changements ou des perturbations environnementales d'origines naturelles (e.g., climatiques, météorologiques) ou anthropiques. L'ensemble de ces patrons paysagers influent sur les processus éco-évolutifs et spatio-temporels que sont les flux de gènes et la dérive génétique, qui à leur tour conditionnent le patron de biodiversité auquel nous nous sommes ici intéressés : la distribution spatiale et temporelle de la variabilité génétique. De plus, ces processus éco-évolutifs n'affectent pas de la même façon toutes les espèces (Figure 6.4). Par exemple, en fonction de leurs traits écologiques, tels que leur capacité de dispersion ou leur niveau de spécialisation, les espèces ne seront pas sujettes au même équilibre entre migration et dérive, ou ne seront pas soumises au même niveau de compétition interspécifique. Ces traits écologiques conditionnent donc la sensibilité des espèces à la connectivité du paysage, et l'échelle spatiale et temporelle à laquelle ils interagissent avec ce paysage.

En résumé, l'effet de l'hétérogénéité du paysage sur la distribution de la variabilité génétique présente une dépendance forte aux paramètres d'échelles spatiales et temporelles, et la forme de cette dépendance n'est pas identique pour toutes les espèces (Figure 6.4). Les modèles étudiant cet effet du paysage sur la structure génétique des populations sont donc susceptibles de présenter un domaine de validité multifactoriel, défini par les espèces, le temps et l'espace. Dans cette partie, nous nous intéresserons à la définition de cet espace de validité, et à la prise en compte explicite ou implicite des effets spatio-temporels et multi-spécifiques dans les inférences de génétique du paysage.



**FIGURE 6.4 :** Focus depuis la figure 6.1 sur les dynamiques spatiales, temporelles et multi-spécifiques de l'interaction entre patron et processus, contraignant le domaine de validité des modèles de connectivité. Cette figure illustre les possibilités de prise en compte des effets spatio-temporels et spécifiques par ces modèles de connectivité.

## 6.2.1. Considération de l'échelle spatiale dans les modèles de connectivité

### 6.2.1.1. Résolution thématique

Un premier niveau de considération de l'échelle spatiale réside dans les choix de modélisation des patrons paysagers (Figure 6.4). Ainsi Cushman et Landguth (2010) ont mis en évidence les effets de la résolution spatiale et thématique sur les relations de génétique du paysage. En particulier, le passage d'une surface continue à une surface catégorielle à la résolution dégradée conduisait, selon leurs simulations, à une chute drastique de la corrélation entre distances génétiques et écologiques. Dans le cadre d'études préliminaires au paramétrage des surfaces de résistance (Ch. 4), nous avons effleuré cette problématique en observant une relation plus forte entre les distances-coût optimisées et distances génétiques en affinant la résolution thématique.

### 6.2.1.2. 'Scale of effect' et seuil de distances

Plusieurs études de génétique du paysage abordent également les effets d'échelles en reproduisant les analyses selon des emprises paysagères de plus en plus étendues, évaluant ainsi l'effet des connectivités à différentes échelles locales, subrégionales, et régionales (Aylward *et al.*, 2020). L'objectif ne réside alors généralement pas dans l'identification d'une échelle spatiale propre à un processus, mais dans une description des facteurs paysagers influençant la variabilité génétique conditionnellement à l'échelle considérée. Toutefois, les relations espèces-paysage pouvant être masquées par l'usage d'une étendue spatiale inadéquate (Fahrig, 2013; Jackson et Fahrig, 2014), certaines études cherchent à intégrer dans leur modèle des variables paysagères calculées à l'échelle spatiale conduisant à la détection optimale de leur effet sur un patron de diversité (*i.e.*, la 'scale of effect') (McGarigal *et al.*, 2016; Moraga *et al.*, 2019; Thompson et McGarigal, 2002) (Figure 6.4). L'utilisation de fenêtres circulaires de diamètre croissant est une méthode couramment utilisée pour l'optimisation de cette 'scale of effect' (Shirk *et al.*, 2014). Dans notre étude sur la validation des graphes paysagers (Ch. 3.2), nous avons optimisé cette échelle spatiale grâce aux propriétés des métriques issues de ces graphes. Ces métriques permettent, comme les fenêtres circulaires, de considérer des étendues paysagères croissantes, tout en intégrant des données sur la capacité de dispersion de l'espèce focale, sur la distribution des taches d'habitat et sur la

résistance de la matrice. La ‘*scale of effect*’ est difficilement prédictible, et semble dépendre de l’interaction entre certains traits spécifiques relatifs à la mobilité et la démographie, ainsi qu’à la nature des variables paysagères et biologiques mises en relation (Miguet *et al.*, 2016; Moraga *et al.*, 2019). Nous avons pu observer cette dépendance dans le chapitre portant sur à l’évaluation de ResistanceGA (Ch. 3.1). En effet, les résultats des simulations génétiques avaient mis en évidence une influence du profil écologique de l’espèce et de ses capacités de dispersion sur l’échelle spatiale conduisant à la plus forte relation entre distances génétiques et paysagères. Par exemple, une espèce généraliste devait présenter de très faibles capacités de dispersion pour que la résistance du paysage soit un facteur limitant du flux génétique à l’échelle de notre fenêtre d’étude de 40 km<sup>2</sup>.

Ces résultats invitent à considérer des distances seuils au-delà desquelles le paysage n’influence plus la structure génétique (Figure 6.4). Dans un paysage homogène, les populations les plus proches spatialement auront tendance à être également les plus proches génétiquement. Un patron de différenciation génétique des populations lié à un isolement par la distance (IBD) peut ainsi être assimilé à une autocorrélation spatiale des fréquences alléliques, liée à l’existence d’un flux génétique dont l’intensité dépend de la distance euclidienne entre les populations. Une autocorrélation nulle ou négative, en revanche, traduit un flux génétique limité ou nul, et une structure génétique reposant majoritairement sur les effets aléatoires de la dérive. Un corrélogramme de Mantel permet de déterminer l’échelle de cette autocorrélation (Palm *et al.*, 2023; Van Strien *et al.*, 2015; Wagner *et al.*, 2005), et ainsi, la taille du paysage local à considérer pour évaluer l’effet du paysage sur le flux génétique. Il est également possible de décliner ce type d’analyses pour les patrons d’isolement par la résistance du paysage, en évaluant l’échelle spatiale correspondant à la corrélation maximale entre distances génétiques et paysagères (Van Strien *et al.*, 2015). Dans notre étude sur la validation empirique du graphe paysager (Ch. 3.2), nous avons adopté une méthode apparentée, basée sur l’évaluation de la relation entre distances-coût et distances paysagères pour différents seuils successifs de distances-coût. Toutefois, la nature des graphes paysagers nous a permis d’explorer d’autres types de partitionnement de données, non conditionnés par une distance seuil ou un niveau d’autocorrélation spatiale. Ces partitionnements reposent sur le processus d’élagage des liens du graphe, qui offre un cadre vaste et flexible pour ajuster l’échelle spatiale d’étude à celle où le paysage affecte la structure génétique (Savary *et al.*, 2021).

### 6.2.1.3. Niveaux d'analyse

Enfin, l'échelle spatiale peut également être modélisée par le niveau de mise en relation des données génétiques et paysagères (Wagner et Fortin, 2013) (Figure 6.4). Kierepka *et al.* (2020) ont ainsi comparé les effets de l'hétérogénéité du paysage sur la diversité et la différenciation génétique des tamias, à l'échelle locale de leur tache d'habitat, et à l'échelle régionale de l'ensemble du paysage échantillonné. Ils ont mis en évidence une variation des effets paysagers sur les composantes de la structure génétique en fonction de l'échelle spatiale, la configuration des habitats influant par exemple sur la diversité génétique à l'échelle régionale uniquement. DiLeo et Wagner (2016) ont discuté de ces interactions entre niveaux d'analyse, dont l'exploration est facilitée par le cadre théorique des graphes paysagers (Savary *et al.*, 2022b).

### 6.2.1.4. L'apport des graphes paysagers pour la prise en compte de l'échelle spatiale dans les inférences de génétique du paysage

#### *Validation multi-niveau de l'effet des connectivités sur la structure génétique*

Les graphes du réseau d'habitat de la Paruline caféïette (Ch. 3.2) nous ont permis d'évaluer les effets de la connectivité à trois niveaux spatiaux, et sur les deux composantes de la structure génétique (*i.e.*, la différenciation et la diversité génétique). Ainsi, un premier apport des graphes paysagers est leur faculté à rassembler sous un seul modèle cinq niveaux potentiels d'analyse : celui de la tache d'habitat (*'node level'*, d'après Wagner et Fortin, 2013), celui du paysage local autour d'une tache d'habitat (*'neighborhood level'*), celui des corridors entre taches d'habitat (*'link level'*), celui d'un ensemble de taches d'habitat constituant un module du réseau (*'boundary level'*), et enfin celui du réseau d'habitat dans son ensemble (niveau global) (Figure 6.4). Dans notre analyse, nous avons montré que la diversité et la différenciation génétique relative d'une population étaient influencées par la capacité des taches d'habitat (*'node level'*) et la quantité d'habitat atteignable au voisinage de cette population (*'neighborhood level'*). Dans un second temps, nous avons également mis en évidence le lien entre distances génétiques et distances-coût le long des liens du graphe (*'link level'*) (Savary *et al.*, 2022b).

### *Vers une validation écologique des modules des graphes paysagers*

Pour explorer la validité de la modélisation au niveau des modules d'un graphe paysager, nous pourrions comparer le partitionnement des taches d'habitat sur le graphe paysager avec le partitionnement des populations en clusters génétiques (par exemple *via* l'approche de *clustering* bayésienne mise en œuvre par l'algorithme STRUCTURE (Pritchard *et al.*, 2000). En effet, des barrières paysagères imperméables au déplacement des individus peuvent empêcher le flux génétique entre les populations se situant de part et d'autre (Gauffre *et al.*, 2015), et ainsi les structurer selon un patron d'Isolement par les Barrières (IBB). Si les modules d'un graphe paysager coïncident avec la partition des populations en clusters génétiques, alors l'étude du graphe permettra d'identifier la nature des éléments paysagers impliqués dans la barrière et conduisant à l'existence de ces clusters génétiques (Savary *et al.*, 2022c). Dans le contexte du réseau des habitats de l'île de Guadeloupe, on pourrait ainsi modéliser le rôle de barrière qu'exerce l'isthme entre Basse-Terre (BT) et Grande-Terre (GT), potentiellement à l'origine des deux principaux clusters génétiques observés. Il s'agirait de tester l'existence d'un scénario de coût conduisant à une partition du graphe paysager en deux modules, concordant avec les taches d'habitat de GT d'une part, et de BT d'autre part.

### *La propriété d'élagage des graphes paysagers*

Au niveau du paysage local, les métriques de connectivité dérivées des graphes paysagers permettent d'adapter l'échelle spatiale considérée en fonction des capacités de dispersion de l'espèce focale, et de sa perception de la résistance du paysage. Au niveau des liens entre populations, l'élagage du graphe permet de contrôler l'échelle spatiale des connexions entre populations (Figure 6.4). L'élagage permet de ne retenir qu'un sous-ensemble de liens entre populations ou taches d'habitat, en fonction d'une hypothèse de départ sur leur connexion. Ainsi, élaguer les liens selon un seuil de distance permet de considérer l'échelle spatiale des flux génétiques. Ces seuils de distance peuvent être inférés à partir d'un corrélogramme de Mantel si des données génétiques sont disponibles, ou à partir de connaissances *a priori* sur les capacités de dispersion de l'espèce. L'élagage peut également reposer sur des critères topologiques. Considérer un élagage au seuil de percolation, par exemple, permet de maximiser la modularité d'un réseau. Des analyses préliminaires sur les simulations démo-génétiques conduites dans le chapitre 3.1 (non présentées dans ce manuscrit), ont montré que cet élagage conduisait à la délimitation de modules paysagers concordants avec ceux d'un graphe représentant les événements de dispersion simulés. Le niveau de concordance était trop fluctuant en fonction des paysages et



des paramètres relatifs à la dispersion pour que nous retenions cet élagage, mais une étude complémentaire serait nécessaire pour améliorer les connaissances sur son potentiel.

Dans nos simulations (Ch. 3.1), nous avons élagué les graphes génétiques sur la base d'un critère statistique appelé élagage par indépendance conditionnelle (Dyer, 2015). Cette méthode évalue la corrélation du patrimoine génétique de deux populations au sein d'un réseau, en tenant compte de la corrélation de chacune d'elles avec l'ensemble des autres populations. Des analyses préliminaires semblent attester de la validité écologique de cet élagage pour les réseaux très faiblement connectés. Dans les autres cas, elle semble réduire trop sévèrement le nombre de connexions (Savary *et al.*, 2021). Néanmoins, lorsque la dispersion était un facteur limitant, les graphes élagués avec cette méthode présentaient un bon niveau de coïncidence avec les graphes modélisant les événements de dispersion simulés, témoignant ainsi de la capacité de cet élagage à représenter les liens de dispersion effective entre les populations.

Enfin, l'élagage peut reposer sur d'autres critères tels que la topographie du paysage d'étude ou l'histoire ancestrale des populations. Ainsi, dans le cas du graphe paysager de la Paruline caféïette (Ch. 3.2), l'élagage basé sur la topographie de la Guadeloupe séparant distinctement les deux entités de BT et GT s'est révélé très adapté. Par ailleurs, si les analyses sur l'histoire phylogéographique de cette espèce (V-A) amenaient à conclure à l'existence de deux lignées évolutives distinctes, élaguer les graphes génétiques et paysagers de façon à dissocier les populations de ces deux lignées serait alors pertinent.

Quelle que soit la méthode utilisée, l'intérêt de l'élagage réside en la modélisation de l'échelle spatiale à laquelle le flux génétique à travers un paysage hétérogène influence de façon prédominante le patron de variabilité génétique. La prise en compte de cette échelle dans les modélisations permet d'améliorer la robustesse des inférences. En effet, bien que passer par la réduction d'un jeu de données pour améliorer les inférences semble contre-intuitif, nous avons empiriquement démontré que la relation entre la structure génétique de la Paruline caféïette et la connectivité de son habitat forestier était renforcée par la considération d'un nombre restreint de paires de populations (Ch. 3.2). L'élagage permet en effet de retirer de l'analyse les paires de populations dont la différenciation génétique n'est pas expliquée par l'intensité du flux génétique, mais par d'autres facteurs comme la dérive génétique ou l'histoire phylogéographique. Ces paires de populations brulent le signal de l'influence du paysage sur les flux de gènes, et les extraire améliore ainsi la qualité des inférences.

### *Intégration du concept d'élagage à l'algorithme ResistanceGA*

Ce constat nous a conduit à transférer ce principe d'élagage au processus d'optimisation des surfaces de résistance basé sur l'algorithme ResistanceGA (Ch. 3.1). Lors de cette optimisation de distances-coût à l'aide de distances génétiques, toutes les paires de populations échantillonnées étaient jusqu'alors considérées dans les analyses. Toutefois, ajuster une distance-coût à une distance génétique résultant davantage des effets de la dérive ou du maintien d'un polymorphisme ancestral que de la résistance du paysage ne peut conduire à des inférences valides. Ainsi, nous avons fait l'hypothèse que sélectionner les paires de populations connectées à l'échelle spatiale où le paysage est susceptible d'influer sur le flux génétique renforcerait la qualité de l'optimisation. Pour vérifier cette hypothèse, nous avons modifié les fonctions d'optimisation pour leur faire intégrer en paramètre d'entrée un graphe de populations élagué, indiquant les paires de populations à considérer pour l'optimisation. Nous avons pu constater qu'élaguer les paires de populations selon le principe d'indépendance conditionnelle permettait d'améliorer la qualité de l'optimisation, en particulier lorsque le réseau de population échantillonné était très faiblement connecté (*e.g.*, dans le cas d'une espèce spécialiste avec une faible capacité de dispersion). Ainsi, l'élagage de réseaux de populations ou d'habitats est un outil très versatile permettant une considération efficace des dépendances spatiales pour les inférences de génétique du paysage.

Bauder *et al.* (2021) ont utilisé ResistanceGA pour optimiser et comparer le support empirique des surfaces de résistance considérant une ou plusieurs échelles spatiales et issues de différentes méthodes de modélisation. Ils ont montré que les surfaces de résistance optimisées à partir de variables paysagères mesurées à différentes échelles spatiales étaient plus performantes que celles modélisées à partir de variables mesurées à une échelle unique. Nous pourrions ainsi imaginer utiliser l'algorithme d'optimisation RGA pour optimiser l'échelle spatiale au même titre que les valeurs de résistance des éléments paysagers. L'échelle optimisée pourrait alors être celle des variables paysagères comme pour Bauder *et al.* (2021), ou encore celle des connexions entre paires de populations. Nous pourrions donc imaginer que l'algorithme soit capable d'optimiser un couple 'scénario de coût / topologie du réseau de populations', afin de proposer un modèle de résistance intégrant les effets de dépendance spatiale dans la relation entre flux de gènes et éléments paysagers.

## 6.2.2. Considération de l'échelle temporelle dans les modèles de connectivité

### 6.2.2.1. Dynamique temporelle des paysages et des déplacements d'individus

La structure d'un paysage n'est cependant pas figée dans le temps. Des changements paysagers s'opèrent à différentes échelles temporelles, liés à des modifications à très long terme de l'environnement (*e.g.*, cycles climatiques, phénomènes géologiques), à des perturbations anthropiques (changements d'occupation des sols), à des variations saisonnières ou encore à des successions écologiques (Wagner et Fortin, 2013; Zeller *et al.*, 2020) (Figure 6.4). Tappeiner *et al.* (2021) ont exposé le concept d'une dépendance de l'évolution d'un paysage à une trajectoire donnée. Cette dépendance prendrait sa source dans l'inertie et les blocages engendrés par les contextes culturels, politiques et économiques. Identifier et comprendre les trajectoires des paysages à travers le temps permettrait une meilleure projection des interactions futures entre paysage et diversité, « les décisions d'aujourd'hui en matière d'aménagement du territoire générant les héritages de demain », (Tappeiner *et al.* 2021).

Par ailleurs, il existe parallèlement à la dynamique du paysage une dynamique spatio-temporelle des déplacements d'individus (Figure 6.4). Certains se réalisent à petite échelle, comme ceux liés à la recherche quotidienne de nourriture, tandis que d'autres se déroulent à plus grande échelle comme ceux liés à la dispersion, ou présentent une cyclicité, comme les migrations saisonnières.

### 6.2.2.2. Délais de réponse de la variabilité génétique à une perturbation paysagère et héritage des paysages passés

Si un évènement de dispersion s'inscrit à l'échelle temporelle d'une génération, le flux génétique, en revanche, est multigénérationnel et peut connecter indirectement deux populations par 'stepping-stone' temporel (Boulangier *et al.*, 2020). Il est donc possible d'observer un intervalle de temps entre un changement paysager et la détection de ses conséquences dans la structure génétique (Etersson *et al.*, 2010). Ce temps de latence dépend de l'interaction de nombreux facteurs, tels que le temps de génération, la taille de la population, le niveau de spécialisation de l'espèce et ses capacités de dispersion (Epps et Keyghobadi, 2015), mais aussi de l'ampleur du changement paysager et de sa dynamique

temporelle (Figure 6.4). Il dépend également de la nature de la variable génétique. En effet, il est attendu que la diversité et la différenciation génétique atteignent un état d'équilibre à des rythmes différents, la différenciation génétique répondant plus rapidement à une perturbation (Keyghobadi *et al.*, 2005; Varvio *et al.*, 1986). Par ailleurs, si les changements paysagers surviennent très fréquemment, il est possible que l'équilibre entre flux et dérive ne soit jamais atteint et que la structure génétique ne soit jamais le reflet du paysage contemporain (Epps et Keyghobadi, 2015). Certaines études empiriques ont ainsi observé un retard de réponse de la structure génétique par rapport au paysage contemporain (Holzhauer *et al.*, 2006), tandis que d'autres ont identifié une réponse très rapide (Zellmer et Knowles, 2009). Dans le chapitre 5.2, nous avons montré que la structure génétique de la Paruline caféïette était plus fortement liée au paysage contemporain de la Guadeloupe qu'à celui des années 50. L'intense changement d'usage des sols au cours des 60 dernières années, avec une forte augmentation des surfaces urbanisées, semble déjà avoir impacté la structure génétique de ce passereau forestier. Toutefois, il faut aussi noter que la structure du paysage de 1950, caractérisée par la distribution stable de l'habitat forestier depuis la phase de déforestation à but agricole de la fin du XVI<sup>e</sup> siècle, explique toujours une part significative de la structure génétique contemporaine.

Intégrer une dimension temporelle aux études de génétique du paysage peut avoir plusieurs intérêts. Tout d'abord, s'assurer de la concordance entre la structure génétique et le paysage étudié permet d'améliorer la robustesse des inférences. De plus, identifier ces effets temporels permet d'améliorer la compréhension de l'histoire évolutive de l'espèce étudiée. Nos résultats tendent ainsi à confirmer le rôle de barrière exercé par l'isthme urbain entre BT et GT. Ils suggèrent aussi qu'avant cette expansion urbaine, les populations étaient déjà structurées par la distribution du couvert forestier à l'échelle de l'île. Il est donc admis qu'une part de la variabilité génétique peut être l'héritage d'un ancien usage des terres, potentiellement non documenté. On parle alors d'un effet de paysage fantôme (Harding *et al.*, 1998).

### 6.2.2.3. Intégrer les effets phylogéographiques dans les modèles de génétique du paysage

Le patron contemporain de diversité génétique peut également porter la marque de conditions et d'événements lointains, remontant à des milliers d'années. En particulier, les grands événements de colonisation et d'extinction liés aux cycles climatiques et aux dynamiques géologiques ont laissé d'importants héritages sur la structuration génétique contemporaine (Figure 6.4). Plusieurs études ont ainsi évalué l'impact des événements démographiques liés aux épisodes de glaciation du Quaternaire sur la distribution contemporaine de la variabilité génétique (Capurucho *et al.*, 2013; Frantz *et al.*, 2014). Ces événements climatiques et paysagers se situent à une échelle spatio-temporelle de bien plus grande ampleur que les changements ponctuels d'usage des sols ou que les variations saisonnières de la structure paysagère (Zeller *et al.*, 2020). Cette différence d'échelle entre dynamiques paysagères induit une différence d'impact sur l'histoire évolutive des populations. Alors que les événements de grandes envergures tels que les cycles climatiques influent à l'échelle macro-évolutive, les dynamiques de petites échelles relèvent davantage de la microévolution. Cette dissociation se traduit en l'existence de deux champs disciplinaires distincts mais complémentaires. La génétique du paysage s'intéresse ainsi à l'effet de l'hétérogénéité du paysage contemporain sur les processus micro-évolutifs dirigés par la migration et la dérive, tandis que la phylogéographie évalue les processus historiques à l'origine de la distribution des lignées évolutives à large échelle spatiale (Manel *et al.*, 2003; Wang, 2010).

Les études de génétique du paysage s'intéressent donc à la part de variabilité génétique expliquée par la structure et la dynamique du paysage contemporain. La variabilité liée à l'histoire ancestrale des populations, en particulier liée aux événements majeurs d'extinction et de colonisation, constitue de fait un bruit résiduel des modèles. Dyer *et al.* (2010) ont montré que retirer les effets phylogéographiques des modèles de génétique du paysage conduisait à une amélioration significative de la détection des effets paysagers. Selon une approche différente, Kebaïli *et al.* (2023) ont étudié la réponse démographique de trois espèces de papillons aux changements climatiques passés, à partir de données génomiques et de modèles de distribution d'espèces. Elles ont montré qu'une meilleure connaissance de l'histoire évolutive permettait d'améliorer la compréhension de la structure génétique contemporaine d'une espèce et d'évaluer sa vulnérabilité au changement climatique actuel.

Bien que nous ayons manqué de temps pour conduire à terme les analyses du chapitre 5.1, notre objectif serait d'inférer l'histoire de colonisation de la Guadeloupe pour la Paruline caféïette. S'il s'avérait que la structure actuelle résultait d'un contact secondaire entre deux lignées évolutives distinctes, la prise en compte de cet effet permettrait une évaluation plus robuste des effets résiduels du paysage contemporain sur la structure génétique, et justifierait les bonnes performances de l'élagage entre les populations de BT et de GT. La prise en compte des facteurs phylogéographiques pourrait en effet consister à élaguer le graphe de populations selon l'assignation aux lignées évolutives, ou encore à modéliser cette appartenance par un effet aléatoire dans les modèles MLPE.

### 6.2.3. Modèles de connectivité multi-spécifique

Nos résultats concernant l'approche multi-spécifique en génétique du paysage (Ch. 4) s'accordent avec de nombreux autres (Amos *et al.*, 2014; Delord *et al.*, 2023; DiLeo et Wagner, 2016; Engler *et al.*, 2014; Mulvaney *et al.*, 2021), affirmant que différentes espèces peuvent répondre de manière contrastée à un même paysage. Comparer ainsi la réponse de différentes espèces permet d'améliorer les connaissances sur les processus démographiques et évolutifs à l'origine de ces contrastes (Kelly et Palumbi, 2010) (Figure 6.4). Manel *et al.* (2012) ont ainsi démontré que la distribution de la diversité adaptative de 13 espèces de plantes à l'échelle globale du biome de l'arc alpin était dirigée par deux facteurs environnementaux communs, la température et les précipitations. Cet exemple montre qu'il est ainsi possible d'identifier des motifs généraux concernant la distribution de la variabilité à large échelle spatiale, pour des groupes d'espèces précautionneusement choisis en fonction de leurs traits écologiques. Dans notre étude (Ch. 4), nous avons montré que des espèces présentant des niveaux similaires de spécialisation pour l'habitat forestier présentaient des réponses génétiques similaires à la structure du paysage

Le domaine de la conservation aurait également à gagner d'un essor des études multi-spécifiques (Figure 6.4). En effet, les gestionnaires attendent des inférences de génétique du paysage qu'elles prédisent l'impact de changements paysagers et environnementaux sur la démographie et le potentiel adaptatif des espèces à l'échelle de leur zone d'action. Koen *et al.* (2014) ont proposé une démarche visant à prédire les déplacements d'un cortège d'espèces dans l'objectif d'aménager des couloirs de dispersion. Pour cela, ils ont élargi le principe de

connectivité fonctionnelle pour pouvoir l'appliquer à un ensemble d'espèces, et paramétrer la résistance des éléments paysagers en fonction de leur seul caractère anthropisé ou naturel. Cette approche rejoint celle de Keeley *et al.* (2021), qui soulignent l'intérêt des métriques de connectivité structurelle tenant simplement compte de l'empreinte humaine pour modéliser simultanément l'effet du paysage sur les événements démographiques de plusieurs espèces.

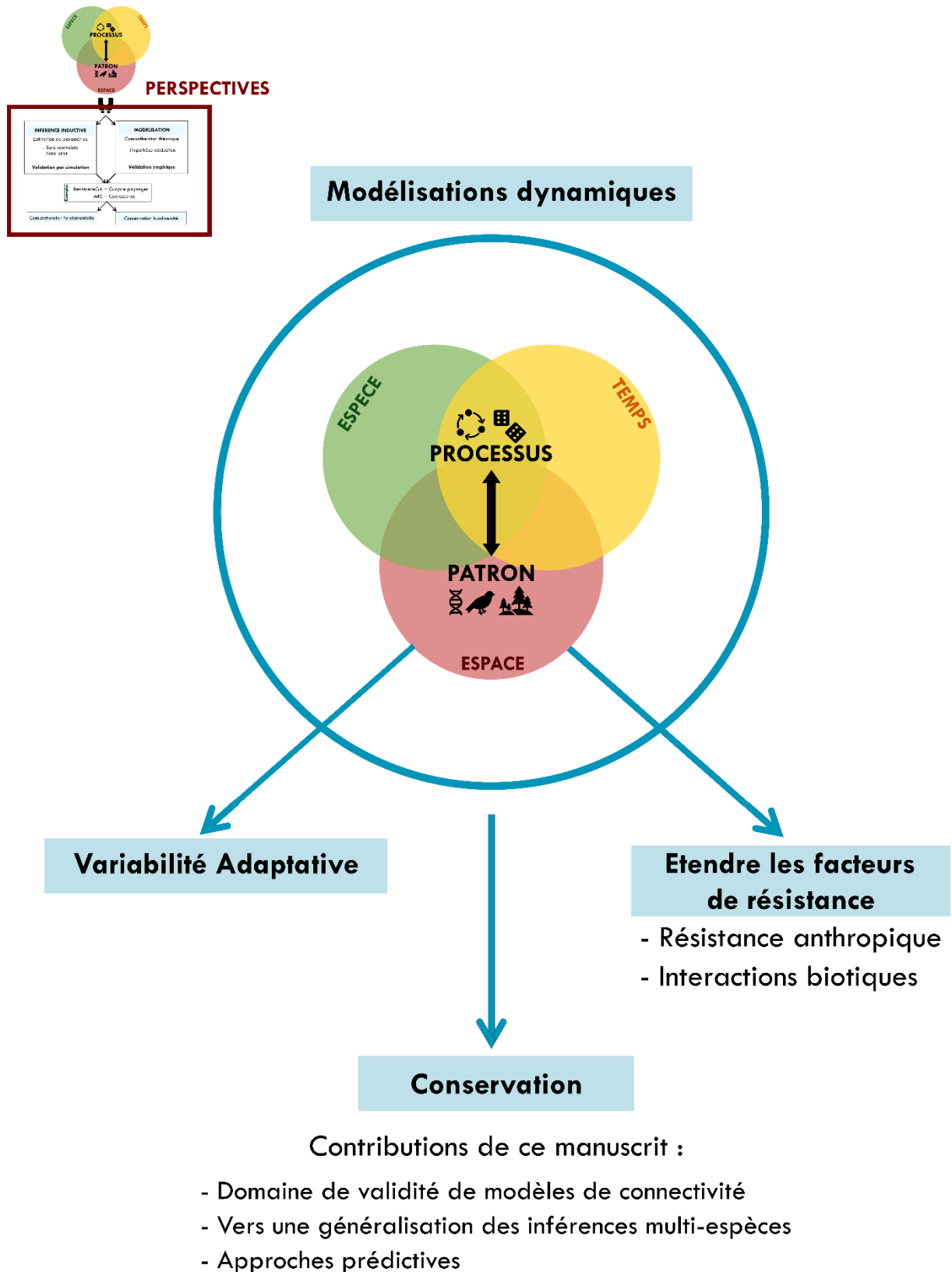
Enfin, bien que diversité spécifique et diversité génétique soient souvent cantonnées aux domaines respectifs de l'écologie des communautés et de la génétique des populations, ces deux niveaux de diversité sont susceptibles de présenter des covariations spatiales (Vellend et Geber, 2005). Les diversité génétique et spécifique peuvent interagir selon trois manières. Tout d'abord, elles peuvent être affectées par les effets de la dérive, de la migration et de la sélection, en fonction des conditions environnementales et paysagères qu'elles partagent. Par ailleurs, la diversité génétique, en influant sur les événements démographiques et adaptatifs, peut impacter la composition des communautés. A l'inverse, la composition d'une communauté, à travers les interactions interspécifiques et les pressions de sélection, est susceptible d'influencer la diversité génétique (Vellend et Geber, 2005). Ainsi, évaluer les corrélations entre diversité génétique et spécifique permet une compréhension plus intégrée de ces deux niveaux fondamentaux de la biodiversité, réunissant conjointement des perspectives écologiques et évolutives. Cette compréhension est particulièrement précieuse pour apporter un éclairage sur l'interaction entre ces deux niveaux de diversité dans le cadre des phénomènes de vortex d'extinction (Fagan et Holmes, 2006). Bertin *et al.* (2017) ont adopté une approche de '*community genetics*' pour évaluer la corrélation entre la diversité génétique et spécifique de plantes dans les zones humides des Etes. Ils ont observé des corrélations positives entre ces deux niveaux de diversité, liés à l'action parallèle de processus neutres, tels que le flux de gène et la dérive, sur la diversité génétique et spécifique. Une telle corrélation positive est notamment attendue en milieu insulaire et fragmenté (Bertin *et al.*, 2017; Vellend et Geber, 2005). Les données du projet CANON, sur lequel reposent les analyses de cette thèse, nous permettraient de vérifier cette hypothèse. En effet, nous disposons de données génétiques pour 5 espèces de passereaux à l'échelle de la Guadeloupe. De plus, des données de communautés (abondance et richesse) ont également été récoltées au niveau des filets utilisés pour l'échantillonnage des individus. Nous pourrions donc imaginer une étude multi-spécifique évaluant la congruence des patrons de diversité génétique et spécifique à l'échelle de l'île.

Jusqu'ici, nous avons traité des objets à valider et de leur méthode de validation dans le cadre de la modélisation de la connectivité des habitats et des populations. Nous avons également considéré les échelles spatiales, temporelles et multi-spécifiques influençant les domaines de validité de ces modèles. Nous aborderons maintenant le dernier point concernant les perspectives d'amélioration et d'application de ces modèles.

### **6.3. Perspectives pour des modèles de connectivité valides**

Nous verrons dans cette partie que plusieurs pistes sont actuellement explorées pour améliorer le réalisme des modèles de connectivité. Ces développements visent notamment une meilleure représentation de la complexité des processus éco-évolutifs structurant le patron de diversité, en considérant la dynamique de ces patrons, en intégrant d'autres sources de résistance au déplacement que les éléments paysagers ou climatiques, et en s'intéressant aux effets de la connectivité sur les patrons de diversité adaptative. Enfin, nous aborderons des perspectives d'applications de ces modèles de connectivité pour la conservation de la biodiversité (Figure 6.5).





**FIGURE 6.5 :** Focus depuis la figure 6.1 sur les perspectives d'amélioration et d'application des modèles de connectivité en génétique du paysage. Ces perspectives regroupent notamment la prise en compte d'un spectre plus étendu de variables explicatives de la connectivité, l'intégration de dynamiques spatio-temporelles et l'étude des patrons de variabilité adaptative. Cette figure résume enfin la contribution de ce travail de thèse au domaine de l'écologie de la conservation.

### 6.3.1. Vers une vision dynamique des modèles de connectivité

Développer des modèles de connectivité intégrant explicitement les dynamiques spatio-temporelles des paysages et des processus écologiques constituerait une grande avancée, en augmentant la pertinence écologique et la précision des inférences de connectivité (Dutta *et al.*, 2022; Zeller *et al.*, 2020) (Figure 6.5). En dépit de la reconnaissance du caractère labile des paysages et des flux les parcourant, une grande partie des études de génétique du paysage confronte une structure paysagère fixe à une structure génétique à un instant  $t$ . En effet, l'incertitude sur le temps de latence de la réponse génétique, associée au coût de la collecte de données à plusieurs pas temporels, et à la faible disponibilité des données historiques, rendent ardue la considération des effets d'échelles dans les modélisations. Toutefois, face à la plus-value qu'apporterait une modélisation de la connectivité dynamique fonctionnelle, de plus en plus d'approches émergent.

Parmi les études pionnières, sensibles aux dynamiques des populations et des paysages, nous pouvons citer celles de Foltête et Giraudoux (2012) et Shirk *et al.* (2014). Les premiers ont utilisé des données diachroniques pour pondérer un graphe d'habitat statique, et ainsi modéliser la dynamique saisonnière des pullulations de campagnols dans le Jura. Shirk *et al.* (2014), quant à eux, ont représenté les variations de sélection de ressources saisonnières chez la Martre d'Amérique selon deux types d'habitats. En utilisant deux modèles distincts, ils ont montré que la sélection de ressources était déterminée par des facteurs différents selon les deux saisons et milieux.

Progressivement, d'autres méthodes se sont développées pour modéliser conjointement les dynamiques spatio-temporelles des paysages et de processus écologiques (Fenderson *et al.*, 2020). Ainsi, Bishop-Taylor *et al.*, (2018) ont comparé des modèles de connectivité statiques avec un modèle de connectivité dynamique constitué d'une série temporelle de 99 graphes, représentant des pas saisonniers basés sur le suivi par télédétection de surfaces en eau durant 25 ans. Les divergences importantes observées entre les deux types de modèles, en particulier dans les régions présentant une forte variabilité spatio-temporelle, suggèrent une plus faible validité écologique des modèles de connectivité statiques. Martensen *et al.* (2017) sont parvenus à des conclusions similaires, en constatant une augmentation de la connectivité de 30 % lorsqu'elle était modélisée en considérant les connexions en 'stepping-stone' temporel (*i.e.*, liées à l'apparition et/ou à la disparition de taches d'habitat au cours du temps). Enfin, Watts *et al.* (2015) ont utilisé des modèles

gravitaires pour démontrer que les variations spatio-temporelles de l'accumulation de neige étaient un facteur important pour la connectivité génétique des amphibiens, en raison de son influence sur les conditions hydrologiques et la qualité des habitats aquatiques.

Ces différents modèles de connectivité dynamique divergent des modèles statiques couramment utilisés, et des études de comparaison et de validation à partir de données empiriques seraient nécessaires pour attester de leur gain coût/bénéfice en termes de pertinence écologique.

### 6.3.2. Vers une diversification des variables explicatives de la connectivité

Outre la considération des dynamiques spatio-temporelles des populations et des paysages, les modèles de connectivité gagneraient certainement à intégrer d'autres prédicteurs que des patrons spatiaux (Ghoddousi *et al.*, 2021) (Figure 6.5). La modélisation des surfaces de résistance pourrait intégrer d'autres processus écologiques que la dispersion, telles que les interactions biotiques intraspécifiques ou interspécifiques. Pour des espèces territoriales ou grégaires, une résistance sociale peut s'exercer sur un individu cherchant à traverser ou s'installer sur un territoire déjà occupé. Khattar *et al.*, (2024) ont ainsi démontré que les stratégies de dispersion au sein des métacommunautés résultaient en partie de l'interaction entre les processus de compétition inter et intraspécifiques et les caractéristiques paysagères et environnementales.

Un autre exemple serait la considération des interactions entre humains et faune sauvage, pouvant être conflictuelles au-delà de la seule caractéristique spatiale de l'occupation du sol. Kimmig *et al.* (2020) ont montré qu'alors que des structures anthropiques comme les routes peuvent servir de couloir de dispersion à des populations urbaines de renards, les zones condensant une forte activité humaine étaient largement évitées. Selon eux, la distinction entre zone naturelle et zone anthropisée n'est parfois pas pertinente pour modéliser le déplacement des individus. La considération d'un « *paysage de peur* » (Laundre *et al.*, 2010) le serait parfois davantage. Ce paysage cartographie le niveau de peur expérimenté par un individu en fonction de sa perception des risques, par exemple liés à la prédation, qu'ils soient avérés ou non. Dans l'étude de Kimmig *et al.* (2020), la présence de l'Homme serait associée au risque de chasse, et elle serait donc évitée. Cet exemple nous permet d'introduire le concept de *Résistance anthropique* (Ghoddousi *et al.*, 2021), qui

caractérise l'impact des comportements humains sur le déplacement des animaux à travers le paysage. Sa modélisation requiert une approche multidisciplinaire, évaluant et spatialisant les facteurs politico-socio-économiques et psychologiques déterminant les comportements humains et leurs conséquences sur la connectivité. Un exemple courant est celui de la pression de prédation exercée par les éleveurs et propriétaires terriens sur les grands carnivores aux États-Unis (Ghoddousi *et al.*, 2021). La connectivité effective entre les populations peut ainsi différer de celle prédite par les seules caractéristiques du paysage (Fletcher *et al.*, 2019). La compréhension des niveaux de résistance anthropique et leur intégration dans les modèles de génétiques du paysage permettrait certainement de meilleures inférences des effets de la connectivité. L'enjeu actuel pour le paramétrage de surfaces de résistance anthropique réside dans la formalisation de ce concept, et dans le développement d'approximations spatiales de ces facteurs anthropiques (Bowman *et al.*, 2004; Kretser *et al.*, 2009).

### 6.3.3. Modéliser l'influence du paysage sur la diversité adaptative

#### 6.3.3.1. Interactions entre flux génétique et adaptation locale

Le flux génétique entre deux populations dépend également de la sélection divergente. Cette dernière est susceptible de rendre des immigrants et leurs descendants mal adaptés aux conditions locales, limitant ainsi le transfert de matériel génétique entre deux populations soumises à des pressions de sélection différentes (Crispo *et al.*, 2006; Spear *et al.*, 2010). On parle alors de processus de *spéciation écologique* (Rundle et Nosil, 2005). Toutefois, Crispo *et al.*, (2006) ont montré que ce processus pouvait être de plus faible importance sur le flux génétique que l'effet de l'hétérogénéité du paysage, notamment en raison de son interaction avec d'autres stratégies évolutives telles que les comportements de reproduction. Ce processus de spéciation écologique est souvent moins évalué que sa réciproque, évaluant l'influence du flux génétique sur le potentiel adaptatif des populations. En effet, si le flux génétique est plus important que l'effet de sélection diversifiante, la divergence adaptative des populations sera limitée. À l'inverse, les populations isolées (*e.g.*, sous l'effet de la fragmentation du paysage) sont susceptibles de présenter une réduction de fitness liée à une dépression de consanguinité sous l'effet dominant de la dérive génétique, et ainsi de faire face à une réduction de leur potentiel d'adaptation (Crispo *et al.*, 2006; Wagner et Fortin, 2013). Ainsi, les facteurs démographiques, déterminés par la composition et la configuration du

paysage, influent non seulement sur la variabilité génétique neutre, mais également sur la variabilité adaptative.

### 6.3.3.2. Comparaison de la variabilité neutre et adaptative

La variabilité génétique neutre n'est pas un bon prédicteur de la variabilité génétique adaptative, et son étude ne permet donc pas de prédire la persistance et le potentiel évolutif des populations (Holderegger *et al.*, 2006). Ainsi, en prenant l'exemple d'un oiseau forestier de Guyane, Perrin *et al.* (2021) ont montré que des *locus* neutres ne répondaient pas de la même manière à la fragmentation des habitats que les *locus* sous sélection. Ils suggèrent que les pressions de sélection exercées en milieu fragmenté, liées à l'augmentation de la prévalence de pathogènes dans ces milieux, conduisent au maintien d'une diversité des gènes de l'immunité par sélection diversifiante, tandis que la diversité neutre s'érode par dérive génétique et isolement géographique. Teixeira et Huber (2021) soutiennent également qu'il n'existe pas de relation évidente entre le niveau de diversité génétique neutre et le risque d'extinction d'une population. Toutefois, l'étude de l'effet du paysage sur la diversité génétique neutre est primordiale à l'étude de ce même effet sur la diversité génétique adaptative. En effet, la diversité adaptative comme la diversité neutre sont soumises aux effets de la dérive et du flux génétique. Les modèles étudiant l'effet de ces processus démographiques sur la diversité génétique neutre constituent donc une forme de modèles nuls pour les modèles portant sur l'effet de la sélection environnementale sur la diversité adaptative.

### 6.3.3.3. Association entre gradient environnemental et variabilité adaptative

La cartographie de la variation génétique adaptative et l'étude de son association avec des variables environnementales pourrait permettre d'identifier les populations porteuses de génotypes potentiellement adaptés à un stress environnemental (Manel *et al.*, 2012; Selmoni *et al.*, 2020) (Figure 6.5). Il a également été montré que le couplage de cette identification des variants génétiques adaptés à une pression sélective avec des analyses de connectivité permettait d'évaluer la dispersion de ces variants à plus large échelle, et ainsi d'améliorer l'efficacité des mesures de conservation (Selmoni *et al.*, 2020). Les populations abritant une forte diversité, garante d'un potentiel d'adaptation, pourraient faire l'objet de mesures de conservation prioritaires ou servir de réservoir pour des mesures d'introduction contrôlée dans le cadre de sauvetage génétique. La compréhension du potentiel adaptatif d'une population et

sa préservation requiert donc l'étude combinée des propriétés de la diversité génétique adaptative, et des facteurs environnementaux, paysagers, démographiques, et écologiques (Teixeira et Huber, 2021).

#### 6.3.3.4. Perspectives de recherche en Guadeloupe

En Guadeloupe, la BT et la GT présentent des conditions environnementales très contrastées, notamment en termes de gradient d'altitude, de température et d'humidité. De plus, la topologie de la Guadeloupe, la distribution des habitats et la résistance exercée par le paysage entraînent pour certaines espèces une réduction du flux génétique entre les populations de BT et GT. Nous pouvons donc faire l'hypothèse que ces populations présentent des adaptations génétiques à ces deux environnements distincts. Si tel est le cas, la restauration d'un flux génétique entre la BT et la GT ne serait pas bénéfique à la *fitness* des individus localement adaptés. Par ailleurs, comme mis en évidence par Perrin *et al.*, (2021), des micro-adaptations peuvent également exister à l'échelle de la BT ou de la GT entre les populations situées en milieu fragmenté et celles situées en milieu plus continu. Étudier la distribution de la variabilité adaptative sur ces territoires pourrait permettre une meilleure compréhension des enjeux de conservation.

Dans cet objectif, nous avons échantillonné 114 individus de Paruline caféïette dans 6 populations réparties sur BT et GT (3 et 3 en BT et GT respectivement). Nous avons ensuite produit des données de séquençage RADseq pour ces individus. Par la suite, notre objectif sera d'identifier des *locus* potentiellement sous sélection, et d'évaluer l'association entre la distribution de ces *locus* et la localisation sur BT ou GT des populations, ou entre ces *locus* et des gradients environnementaux de température et d'humidité. Cela nous permettra d'évaluer s'il existe une potentielle adaptation des individus aux pressions environnementales distinctes exercées sur BT et GT.

Par ailleurs, nous avons également échantillonné 133 individus de Saltator Gros-Bec (*Saltator albicollis*), répartis entre 4 couples de populations (2 couples sur BT et 2 couples sur GT). Chaque couple est constitué d'une population bénéficiant d'une grande quantité d'habitat disponible, et d'une population voisine située dans un fragment d'habitat restreint et isolé au sein d'une matrice agricole. Nous avons également produit des données de séquençage RADseq pour ces individus. Notre objectif était d'évaluer comment la structure

du paysage peut influencer sur la diversité adaptative. Nous pouvons imaginer mettre en relation des indices de diversité et de différenciation relative calculés à partir de marqueurs sous sélection, avec des métriques de connectivité calculées au niveau des nœuds d'un graphe paysager.

### 6.3.4. Traduction des inférences en mesures de conservation

#### 6.3.4.1. Les apports de cette thèse à la conservation

##### *Valider les modèles et les métriques de connectivité*

Les enjeux autour de la conservation des connectivités fonctionnelles sont grandissants, notamment dans un contexte de changements globaux et de redistribution spatiale des assemblages d'espèces. Ainsi, le cadre mondial de la biodiversité de Kunming-Montréal (2023) cible l'objectif d'avoir 30% des zones terrestres, et notamment celles d'importance pour la biodiversité, 'dument conservées et connectées' d'ici 2030. Dans ce contexte, les modèles de connectivité sont de plus en plus largement utilisés à des fins de conservation. L'acquisition facilitée de données biologiques et environnementales de grandes ampleurs ainsi que le développement de méthodes statistiques et l'essor de l'intelligence artificielle, ont conduit à l'émergence de nombreux modèles de connectivité. Cette multitude de modèles appelle à des synthèses sur leur utilisation en fonction de l'usage projeté et des conditions d'application (voir par exemple Keeley *et al.*, 2021), mais également sur leur niveau de validité.

Dans cette thèse, nous avons validé la capacité des graphes paysagers à modéliser les effets de la connectivité sur la structure génétique d'une espèce d'oiseau forestier en Guadeloupe (Ch. 3.2, Figure 6.5). Bien que des travaux soient nécessaires sur d'autres espèces dans d'autres paysages pour entériner et généraliser ce résultat, notre étude tend à confirmer la fiabilité des inférences de connectivité issues des graphes paysagers. Pour une espèce spécialiste de l'habitat forestier, la conservation prioritaire des taches d'habitat associées à une forte quantité d'habitat atteignable et à un poids important pour la traversabilité du réseau devrait efficacement conduire à la préservation de la connectivité génétique des populations de cette espèce. Dans un deuxième temps, nous avons exploré la validité de l'algorithme RGA couramment utilisé pour paramétrer des surfaces de résistance

au déplacement des individus (Ch. 3.1, Figure 6.5). Nous avons ici démontré que l'utilisation de cet algorithme pour inférer les éléments paysagers faisant barrière aux flux génétiques était peu fiable. Nous avons ainsi appelé à une grande prudence lors de l'utilisation des inférences de coûts de RGA à des fins de conservation. En revanche, nous avons mis en avant une capacité prometteuse mais rarement explorée de cet algorithme, consistant à prédire avec précision les différenciations génétiques entre populations.

#### *Un pas vers la généralisation des inférences*

Si les études de génétique du paysage se multiplient, un frein à leur pertinence pour le domaine de la conservation réside dans la portée de leurs inférences, majoritairement focalisées sur un unique couple espèce-paysage (Keller *et al.*, 2015) pour une revue sur le sujet). La généralisation des inférences de génétique du paysage à différents territoires et à plusieurs espèces est un objectif déterminant pour la contribution de ce champ disciplinaire aux enjeux conservation. Dans cette thèse, nous avons donc essayé de nous extraire, en partie, du contexte espèce et territoire-spécifique pour aborder la question de l'extrapolation des inférences des modèles de génétique du paysage à diverses espèces et territoires (Ch. 3, 4, Figure 6.5).

Tout d'abord, l'approche par simulation conduite dans le cadre de l'évaluation de ResistanceGA nous a permis de modéliser les processus de dispersion pour 14 espèces virtuelles (caractérisées par 7 capacités de dispersion et 2 niveaux de spécialisation) à travers 30 paysages différents. Bien que nous n'ayons pas obtenu de résultats tranchés concernant l'association d'un patron de connectivité à un type de structure paysagère ou à un profil écologique, cette étude nous a permis de mettre en évidence i) le caractère déterminant du choix de l'échelle spatiale dans les modèles en fonction du profil écologique, ii) la variation importante des performances des modèles entre des paysages différents pour un même profil. Cela appelle donc à des études complémentaires, concernant notamment l'influence de la structure paysagère sur la répliquabilité des inférences de génétique du paysage. De plus, dans le cadre de cette étude, nous avons explicitement évalué la capacité d'extrapolation à d'autres paysages des inférences de coûts optimisées par ResistanceGA. Nos résultats ont montré que cet outil ne pourra pas être utilisé pour prédire la connectivité d'une espèce sur l'ensemble de son aire de répartition (si celle-ci n'a pas été entièrement échantillonnée), ni pour prédire la connectivité pour une espèce selon différents scénarios de changements paysagers, par exemple dans le cadre d'études évaluant les conséquences futures du changement climatique.



Enfin, nous avons également exploré de façon empirique la réponse de différentes espèces à un même paysage. Cette étude multi-spécifique nous a permis de démontrer que l'effet d'un même paysage sur la structure génétique des populations pouvait fortement varier entre différentes espèces. La généralisation des inférences de résistance du paysage d'une espèce à une autre doit donc être réalisée avec prudence. En particulier, nous avons montré que cette extrapolation devait cibler des espèces partageant le même comportement de dispersion et les mêmes préférences d'habitat pour conduire à des mesures de conservation multi-spécifiques fiables et efficaces.

#### *Développer les approches prédictives*

Enfin, dans le contexte de changement climatique actuel, des appels ont été lancés pour que la recherche en génétique du paysage contribue à la prédiction de la réponse génétique aux changements paysagers (Van Strien *et al.*, 2014). En dépit de cela, et bien que les attitudes commencent à évoluer (Palm *et al.*, 2023), peu d'études explorent les capacités de prédiction des inférences de coût. Nous nous sommes donc attachés, dans le cadre de cette thèse, à vérifier les capacités prédictives des modèles de connectivité évalués (Figures 3, 4). Nous avons employé le critère statistique du  $R^2$  de validation, basé sur le calcul de l'erreur moyenne de prédiction et issu d'une approche adaptée de validation croisée de type '*Leave One Out Cross-Validation*'. Cette approche nous a permis d'affirmer la capacité des distances-coût le long des liens d'un graphe paysager à prédire des distances génétiques. Cette capacité restant néanmoins dépendante de l'échelle spatiale et du paramétrage d'une surface de résistance, nous avons également évalué la capacité prédictive d'une surface de résistance optimisée par ResistanceGA. Nos résultats ont mis en évidence une excellente capacité de prédiction des distances-coût optimisées, et encouragent l'évaluation de l'utilisation conjointe de scénarios de coûts optimisés et des modèles de graphes paysagers. Enfin, dans notre étude multi-spécifique, nous avons évalué la capacité d'un modèle de connectivité calibré pour une espèce focale, à prédire les distances génétiques d'une espèce indépendante sur le même territoire, selon une approche de validation croisée interspécifique. L'intérêt de la démarche ne résidait pas dans l'évaluation des performances prédictives d'un modèle ni de son potentiel de généralisation, mais plutôt dans la prédiction de la réponse de la structure génétique à la connectivité des habitats à l'échelle d'une communauté d'espèces. Ces prédictions pourraient permettre d'améliorer les connaissances sur les cibles potentielles des mesures de conservation, et les rendre ainsi plus efficaces.

#### 6.3.4.2. La génétique du paysage face aux changements globaux

Face au contexte de changements globaux, de nombreuses attentes reposent sur la génétique du paysage. Modéliser les distributions futures des niches écologiques sous divers scénarios de changements climatiques ne sera pas suffisant pour prédire l'évolution parallèle de la distribution des espèces. Une compréhension et une modélisation fine des capacités de dispersion des espèces, en fonction de leurs traits écologiques et des potentielles résistances du paysage, sont essentielles à la production de projections spatio-temporelles fiables (Kebaïli *et al.*, 2023; Manel, 2013; Pauls *et al.*, 2013; Wasserman *et al.*, 2012). Les modèles de génétique du paysage proposent un cadre prometteur pour modéliser ces événements de dispersion, et prédire la circulation des flux génétiques à large échelle spatiale (Wasserman *et al.*, 2012). Par ailleurs, si la préservation de la connectivité fonctionnelle est un préalable essentiel au déplacement (parallèle) des populations et de leurs niches écologiques, la préservation de la diversité génétique adaptative des populations est également nécessaire pour garantir leur résilience face aux perturbations environnementales. Pour relever ces enjeux, les modèles de génétique du paysage doivent donc tendre à une évaluation conjointe des diversités neutre et adaptative, à l'échelle des communautés. Par ailleurs, leur performance repose sur leur capacité à intégrer les dynamiques spatio-temporelles des changements environnementaux. Les cycles glaciaires, et en particulier le dernier maximum glaciaire au Pléistocène, correspondent à des changements climatiques majeurs, de l'ampleur de celui auquel nous faisons actuellement face. Le passé peut-il permettre de prédire le futur ? Plusieurs études s'intéressent ainsi à la reconstruction des événements démographiques et des réponses de la diversité génétique aux cycles climatiques passés, pour ensuite inférer la vulnérabilité des populations aux changements climatiques contemporains (Kebaïli *et al.*, 2023; Velo-Antón *et al.*, 2013). Par ailleurs, les gestionnaires de la biodiversité attendent des recherches menées en écologie de la conservation et la mise en place de plans régionaux de préservation et de restauration des connectivités, favorables à l'ensemble des communautés d'espèces, et robustes aux changements climatiques et d'usage des terres contemporains et à venir (Beier *et al.*, 2008; Rudnick *et al.*, 2012). Si ce défi n'est pas encore résolu, les approches méthodologiques se développent pour proposer aux gestionnaires des modèles adaptés à ces enjeux. Ainsi, Albert *et al.* (2017) proposent une méthode pour prioriser la conservation des habitats, selon des critères de qualité et de connectivité. Cette méthode, dérivée de la théorie des graphes, a été conçue pour être robuste face aux conséquences des changements climatiques et adaptée à un large corpus d'espèces. Schoen *et al.* (2022) ont

quant à eux proposé une méthode visant à synthétiser des résultats de plusieurs études de connectivité, pour aboutir à un ligne directionnelle consensus sur les actions de conservation à mettre en place.





# CHAPITRE 7

## CONCLUSIONS GÉNÉRALES

---

La génétique du paysage étudie la complexité des interactions entre des processus éco-évolutifs et des patrons spatiaux paysagers et génétiques, à travers l'espace et le temps. Un modèle de génétique du paysage correspond à la formalisation de cette interaction. Sa construction repose sur une conceptualisation des processus et sur une approximation de leurs paramètres. La validation d'un modèle doit donc porter à la fois sur la validité de notre compréhension des processus et sur le degré de réalisme des paramètres utilisés pour les modéliser. Dans cette thèse, nous avons tâché de valider ces deux composantes essentielles de la construction d'un modèle spatial, et plus particulièrement de ceux dédiés à la connectivité fonctionnelle des habitats.

L'approche hypothético-déductive est couramment employée dans la construction des modèles spatiaux. Les modèles fondés sur cette approche proposent une traduction explicite d'hypothèses reflétant notre compréhension d'un processus éco-évolutif. Ces hypothèses concernent la façon dont le processus se déroule dans l'espace, en fonction notamment de propriétés (éléments paysagers) et de contraintes (isolement relatif) spatiales. Les hypothèses sont enfin testées, en confrontant les résultats de la modélisation avec un patron empirique supposé refléter les conséquences écologiques du processus modélisé. Les conclusions tirées de cette confrontation valident ou invalident enfin notre compréhension théorique de l'interaction 'patron-processus' étudiée.

Dans cette thèse, nous nous sommes appuyés sur cette approche pour valider des modèles théoriques de connectivité et leur métriques dérivées. Nous avons ainsi validé empiriquement la capacité du modèle théorique des graphes paysagers à représenter l'influence de la connectivité des habitats sur le patron de variabilité génétique d'une espèce de passereau forestier en contexte insulaire et fragmenté. Nous avons également exploré le domaine de validité temporel du modèle théorique des distances de moindre coût, en confrontant un patron empirique contemporain de différenciation génétique à des résultats de modélisation diachroniques.

Par ailleurs, si la validation empirique d'un modèle théorique dépend de la justesse de notre compréhension de l'interaction 'patron-processus', elle dépend également de la validité intrinsèque des outils d'analyse permettant d'en estimer les paramètres. Les méthodes d'inférences sont couramment utilisées pour cet objectif de paramétrage des modèles de connectivité. A l'ère du *Big data* et de l'essor de l'Intelligence Artificielle, on assiste au développement de méthodes '*data-driven*' d'apprentissage automatiques, capables d'extraire de l'information à partir de systèmes complexes en contexte de forte incertitude et de non

stationnarité. Leur application aux systèmes vivants, complexes par nature, est donc prometteuse. L'étude du domaine de validité des solutions obtenues via ces algorithmes est cruciale à l'heure où ces méthodes émergent dans le domaine de l'écologie. Nous avons ainsi évalué par simulation la validité et les conditions d'application de ResistanceGA, une méthode d'inférence fondée sur un algorithme génétique. Nous avons démontré ses bonnes capacités de prédiction des différenciations génétiques entre populations à partir de données génétiques et paysagères. Nous avons également identifié ses limites, prévenant notamment l'extrapolation des inférences à d'autres paysages et espèces que ceux faisant l'objet de l'analyse.

L'approche de validation permet donc d'identifier les limites du modèle ou de l'outil évalué. Dans le cas de l'approche hypothético-déductive, le choix des processus à inclure dans la modélisation et la part de stochasticité du patron étudié déterminent le niveau de validité du modèle considéré. Dans le cas de l'approche d'inférence inductive, le phénomène de surajustement et le manque de compréhension mécaniste peuvent limiter l'usage valide des outils considérés. Nous avons ainsi mis en évidence l'incapacité des graphes paysagers à prédire correctement les valeurs de différenciation génétique pour des populations localisées dans certaines zones paysagères, laissant présupposer l'existence d'autres processus à l'œuvre dans ces régions. Nous avons également montré que ResistanceGA ne permettait pas d'inférer avec fiabilité le mécanisme d'action de la structure paysagère sur le flux génétique (i.e., le caractère perméable d'un élément paysager au flux génétique), conformément aux propriétés des méthodes d'inférences basées sur la donnée.

Aussi, il est parfois bénéfique de prendre le contre-pied de l'approche objective des méthodes d'apprentissage automatique, en explorant des méthodes d'inférences basées sur des *priors* définis à partir de connaissances et d'observations empiriques. Par exemple, l'inférence bayésienne couplée aux méthodes de coalescence permet d'inférer les paramètres démogénétiques de l'histoire évolutive d'une population. Notre objectif serait d'utiliser cette approche pour inférer le scénario de colonisation de la Paruline caféïette en Guadeloupe, et à terme de prendre en compte ces effets phylogéographiques dans les modèles de génétique du paysage. En effet, l'association de ces différentes méthodes de modélisation devrait permettre une meilleure désintringation des différents processus façonnant la distribution spatiale de la variabilité génétique neutre.

En particulier, l'étude de l'influence du flux de gènes à travers un paysage hétérogène sur la variabilité génétique renseigne sur la connectivité fonctionnelle des habitats et des



populations. Si la préservation des habitats est fondamentale pour la conservation de la biodiversité, le maintien de leur connexion est également un enjeu crucial pour garantir le fonctionnement dynamique et la résilience des écosystèmes. Préserver la diversité fonctionnelle à l'échelle des écosystèmes implique que cet enjeu soit abordé dans une perspective multi-espèces. Pour contribuer à la recherche d'une solution opérationnelle à cet enjeu, notamment dans l'objectif de définir des cortèges optimaux d'espèces pour le protocole des mesures de conservation, nous avons montré que les espèces présentant un patron spatial de déplacement et des habitats similaires, partageaient également une perception similaire de l'effet de l'hétérogénéité du paysage sur le flux génétique. Nous espérons ainsi que les résultats de ce travail de recherche doctoral contribueront à la progression du champ disciplinaire de la génétique du paysage, tant sur le plan méthodologique qu'à des fins de conservation de la biodiversité et de la connectivité fonctionnelle des habitats qui la maintient.





# RÉFÉRENCES

- Adamack, A.T., Gruber, B., 2014. POPGENREPORT: simplifying basic population genetic analyses in R. *Methods Ecol Evol* 5, 384–387. <https://doi.org/10.1111/2041-210X.12158>
- Adler, G.H., Levins, R., 1994. The Island Syndrome in Rodent Populations. *The Quarterly Review of Biology* 69, 473–490. <https://doi.org/10.1086/418744>
- Adriaensen, F., Chardon, J.P., De Blust, G., Swinnen, E., Villalba, S., Gulinck, H., Matthysen, E., 2003. The application of ‘least-cost’ modelling as a functional landscape model. *Landscape and Urban Planning* 64, 233–247. [https://doi.org/10.1016/S0169-2046\(02\)00242-6](https://doi.org/10.1016/S0169-2046(02)00242-6)
- Aiello, C., Galloway, N., Prentice, P., Darby, N., Hughson, D., Epps, C., 2023. Movement models and simulation reveal highway impacts and mitigation opportunities for a metapopulation-distributed species. *Landscape Ecology* 38, 1085–1103. <https://doi.org/10.1007/s10980-023-01600-6>
- Albert, C.H., Rayfield, B., Dumitru, M., Gonzalez, A., 2017. Applying network theory to prioritize multispecies habitat networks that are robust to climate and land-use change. *Conservation Biology* 31, 1383–1396. <https://doi.org/10.1111/cobi.12943>
- Amos, J.N., Harrison, 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, 1556–1568. <https://doi.org/10.1890/13-1328.1>
- Anderson, C.D., Epperson, B.K., Fortin, M., Holderegger, R., James, P.M.A., Rosenberg, M.S., Scribner, K.T., Spear, S., 2010. Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Molecular Ecology* 19, 3565–3575. <https://doi.org/10.1111/j.1365-294X.2010.04757.x>
- Antunes, B., Figueiredo-Vázquez, C., Dudek, K., Liana, M., Pabijan, M., Zieliński, P., Babik, W., 2023. Landscape genetics reveals contrasting patterns of connectivity in two newt species (*Lissotriton montandoni* and *L. vulgaris*). *Molecular Ecology* 32, 4515–4530. <https://doi.org/10.1111/mec.16543>
- Arnoux, E., Eraud, C., Navarro, N., Tougard, C., Thomas, A., Cavallo, F., Vetter, N., Faivre, B., Garnier, S., 2014. Morphology and genetics reveal an intriguing pattern of differentiation at a very small geographic scale in a bird species, the forest thrush *Turdus lherminieri*. *Heredity* 113, 514–525. <https://doi.org/10.1038/hdy.2014.56>
- Atzeni, L., Wang, J., Riordan, P., Shi, K., Cushman, S.A., 2023. Landscape resistance to gene flow in a snow leopard population from Qilianshan National Park, Gansu, China. *Landscape Ecology* 38, 1847–1868. <https://doi.org/10.1007/s10980-023-01660-8>
- Aylward, C.M., Murdoch, J.D., Kilpatrick, C.W., 2020. Multiscale landscape genetics of American marten at their southern range periphery. *Heredity* 124, 550–561. <https://doi.org/10.1038/s41437-020-0295-y>
- Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A., 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE* 3, e3376. <https://doi.org/10.1371/journal.pone.0003376>

- Balkenhol, N., Cushman, S.A., Storfer, A.T., Waits, L.P. (Eds.), 2016. *Landscape Genetics*, 1st ed. Wiley. <https://doi.org/10.1002/9781118525258>
- Balkenhol, N., Gugerli, F., Cushman, S.A., Waits, L.P., Coulon, A., Arntzen, J.W., Holderegger, R., Wagner, H.H., Participants of the Landscape Genetics Research Agenda Workshop 2007, 2009. Identifying future research needs in landscape genetics: where to from here? *Landscape Ecology* 24, 455–463. <https://doi.org/10.1007/s10980-009-9334-z>
- 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, 1301–1310. <https://doi.org/10.1016/j.ecolind.2011.02.003>
- Barbosa, S., Paupério, J., Herman, J.S., Ferreira, C.M., Pita, R., Vale-Gonçalves, H.M., Cabral, J.A., Garrido-García, J.A., Soriguer, R.C., Beja, P., Mira, A., Alves, P.C., Searle, J.B., 2017. Endemic species may have complex histories: within-refugium phylogeography of an endangered Iberian vole. *Molecular Ecology* 26, 951–967. <https://doi.org/10.1111/mec.13994>
- Barr, K.R., Kus, B.E., Preston, K.L., Howell, S., Perkins, E., Vandergast, A.G., 2015. Habitat fragmentation in coastal southern California disrupts genetic connectivity in the cactus wren (*Campylorhynchus brunneicapillus*). *Molecular Ecology* 24, 2349–2363. <https://doi.org/10.1111/mec.13176>
- Bartoń, K., 2013. MuMIn: Multi-model inference, R package version 1.10.0.
- Bauder, J.M., Peterman, W.E., Spear, S.F., Jenkins, C.L., Whiteley, A.R., McGarigal, K., 2021. Multiscale assessment of functional connectivity: Landscape genetics of eastern indigo snakes in an anthropogenically fragmented landscape in central Florida. *Molecular Ecology* 30, 3422–3438. <https://doi.org/10.1111/mec.15979>
- Beaumont, M.A., Zhang, W., Balding, D.J., 2002. Approximate Bayesian Computation in Population Genetics. *Genetics* 162, 2025–2035. <https://doi.org/10.1093/genetics/162.4.2025>
- Beier, P., Majka, D.R., Newell, S.L., 2009. Uncertainty analysis of least-cost modeling for designing wildlife linkages. *Ecological Applications* 19, 2067–2077. <https://doi.org/10.1890/08-1898.1>
- Beier, P., Majka, D.R., Spencer, W.D., 2008. Forks in the Road: Choices in Procedures for Designing Wildland Linkages: *Design of Wildlife Linkages*. *Conservation Biology* 22, 836–851. <https://doi.org/10.1111/j.1523-1739.2008.00942.x>
- Beier, P., Penrod, K., Luke, C., Spencer, W., Cabanero, C. 2006. South Coast Missing Linkages: restoring connectivity to wildlands in the largest metropolitan area in the United States. *Connectivity and conservation*, 555-586.
- Bellil, A., 2023. Des discours aux cartographies de la Trame Verte et Bleue: représentations et prises en compte de la connectivité écologique dans les Schémas de Cohérence Territoriale en région Bourgogne Franche-Comté (PhD Thesis). Bourgogne Franche-Comté.
- Beninde, J., Wittische, J., Frantz, A.C., 2023. Quantifying uncertainty in inferences of landscape genetic resistance due to choice of individual-based genetic distance metric. *Molecular Ecology Resources* 1755–0998.13831. <https://doi.org/10.1111/1755-0998.13831>

- 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, 431–443. <https://doi.org/10.1111/mec.13923>
- Bertorelle, G., Benazzo, A., Mona, S., 2010. ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular Ecology* 19, 2609–2625. <https://doi.org/10.1111/j.1365-294X.2010.04690.x>
- Betts, M.G., Wolf, C., Pfeifer, M., Banks-Leite, C., Arroyo-Rodríguez, V., Ribeiro, D.B., Barlow, J., Eigenbrod, F., Faria, D., Fletcher, R.J., Hadley, A.S., Hawes, J.E., Holt, R.D., Klingbeil, B., Kormann, U., Lens, L., Levi, T., Medina-Rangel, G.F., Melles, S.L., Mezger, D., Morante-Filho, J.C., Orme, C.D.L., Peres, C.A., Phalan, B.T., Pidgeon, A., Possingham, H., Ripple, W.J., Slade, E.M., Somarriba, E., Tobias, J.A., Tylianakis, J.M., Urbina-Cardona, J.N., Valente, J.J., Watling, J.I., Wells, K., Wearn, O.R., Wood, E., Young, R., Ewers, R.M., 2019. Extinction filters mediate the global effects of habitat fragmentation on animals. *Science* 366, 1236–1239. <https://doi.org/10.1126/science.aax9387>
- Biek, R., Real, L.A., 2010. The landscape genetics of infectious disease emergence and spread. *Molecular Ecology* 19, 3515–3531. <https://doi.org/10.1111/j.1365-294X.2010.04679.x>
- Bishop-Taylor, R., Tulbure, M.G., Broich, M., 2018. Evaluating static and dynamic landscape connectivity modelling using a 25-year remote sensing time series. *Landscape Ecology* 33, 625–640. <https://doi.org/10.1007/s10980-018-0624-1>
- Blair, C., Jiménez Arcos, V.H., De La Cruz, F.R.M., Murphy, R.W., 2015. Historical and contemporary demography of leaf-toed geckos (Phyllodactylidae: *Phyllodactylus tuberculatus saxatilis*) in the Mexican dry forest. *Conservation Genetics* 16, 419–429. <https://doi.org/10.1007/s10592-014-0668-y>
- Bohling, J.H., 2016. Strategies to address the conservation threats posed by hybridization and genetic introgression. *Biological Conservation* 203, 321–327. <https://doi.org/10.1016/j.biocon.2016.10.011>
- Bolliger, J., Lander, T., Balkenhol, N., 2014. Landscape genetics since 2003: status, challenges and future directions. *Landscape Ecology* 29, 361–366. <https://doi.org/10.1007/s10980-013-9982-x>
- Bond, M.L., Bradley, C.M., Kiffner, C., Morrison, T.A., Lee, D.E., 2017. A multi-method approach to delineate and validate migratory corridors. *Landscape Ecology* 32, 1705–1721. <https://doi.org/10.1007/s10980-017-0537-4>
- Boulanger, E., Dalongeville, A., Andrello, M., Mouillot, D., Manel, S., 2020. Spatial graphs highlight how multi-generational dispersal shapes landscape genetic patterns. *Ecography* 43, 1167–1179. <https://doi.org/10.1111/ecog.05024>
- Bourdouxhe, A., Dufnot, R., Radoux, J., Dufrêne, 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. <https://doi.org/10.1016/j.jnc.2020.125901>
- Bowcock, A.M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J.R., Cavalli-Sforza, L.L., 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 368, 455–457. <https://doi.org/10.1038/368455a0>

- Bowman, J.L., Greenhorn, J.E., Marrotte, R.R., McKay, M.M., Morris, K.Y., Prentice, M.B., Wehtje, M., 2016. On applications of landscape genetics. *Conservation Genetics* 17, 753–760. <https://doi.org/10.1007/s10592-016-0834-5>
- Bowman, J.L., Leopold, B.D., Vilella, F.J., Gill, D.A., 2004. A spatially explicit model, derived from demographic variables, to predict attitudes toward black bear restoration. *Journal of Wildlife Management* 68, 223–232. [https://doi.org/10.2193/0022-541X\(2004\)068\[0223:ASEMDF\]2.0.CO;2](https://doi.org/10.2193/0022-541X(2004)068[0223:ASEMDF]2.0.CO;2)
- Brooks, M., Bolker, B., Kristensen, K., Mächler, M., Magnusson, A., Skaug, H., Nielsen, A., Berg, C., van Benthem, K., 2017. *glmmTMB: Generalized linear mixed models using Template Model Builder*.
- Bunn, A.G., Urban, D.L., Keitt, T.H., 2000. Landscape connectivity: A conservation application of graph theory. *Journal of Environmental Management* 59, 265–278. <https://doi.org/10.1006/jema.2000.0373>
- Burel, F., 2003. *Landscape ecology: concepts, methods, and applications*. CRC Press.
- Burgess, S.M., Garrick, R.C., 2020. Regional replication of landscape genetics analyses of the Mississippi slimy salamander, *Plethodon mississippi*. *Landscape Ecology* 35, 337–351. <https://doi.org/10.1007/s10980-019-00949-x>
- Burkhardt, J.J., Peterman, W.E., Brocato, E.R., Romine, K.M., Willis, M.M.S., Ousterhout, B.H., Anderson, T.L., Drake, D.L., Rowland, F.E., Semlitsch, R.D., Eggert, L.S., 2017. The influence of breeding phenology on the genetic structure of four pond-breeding salamanders. *Ecology and Evolution* 7, 4670–4681. <https://doi.org/10.1002/ece3.3060>
- Butchart, S.H.M., Walpole, M., Collen, B., Van Strien, A., Scharlemann, J.P.W., Almond, R.E.A., Baillie, J.E.M., Bomhard, B., Brown, C., Bruno, J., Carpenter, K.E., Carr, G.M., Chanson, J., Chenery, A.M., Csirke, J., Davidson, N.C., Dentener, F., Foster, M., Galli, A., Galloway, J.N., Genovesi, P., Gregory, R.D., Hockings, M., Kapos, V., Lamarque, J.-F., Leverington, F., Loh, J., McGeoch, M.A., McRae, L., Minasyan, A., Morcillo, M.H., Oldfield, T.E.E., Pauly, D., Quader, S., Revenga, C., Sauer, J.R., Skolnik, B., Spear, D., Stanwell-Smith, D., Stuart, S.N., Symes, A., Tierney, M., Tyrrell, T.D., Vié, J.-C., Watson, R., 2010. Global Biodiversity: Indicators of Recent Declines. *Science* 328, 1164–1168. <https://doi.org/10.1126/science.1187512>
- Cabrera, A.A., Palsbøll, P.J., 2017. Inferring past demographic changes from contemporary genetic data: A simulation-based evaluation of the ABC methods implemented in DIY ABC. *Molecular Ecology Resources* 17. <https://doi.org/10.1111/1755-0998.12696>
- Calabrese, J.M., Fagan, W.F., 2004. A comparison-shopper’s guide to connectivity metrics. *Frontiers in Ecology and the Environment* 2, 529–536. [https://doi.org/10.1890/15409295\(2004\)002\[0529:ACGTTCM\]2.0.CO;2](https://doi.org/10.1890/15409295(2004)002[0529:ACGTTCM]2.0.CO;2)
- Capurro, J.M.G., Cornelius, C., Borges, S.H., Cohn-Haft, M., Aleixo, A., Metzger, J.P., Ribas, C.C., 2013. Combining phylogeography and landscape genetics of *Xenopipo atronitens* (Aves: Pipridae), a white sand *campina* specialist, to understand Pleistocene landscape evolution in Amazonia: Pleistocene Landscape Evolution. *Biological Journal of Linnean Society London* 110, 60–76. <https://doi.org/10.1111/bij.12102>
- Carroll, C., McRAE, B.H., 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, 10.

- Carroll, K.A., Hansen, A.J., Inman, R.M., Lawrence, R.L., Hoegh, A.B., 2020. Testing landscape resistance layers and modeling connectivity for wolverines in the western United States. *Global Ecology and Conservation* 23, e01125. <https://doi.org/10.1016/j.gecco.2020.e01125>
- Ceballos, G., Ehrlich, P.R., Barnosky, A.D., García, A., Pringle, R.M., Palmer, T.M., 2015. Accelerated modern human-induced species losses: Entering the sixth mass extinction. *Science Advances* 1, e1400253. <https://doi.org/10.1126/sciadv.1400253>
- Ceballos, G., Ehrlich, P.R., Dirzo, R., 2017. Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proceedings of the national academy of sciences. U.S.A.* 114. <https://doi.org/10.1073/pnas.1704949114>
- Clarke, R.T., Rothery, P., Raybould, A.F., 2002. Confidence limits for regression relationships between distance matrices: Estimating gene flow with distance. *JABES* 7, 361–372. <https://doi.org/10.1198/108571102320>
- Clauzel, C., Godet, C., 2020. Combining spatial modeling tools and biological data for improved multispecies assessment in restoration areas. *Biological Conservation* 250, 108713. <https://doi.org/10.1016/j.biocon.2020.108713>
- Clobert, J., 2012. *Dispersal ecology and evolution*. Oxford University Press.
- Cornuet, J.-M., Santos, F., Beaumont, M.A., Robert, C.P., Marin, J.-M., Balding, D.J., Guillemaud, T., Estoup, A., 2008. Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics*, 24(23), 2713-2719.
- 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: Earth and Environment* 40, 7–37. <https://doi.org/10.1177/0309133315598713>
- Coulon, A., Cosson, J.F., Angibault, J.M., Cargnelutti, B., Galan, M., Morellet, N., Petit, E., Aulagnier, S., Hewison, A.J.M., 2004. Landscape connectivity influences gene flow in a roe deer population inhabiting a fragmented landscape: an individual-based approach. *Molecular Ecology* 13, 2841–2850. <https://doi.org/10.1111/j.1365-294X.2004.02253.x>
- Cowie, R.H., Bouchet, P., Fontaine, B., 2022. The Sixth Mass Extinction: fact, fiction or speculation? *Biological Reviews* 97, 640–663. <https://doi.org/10.1111/brv.12816>
- Crispo, E., Bentzen, P., Reznick, D.N., Kinnison, M.T., Hendry, A.P., 2006. The relative influence of natural selection and geography on gene flow in guppies. *Molecular Ecology* 15, 49–62. <https://doi.org/10.1111/j.1365-294X.2005.02764.x>
- Crispo, E., Moore, J.-S., Lee-Yaw, J.A., Gray, S.M., Haller, B.C., 2011. Broken barriers: Human-induced changes to gene flow and introgression in animals: An examination of the ways in which humans increase genetic exchange among populations and species and the consequences for biodiversity. *Bioessays* 33, 508–518. <https://doi.org/10.1002/bies.201000154>
- Crooks, K.R., Burdett, C.L., Theobald, D.M., King, S.R.B., Di Marco, M., Rondinini, C., Boitani, L., 2017. Quantification of habitat fragmentation reveals extinction risk in terrestrial mammals. *Proceedings of the national academy of sciences. U.S.A.* 114, 7635–7640. <https://doi.org/10.1073/pnas.1705769114>



- Crow, J.F., Aoki, K., 1984. Group selection for a polygenic behavioral trait: estimating the degree of population subdivision. *Proceedings of the national academy of sciences*. U.S.A. 81, 6073–6077. <https://doi.org/10.1073/pnas.81.19.6073>
- Cushman, S.A., Evans, J.S., McGarigal, K., 2010. Landscape Ecology: Past, Present, and Future, in: Cushman, S.A., Huettmann, F. (Eds.), *Spatial Complexity, Informatics, and Wildlife Conservation*. Springer Japan, Tokyo, pp. 65–82. [https://doi.org/10.1007/978-4-431-87771-4\\_4](https://doi.org/10.1007/978-4-431-87771-4_4)
- Cushman, S.A., Landguth, E.L., 2012. Multi-taxa population connectivity in the Northern Rocky Mountains. *Ecological Modelling* 231, 101–112. <https://doi.org/10.1016/j.ecolmodel.2012.02.011>
- Cushman, S.A., Landguth, E.L., 2010. Scale dependent inference in landscape genetics. *Landscape Ecology* 25, 967–979. <https://doi.org/10.1007/s10980-010-9467-0>
- 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, 486–499. <https://doi.org/10.1086/506976>
- Cushman, S.A., McRae, B., Adriaensen, F., Beier, P., Shirley, M., Zeller, K., 2013. Biological corridors and connectivity, in: Macdonald, D.W., Willis, K.J. (Eds.), *Key Topics in Conservation Biology 2*. Wiley, pp. 384–404. <https://doi.org/10.1002/9781118520178.ch21>
- Cushman, S.A., Shirk, A., Landguth, E.L., 2012. Separating the effects of habitat area, fragmentation and matrix resistance on genetic differentiation in complex landscapes. *Landscape Ecology* 27, 369–380. <https://doi.org/10.1007/s10980-011-9693-0>
- Daniel, A., Savary, P., Foltête, J., 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* 37, e14047. <https://doi.org/10.1111/cobi.14047>
- Delord, C., Petit, E.J., Blanchet, S., Longin, G., Rinaldo, R., Vigouroux, R., Roussel, J.-M., Le Bail, P.-Y., Launey, S., 2023. Contrasts in riverscape patterns of intraspecific genetic variation in a diverse Neotropical fish community of high conservation value. *Heredity* 131, 1–14. <https://doi.org/10.1038/s41437-023-00616-7>
- DiLeo, M.F., Wagner, H.H., 2016. A Landscape Ecologist’s Agenda for Landscape Genetics. *Current Landscape Ecology Reports* 1, 115–126. <https://doi.org/10.1007/s40823-016-0013-x>
- Diniz, M.F., Cushman, S.A., Machado, R.B., De Marco Júnior, P., 2020. Landscape connectivity modeling from the perspective of animal dispersal. *Landscape Ecology* 35, 41–58. <https://doi.org/10.1007/s10980-019-00935-3>
- Draheim, H.M., Moore, J.A., Fortin, M., Scribner, K.T., 2018. Beyond the snapshot: Landscape genetic analysis of time series data reveal responses of American black bears to landscape change. *Evolutionary Applications* 11, 1219–1230. <https://doi.org/10.1111/eva.12617>
- Draheim, H.M., Moore, J.A., Winterstein, S.R., Scribner, K.T., 2021. Spatial genetic structure and landscape connectivity in black bears: Investigating the significance of using different land cover datasets and classifications in landscape genetics analyses. *Ecology and Evolution* 11, 978–989
- Driezen, K., Adriaensen, F., Rondinini, C., Doncaster, C.P., Matthysen, E., 2007. Evaluating least-cost model predictions with empirical dispersal data: A case-study using radiotracking data of hedgehogs

- (*Erinaceus europaeus*). *Ecological Modelling* 209, 314–322. <https://doi.org/10.1016/j.ecolmodel.2007.07.002>
- Dutta, T., Sharma, S., Meyer, N.F.V., Larroque, J., Balkenhol, N., 2022. An overview of computational tools for preparing, constructing and using resistance surfaces in connectivity research. *Landscape Ecology* 37, 2195–2224. <https://doi.org/10.1007/s10980-022-01469-x>
- Dyer, R.J., 2015. Population Graphs and Landscape Genetics. *Annual Review of Ecology, Evolution and Systematics* 46, 327–342. <https://doi.org/10.1146/annurev-ecolsys-112414-054150>
- Dyer, R.J., Nason, J.D., 2004. Population Graphs: the graph theoretic shape of genetic structure. *Molecular Ecology* 13, 1713–1727. <https://doi.org/10.1111/j.1365-294X.2004.02177.x>
- Dyer, R.J., Nason, J.D., Garrick, R.C., 2010a. Landscape modelling of gene flow: improved power using conditional genetic distance derived from the topology of population networks. *Molecular Ecology* 19, 3746–3759. <https://doi.org/10.1111/j.1365-294X.2010.04748.x>
- Emel, S.L., Olson, D.H., Knowles, L.L., Storfer, A., 2019. Comparative landscape genetics of two endemic torrent salamander species, *Rhyacotriton kezeri* and *R. variegatus*: implications for forest management and species conservation. *Conservation Genetics* 20, 801–815. <https://doi.org/10.1007/s10592-019-01172-6>
- Engler, J.O., Balkenhol, N., Filz, K.J., Habel, J.C., Rödder, D., 2014. Comparative Landscape Genetics of Three Closely Related Sympatric Hesperid Butterflies with Diverging Ecological Traits. *PLoS ONE* 9, e106526. <https://doi.org/10.1371/journal.pone.0106526>
- Epps, C.W., Keyghobadi, N., 2015. Landscape genetics in a changing world: disentangling historical and contemporary influences and inferring change. *Molecular Ecology* 24, 6021–6040. <https://doi.org/10.1111/mec.13454>
- Epps, C.W., Wasser, S.K., Keim, J.L., Mutayoba, B.M., Brashares, J.S., 2013. Quantifying past and present connectivity illuminates a rapidly changing landscape for the African elephant. *Molecular Ecology* 22, 1574–1588. <https://doi.org/10.1111/mec.12198>
- Eraud, C., Arnoux, E., Levesque, A., Van Laere, G., Magnin, H., 2012. Biologie des populations et statut de conservation des oiseaux endémiques des Antilles en Guadeloupe. Rapport d'étude ONCFS-Parc National de la Guadeloupe 107.
- Fagan, W.F., Calabrese, J.M., 2006. Quantifying connectivity: balancing metric performance with data requirements, in: Crooks, K.R., Sanjayan, M. (Eds.), *Connectivity Conservation*. Cambridge University Press, pp. 297–317. <https://doi.org/10.1017/CBO9780511754821.014>
- Fagan, W.F., Holmes, E.E., 2006. Quantifying the extinction vortex. *Ecology Letters* 9, 51–60. <https://doi.org/10.1111/j.1461-0248.2005.00845.x>
- Fahrig, L., 2013. Rethinking patch size and isolation effects: the habitat amount hypothesis. *J. Biogeogr.* 40, 1649–1663. <https://doi.org/10.1111/jbi.12130>
- Fahrig, L., 1997. Relative Effects of Habitat Loss and Fragmentation on Population Extinction. *The Journal of Wildlife Management* 61, 603. <https://doi.org/10.2307/3802168>

- Fahrig, L., Watling, J., Arnillas, C., Arroyo-Rodríguez, V., Jörger-Hickfang, T., Müller, J., Pereira, H., Riva, F., Rösch, V., Seibold, S., Tschardtke, T., May, F., 2022. Resolving the SLOSS dilemma for biodiversity conservation: a research agenda (preprint). *Biological Reviews*, 97(1), 99-114. <https://doi.org/10.22541/au.160821552.23263247/v1>
- Fenderson, L.E., Kovach, A.I., Llamas, B., 2020. Spatiotemporal landscape genetics: Investigating ecology and evolution through space and time. *Molecular Ecology* 29, 218–246. <https://doi.org/10.1111/mec.15315>
- Fischer, J., Lindenmayer, D.B., 2007. Landscape modification and habitat fragmentation: a synthesis. *Global Ecology and Biogeography* 16, 265–280. <https://doi.org/10.1111/j.1466-8238.2007.00287.x>
- Flavenot, T., Fellous, S., Abdelkrim, J., Bague, M., Coulon, A., 2015. Impact of quarrying on genetic diversity: an approach across landscapes and over time. *Conservation Genetics* 16, 181–194. <https://doi.org/10.1007/s10592-014-0650-8>
- Fletcher, R.J., Didham, R.K., Banks-Leite, C., Barlow, J., Ewers, R.M., Rosindell, J., Holt, R.D., Gonzalez, A., Pardini, R., Damschen, E.I., Melo, F.P.L., Ries, L., Prevedello, J.A., Tschardtke, T., Laurance, W.F., Lovejoy, T., Haddad, N.M., 2018. Is habitat fragmentation good for biodiversity? *Biological Conservation* 226, 9–15. <https://doi.org/10.1016/j.biocon.2018.07.022>
- Fletcher, R.J., Sefair, J.A., Wang, C., Poli, C.L., Smith, T.A.H., Bruna, E.M., Holt, R.D., Barfield, M., Marx, A.J., Acevedo, M.A., 2019. Towards a unified framework for connectivity that disentangles movement and mortality in space and time. *Ecology Letters* 22, 1680–1689. <https://doi.org/10.1111/ele.13333>
- Foltête, J.-C., Clauzel, C., Vuidel, G., 2012. A software tool dedicated to the modelling of landscape networks. *Environmental Modelling & Software* 38, 316–327. <https://doi.org/10.1016/j.envsoft.2012.07.002>
- 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. <https://doi.org/10.1016/j.landurbplan.2013.12.012>
- Foltête, J.-C., Giraudoux, P., 2012. A graph-based approach to investigating the influence of the landscape on population spread processes. *Ecological Indicators* 18, 684–692. <https://doi.org/10.1016/j.ecolind.2012.01.011>
- 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, 1035–1052. <https://doi.org/10.1007/s10980-020-00998-7>
- Foltête, J.-C., Vuidel, G., Savary, P., Clauzel, C., Sahraoui, Y., Girardet, X., Bourgeois, M., 2021. Graphab: An application for modeling and managing ecological habitat networks. *Software Impacts* 8, 100065. <https://doi.org/10.1016/j.simpa.2021.100065>
- Forman, R.T.T., Godron, M., 1986. *Landscape Ecology*. Wiley.
- Foster, D., Swanson, F., Aber, J., Burke, I., Brokaw, N., Tilman, D., Knapp, A., 2003. The Importance of Land-Use Legacies to Ecology and Conservation. *BioScience* 53, 77-88. [https://doi.org/10.1641/0006-3568\(2003\)053\[0077:TIOLUL\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2003)053[0077:TIOLUL]2.0.CO;2)
- Frankham, R., 2015. Genetic rescue of small inbred populations: meta-analysis reveals large and consistent benefits of gene flow. *Molecular Ecology* 24, 2610–2618. <https://doi.org/10.1111/mec.13139>

- Frankham, R., Ballou, J.D., Ralls, K., Eldridge, M., Dudash, M.R., Fenster, C.B., Lacy, R.C., Sunnucks, P., 2017. Genetic Management of Fragmented Animal and Plant Populations. Oxford University Press. <https://doi.org/10.1093/oso/9780198783398.001.0001>
- Frantz, A.C., McDevitt, A.D., Pope, L.C., Kochan, J., Davison, J., Clements, C.F., Elmeros, M., Molina-Vacas, G., Ruiz-Gonzalez, A., Balestrieri, A., Van Den Berge, K., Breyne, P., Do Linh San, E., Ågren, E.O., Suchentrunk, F., Schley, L., Kowalczyk, R., Kostka, B.I., Ćirović, D., Šprem, N., Colyn, M., Ghirardi, M., Racheva, V., Braun, C., Oliveira, R., Lanszki, J., Stubbe, A., Stubbe, M., Stier, N., Burke, T., 2014. Revisiting the phylogeography and demography of European badgers (*Meles meles*) based on broad sampling, multiple markers and simulations. *Heredity* 113, 443–453. <https://doi.org/10.1038/hdy.2014.45>
- Fu, Y.X., Li, W.H., 1997. Estimating the age of the common ancestor of a sample of DNA sequences. *Molecular Biology and Evolution* 14, 195–199. <https://doi.org/10.1093/oxfordjournals.molbev.a025753>
- Gaggiotti, O.E., Foll, M., 2010. Quantifying population structure using the F-model. *Molecular Ecology Resources* 10, 821–830. <https://doi.org/10.1111/j.1755-0998.2010.02873.x>
- Galpern, P., 2011. Patch-based graphs of landscape connectivity: A guide to construction, analysis and application for conservation. *Biological Conservation* 144, 44-55. <https://doi.org/10.1016/j.biocon.2010.09.002>
- Galpern, P., Peres-Neto, P.R., Polfus, J., Manseau, M., 2014. MEMGENE : Spatial pattern detection in genetic distance data. *Methods in Ecology and Evolution* 5, 1116–1120. <https://doi.org/10.1111/2041-210X.12240>
- Garg, K.M., Chattopadhyay, B., Koane, B., Sam, K., Rheindt, F.E., 2020. Last Glacial Maximum led to community-wide population expansion in a montane songbird radiation in highland Papua New Guinea. *BMC Evolutionary Biology* 20, 1-10, 82. <https://doi.org/10.1186/s12862-020-01646-z>
- Garnier, S., Alibert, P., Audiot, P., Prieur, B., Rasplus, J. -Y., 2004. Isolation by distance and sharp discontinuities in gene frequencies: implications for the phylogeography of an alpine insect species, *Carabus solieri*. *Molecular Ecology* 13, 1883–1897. <https://doi.org/10.1111/j.1365-294X.2004.02212.x>
- Gauffre, B., Mallez, S., Chapuis, M., Leblois, R., Litrico, I., Delaunay, S., Badenhauer, I., 2015. Spatial heterogeneity in landscape structure influences dispersal and genetic structure: empirical evidence from a grasshopper in an agricultural landscape. *Molecular Ecology* 24, 1713–1728. <https://doi.org/10.1111/mec.13152>
- Ghoddousi, A., Buchholtz, E.K., Dietsch, A.M., Williamson, M.A., Sharma, S., Balkenhol, N., Kuemmerle, T., Dutta, T., 2021. Anthropogenic resistance: accounting for human behavior in wildlife connectivity planning. *One Earth* 4, 39–48. <https://doi.org/10.1016/j.oneear.2020.12.003>
- Gilpin, M.E., 1986. Minimum viable populations: processes of species extinction. *Conservation biology: the science of scarcity and diversity* 19–34.
- Girardet, X., Clauzel, C., 2018. Graphab. 14 réalisations à découvrir. Actes de la journée" Retour d'expérience sur Graphab" du 27 juin 2017.

- Gutiérrez-Rodríguez, J., Gonçalves, J., Civantos, E., Martínez-Solano, I., 2017. Comparative landscape genetics of pond-breeding amphibians in Mediterranean temporal wetlands: The positive role of structural heterogeneity in promoting gene flow. *Molecular Ecology* 26, 5407–5420. <https://doi.org/10.1111/mec.14272>
- Haddad, N.M., Brudvig, L.A., Clobert, J., Davies, K.F., Gonzalez, A., Holt, R.D., Lovejoy, T.E., Sexton, J.O., Austin, M.P., Collins, C.D., Cook, W.M., Damschen, E.I., Ewers, R.M., Foster, B.L., Jenkins, C.N., King, A.J., Laurance, W.F., Levey, D.J., Margules, C.R., Melbourne, B.A., Nicholls, A.O., Orrock, J.L., Song, D.-X., Townshend, J.R., 2015. Habitat fragmentation and its lasting impact on Earth's ecosystems. *Science Advances*. 1, e1500052. <https://doi.org/10.1126/sciadv.1500052>
- Haig, S.M., Bronaugh, W.M., Crowhurst, R.S., D'Elia, J., Eagles-Smith, C.A., Epps, C.W., Knaus, B., Miller, M.P., Moses, M.L., Oyler-McCance, S., Robinson, W.D., Sidlauskas, B., 2011. Genetic Applications in Avian Conservation. *The Auk* 128, 205–229. <https://doi.org/10.1525/auk.2011.128.2.205>
- Halley, J.M., Monokrousos, N., Mazaris, A.D., Newmark, W.D., Vokou, D., 2016. Dynamics of extinction debt across five taxonomic groups. *Nature Communications* 7, 12283. <https://doi.org/10.1038/ncomms12283>
- Hanski, I., 1994. A Practical Model of Metapopulation Dynamics. *The Journal of Animal Ecology* 63, 151. <https://doi.org/10.2307/5591>
- Harding, J.S., Benfield, E.F., Bolstad, P.V., Helfman, G.S., Jones, E.B.D., 1998. Stream biodiversity: The ghost of land use past. *Proc. Natl. Acad. Sci. U.S.A.* 95, 14843–14847. <https://doi.org/10.1073/pnas.95.25.14843>
- Harris, R.J., Reed, J.M., 2002. Behavioral barriers to non-migratory movements of birds. *Annales Zoologici Fennici* 39, 275–290.
- Hedrick, P.W., 2005. A standardized genetic differentiation measure. *Evolution* 59, 1633–1638. <https://doi.org/10.1111/j.0014-3820.2005.tb01814.x>
- Heller, N.E., Zavaleta, E.S., 2009. Biodiversity management in the face of climate change: A review of 22 years of recommendations. *Biological Conservation* 142, 14–32. <https://doi.org/10.1016/j.biocon.2008.10.006>
- Hewitt, G.M., 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London* 359, 183–195. <https://doi.org/10.1098/rstb.2003.1388>
- Holderegger, R., Kamm, U., Gugerli, F., 2006. Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecology* 21, 797–807. <https://doi.org/10.1007/s10980-005-5245-9>
- Holderegger, R., Wagner, H.H., 2008. Landscape Genetics. *BioScience* 58, 199–207. <https://doi.org/10.1641/B580306>
- Holzhauser, S.I.J., Ekschmitt, K., Sander, A.-C., Dauber, J., Wolters, V., 2006. Effect of historic landscape change on the genetic structure of the bush-cricket *Metrioptera roeseli*. *Landscape Ecology* 21, 891–899. <https://doi.org/10.1007/s10980-005-0438-9>
- Honnay, O., Coart, E., Butaye, J., Adriaens, D., Van Glabeke, S., Roldán-Ruiz, I., 2006. Low impact of present and historical landscape configuration on the genetics of fragmented *Anthyllis vulneraria* populations. *Biological Conservation* 127, 411–419. <https://doi.org/10.1016/j.biocon.2005.09.006>

- 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, 1898–1914. <https://doi.org/10.1111/j.1558-5646.1999.tb04571.x>
- Ingvarsson, P.K., 2001. Restoration of genetic variation lost – the genetic rescue hypothesis. *Trends in Ecology & Evolution* 16, 62–63. [https://doi.org/10.1016/S0169-5347\(00\)02065-6](https://doi.org/10.1016/S0169-5347(00)02065-6)
- IPBES, 2019. Summary for policymakers of the global assessment report on biodiversity and ecosystem services. Zenodo. <https://doi.org/10.5281/ZENODO.10413114>
- Jackson, H., Fahrig, L., 2014. Are ecologists conducting research at the optimal scale? *Global Ecology and Biogeography* 24, 52–63. <https://doi.org/10.1111/geb.12233>
- Jackson, N.D., Fahrig, L., 2014. Landscape context affects genetic diversity at a much larger spatial extent than population abundance. *Ecology* 95, 871–881. <https://doi.org/10.1890/13-0388.1>
- Jeffries, D.L., Copp, G.H., Lawson Handley, L., Olsén, K.H., Sayer, C.D., Hänfling, B., 2016. Comparing RADseq and microsatellites to infer complex phylogeographic patterns, an empirical perspective in the Crucian carp, *Carassius carassius*. *Molecular Ecology* 25, 2997–3018. <https://doi.org/10.1111/mec.13613>
- 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, 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>
- Kebaïli, C., Sherpa, S., Guéguen, M., Renaud, J., Rioux, D., Després, L., 2023. Comparative genetic and demographic responses to climate change in three peatland butterflies in the Jura massif. *Biological Conservation* 287, 110332. <https://doi.org/10.1016/j.biocon.2023.110332>
- Keeley, A.T.H., Beier, P., Gagnon, J.W., 2016. Estimating landscape resistance from habitat suitability: effects of data source and nonlinearities. *Landscape Ecology* 31, 2151–2162. <https://doi.org/10.1007/s10980-016-0387-5>
- Keeley, A.T.H., Beier, P., Jenness, J.S., 2021. Connectivity metrics for conservation planning and monitoring. *Biological Conservation* 255, 109008. <https://doi.org/10.1016/j.biocon.2021.109008>
- Keeley, A.T.H., Beier, P., Keeley, B.W., Fagan, M.E., 2017. Habitat suitability is a poor proxy for landscape connectivity during dispersal and mating movements. *Landscape and Urban Planning* 161, 90–102. <https://doi.org/10.1016/j.landurbplan.2017.01.007>
- Keller, D., Holderegger, R., Van Strien, M.J., 2013. Spatial scale affects landscape genetic analysis of a wetland grasshopper. *Molecular Ecology* 22, 2467–2482. <https://doi.org/10.1111/mec.12265>
- Keller, D., Holderegger, R., van Strien, M.J., Bolliger, J., 2015. How to make landscape genetics beneficial for conservation management? *Conservation Genetics* 16, 503–512. <https://doi.org/10.1007/s10592-014-0684-y>
- Kelly, R.P., Palumbi, S.R., 2010. Genetic Structure Among 50 Species of the Northeastern Pacific Rocky Intertidal Community. *PLoS ONE* 5, e8594. <https://doi.org/10.1371/journal.pone.0008594>

- Keyghobadi, N., 2007. The genetic implications of habitat fragmentation for animals. *Can. J. Zool.* 85, 1049–1064. <https://doi.org/10.1139/Z07-095>
- Keyghobadi, N., Roland, J., Matter, S.F., Strobeck, C., 2005. Among- and within-patch components of genetic diversity respond at different rates to habitat fragmentation: an empirical demonstration. *Proceedings of the Royal Society B: Biological Sciences* 272, 553–560. <https://doi.org/10.1098/rspb.2004.2976>
- Khattar, G., Savary, P., Peres-Neto, P.R., 2024. The biotic and abiotic contexts of ecological selection mediate the dominance of distinct dispersal strategies in competitive metacommunities. *Proceedings of the national academy of sciences* 379, 20230132. <https://doi.org/10.1098/rstb.2023.0132>
- Khimoun, A., Arnoux, E., Martel, G., Pot, A., Eraud, C., Condé, B., Loubon, M., Théron, F., Covas, R., Faivre, B., Garnier, S., 2016a. Contrasted patterns of genetic differentiation across eight bird species in the Lesser Antilles. *Genetica* 144, 125–138. <https://doi.org/10.1007/s10709-016-9883-4>
- Khimoun, A., Arnoux, E., Martel, G., Pot, A., Eraud, C., Condé, B., Loubon, M., Théron, F., Covas, R., Faivre, B., Garnier, S., 2016b. Contrasted patterns of genetic differentiation across eight bird species in the Lesser Antilles. *Genetica* 144, 125–138. <https://doi.org/10.1007/s10709-016-9883-4>
- 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., 2016c. Habitat specialization predicts genetic response to fragmentation in tropical birds. *Molecular Ecology* 25, 3831–3844. <https://doi.org/10.1111/mec.13733>
- 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, 4906–4919. <https://doi.org/10.1111/mec.14233>
- Kierepka, E.M., Anderson, S.J., Swihart, R.K., Rhodes, O.E., 2020. Differing, multiscale landscape effects on genetic diversity and differentiation in eastern chipmunks. *Heredity* 124, 457–468. <https://doi.org/10.1038/s41437-020-0293-0>
- Kierepka, E.M., Anderson, S.J., Swihart, R.K., Rhodes, O.E., 2016. Evaluating the influence of life-history characteristics on genetic structure: a comparison of small mammals inhabiting complex agricultural landscapes. *Ecology and Evolution* 6, 6376–6396. <https://doi.org/10.1002/ece3.2269>
- Kimmig, S.E., Beninde, J., Brandt, M., Schleimer, A., Kramer-Schadt, S., Hofer, H., Börner, K., Schulze, C., Wittstatt, U., Heddergott, M., Halczok, T., Staubach, C., Frantz, A.C., 2020. Beyond the landscape: Resistance modelling infers physical and behavioural gene flow barriers to a mobile carnivore across a metropolitan area. *Molecular Ecology* 29, 466–484. <https://doi.org/10.1111/mec.15345>
- Kimura, M., others, 1968. Evolutionary rate at the molecular level. *Nature* 217, 624–626.
- Kingman, J.F.C., 1982. The coalescent. *Stochastic Processes and their Applications* 13, 235–248. [https://doi.org/10.1016/0304-4149\(82\)90011-4](https://doi.org/10.1016/0304-4149(82)90011-4)
- Kittlein, M.J., Mora, M.S., Mapelli, F.J., Austrich, A., Gaggiotti, O.E., 2022. Deep learning and satellite imagery predict genetic diversity and differentiation. *Methods in Ecology and Evolution* 13, 711–721. <https://doi.org/10.1111/2041-210X.13775>

- Koen, E.L., Bowman, J., 2015. Node-based measures of connectivity in genetic networks. *Molecular Ecology Resources*, 16, 69–79. <https://doi.org/10.1111/1755-0998.12423>
- Koen, E.L., Bowman, J., Sadowski, C., Walpole, A.A., 2014. Landscape connectivity for wildlife: development and validation of multispecies linkage maps. *Methods in Ecology and Evolution* 5, 626–633. <https://doi.org/10.1111/2041-210X.12197>
- Kozakiewicz, C.P., Ricci, L., Patton, A.H., Stahlke, A.R., Hendricks, S.A., Margres, M.J., Ruiz-Aravena, M., Hamilton, D.G., Hamede, R., McCallum, H., Jones, M.E., Hohenlohe, P.A., Storfer, A., 2020. Comparative landscape genetics reveals differential effects of environment on host and pathogen genetic structure in Tasmanian devils (*Sarcophilus harrisi*) and their transmissible tumour. *Molecular Ecology* 29, 3217–3233. <https://doi.org/10.1111/mec.15558>
- Kretser, H.E., Curtis, P.D., Knuth, B.A., 2009. Landscape, Social, and Spatial Influences on Perceptions of Human–Black Bear Interactions in the Adirondack Park, NY. *Human Dimensions of Wildlife* 14, 393–406. <https://doi.org/10.1080/10871200903055318>
- Laita, A., Kotiaho, J.S., Monkkonen, M., 2011. Graph-theoretic connectivity measures: what do they tell us about connectivity? *Landscape Ecology* 17, 951–967. <https://doi.org/10.1007/s10980-011-9620-4>
- Laliberté, J., St-Laurent, M.-H., 2020. Validation of functional connectivity modeling: The Achilles’ heel of landscape connectivity mapping. *Landscape and Urban Planning* 202, 103878. <https://doi.org/10.1016/j.landurbplan.2020.103878>
- Landguth, E.L., Cushman, S.A., Schwartz, M.K., McKelvey, K.S., Murphy, M., Luikart, G., 2010. Quantifying the lag time to detect barriers in landscape genetics. *Molecular Ecology* 19, 4179–4191. <https://doi.org/10.1111/j.1365-294X.2010.04808.x>
- Langhammer, P.F., Bull, J.W., Bicknell, J.E., Oakley, J.L., Brown, M.H., Bruford, M.W., Butchart, S.H.M., Carr, J.A., Church, D., Cooney, R., Cutajar, S., Foden, W., Foster, M.N., Gascon, C., Geldmann, J., Genovesi, P., Hoffmann, M., Howard-McCombe, J., Lewis, T., Macfarlane, N.B.W., Melvin, Z.E., Merizalde, R.S., Morehouse, M.G., Pagad, S., Polidoro, B., Sechrest, W., Segelbacher, G., Smith, K.G., Steadman, J., Strongin, K., Williams, J., Woodley, S., Brooks, T.M., 2024. The positive impact of conservation action. *Science* 384, 453–458. <https://doi.org/10.1126/science.adj6598>
- Latta, R.G., Mitton, J.B., 1997. A comparison of population differentiation across four classes of gene marker in limber pine (*Pinus flexilis* James). *Genetics* 146, 1153–1163. <https://doi.org/10.1093/genetics/146.3.1153>
- Latter, B.D.H., 1973. The island model of population differentiation: a general solution. *Genetics* 73, 147–157. <https://doi.org/10.1093/genetics/73.1.147>
- Laundre, J.W., Hernandez, L., Ripple, W.J., 2010. The landscape of fear: ecological implications of being afraid. *The open ecology journal*, 3(1) <https://doi.org/10.2174/1874213001003030001>
- Legendre, P., Fortin, M., 2010. Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources* 10, 831–844. <https://doi.org/10.1111/j.1755-0998.2010.02866.x>
- Legendre, P., Legendre, L., 2012. *Numerical ecology*. Elsevier.



- Lenormand, T., 2002. Gene flow and the limits to natural selection. *Trends in Ecology & Evolution* 17, 183–189. [https://doi.org/10.1016/S0169-5347\(02\)02497-7](https://doi.org/10.1016/S0169-5347(02)02497-7)
- Levesque, Mathurin, Birdlife International, Association AMAZONA, 2008. Les Zones importantes pour la conservation des oiseaux en Guadeloupe.
- Levins, R., 1969. Some Demographic and Genetic Consequences of Environmental Heterogeneity for Biological Control. *Bulletin of the Entomological Society of America* 15, 237–240. <https://doi.org/10.1093/besa/15.3.237>
- Lopes, J.S., Beaumont, M.A., 2010. ABC: A useful Bayesian tool for the analysis of population data. *Infection, Genetics and Evolution*, 10, 825-832. <https://doi.org/10.1016/j.meegid.2009.10.010>
- Lovette, I.J., Bermingham, E., Seutin, G., Ricklefs, R.E., 1998. Evolutionary Differentiation in Three Endemic West Indian Warblers. *The Auk* 115, 890–903. <https://doi.org/10.2307/4089508>
- Lowe, W.H., Allendorf, F.W., 2010. What can genetics tell us about population connectivity. *Molecular Ecology* 19, 3038–3051. <https://doi.org/10.1111/j.1365-294X.2010.04688.x>
- Lucas, T.C.D., 2020. A translucent box: interpretable machine learning in ecology. *Ecological Monographs* 90, e01422. <https://doi.org/10.1002/ecm.1422>
- Macarthur, R.H., Wilson, E.O., 1967. *The Theory of Island Biogeography*, REV-Revised. ed. Princeton University Press.
- Manel, S., Gugerli, F., Thuiller, W., Alvarez, N., Legendre, P., Holderegger, R., Gielly, L., Taberlet, P., IntraBioDiv Consortium, 2012. Broad-scale adaptive genetic variation in alpine plants is driven by temperature and precipitation. *Molecular Ecology* 21, 3729–3738. <https://doi.org/10.1111/j.1365-294X.2012.05656.x>
- Manel, S., Holderegger, R., 2013. Ten years of landscape genetics. *Trends in Ecology & Evolution* 28, 614–621. <https://doi.org/10.1016/j.tree.2013.05.012>
- Manel, S., Schwartz, M.K., Luikart, G., Taberlet, P., 2003. Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution* 18, 189–197. [https://doi.org/10.1016/S0169-5347\(03\)00008-9](https://doi.org/10.1016/S0169-5347(03)00008-9)
- Mapelli, F.J., Boston, E.S.M., Fameli, A., Gómez Fernández, M.J., Kittlein, M.J., Mirol, P.M., 2020. Fragmenting fragments: landscape genetics of a subterranean rodent (Mammalia, Ctenomyidae) living in a human-impacted wetland. *Landscape Ecology* 35, 1089–1106. <https://doi.org/10.1007/s10980-020-01001-z>
- Marrotte, R.R., Bowman, J., 2017. The relationship between least-cost and resistance distance. *PLoS One* 12, e0174212. <https://doi.org/10.1371/journal.pone.0174212>
- Martensen, A.C., Saura, S., Fortin, M., 2017. Spatio-temporal connectivity: assessing the amount of reachable habitat in dynamic landscapes. *Methods in Ecology and Evolution* 8, 1253–1264. <https://doi.org/10.1111/2041-210X.12799>

- Martin, S.A., Peterman, W.E., Lipps, G.J., Gibbs, H.L., 2023. Inferring population connectivity in eastern massasauga rattlesnakes (*Sistrurus Catenatus*) using landscape genetics. *Ecological Applications* 33, e2793. <https://doi.org/10.1002/eap.2793>
- McClure, M.L., Hansen, A.J., Inman, R.M., 2016. Connecting models to movements: testing connectivity model predictions against empirical migration and dispersal data. *Landscape Ecology* 31, 1419–1432. <https://doi.org/10.1007/s10980-016-0347-0>
- McCluskey, E.M., Lulla, V., Peterman, W.E., Stryszowska-Hill, K.M., Denton, R.D., Fries, A.C., Langen, T.A., Johnson, G., Mockford, S.W., Gonser, R.A., 2022. Linking genetic structure, landscape genetics, and species distribution modeling for regional conservation of a threatened freshwater turtle. *Landscape Ecology* 37, 1017–1034. <https://doi.org/10.1007/s10980-022-01420-0>
- McGarigal, K., Marks, B.J., 1995. FRAGSTATS: spatial pattern analysis program for quantifying landscape structure. U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. <https://doi.org/10.2737/pnw-gtr-351>
- McGarigal, K., Wan, H.Y., Zeller, K.A., Timm, B.C., Cushman, S.A., 2016. Multi-scale habitat selection modeling: a review and outlook. *Landscape Ecology* 31, 1161–1175. <https://doi.org/10.1007/s10980-016-0374-x>
- McRae, B.H., 2006. Isolation by resistance. *Evolution* 60, 1551–1561. <https://doi.org/10.1111/j.0014-3820.2006.tb00500.x>
- Meirmans, P.G., Hedrick, P.W., 2011. Assessing population structure: FST and related measures. *Molecular Ecology Resources* 11, 5–18. <https://doi.org/10.1111/j.1755-0998.2010.02927.x>
- Meurant, M., Gonzalez, A., Doxa, A., Albert, C.H., 2018. Selecting surrogate species for connectivity conservation. *Biological Conservation* 227, 326–334. <https://doi.org/10.1016/j.biocon.2018.09.028>
- Miguet, P., Jackson, H.B., Jackson, N.D., Martin, A.E., Fahrig, L., 2016. What determines the spatial extent of landscape effects on species? *Landscape Ecology* 31, 1177–1194. <https://doi.org/10.1007/s10980-015-0314-1>
- Miller, M.P., Bianchi, C.A., Mullins, T.D., Haig, S.M., 2013. Associations between forest fragmentation patterns and genetic structure in Pfrimer's Parakeet (*Pyrrhura pfrimeri*), an endangered endemic to central Brazil's dry forests. *Conservation Genetics* 14, 333–343. <https://doi.org/10.1007/s10592-012-0420-4>
- Millette, K.L., Keyghobadi, N., 2015. The relative influence of habitat amount and configuration on genetic structure across multiple spatial scales. *Ecology and Evolution* 5, 73–86. <https://doi.org/10.1002/ece3.1325>
- Mills, L.S., Allendorf, F.W., 1996. The One-Migrant-per-Generation Rule in Conservation and Management. *Conservation Biology* 10, 1509–1518. <https://doi.org/10.1046/j.1523-1739.1996.10061509.x>
- Mims, M.C., Phillipsen, I.C., Lytle, D.A., Kirk, E.E.H., Olden, J.D., 2015. Ecological strategies predict associations between aquatic and genetic connectivity for dryland amphibians. *Ecology* 96, 1371–1382. <https://doi.org/10.1890/14-0490.1>
- Mitchell, M., 1995. Genetic algorithms: An overview. *Complexity* 1, 31–39. <https://doi.org/10.1002/cplx.6130010108>

- Moilanen, A., 2011. On the limitations of graph-theoretic connectivity in spatial ecology and conservation. *Journal of Applied Ecology* 48, 1543-1547.
- Moraga, A., Martin, A., Fahrig, L., 2019. The scale of effect of landscape context varies with the species' response variable measured. *Landscape Ecology* 34, 703-715. <https://doi.org/10.1007/s10980-019-00808-9>
- Mulvaney, J.M., Matthee, C.A., Cherry, M.I., 2021. Species–landscape interactions drive divergent population trajectories in four forest-dependent Afromontane forest songbird species within a biodiversity hotspot in South Africa. *Evolutionary Applications* 14, 2680–2697. <https://doi.org/10.1111/eva.13306>
- Münzbergová, Z., Cousins, S.A.O., Herben, T., Plačková, I., Mildén, M., Ehrlén, J., 2013. Historical habitat connectivity affects current genetic structure in a grassland species: Genetic diversity and past landscape structure. *Plant Biology* 15, 195–202. <https://doi.org/10.1111/j.1438-8677.2012.00601.x>
- Murphy, M. A., Dezzani, R., Pilliod, D.S., Storfer, A., 2010a. Landscape genetics of high mountain frog metapopulations. *Molecular Ecology* 19, 3634–3649. <https://doi.org/10.1111/j.1365-294X.2010.04723.x>
- Murphy, M.A., Evans, J.S., Cushman, S.A., Storfer, A., 2008. Representing genetic variation as continuous surfaces: an approach for identifying spatial dependency in landscape genetic studies. *Ecography* 31, 685–697. <https://doi.org/10.1111/j.1600-0587.2008.05428.x>
- Murphy, Melanie A., Evans, J.S., Storfer, A., 2010. Quantifying *Bufo boreas* connectivity in Yellowstone National Park with landscape genetics. *Ecology* 91, 252–261. <https://doi.org/10.1890/08-0879.1>
- Naujokaitis-Lewis, I.R., Rico, Y., Lovell, J., Fortin, M.-J., Murphy, M.A., 2013. Implications of incomplete networks on estimation of landscape genetic connectivity. *Conservation Genetics* 14, 287–298. <https://doi.org/10.1007/s10592-012-0385-3>
- Newmark, W.D., Halley, J.M., Beier, P., Cushman, S.A., McNeally, P.B., Soulé, M.E., 2023. Enhanced regional connectivity between western North American national parks will increase persistence of mammal species diversity. *Sci Rep* 13, 474. <https://doi.org/10.1038/s41598-022-26428-z>
- Nowakowski, A.J., DeWoody, J.A., Fagan, M.E., Willoughby, J.R., Donnelly, M.A., 2015. Mechanistic insights into landscape genetic structure of two tropical amphibians using field-derived resistance surfaces. *Molecular Ecology* 24, 580–595. <https://doi.org/10.1111/mec.13052>
- Ohta, T., 1992. The nearly neutral theory of molecular evolution. *Annual review of ecology and systematics* 263–286.
- Ollagnon, H., 2006. La gestion de la biodiversité : quelles stratégies patrimoniales, in : *Annales Des Mines*. pp. 50–57.
- Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J., De Meester, L., 2013. Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology* 22, 5983–5999. <https://doi.org/10.1111/mec.12561>

- Palm, E.C., Landguth, E.L., Holden, Z.A., Day, C.C., Lamb, C.T., Frame, P.F., Morehouse, A.T., Mowat, G., Proctor, M.F., Sawaya, M.A., Stenhouse, G., Whittington, J., Zeller, K.A., 2023. Corridor-based approach with spatial cross-validation reveals scale-dependent effects of geographic distance, human footprint and canopy cover on grizzly bear genetic connectivity. *Molecular Ecology* 32, 5211–5227. <https://doi.org/10.1111/mec.17098>
- Pauls, S.U., Nowak, C., Bálint, M., Pfenninger, M., 2013. The impact of global climate change on genetic diversity within populations and species. *Molecular Ecology* 22, 925–946. <https://doi.org/10.1111/mec.12152>
- Pavlacky Jr, D.C., Goldizen, A.W., Prentis, P.J., Nicholls, J.A., Lowe, A.J., 2009. A landscape genetics approach for quantifying the relative influence of historic and contemporary habitat heterogeneity on the genetic connectivity of a rainforest bird. *Molecular Ecology* 18, 2945–2960. <https://doi.org/10.1111/j.1365-294X.2009.04226.x>
- Pease, K.M., Freedman, A.H., Pollinger, J.P., McCormack, J.E., Buermann, W., Rodzen, J., Banks, J., Meredith, E., Bleich, V.C., Schaefer, R.J., Jones, K., Wayne, R.K., 2009. Landscape genetics of California mule deer (*Odocoileus hemionus*): the roles of ecological and historical factors in generating differentiation. *Molecular Ecology* 18, 1848–1862. <https://doi.org/10.1111/j.1365-294X.2009.04112.x>
- Pelletier, T.A., Carstens, B.C., 2014. Model choice for phylogeographic inference using a large set of models. *Molecular Ecology* 23, 3028–3043. <https://doi.org/10.1111/mec.12722>
- Penjor, U., Kaszta, Z.M., Macdonald, D.W., Cushman, S.A., 2024. Identifying umbrella and indicator species to support multispecies population connectivity in a Himalayan biodiversity hotspot. *Frontiers in Conservation Sciences*. 5, 1306051. <https://doi.org/10.3389/fcosc.2024.1306051>
- Perrin, A., 2021. Fragmentation des habitats et interactions hôtes-parasites (PhD Thesis).
- Perrin, A., Khimoun, A., Faivre, B., Ollivier, A., De Pracontal, N., Théron, F., Loubon, M., Leblond, G., Duron, O., Garnier, S., 2021. Habitat fragmentation differentially shapes neutral and immune gene variation in a tropical bird species. *Heredity* 126, 148–162. <https://doi.org/10.1038/s41437-020-00366-w>
- Perrin, A., Khimoun, A., Ollivier, A., Richard, Y., Pérez-Rodríguez, A., Faivre, B., Garnier, S., 2023. Habitat fragmentation matters more than habitat loss: The case of host–parasite interactions. *Molecular Ecology* 32, 951–969. <https://doi.org/10.1111/mec.16807>
- Peterman, W.E., 2018. ResistanceGA: An R package for the optimization of resistance surfaces using genetic algorithms. *Methods in Ecology and Evolution* 9, 1638–1647. <https://doi.org/10.1111/2041-210X.12984>
- 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, 59–69. <https://doi.org/10.1007/s10592-014-0640-x>
- Peterman, W.E., Pope, N.S., 2021. The use and misuse of regression models in landscape genetic analyses. *Molecular Ecology* 30, 37–47. <https://doi.org/10.1111/mec.15716>
- Peterman, W.E., Winiarski, K.J., Moore, C.E., Carvalho, C. da S., Gilbert, A.L., Spear, S.F., 2019. A comparison of popular approaches to optimize landscape resistance surfaces. *Landscape Ecology* 34, 2197–2208. <https://doi.org/10.1007/s10980-019-00870-3>

- Pflüger, F.J., Balkenhol, N., 2014. A plea for simultaneously considering matrix quality and local environmental conditions when analysing landscape impacts on effective dispersal. *Molecular Ecology* 23, 2146–2156. <https://doi.org/10.1111/mec.12712>
- Pless, E., Saarman, N.P., Powell, J.R., Caccone, A., Amatulli, G., 2021. A machine-learning approach to map landscape connectivity in *Aedes aegypti* with genetic and environmental data. *Proceedings of the national academy of sciences. U.S.A.* 118, e2003201118. <https://doi.org/10.1073/pnas.2003201118>
- Poli, C., Hightower, J., Fletcher, R.J., 2020. Validating network connectivity with observed movement in experimental landscapes undergoing habitat destruction. *Journal of Applied Ecology* 57, 1426–1437. <https://doi.org/10.1111/1365-2664.13624>
- Pritchard, J.K., Seielstad, M.T., Perez-Lezaun, A., Feldman, M.W., 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution* 16, 1791–1798. <https://doi.org/10.1093/oxfordjournals.molbev.a026091>
- Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155, 945–959. <https://doi.org/10.1093/genetics/155.2.945>
- Prunier, J.G., Dubut, V., Chikhi, L., 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, 1866–1877. <https://doi.org/10.1111/2041-210X.12820>
- Prunier, J.G., Kaufmann, B., Fenet, S., Picard, D., Pompanon, F., Joly, P., Lena, J.P., 2013. Optimizing the trade-off between spatial and genetic sampling efforts in patchy populations: towards a better assessment of functional connectivity using an individual-based sampling scheme. *Molecular Ecology* 22, 5516–5530. <https://doi.org/10.1111/mec.12499>
- Radford, J.Q., Amos, N., Harrison, K., Sunnucks, P., Pavlova, A., 2021. Functional connectivity and population persistence in woodland birds: insights for management from a multi-species conservation genetics study. *Emu - Austral Ornithology* 121, 147–159. <https://doi.org/10.1080/01584197.2021.1903331>
- Raufaste, N., Rousset, F., 2001. Are partial mantel tests adequate? *Evolution* 55, 1703–1705. <https://doi.org/10.1111/j.0014-3820.2001.tb00689.x>
- Rayfield, B., Fortin, M.-J., Fall, A., 2011. Connectivity for conservation: a framework to classify network measures. *Ecology* 92, 847–858. <https://doi.org/10.1890/09-2190.1>
- Reyne, M.I., Dicks, K., Flanagan, J., Nolan, P., Twining, J.P., Aubry, A., Emmerson, M., Marnell, F., Helyar, S., Reid, N., 2023. Landscape genetics identifies barriers to Natterjack toad metapopulation dispersal. *Conservation Genetics* 24, 375–390. <https://doi.org/10.1007/s10592-023-01507-4>
- Richardson, J.L., 2012. Divergent landscape effects on population connectivity in two co-occurring amphibian species. *Molecular Ecology* 21, 4437–4451. <https://doi.org/10.1111/j.1365-294X.2012.05708.x>
- Richardson, J.L., Brady, S.P., Wang, I.J., Spear, S.F., 2016. Navigating the pitfalls and promise of landscape genetics. *Molecular Ecology* 25, 849–863. <https://doi.org/10.1111/mec.13527>
- Ricklefs, R., Bermingham, E., 2008. The West Indies as a laboratory of biogeography and evolution. *Philosophical Transactions of the Royal Society B* 363, 2393–2413. <https://doi.org/10.1098/rstb.2007.2068>

- Ricklefs, R.E., 2010. Colonization of the Lesser Antilles by land birds. *Ecology* 91, 1811–1821. <https://doi.org/10.1890/09-0682.1>
- Ricklefs, R.E., Bermingham, E., 2007. The Causes of Evolutionary Radiations in Archipelagoes: Passerine Birds in the Lesser Antilles. *The American Naturalist* 169, 285–297. <https://doi.org/10.1086/510730>
- Ricklefs, R.E., Bermingham, E., 2004. History and the Species-Area Relationship in Lesser Antillean Birds. *The American Naturalist* 163, 227–239. <https://doi.org/10.1086/381002>
- Ricklefs, R.E., Bermingham, E., 2001. Nonequilibrium Diversity Dynamics of the Lesser Antillean Avifauna. *Science* 294, 1522–1524. <https://doi.org/10.1126/science.1065005>
- Ricklefs, R.E., Cox, G.W., 1972. Taxon Cycles in the West Indian Avifauna. *The American Naturalist* 106, 195–219. <https://doi.org/10.1086/282762>
- Rissler, L.J., 2016. Union of phylogeography and landscape genetics. *Proceedings of the national academy of sciences. U.S.A.* 113, 8079–8086. <https://doi.org/10.1073/pnas.1601073113>
- Rousset, F., 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* 8, 103–106. <https://doi.org/10.1111/j.1471-8286.2007.01931.x>
- Rudnick, D., Ryan, S., Beier, P., Cushman, S., Dieffenbach, F., Epps, C., Gerber, L., Hartter, J., Jenness, J., Kintsch, J., Merenlender, A., Perkl, R., Preziosi, D., Trombulak, S., 2012. The Role of Landscape Connectivity in Planning and Implementing Conservation and Restoration Priorities. *Issues in Ecology* 16, 1–23.
- Ruiz-Lopez, M.J., Barelli, C., Rovero, F., Hodges, K., Roos, C., Peterman, W.E., Ting, N., 2016. A novel landscape genetic approach demonstrates the effects of human disturbance on the Udzungwa red colobus monkey (*Procolobus gordonorum*). *Heredity* 116, 167–176. <https://doi.org/10.1038/hdy.2015.82>
- Rundle, H.D., Nosil, P., 2005. Ecological speciation. *Ecology Letters* 8, 336–352. <https://doi.org/10.1111/j.1461-0248.2004.00715.x>
- Sacks, B.N., Brazeal, J.L., Lewis, J.C., 2016. Landscape genetics of the nonnative red fox of California. *Ecology and Evolution* 6, 4775–4791. <https://doi.org/10.1002/ece3.2229>
- Santini, L., Saura, S., Rondinini, C., 2016. A Composite Network Approach for Assessing Multi-Species Connectivity: An Application to Road Defragmentation Prioritisation. *PLoS ONE* 11, e0164794. <https://doi.org/10.1371/journal.pone.0164794>
- Saura, S., 2020. The Habitat Amount Hypothesis implies negative effects of habitat fragmentation on species richness and occurrence (preprint). *Ecology*. <https://doi.org/10.1101/2020.02.02.930784>
- Savary, P., 2021a. Utilisation conjointe de graphes génétiques et paysagers pour l'analyse de la connectivité écologique des habitats (PhD Thesis). Université Bourgogne Franche-Comté.
- Savary, P., Foltête, J., Moal, H., Vuidel, G., Garnier, S., 2023a. Inferring landscape resistance to gene flow when genetic drift is spatially heterogeneous. *Molecular Ecology Resources* 23, 1574–1588. <https://doi.org/10.1111/1755-0998.13821>

- Savary, P., Foltête, J., Moal, H., Vuidel, G., Garnier, S., 2021a. graph4lg: A package for constructing and analysing graphs for landscape genetics in R. *Methods in Ecology and Evolution* 12, 539–547. <https://doi.org/10.1111/2041-210X.13530>
- Savary, P., Foltête, J., Moal, H., Vuidel, G., Garnier, S., 2021b. Analysing landscape effects on dispersal networks and gene flow with genetic graphs. *Molecular Ecology Resources* 21, 1167–1185. <https://doi.org/10.1111/1755-0998.13333>
- Savary, P., Foltête, J.C., Garnier, S., 2022a. Cost distances and least cost paths respond differently to cost scenario variations: a sensitivity analysis of ecological connectivity modeling. *International Journal of Geographical Information Science* 36, 1652–1676. <https://doi.org/10.1080/13658816.2021.2014852>
- Savary, P., Foltête, J.-C., Moal, H., Garnier, S., 2022b. Combining landscape and genetic graphs to address key issues in landscape genetics. *Landscape Ecology* 37, 2293–2309. <https://doi.org/10.1007/s10980-022-01489-7>
- Savary, P., Foltête, J.-C., van Strien, M.J., Moal, H., Vuidel, G., Garnier, S., 2022c. Assessing the influence of the amount of reachable habitat on genetic structure using landscape and genetic graphs. *Heredity* 128, 120–131. <https://doi.org/10.1038/s41437-021-00495-w>
- Savary, P., Lessard, J.-P., Peres-Neto, P.R., 2023b. Heterogeneous dispersal networks to improve biodiversity science. *Trends in Ecology & Evolution* S0169534723002677. <https://doi.org/10.1016/j.tree.2023.10.002>
- 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, 668–678. <https://doi.org/10.1111/j.1365-2664.2011.01970.x>
- Schlägel, U.E., Grimm, V., Blaum, N., Colangeli, P., Dammhahn, M., Eccard, J.A., Hausmann, S.L., Herde, A., Hofer, H., Joshi, J., Kramer-Schadt, S., Litwin, M., Lozada-Gobilard, S.D., Müller, M.E.H., Müller, T., Nathan, R., Petermann, J.S., Pirhofer-Walzl, K., Radchuk, V., Rillig, M.C., Roeleke, M., Schäfer, M., Scherer, C., Schiro, G., Scholz, C., Teckentrup, L., Tiedemann, R., Ullmann, W., Voigt, C.C., Weithoff, G., Jeltsch, F., 2020. Movement-mediated community assembly and coexistence. *Biological Reviews* 95, 1073–1096. <https://doi.org/10.1111/brv.12600>
- Schoen, J.M., Neelakantan, A., Cushman, S.A., Dutta, T., Habib, B., Jhala, Y.V., Mondal, I., Ramakrishnan, U., Reddy, P.A., Saini, S., Sharma, S., Thatte, P., Yumnam, B., DeFries, R., 2022. Synthesizing habitat connectivity analyses of a globally important human-dominated tiger-conservation landscape. *Conservation Biology* 36, e13909. <https://doi.org/10.1111/cobi.13909>
- Schuelke, M., 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* 18, 233–234. <https://doi.org/10.1038/72708>
- Scrucca, L., 2013. GA : A Package for Genetic Algorithms in R. *Journal of Statistical. Software* 53. <https://doi.org/10.18637/jss.v053.i04>
- Segelbacher, G., Cushman, S.A., Epperson, B.K., Fortin, M.-J., Francois, O., Hardy, O.J., Holderegger, R., Taberlet, P., Waits, L.P., Manel, S., 2010. Applications of landscape genetics in conservation biology: concepts and challenges. *Conservation Genetics* 11, 375–385. <https://doi.org/10.1007/s10592-009-0044-5>

- Selmoni, O., Rochat, E., Lecellier, G., Berteaux-Lecellier, V., Joost, S., 2020. Seascape genomics as a new tool to empower coral reef conservation strategies: An example on north-western Pacific *Acropora digitifera*. *Evolutionary Applications* 13, 1923–1938. <https://doi.org/10.1111/eva.12944>
- Senczuk, G., Mastrangelo, S., Ajmone-Marsan, P., Becskei, Z., Colangelo, P., Colli, L., Ferretti, L., Karsli, T., Lancioni, H., Lasagna, E., Marletta, D., Persichilli, C., Portolano, B., Sarti, F.M., Ciani, E., Pilla, F., 2021. On the origin and diversification of Podolian cattle breeds: testing scenarios of European colonization using genome-wide SNP data. *Genetics Selection Evolution* 53, 48. <https://doi.org/10.1186/s12711-021-00639-w>
- Serrouya, R., Paetkau, D., McLellan, B.N., Boutin, S., Campbell, M., Jenkins, D.A., 2012. Population size and major valleys explain microsatellite variation better than taxonomic units for caribou in western Canada. *Molecular Ecology* 21, 2588–2601. <https://doi.org/10.1111/j.1365-294X.2012.05570.x>
- Shah, V., McRae, B., 2008. Circuitscape: A Tool for Landscape Ecology. *Proceedings of the 7th Python in Science Conference* 62–66.
- Shirk, A.J., Landguth, E.L., Cushman, S.A., 2018. A comparison of regression methods for model selection in individual-based landscape genetic analysis. *Molecular Ecology Resources* 18, 55–67. <https://doi.org/10.1111/1755-0998.12709>
- Shirk, A.J., Raphael, M.G., Cushman, S.A., 2014. Spatiotemporal variation in resource selection: insights from the American marten (*Martes americana*). *Ecological Applications* 24, 1434–1444. <https://doi.org/10.1890/13-1510.1>
- Shmueli, G., 2010. To Explain or to Predict? *Statistics Science*. 25. <https://doi.org/10.1214/10-STS330>
- Short Bull, R.A., Cushman, S.A., Mace, R., Chilton, T., Kendall, K.C., Landguth, E.L., Schwartz, M.K., Mckelvey, K., Allendorf, F.W., Luikart, G., 2011. Why replication is important in landscape genetics: American black bear in the Rocky Mountains. *Molecular Ecology* 20, 1092–1107. <https://doi.org/10.1111/j.1365-294X.2010.04944.x>
- Simpkins, C.E., Dennis, T.E., Etherington, T.R., Perry, G.L.W., 2018. Assessing the performance of common landscape connectivity metrics using a virtual ecologist approach. *Ecological Modelling* 367, 13–23. <https://doi.org/10.1016/j.ecolmodel.2017.11.001>
- Slatkin, M., 1993. Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* 47, 264–279. <https://doi.org/10.1111/j.1558-5646.1993.tb01215.x>
- Spear, S.F., Balkenhol, N., Fortin, M.-J., Mcrae, B.H., Scribner, K., 2010. Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. *Molecular Ecology* 19, 3576–3591. <https://doi.org/10.1111/j.1365-294X.2010.04657.x>
- Spear, S.F., Cushman, S.A., McRae, B.H., 2015. Resistance Surface Modeling in Landscape Genetics, in: Balkenhol, N., Cushman, S.A., Storer, A.T., Waits, L.P. (Eds.), *Landscape Genetics*. John Wiley & Sons, Ltd, Chichester, UK, pp. 129–148. <https://doi.org/10.1002/9781118525258.ch08>
- Spielman, D., Brook, B.W., Frankham, R., 2004. Most species are not driven to extinction before genetic factors impact them. *Proceedings of the national academy of sciences U.S.A.* 101, 15261–15264. <https://doi.org/10.1073/pnas.0403809101>



- Storfer, A., Antolin, M.F., Manel, S., Epperson, B.K., Scribner, K.T., 2015. Genomic Approaches in Landscape Genetics, in: Balkenhol, N., Cushman, S.A., Storfer, A.T., Waits, L.P. (Eds.), *Landscape Genetics*. Wiley, pp. 149–164. <https://doi.org/10.1002/9781118525258.ch09>
- Storfer, A., Murphy, M.A., Evans, J.S., Goldberg, C.S., Robinson, S., Spear, S.F., Dezzani, R., Delmelle, E., Vierling, L., Waits, L.P., 2007. Putting the ‘landscape’ in landscape genetics. *Heredity* 98, 128–142. <https://doi.org/10.1038/sj.hdy.6800917>
- Storfer, A., Murphy, M.A., Spear, S.F., Holderegger, R., Waits, L.P., 2010. Landscape genetics: where are we now? *Molecular Ecology* 19, 3496–3514. <https://doi.org/10.1111/j.1365-294X.2010.04691.x>
- Stutchbury, B.J., Morton, E.S., 2022. *Behavioral ecology of tropical birds*. Academic press.
- Takeuchi, T., Masaoka, T., Aoki, H., Koyanagi, R., Fujie, M., Satoh, N., 2020. Divergent northern and southern populations and demographic history of the pearl oyster in the western Pacific revealed with genomic SNPs. *Evolutionary Applications* 13, 837–853. <https://doi.org/10.1111/eva.12905>
- Tappeiner, U., Leitinger, G., Zariņa, A., Bürgi, M., 2021. How to consider history in landscape ecology: patterns, processes, and pathways. *Landscape Ecology* 36, 2317–2328. <https://doi.org/10.1007/s10980-020-01163-w>
- Tavaré, S., Balding, D.J., Griffiths, R.C., Donnelly, P., 1997. Inferring coalescence times from DNA Sequence Data. *Genetics* 145, 505–518. <https://doi.org/10.1093/genetics/145.2.505>
- Taylor, P.D., Fahrig, L., Henein, K., Merriam, G., 1993. Connectivity is a vital element of landscape structure. *Oikos* 68, 571–573.
- Teixeira, J.C., Huber, C.D., 2021. The inflated significance of neutral genetic diversity in conservation genetics. *Proceedings of the national academy of sciences*. U.S.A. 118, e2015096118. <https://doi.org/10.1073/pnas.2015096118>
- Templeton, A.R., Shaw, K., Routman, E., Davis, S.K., 1990. The Genetic Consequences of Habitat Fragmentation. *Annals of the Missouri Botanical Garden* 77, 13-27. <https://doi.org/10.2307/2399621>
- Thompson, C.M., McGarigal, K., 2002. The influence of research scale on bald eagle habitat selection along the lower Hudson River, New York (USA). *Landscape Ecology* 17, 569–586. <https://doi.org/10.1023/A:1021501231182>
- Turner, M.G., 1989. Landscape ecology: the effect of pattern on process. *Annual review of ecology and systematics* 20, 171–197.
- Union européenne – SOeS, CORINE Land Cover [WWW Document], 2012. URL <https://www.statistiques.developpement-durable.gouv.fr/corine-land-cover-0> (accessed 2.20.24).
- Urban, D., Keitt, T., 2001. Landscape connectivity: a graph-theoretic perspective 82, 1205-1218. [https://doi.org/10.1890/0012-9658\(2001\)082\[1205:LCAGTP\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2001)082[1205:LCAGTP]2.0.CO;2)
- Van Strien, M.J., Holderegger, R., Van Heck, H.J., 2015. Isolation-by-distance in landscapes: considerations for landscape genetics. *Heredity* 114, 27–37. <https://doi.org/10.1038/hdy.2014.62>

- Van Strien, M.J., Keller, D., Holderegger, R., 2012. A new analytical approach to landscape genetic modelling: least-cost transect analysis and linear mixed models. *Molecular Ecology* 21, 4010–4023. <https://doi.org/10.1111/j.1365-294X.2012.05687.x>
- Van Strien, M.J., Keller, D., Holderegger, R., Ghazoul, J., Kienast, F., Bolliger, J., 2014. Landscape genetics as a tool for conservation planning: predicting the effects of landscape change on gene flow. *Ecological Applications* 24, 327–339. <https://doi.org/10.1890/13-0442.1>
- Van Strien, M.J., 2017. Consequences of population topology for studying gene flow using link-based landscape genetic methods. *Ecology and Evolution* 7, 5070–5081. <https://doi.org/10.1002/ece3.3075>
- Vandergast, A.G., Bohonak, A.J., Weissman, D.B., Fisher, R.N., 2007. Understanding the genetic effects of recent habitat fragmentation in the context of evolutionary history: phylogeography and landscape genetics of a southern California endemic Jerusalem cricket (Orthoptera: Stenopelmatidae: *Stenopelmatus*). *Molecular Ecology* 16, 977–992. <https://doi.org/10.1111/j.1365-294X.2006.03216.x>
- Vanhove, M., Launey, S., 2023. Estimating resistance surfaces using gradient forest and allelic frequencies. *Molecular Ecology Resources* 1755–0998.13778. <https://doi.org/10.1111/1755-0998.13778>
- Varvio, S.-L., Chakraborty, R., Nei, M., 1986. Genetic variation in subdivided populations and conservation genetics. *Heredity* 57, 189–198. <https://doi.org/10.1038/hdy.1986.109>
- Vellend, M., Geber, M.A., 2005. Connections between species diversity and genetic diversity. *Ecology Letters* 8, 767–781. <https://doi.org/10.1111/j.1461-0248.2005.00775.x>
- Velo-Antón, G., Parra, J.L., Parra-Olea, G., Zamudio, K.R., 2013. Tracking climate change in a dispersal-limited species: reduced spatial and genetic connectivity in a montane salamander. *Molecular Ecology* 22, 3261–3278. <https://doi.org/10.1111/mec.12310>
- Vergara, M., Cushman, S.A., Ruiz-González, A., 2017. Ecological differences and limiting factors in different regional contexts: landscape genetics of the stone marten in the Iberian Peninsula. *Landscape Ecology* 32, 1269–1283. <https://doi.org/10.1007/s10980-017-0512-0>
- Voeten, C.C., 2020. buildmer: Stepwise elimination and term reordering for mixed-effects regression. R package version 1.
- Wagner, H.H., Fortin, M.-J., 2013. A conceptual framework for the spatial analysis of landscape genetic data. *Conservation Genetics* 14, 253–261. <https://doi.org/10.1007/s10592-012-0391-5>
- Wagner, H.H., Holderegger, R., Werth, S., Gugerli, F., Hoesbee, S.E., Scheidegger, C., 2005. Variogram Analysis of the Spatial Genetic Structure of Continuous Populations Using Multilocus Microsatellite Data. *Genetics* 169, 1739–1752. <https://doi.org/10.1534/genetics.104.036038>
- Waits, L.P., Storfer, A., 2015. Basics of population genetics: quantifying neutral and adaptive genetic variation for landscape genetic studies. *Landscape genetics* 35–57. <https://doi.org/10.1002/9781118525258.ch03>
- Wang, I.J., 2010. Recognizing the temporal distinctions between landscape genetics and phylogeography. *Molecular Ecology* 19, 2605–2608. <https://doi.org/10.1111/j.1365-294X.2010.04715.x>

- Wasserman, T.N., Cushman, S.A., Shirk, A.S., Landguth, E.L., Littell, J.S., 2012. Simulating the effects of climate change on population connectivity of American marten (*Martes americana*) in the northern Rocky Mountains, USA. *Landscape Ecology* 27, 211–225. <https://doi.org/10.1007/s10980-011-9653-8>
- Watts, A.G., Schlichting, P.E., Billerman, S.M., Jesmer, B.R., Micheletti, S., Fortin, M.-J., Funk, W.C., Hapeman, P., Muths, E., Murphy, M.A., 2015. How spatio-temporal habitat connectivity affects amphibian genetic structure. *Frontiers in Genetics*. 6. <https://doi.org/10.3389/fgene.2015.00275>
- Weckworth, B.V., Musiani, M., DeCesare, N.J., McDevitt, A.D., Hebblewhite, M., Mariani, S., 2013. Preferred habitat and effective population size drive landscape genetic patterns in an endangered species. *Proceedings of the Royal Society B: Biological Sciences* 280, 20131756. <https://doi.org/10.1098/rspb.2013.1756>
- Weir, B.S., Cockerham, C.C., 1984. Estimating F-Statistics for the Analysis of Population Structure. *Evolution* 38, 1358–1370. <https://doi.org/10.2307/2408641>
- Weiss, G., Von Haeseler, A., 1998. Inference of Population History Using a Likelihood Approach. *Genetics* 149, 1539–1546. <https://doi.org/10.1093/genetics/149.3.1539>
- Winiarski, K.J., Peterman, W.E., McGarigal, K., 2020. Evaluation of the R package ‘RESISTANCEGA’: A promising approach towards the accurate optimization of landscape resistance surfaces. *Molecular Ecology Resources* 20, 1583–1596. <https://doi.org/10.1111/1755-0998.13217>
- Wright, S., 1931. Evolution in mendelian populations. *Genetics* 16, 97–159. <https://doi.org/10.1093/genetics/16.2.97>
- Yates, K.L., Bouchet, P.J., Caley, M.J., Mengersen, K., Randin, C.F., Parnell, S., Fielding, A.H., Bamford, A.J., Ban, S., Barbosa, A.M., Dormann, C.F., Elith, J., Embling, C.B., Ervin, G.N., Fisher, R., Gould, S., Graf, R.F., Gregr, E.J., Halpin, P.N., Heikkinen, R.K., Heinänen, S., Jones, A.R., Krishnakumar, P.K., Lauria, V., Lozano-Montes, H., Mannocci, L., Mellin, C., Mesgaran, M.B., Moreno-Amat, E., Mormede, S., Novaczek, E., Opper, S., Ortuño Crespo, G., Peterson, A.T., Rapacciuolo, G., Roberts, J.J., Ross, R.E., Scales, K.L., Schoeman, D., Snelgrove, P., Sundblad, G., Thuiller, W., Torres, L.G., Verbruggen, H., Wang, L., Wenger, S., Whittingham, M.J., Zharikov, Y., Zurell, D., Sequeira, A.M.M., 2018. Outstanding Challenges in the Transferability of Ecological Models. *Trends in Ecology & Evolution* 33, 790–802. <https://doi.org/10.1016/j.tree.2018.08.001>
- Zeller, K., Lewison, R., Fletcher, R., Tulbure, M., Jennings, M., 2020. Understanding the Importance of Dynamic Landscape Connectivity. *Land* 9, 303. <https://doi.org/10.3390/land9090303>
- 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, 868–879. <https://doi.org/10.1111/ddi.12742>
- Zeller, K.A., McGarigal, K., Whiteley, A.R., 2012. Estimating landscape resistance to movement: a review. *Landscape Ecology* 27, 777–797. <https://doi.org/10.1007/s10980-012-9737-0>
- Zeller, K.A., Vickers, T.W., Ernest, H.B., Boyce, W.M., 2017. Multi-level, multi-scale resource selection functions and resistance surfaces for conservation planning: Pumas as a case study. *PLoS One* 12, e0179570. <https://doi.org/10.1371/journal.pone.0179570>

Zeller, K.A., Wulsch, C., Welfelt, L.S., Beausoleil, R.A., Landguth, E.L., 2023. Accounting for sex-specific differences in gene flow and functional connectivity for cougars and implications for management. *Landscape Ecology* 38, 223–237. <https://doi.org/10.1007/s10980-022-01556-z>

Zellmer, A.J., Knowles, L.L., 2009. Disentangling the effects of historic vs. contemporary landscape structure on population genetic divergence. *Molecular Ecology* 18, 3593–3602. <https://doi.org/10.1111/j.1365-294X.2009.04305.x>

Zero, V.H., Barocas, A., Jochimsen, D.M., Pelletier, A., Giroux-Bougard, X., Trumbo, D.R., Castillo, J.A., Evans Mack, D., Linnell, M.A., Pigg, R.M., Hoisington-Lopez, J., Spear, S.F., Murphy, M.A., Waits, L.P., 2017. Complementary Network-Based Approaches for Exploring Genetic Structure and Functional Connectivity in Two Vulnerable, Endemic Ground Squirrels. *Frontiers in Genetics* 8, 81. <https://doi.org/10.3389/fgene.2017.00081>



# ANNEXES



<b>Annexe 1</b> : Analyses moléculaires	<b>Page 270</b>
<b>Annexe 2</b> : Supplemental Information for: “ <i>What can optimized cost distances based on genetic distances offer? A simulation study on the use and misuse of ResistanceGA</i> ”	<b>Page 275</b>
<b>Annexe 3</b> : Supplemental Information for: “ <i>Is landscape effect on gene flow shared across species? A multispecies cross-validation landscape genetic study</i> ”	<b>Page 291</b>
<b>Annexe 4</b> : Supplemental Information for: “ <i>Disentangling past landscape legacies from the contemporary genetic structure of a forest bird using a historical map</i> ”	<b>Page 299</b>
<b>Annexe 5</b> : Rapport « Etude de la connectivité des habitats forestiers entre la Grande-Terre et la Basse-Terre »	<b>Page 305</b>

## **Annexe 1 : Analyses moléculaires**

## S2.1 : Loci microsatellites et conditions d'amplifications.

Espèce	Fluorochrome	Locus	Référence	T° Hybridation	[MgCl2]
SR	FAM	TG1-147	Dawson et al. 2010	51	2
	VIC	CAM-10	Dawson et al. 2013	54	2
	VIC	CAM-02	Dawson et al. 2013	51	1,5
	NED	CAM-15	Dawson et al. 2013	51	2,5
	PET	CAM-05	Dawson et al. 2013	51	2
	FAM	CAM-17	Dawson et al. 2013	51	1,5
	FAM	TG7-22	Dawson et al. 2010	51	2
	VIC	TG13-17	Dawson et al. 2010	51	1,5
	NED	CAM-18	Dawson et al. 2013	54	1,5
	PET	CAM-19	Dawson et al. 2013	51	2
MG	VIC	TG4-4	Dawson et al. 2010	2,5	51
	PET	CAM-10	Dawson et al. 2013	2	51
	FAM	TG3-98	Dawson et al. 2010	2,5	51
	NED	CAM-18	Dawson et al. 2013	2	54
	PET	TG12-15	Dawson et al. 2010	2,5	54
	FAM	TG22-1	Dawson et al. 2010	2,5	51
	VIC	TG4-12	Dawson et al. 2010	2,5	51
	NED	CAM-20	Dawson et al. 2013	2	51
	VIC	TG11-11	Dawson et al. 2010	2,5	51
	FAM	TG1-86*	Dawson et al. 2010	2,5	54
GJ	FAM	Ase40	Richardson et al. 2000	60	1,5
	VIC	Ase64	Richardson et al. 2000	52	2
	NED	Asu15	Bulgin et al. 2003	49	2
	PET	Cuu5	Gibbs et al. 1999	48	2,5
	PET	Cuu28	Gibbs et al. 1999	48	2
	VIC	DkiB119	Kings et al. 2005	50	1,5
	PET	PatMP2-43	Otter et al. 1998	52	1,5
	VIC	Tgu02	Replogle et al. 2008 ; Slate et al. 2000	50	2
	PET	Tgu07	Replogle et al. 2008 ; Slate et al. 2000	50	2
	NED	TG4-12	Replogle et al. 2008	49	2
PC	NED	CAM19	Dawson et al. 2013	51	1,5
	FAM	DkiD124	King et al. 2005	51	1,5
	PET	DkiD102	King et al. 2005	54	1,5
	VIC	DkiD104	King et al. 2005	54	1,5
	FAM	DkiD128	King et al. 2005	51	1,5
	PET	DkiC118	King et al. 2005	57	1,5
	NED	Lswμ7	Winker et al. 1999	54	1,5
	VIC	DkiB12	King et al. 2005	54	1,5
	NED	TG12-15	Dawson et al. 2010	54	1,5
	PET	DkiD12	King et al. 2005	54	1,5



	VIC	DkiD109	King et al. 2005	51	1,5
	FAM	Lswμ5	Winker et al. 1999	54	1,5
SB	FAM	TG1_114	Ferreira et al., 2015	51	2
	VIC	PD02	Ferreira et al., 2015	54	2
	NED	CB03	Ferreira et al., 2015	54	1,5
	PET	SF03	Ferreira et al., 2015	54	1,5
	PET	CB04	Ferreira et al., 2015	54	1,5
	VIC	TG1_147	Dawson et al. 2010	2,5	51
	NED	CAM18	Dawson et al. 2013	2	57
	FAM	CAM01	Dawson et al. 2013	2	54
	PET	TG11_11	Dawson et al. 2010	2	51
	NED	CAM19	Dawson et al. 2013	2	51
	VIC	CAM_05	Dawson et al. 2013	2	51

[MgCl<sub>2</sub>] : concentration en MgCl<sub>2</sub> optimisée en mM

## Références :

- Bulgin NL, Gibbs HL, Vickery P et al. (2003) Ancestral polymorphisms in genetic markers obscure detection of evolutionarily distinct populations in the endangered Florida grasshopper sparrow (*Ammodramus sevensanarum floridanus*). *Molecular Ecology*, 12, 831-844
- Dawson. D.A., Ball. A.D., Spurgin. L.G., Martín-Gálvez. D., Stewart. I.R., Horsburgh. G.J., Potter. J., Molina-Morales. M., Bicknell. A.W., Preston. S.A. & others. (2013). High-utility conserved avian microsatellite markers enable parentage and population studies across a wide range of species. *BMC genomics*. **14**. 176.
- Dawson. D.A., Horsburgh. G.J., Küpper. C., Stewart. I.R.K., Ball. A.D., Durrant. K.L., Hansson. B., Bacon. I., Bird. S., Klein. áKos. Krupa. A.P., Lee. J.-W., Martín-Gálvez. D., Simeoni. M., Smith. G., Spurgin. L.G. & Burke. T. (2010). New methods to identify conserved microsatellite loci and develop primer sets of high cross-species utility - as demonstrated for birds. *Molecular Ecology Resources*. **10**. 475–494.
- Ferreira, J.M., Burnham-Curtis, M., Schunck, F., Rocha, M., Miyaki, C.Y., Silveira, L.F., Melo, M., Morgante, J.S., 2015. Development of novel polymorphic microsatellite markers for four bird species exploited by the illegal wildlife trade in Brazil. *Conservation Genet Resour* 7, 435–436. <https://doi.org/10.1007/s12686-014-0389-z>
- Gibbs HL, Tabak LM, Hobson K (1999). Characterization of microsatellite DNA loci for a neotropical migrant songbird, the Swainson's thrush (*Catharus ustulatus*). *Molecular Ecology*, 8, 1551-1561.
- King. T.L., Eackles. M.S., Henderson. A.P., Bocetti. C.I., Currie. D. & Wunderle. J.M. (2005). Microsatellite DNA markers for delineating population structure and kinship among the endangered Kirtland's warbler (*Dendroica kirtlandii*). *Molecular Ecology Notes*. **5**. 569–571.
- Otter K, Ratcliffe L, Michaud D et al. (1998) Do female black-capped chickadees prefer high-ranking males as extra-pair partners? *Behavioural Ecology and Sociobiology*, 43, 25-36

- Replogle K, Arnold AP, Ball GF et al. (2008) The songbird neurogenomics (SoNG) initiative: community-based tools and strategies for study of brain gene function and evolution. *BMC genomics*, 9, 131.
- Richardson DS, Jury FL, Dawson DA et al. (2000) Isolation and characterisation of 50 microsatellite loci from the Seychelles warbler (*Acrocephalus sechellensis*), and their cross-species amplification within the Sylviidae family and a range of passerine birds. *Molecular Ecology*, 9, 2226-2231
- Slate J, Hale MC, Birkhead TR (2007) Simple sequence repeats in zebra finch (*Taenopygia guttata*) expressed sequence tags: a new resource for evolutionary genetic studies of passerines. *BMC genomics*, 8, 52
- Winker. K., Glenn. T.C. & Graves. G.R. (1999). Dinucleotide microsatellite loci in a migratory wood warbler (*Parulidae: Limnothlypis swainsonii*) and amplification among other songbirds. *Molecular Ecology*. **8**. 1553–1556.



## Annexe 2 :

### Supplemental Information for

#### What can optimized cost distances based on genetic distances offer? A simulation study on the use and misuse of ResistanceGA

Daniel, Alexandrine<sup>1</sup>, Savary, Paul<sup>2</sup>, Foltête, Jean-Christophe<sup>3</sup>, Vuidel, Gilles<sup>3</sup>, Faivre, Bruno<sup>1</sup>, Garnier, Stéphane<sup>1,4</sup>, Khimoun, Aurélie<sup>1,4</sup>

S3.1.1. Real landscape sampling	<b>Page 276</b>
S3.1.2. Construction of the true cost scenarios	<b>Page 276</b>
S3.1.3. Parameters for demo-genetic simulations	<b>Page 277</b>
S3.1.4. Building the effective-dispersal scale dataset	<b>Page 278</b>
S3.1.5. ResistanceGA workflow	<b>Page 279</b>
S3.1.6. Landscape sampling results	<b>Page 280</b>
S3.1.7. Distribution of simulated genetic patterns according to specialization profile and dispersal capacity	<b>Page 283</b>
S3.1.8. Spurious patterns of Isolation by Distance for low dispersal capacity	<b>Page 284</b>
S3.1.9. Confusion matrices assessing the optimization sensitivity and specificity to the simulated genetic patterns	<b>Page 285</b>
S3.1.10. Significance of Spearman correlation values for land cover ranking in optimized cost scenarios	<b>Page 287</b>

### S3.1.1. Real landscape sampling

We built a categorical raster surface (resolution: 100m) covering the whole of metropolitan France, using the 2012 CORINE Land Cover data (“Union Européenne – SOeS, CORINE Land Cover” 2012). We reclassified the initial land cover types into four categories: forest, grassland, agricultural, and urban areas. We then randomly subsampled 1000 landscapes of 40 by 40 km and we kept those having the four considered land cover types in significant amounts (over 5% grassland and agricultural areas, over 2% urban areas and over 15% forested areas).

We characterized the composition and structure of each sampled landscape with FRAGSTAT metrics (McGarigal and Marks 1995). We computed two class metrics, *Class Amount* and *Clumpy*, respectively related to the amount and aggregation of each of the four land cover types. We also computed a global *Aggregation Index* assessing the aggregation of land cover summarized over each type for the whole landscape. We then performed a Principal Component Analysis (PCA) of these nine landscape metrics and extracted the first two principal component (PC) axes. We finally ranked the 1000 landscapes according to their coordinates on PC axes and we selected 15 landscapes reflecting the range of variation on each of the two PC axes, resulting in a total of 30 landscapes.

### S3.1.2. Construction of true cost scenarios

We based the construction of the true cost scenarios on the negative exponential function introduced by Keeley et al. (Keeley, Beier, and Gagnon 2016) :

$$R = Cost_{max} - (Cost_{max} - 1) \times \frac{1 - e^{c \times H}}{1 - e^{-c}}$$

Where R is the landscape resistance,  $Cost_{max}$  is the maximum cost value, H is the suitability of the landscape feature for the species, and c determines the shape of the curve. We arbitrarily associated each land cover type to a suitability score to model a forest species profile (0% for the urban areas, 33% for the agricultural areas, 66% for the grassland areas, and 100% for the

forest areas) and set the maximum cost at 1000. We then explored the shape of the relationship for 16 values of  $c$  ranging from -32 to 32. We finally selected two values of  $c$  (4 and -4) leading to contrasted scenarios reflecting two ecological profiles. The first one ( $c=-4$ ) represented the ecology of a species specialist of forest habitat, with a majority of very resistant land cover types except for the forest:  $\{Cost_{forest}=1; Cost_{grassland}= 700; Cost_{agricultural} =900; Cost_{urban}=1000\}$ . The second one ( $c = 4$ ) was fitted for a species much less specialist of forest habitat, with a more continuous range of resistance values and a strong resistance for urban areas only:  $\{Cost_{forest}=1; Cost_{grassland}= 50; Cost_{agricultural} =200; Cost_{urban}=1000\}$ .

### **S3.1.3. Parameters for demo-genetic simulations**

The individual genotypes were simulated for 25 loci with 25 alleles per locus, as a high allelic richness is known to limit the risk of size homoplasy (Estoup, Jarne, and Cornuet 2002). We assumed no selection pressure but mutations could occur, according to a simple Kmax-allele model with a mutation rate  $\mu = 5e-5$ . We first initialize the individual genotypes for one of the 80 populations, by assigning each individual two alleles randomly chosen among the 25 alleles for the 25 loci. The remaining 79 populations were then initialized as clones of this population.

At each generation, populations exchanged individuals according to a constant migration rate (set at 0.125) and a dispersal probability between each population pair. This probability depended on the dispersal capacities of individuals and the cost distance between these populations. These cost distances were calculated as described in Material & Methods – section II for each landscape, sampling design and specialization level. Therefore, the dispersal probability  $p_{ij}$  between population  $i$  and  $j$  was the negative exponential of the cost distance  $CD_{ij}$ , such that  $p_{ij} = e^{-\alpha \cdot CD_{ij}}$ , following Urban and Keitt (2001). We chose seven values of  $\alpha$  modeling seven dispersal capacities, characterized by a 5% probability of dispersing beyond 1,000, 5,000, 10,000, 15,000, 20,000, 30,000 and 40,000 cost units respectively. This dispersal step, conditioning the simulated genetic structure, is thus the reflection of the implemented true cost scenarios described in Material & Methods – section II.

Reproduction occurred after the dispersal stage. The number of offsprings per female followed a Poisson distribution of parameter  $\lambda=3$ . The sex ratio remained constant at 1. Offspring genotypes were constructed by random association of a male and a female allele for each locus. All adults older than three generations died, but mortality could also affect younger adults and juveniles randomly to keep a population size of 25 individuals.

We performed a total of 4,200 simulations (30 landscapes x 2 cost scenarios x 10 distributions of populations x 7 dispersal capacities) conducted over 200 generations to obtain a steady pattern of genetic differentiation.

#### **S3.1.4. Building the effective-dispersal scale dataset**

The effective-dispersal scale dataset was composed of both pruned and unpruned statistical individuals. Indeed, any combination of landscape x specialization profile x dispersal capacity x sampling design that was likely to exhibit a dispersal scale smaller than the sampling spatial scale (i.e., encompassing populations that do not exchange migrants) was pruned, but the others were integrated as such into the effective-dispersal dataset (i.e. considering all population pairs expected to exchange migrants). This means that some statistical individuals were redundant between the sampling scale and the effective-dispersal scale datasets.

To determine if the dispersal scale was smaller than the study scale and if the set of distances required pruning, we performed some preliminary explorations of IBD patterns. They revealed flat trends with low differentiation values for values of dispersal capacity above 1,000 cost units (cu) and 15,000 cu for generalist and specialist profiles respectively, reflecting a situation of high exchange not requiring pruning. Therefore, only parameter combinations with dispersal capacity of less than 5,000 cu for generalist species and 20,000 cu for specialist species were suspected to show a dispersal scale smaller than the study scale. For each of these combinations, we pruned population pairs according to the conditional independence method (Dyer and Nason 2004). We then looked at how many populations were isolated or poorly connected as a result. To ensure the operation of the RGA optimization algorithm, each population must be connected to at least 3 others. Therefore, a population was considered as “too poorly connected” if it was linked by dispersal events to fewer than 3 populations.

The conditional independence pruning is known for retaining very few links (Savary et al. 2021). In particular, it removes indirect links between populations, but these are still informative for our purposes, as populations indirectly linked through multiple dispersal events are still related by indirect gene flow. To avoid losing too much information by pruning too severely, we follow a decision tree to select the statistical individuals to prune. Pruning is necessary when some populations or groups of populations are completely disconnected from others, so genetic differentiation relies only on drift. On the other hand, if all populations are connected, even by indirect links, we could assume that gene flow circulates through the whole network and therefore pruning is not necessary. Thus, we set an arbitrary threshold for the number of population disconnections required to assume that gene flow is not spreading throughout the network, and so prune the set of distances. More than 20% of sampled populations have to be considered as “too poorly connected” to prune the related set of distances. If this was not the case, the set of distances was integrated unchanged into the effective-dispersal scale data set.

It should be noted that these pruning criteria may seem onerous here, but this is due to the large number of parameter combinations to be dealt with in this study (i.e., landscape x ecological profile x dispersal capacity x sampling design x spatial scale). For an empirical study, it would simply be a matter of testing performance with and without pruning and choosing the most satisfactory option.

### **S3.1.5. ResistanceGA workflow**

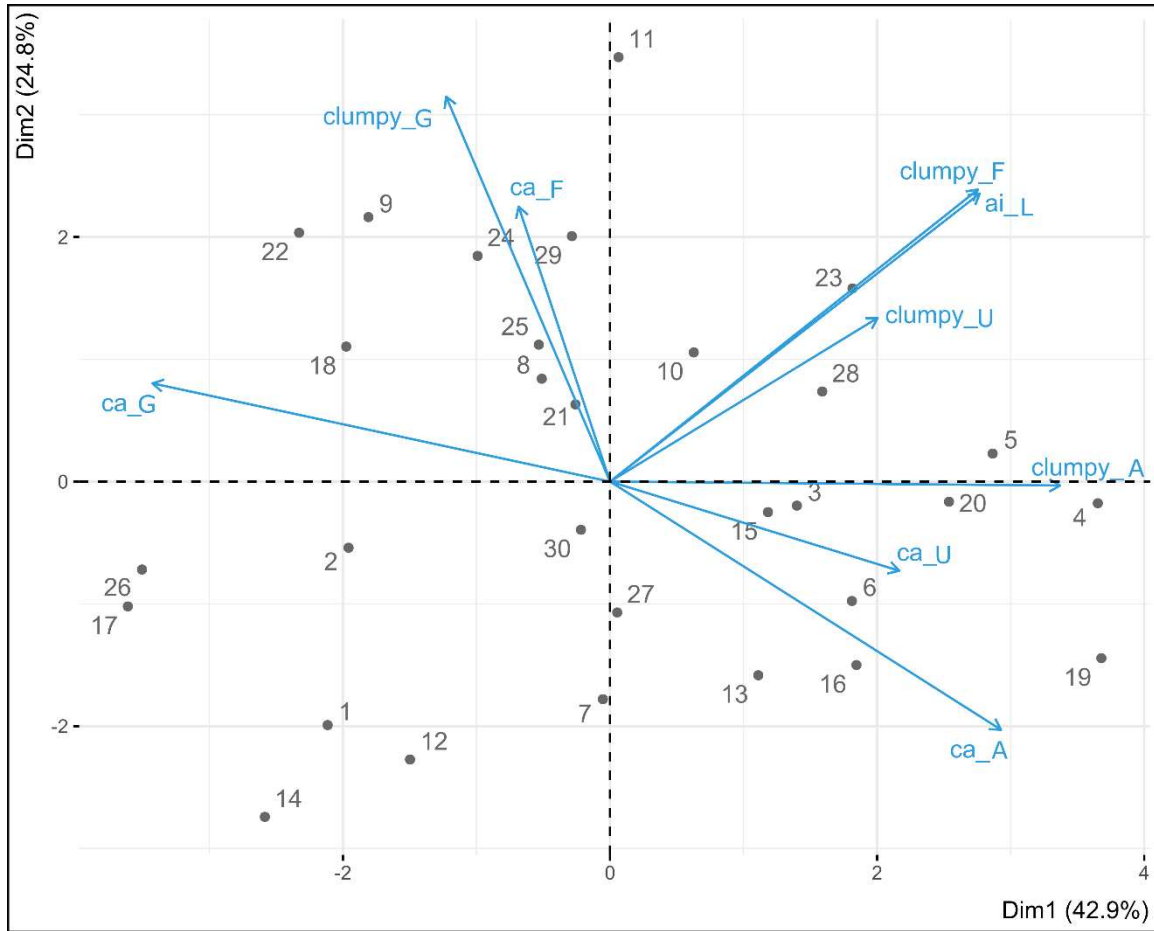
ResistanceGA is an objective framework for resistance surface optimization, based on an algorithm that maximizes the relationship between genetic and cost distances (Peterman 2018). The algorithm randomly generates a set of resistance surfaces associated with a set of parameters, corresponding to the costs related to categorical or continuous landscape features. On each resistance surface and according to these cost parameters, least-cost distances or resistance distances are computed between all pairs of sampled populations. These pairwise ecological distances are then confronted with genetic distances using MLPE models accounting for non-independence among the pairwise data. If the cost parameters are optimal, then the ecological distances would meet the genetic distances and the models would be strongly fitted. Thus, only the best models according to an objective criterion (such as the AIC) are kept and



their related resistance surfaces are selected to constitute the next generation. Then recombination events and mutations occur in the settings to ensure the mixing of parameters and the creation of the ‘offspring’ surfaces. All these steps are then repeated until there is no further improvement in MLPE models. At last, the final model with the optimized cost parameters is ranked according to the AIC criterion against the Euclidean model (in which genetic distances are explained by Euclidean distances), and against the null model (in which genetic distances are explained by a constant, assuming that the pairwise genetic distances are related neither to cost distance nor Euclidian distances).

### **S3.1.6. Landscape sampling results**

A total of 124 landscapes over 1000 remained after filtering according to land-use proportions. The first two axes of the PCA of the nine structural metrics accounted for 61% of the variance between these landscapes (S6. A). In the 30 landscapes sampled along the gradient of the PCA axes, urban areas were poorly represented (5% of the landscape area on average), while other land covers were balanced at around 30% on average. The degree of aggregation of forest areas was slightly less variable (coefficient of variation of 13%) than that of grassland, agricultural and urban areas (coefficient of variation of 24%, 21% and 10% respectively) (S6. B). The proportion of forest and agricultural areas was inversely proportional, meaning that landscapes were either predominantly composed of forest and grassland, or predominantly agricultural. The aggregation of agricultural and forest areas was correlated (Pearson’s coefficient  $r = 0.63^{***}$ ), regardless of their proportion.

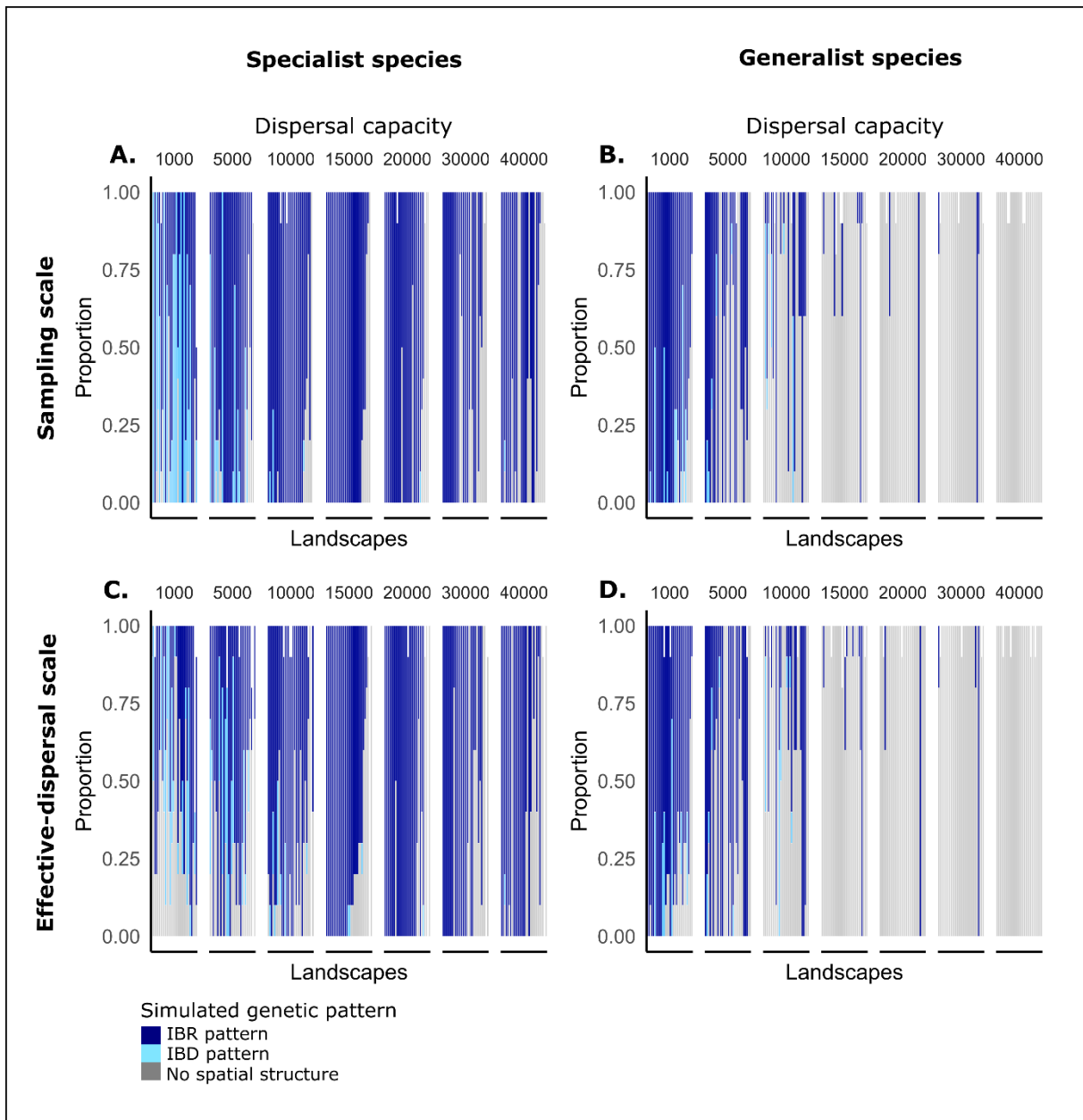


**S3.6. A:** Results of the Principal Component Analysis of the structure (composition and configuration) of the sampled landscapes. Projection of individuals and variables onto the first two components, accounting for 67.7% of the total variance. The landscapes sampled are numbered from 1 to 30. Clumpy\_F: forest aggregation; Clumpy\_G: grassland aggregation; Clumpy\_A: agricultural aggregation; Clumpy\_U: urban aggregation, ai\_L: landscape aggregation; Ca\_F: forest quantity; Ca\_G: grassland quantity; Ca\_A: agricultural quantity; Ca\_U: urban quantity.

<b>Landscape variables</b>	<b>Mean</b>	<b>SD</b>	<b>Min</b>	<b>Max</b>
<b>clumpy_F</b>	0,72	0,09	0,51	0,87
<b>ca_F (ha)</b>	45551	17961	24271	97575
<b>clumpy_G</b>	0,41	0,1	0,27	0,6
<b>ca_G (ha)</b>	47495	23921	14488	107728
<b>clumpy_A</b>	0,57	0,12	0,31	0,74
<b>ca_A (ha)</b>	58544	26721	15517	100194
<b>clumpy_U</b>	0,5	0,1	0,33	0,74
<b>ca_U</b>	8412	7397	3228	43030
<b>ai_L</b>	73,6	5,3	63,4	81
<b>Forest (%)</b>	28	11	15	57
<b>Grassland (%)</b>	30	15	9	67
<b>Agricultural (%)</b>	37	19	10	63
<b>Urban (%)</b>	5	5	2	27

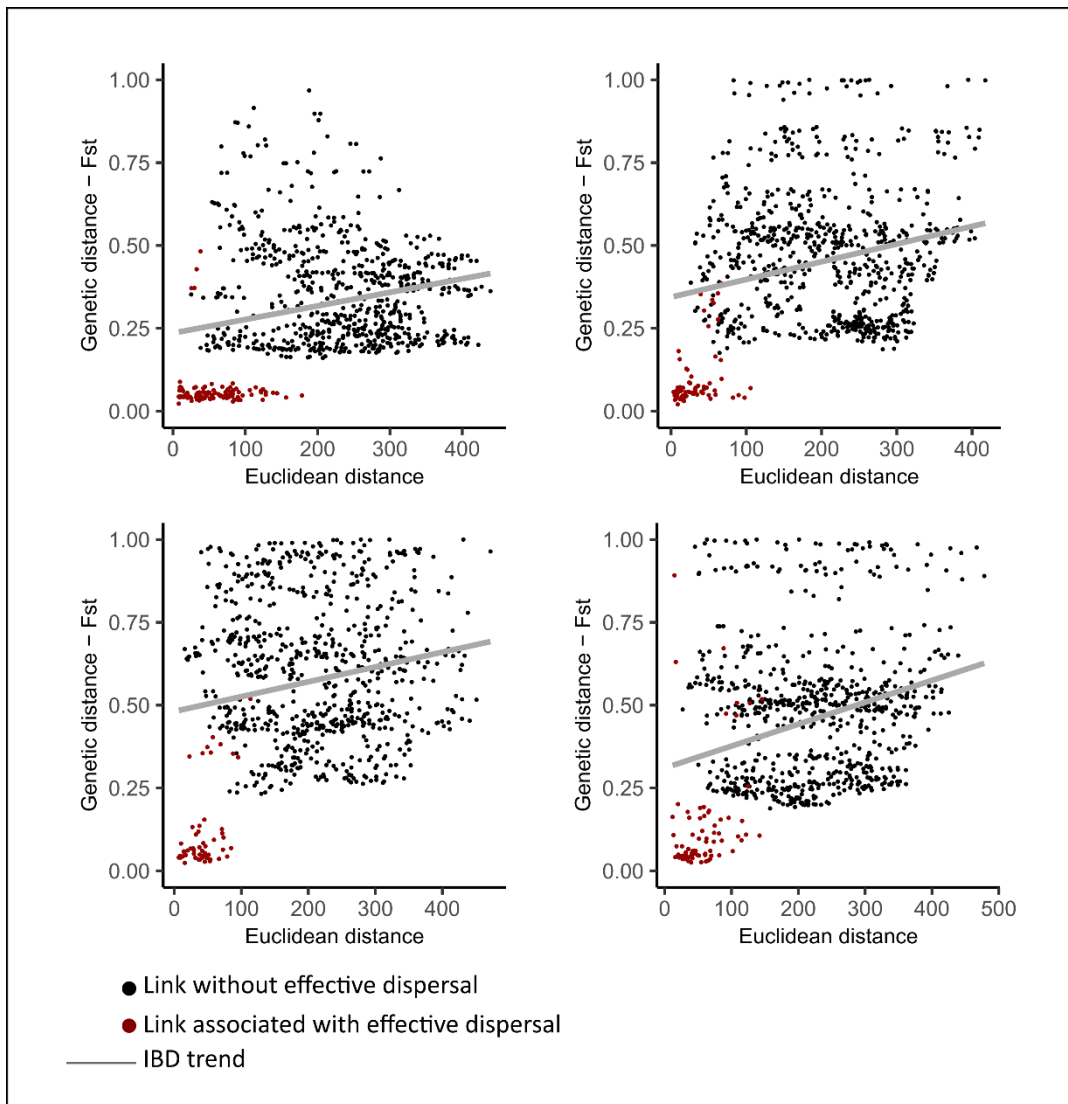
**S3.6. B:** Composition and configuration metrics of the sampled landscapes. Clumpy\_F: forest aggregation; Clumpy\_G: grassland aggregation; Clumpy\_A: agricultural aggregation; Clumpy\_U: urban aggregation, ai\_L: landscape aggregation; Ca\_F: forest quantity; Ca\_G: grassland quantity; Ca\_A: agricultural quantity; Ca\_U: urban quantity in hectares. SD: standard deviation; Min: minimum; Max: maximum.

### S3.1.7. Distribution of simulated genetic patterns according to specialization profile and dispersal capacity



**S3.7:** Distribution of simulated genetic patterns according to specialization profile (specialist versus generalist), dispersal capacity (cost units) and spatial scale (sampling-scale versus effective-dispersal scale). Each vertical bar represents a landscape and the length of coloured segments represents the proportion of the three genetic patterns (IBR, IBD, no spatial structure) among the 10 samples of populations in this landscape, for a given dispersal capacity. This analysis was conducted 4 times, one for each combination of spatial scale (A, B sampling scale versus C, D effective-dispersal scale) and specialization profile (A, C specialist versus B, D generalist).

### S3.1.8. Spurious patterns of Isolation by Distance for low dispersal capacity



**S3.8:** Genetic distances (measured with  $F_{ST}$  index) as a function of Euclidean distances for four sampled landscapes. The red points correspond to links between two populations effectively related by dispersal. On the contrary, dark points correspond to population pairs unlikely to be connected by effective dispersal and whose genetic differentiation may be mainly driven by genetic drift. The grey regression line shows the linear relationship between genetic and Euclidean distances, all points being considered.

### S3.1.9. Confusion matrices assessing the optimization sensitivity and specificity to the simulated genetic patterns

**A.**

	Sampling-scale dataset			Dispersal-scale dataset		
	IBR simulated	IBD simulated	~1 simulated	IBR simulated	IBD simulated	~1 simulated
IBR detected by RGA	1819 <b>True positive</b>	48 <b>False positive</b>	124 <b>False positive</b>	1696 <b>True positive</b>	92 <b>False positive</b>	179 <b>False positive</b>
IBD detected by RGA	178 <b>False negative</b>	155 <b>True negative</b>	1 <b>False positive</b>	117 <b>False negative</b>	57 <b>True negative</b>	1 <b>False positive</b>
~1 detected by RGA	29 <b>False negative</b>	0 <b>False negative</b>	1756 <b>True negative</b>	31 <b>False negative</b>	2 <b>False negative</b>	1924 <b>True negative</b>

**B.**

	Sampling - scale dataset		Dispersal-scale dataset	
	<b>Simulation positive</b>	<b>Simulation negative</b>	<b>Simulation positive</b>	<b>Simulation negative</b>
<b>RGA positive</b>	True Positive 44%	False positive 4%	True Positive 41%	False positive 7%
<b>RGA negative</b>	False negative 5%	True negative 47%	False negative 4%	True negative 48%

**S3.1.9. a:** Confusion matrices assessing the sensitivity and the specificity of RGA algorithm to simulated genetic patterns as a function of the spatial scale. A) Detailed description of the absolute frequencies for all possible combinations of the simulated pattern *versus* the RGA optimized pattern. The combination type name is shown below the frequencies. The left matrix results from the sampling-scale dataset, and the right matrix from the effective-dispersal-scale dataset. B) True confusion matrices resulting from the synthesis of the table A. *RGA positive* corresponds to ‘IBR detected by RGA’ and *RGA negative* includes ‘IBD detected by RGA’ and ‘~1 detected by RGA’. *Simulation positive* corresponds to ‘IBR simulated’ and *Simulation negative* includes ‘IBD simulated’ and ‘~1 simulated’. The matrices show the relative frequencies of each combination type.

**S3.1.9. b:** *Statistical analysis of the difference in sensitivity and specificity between sampling-scale and effective-dispersal-scale datasets*

To assess the effect of the spatial scale consideration on the sensitivity and the specificity of the optimization process, we built a confusion matrix for both the sampling-scale dataset and the effective-dispersal-scale dataset. We then computed the sensitivity (True positives / (True positives + False negatives)) and the specificity (True negatives / (True negatives + False positives)) indices between both datasets.

We then performed two Chi-square tests comparing:

- i) the distribution of the true positives and the false negatives between the sampling-scale and the effective-dispersal-scale dataset (for the sensitivity),
- ii) the distribution of the true negatives and the false positives between the sampling-scale and the effective-dispersal-scale dataset (for the specificity).

For the sensitivity index, the chi-squared test comparing the distribution of true positives and false negatives between the two datasets showed a significantly better performance of the effective-dispersal scale dataset ( $X^2 = 5.3$ ,  $df = 1$ ,  $p=0.02$ ).

For the specificity index, the chi-squared test comparing the distribution of true negatives and false positives between the two datasets found a significantly better performance of the sampling-scale dataset, related to a higher than expected number of false positives in the effective-dispersal-scale dataset ( $X^2 = 16.11$ ,  $df = 1$ ,  $p=5.1e^{-5}$ ).

### S3.1.10 Significance of Spearman correlation values for land cover ranking in optimized cost scenarios

r	Land cover ranking compared to the specialist and generalist scenario	Land cover ranking compared to the 1 <sup>st</sup> variant	Land cover ranking compared to the 2 <sup>nd</sup> variant
1	F, G, A, U	F, A, G, U	F, U, A, G
0,8	F, G, U, A	F, A, U, G	F, U, G, A
0,8	G, F, A, U	A, F, G, U	U, F, A, G
0,8	F, A, G, U	F, G, A, U	F, A, U, G
0,6	G, F, U, A	A, F, U, G	U, F, G, A
0,4	A, F, G, U	G, F, A, U	A, F, U, G
0,4	F, U, G, A	F, U, A, G	F, G, U, A
0,4	F, A, U, G	F, G, U, A	F, A, G, U
0,4	G, A, F, U	A, G, F, U	U, A, F, G
0,2	A, G, F, U	G, A, F, U	A, U, F, G
0,2	F, U, A, G	F, U, G, A	F, G, A, U
0	A, F, U, G	G, F, U, A	A, F, G, U
0	G, U, F, A	A, U, F, G	U, G, F, A
-0,2	G, A, U, F	A, G, U, F	U, A, G, F
-0,2	U, F, G, A	U, F, A, G	G, F, U, A
-0,4	U, F, A, G	U, F, G, A	G, F, A, U
-0,4	G, U, A, F	A, U, G, F	U, G, A, F
-0,4	U, G, F, A	U, A, F, G	G, U, F, A
-0,4	A, G, U, F	G, A, U, F	A, U, G, F
-0,6	A, U, F, G	G, U, F, A	A, G, F, U
-0,8	U, A, F, G	U, G, F, A	G, A, F, U
-0,8	A, U, G, F	G, U, A, F	A, G, U, F
-0,8	U, G, A, F	U, A, G, F	G, U, A, F
-1	U, A, G, F	U, G, A, F	G, A, U, F

**S3.10:** Correspondence table of the Spearman correlation of the ranking consistency between the true and optimized cost scenarios, and the related land cover types ranking (F: forest; G: grassland; A: agricultural; U: urban area), for the four true cost scenarios. r: Spearman's coefficient of correlation; Land cover types are ranked from the least to the most resistant.



## REFERENCES

- Dyer, Rodney J., and John D. Nason. 2004. "Population Graphs: The Graph Theoretic Shape of Genetic Structure: POPULATION GRAPHS." *Molecular Ecology* 13 (7): 1713–27. <https://doi.org/10.1111/j.1365-294X.2004.02177.x>.
- Estoup, Arnaud, Philippe Jarne, and Jean-Marie Cornuet. 2002. "Homoplasy and Mutation Model at Microsatellite Loci and Their Consequences for Population Genetics Analysis." *Molecular Ecology* 11 (9): 1591–1604. <https://doi.org/10.1046/j.1365-294X.2002.01576.x>.
- Keeley, Annika T. H., Paul Beier, and Jeffrey W. Gagnon. 2016. "Estimating Landscape Resistance from Habitat Suitability: Effects of Data Source and Nonlinearities." *Landscape Ecology* 31 (9): 2151–62. <https://doi.org/10.1007/s10980-016-0387-5>.
- McGarigal, Kevin, and Barbara J. Marks. 1995. *FRAGSTATS: Spatial Pattern Analysis Program for Quantifying Landscape Structure*. U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. <https://doi.org/10.2737/pnw-gr-351>.
- Peterman, William E. 2018. "ResistanceGA: An R Package for the Optimization of Resistance Surfaces Using Genetic Algorithms." Edited by Simon Jarman. *Methods in Ecology and Evolution* 9 (6): 1638–47. <https://doi.org/10.1111/2041-210X.12984>.
- Savary, Paul, Jean-Christophe Foltête, Hervé Moal, Gilles Vuidel, and Stéphane Garnier. 2021. "Analysing Landscape Effects on Dispersal Networks and Gene Flow with Genetic Graphs." *Molecular Ecology Resources* 21 (4): 1167–85. <https://doi.org/10.1111/1755-0998.13333>.
- "Union Européenne – SOeS, CORINE Land Cover." 2012. 2012. <https://www.statistiques.developpement-durable.gouv.fr/corine-land-cover-0>.
- Urban, Dean, and Timothy Keitt. 2001. "Landscape connectivity: a graph-theoretic perspective" 82 (5): 14.





## Annexe 3:

### Supplemental Information for:

#### Is landscape effect on gene flow shared across species? A multispecies cross-validation landscape genetic study.

Daniel, Alexandrine, Savary, Paul, Foltête, Jean-Christophe, Vuidel, Gilles, Faivre, Bruno, Garnier, Stéphane, Khimoun, Aurélie.

S4.1 Sampling description	<b>Page 292</b>
S4.2 Deviation for Hardy-Weinberg equilibrium, linkage disequilibrium and genetic markers, temporal differentiation	<b>Page 292</b>
S4.4 Optimization results	<b>Page 292</b>
S4.5 STRUCTURE results	<b>Page 293</b>
S4.5. Validation R2 of cross species model	<b>Page 295</b>
S4.6 Spatial representation of Flux values of habitat patches for each species	<b>Page 297</b>

### S4.1. Sampling description

In this PhD manuscript, see section 2.4 and 2.5.

### S4.2. Deviation for Hardy-Weinberg equilibrium, linkage disequilibrium and genetic markers, temporal differentiation

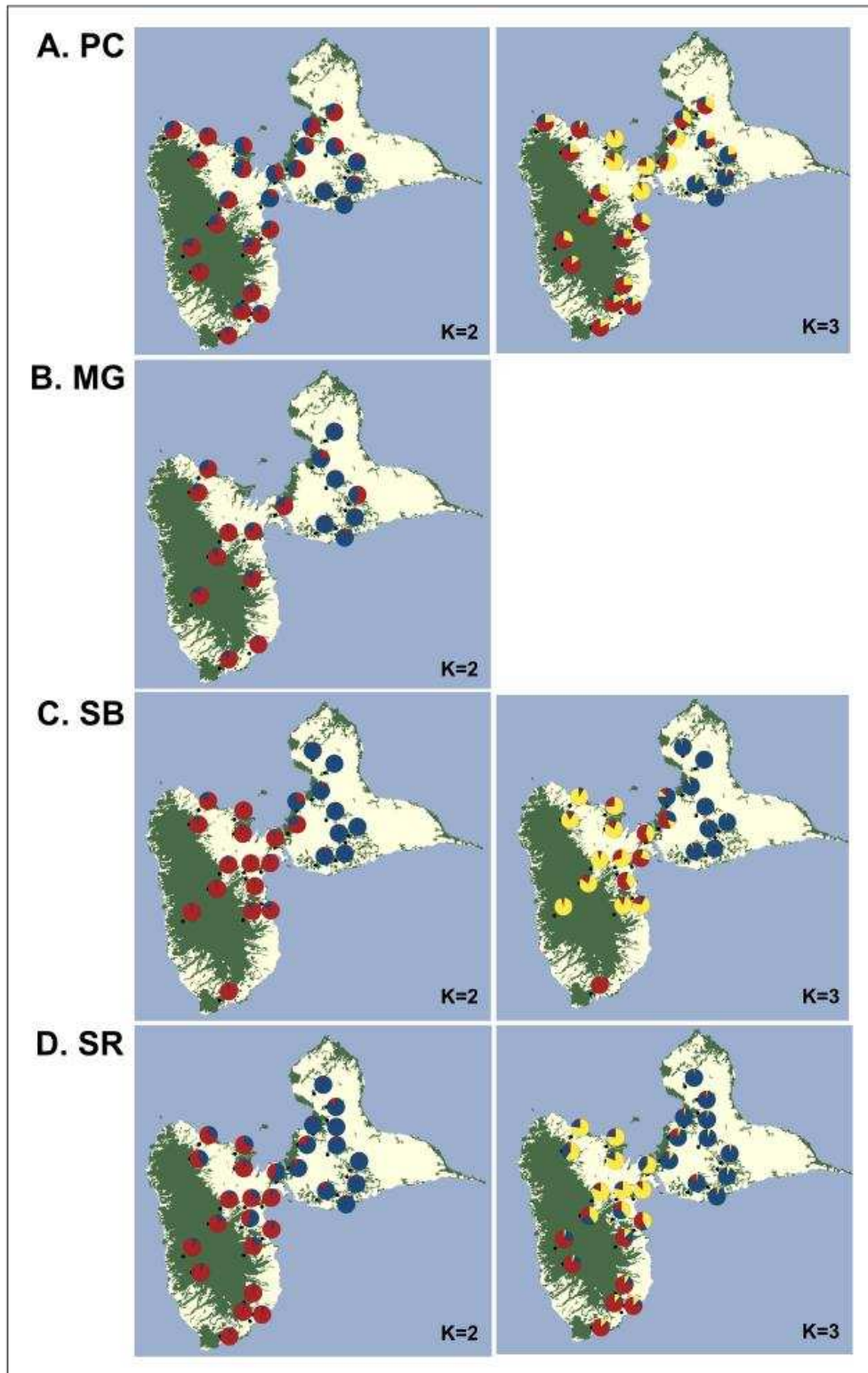
In this PhD manuscript, see section 2.6.1.3 and 2.6.1.4.

### S4.3. Optimization results

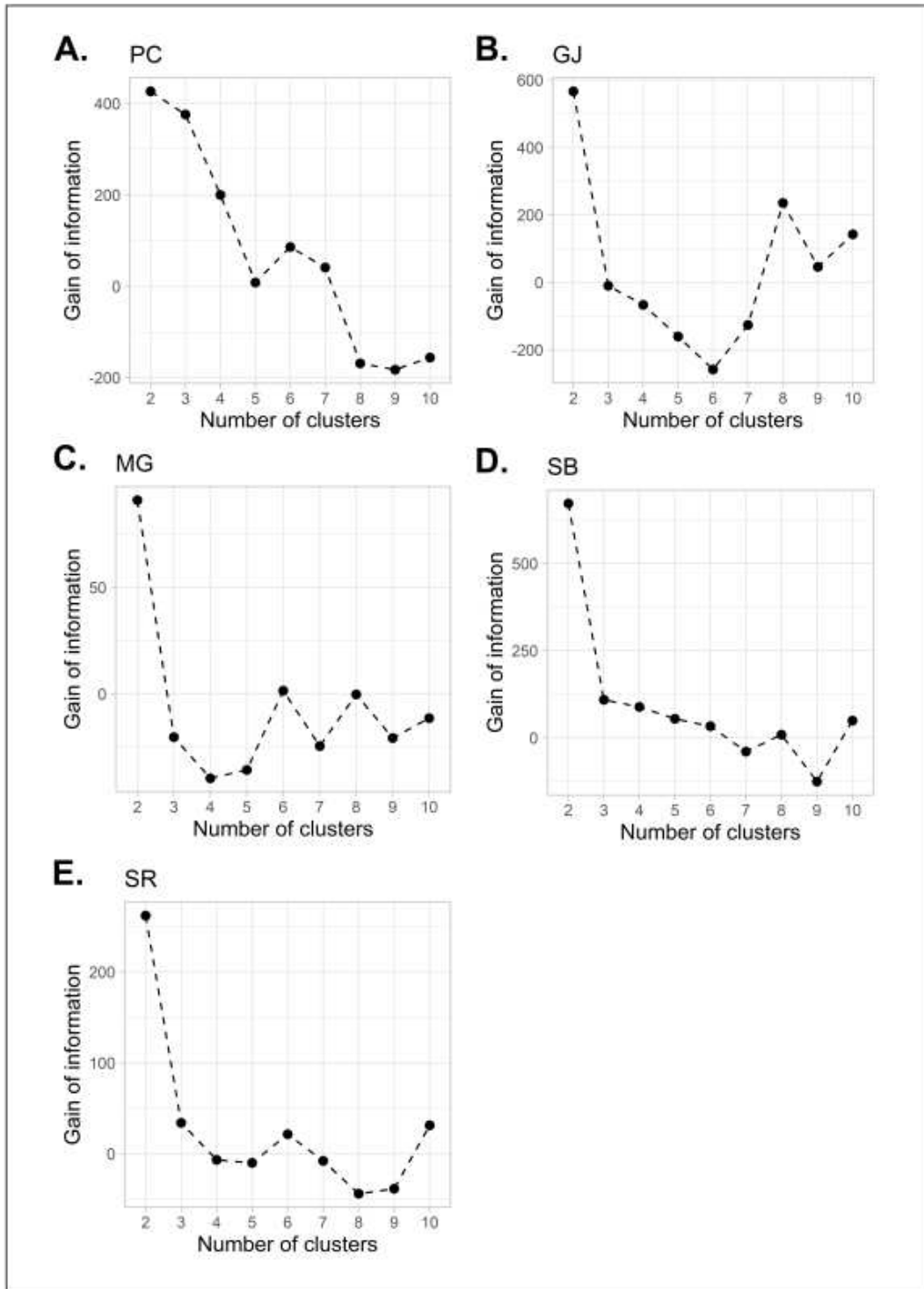
<i>Pruning</i>		<i>Complete scenario</i>					<i>Island scenario</i>				
Land cover	Species	PC	GJ	MG	SB	SR	PC	GJ	MG	SB	SR
Habitat		1	1	1	3	8	1	1	1	2	96
Sub-optimal woody vegetation (Forest)		106	2901	2	4	869	843	4370	3	87	32731
Seminatural (S-N)		64	66	334	10	336	104	43	976	4	822
Semiopen (S-O)		2264	2339	770	46	1	593	111	54	10	1
Open area		360	2	1603	1	3	458	35	2169	1	2
Agricultural		773	349	157	11	82	268	473	290	1	296
Wet areas and inland water (Wet)		9870	3599	379	18	28	24	150	623	3	13
Ocean		9	57	3	58	108	78	1142	7	9	121
Artificial (Artif.)		1356	392	238	1	3	553	1180	217	1	457

**Table S4.3:** Optimized costs obtained from ResistanceGA workflow and assigned to land cover type for the five species and the two spatial scales.

#### S4.4. STRUCTURE results



**S4.4.1:** Spatialization of the results from the clustering analyses with STRUCTURE for each of the five species. Only the lowest levels of structuring are displayed here. See figure 4 for the optimal finest levels. Each pie chart illustrates the genetic composition of a sampled population, according to the number of clusters considered for each species. K represents the optimal number of clusters for the finest genetic structuring.



**S4.4.2:** Successive increases of the posterior probability of the data with increasing number of clusters from STRUCTION clustering analyses, following the approach proposed by Garnier et al. (2004). To infer the optimal number of clusters (as represented in Figure 4), we also examined barplots representing the proportion of each allelic pool associated with each individual (see Annexe 5).

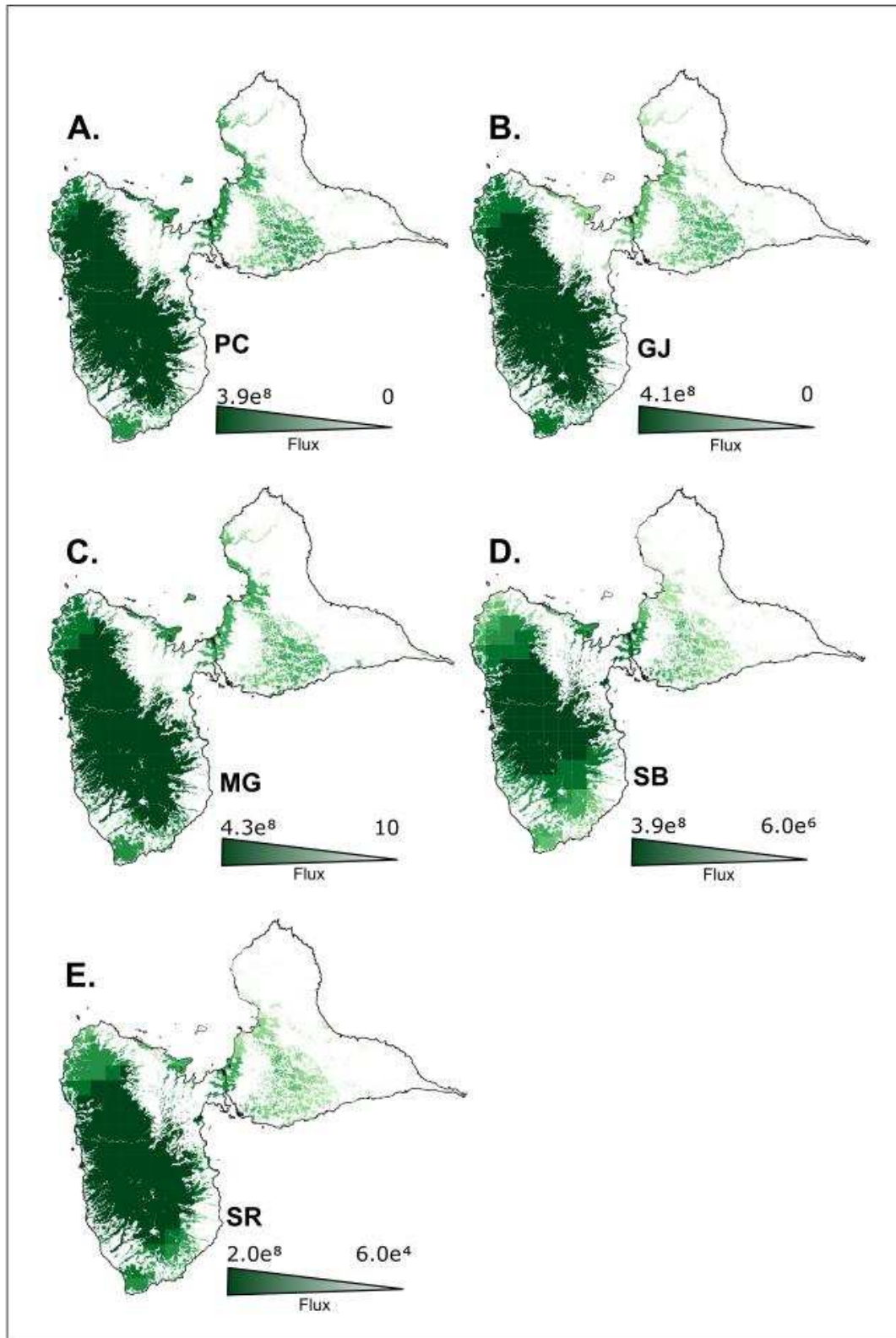
#### S4.5. Validation R<sup>2</sup> of cross species model

Predicted Y	Predictor X	Spatial scale	Cost-distance vR <sup>2</sup>	Euclidean R <sup>2</sup> v	dAIC <sub>Cost-distance-Euclidean distance</sub>
GJ	MG	Island	0	0	-1,4
GJ	SB	Island	0		0,7
GJ	SR	Island	0		7,6
GJ	PC	Island	0,099		-6,2
GJ	MG	Whole	0,716	0,293	-196,6
GJ	SB	Whole	0,041		53,0
GJ	SR	Whole	0,122		44,3
GJ	PC	Whole	0,485		-107,0
MG	PC	Island	0,247	0	-4,3
MG	SB	Island	0		0,4
MG	GJ	Island	0		0,7
MG	SR	Island	0		9,9
MG	PC	Whole	0,319	0,087	-12,8
MG	GJ	Whole	0,066		-14,0
MG	SB	Whole	0		9,0
MG	SR	Whole	0,007		6,5
PC	SB	Island	0	0	-91,1
PC	MG	Island	0,266		-143,6
PC	SR	Island	0		-90,5
PC	GJ	Island	0,293		-124,8
PC	SB	Whole	0	0	-138,6
PC	SR	Whole	0		-141,1
PC	MG	Whole	0,105		-182,4
PC	GJ	Whole	0		-176,2
SB	PC	Island	0	0,216	16,4
SB	MG	Island	0		16,3
SB	SR	Island	0,128		16,2
SB	GJ	Island	0		16,9
SB	PC	Whole	0	0,452	91,0
SB	MG	Whole	0,017		67,0
SB	SR	Whole	0,430		-8,2
SB	GJ	Whole	0,096		46,6
SR	SB	Island	0,220	0,179	-1,9
SR	PC	Island	0		41,9
SR	MG	Island	0		38,0
SR	GJ	Island	0		45,7
SR	PC	Whole	0	0,208	84,8
SR	MG	Whole	0,087		52,7
SR	GJ	Whole	0,167		45,8
SR	SB	Whole	0,278		-25,5



**S4.5:** Statistic indicators of the cross-predictive models based on both Euclidean distances and cost distances, for the five species and the two spatial scales. Predicted Y: ID of the species whose genetic structure was predicted, *i.e.*, the species on the right panel on Figure 5. Predictor X: ID of the species whose genetic structure was used for the calibration of a cost-distance model, *i.e.*, the species on the left panel on Figure 5. Spatial scale: the spatial scale of the analysis, either the within island scale ('island') or the whole Guadeloupe scale ('Whole'). Cost-distance  $vR^2$ : validation  $R^2$  assessing the predictive quality of the cross-predictive model based on cost distances (*i.e.*, the width of the dark blue and chequered arrows on Figure 5). Euclidean  $vR^2$ : validation  $R^2$  assessing the predictive quality of the cross-predictive model based on cost distances (*i.e.*, the width of the light blue arrows on Figure 5).  $dAIC_{\text{Cost-distance-Euclidean distance}}$  : Difference between the AIC of the cost-distance models and the AIC of the Euclidean distance models. If this difference was positive, the arrow of the related cost-distance "cross-prediction" was chequered on Figure 5.

**S4.6. Spatial representation of Flux values of habitat patches for each species**



**S4.6:** Spatial representation of Flux values of forest habitat patches for each species, derived from landscape graphs constructed from species-specific data.



## Annexe 4:

### Supplemental Information for:

#### Disentangling past landscape legacies from the contemporary genetic structure of a forest bird using a historical map

Daniel, Alexandrine, Savary, Paul, Foltête, Jean-Christophe, Roy, Damien, Vuidel, Gilles, Faivre, Bruno, Garnier, Stéphane, Khimoun, Aurélie.

S5.2.1. Formations for the building of landcover maps	<b>Page 300</b>
S5.2.2. Cost scenarios	<b>Page 301</b>
S5.2.3. Table of cost scenarios resulting in the best resistance models according to the AIC criterion for the <i>WITHIN-ISLAND-SCALE MODELS</i> – Continuation	<b>Page 302</b>

### S5.2.1. Formations for the building of landcover maps

Vegetation Formations	Land cover
Altitude thickets	Matrix
Low-growth altitude thickets	Matrix
Low-growth pioneer formations	Matrix
Various silvicultural areas	Forest
Diversified sylvicultural areas	Forest
Limestone lowland forests	Forest
Coastal forest	Forest
Rain forest	Forest
Valley floor forest	Forest
Fallow lands with low-growth woody vegetation	Matrix
Coastal thicket	Matrix
Mangrove	Foret
Swamp forest	Foret
Mahogany forest	Foret
Semi-deciduous forest	Foret
Seasonal evergreen forest	Foret
Forests of agricultural areas	Matrix

**Table S5.1:** Vegetation formations of the 1950 maps and the 2010 maps created by the Conseil Departemental of Guadeloupe, the IGN and the ONF using manually interpreted aerial photographs. The "Land cover" column indicates whether the formation was considered as being part of the forest or the matrix area.

### S5.2.2. Cost scenarios

	Forest	Urban	Matrix	Sea
<b>1</b>	871	972	934	1000
<b>2</b>	708	931	841	1000
<b>3</b>	500	867	708	1000
<b>4</b>	250	750	500	1000
<b>5</b>	63	563	250	1000
<b>6</b>	4	317	63	1000
<b>7</b>	1	101	4	1000
<b>8</b>	1	11	1	1000
<b>9</b>	10	643	314	1000
<b>10</b>	7	556	237	1000
<b>11</b>	5	442	154	1000
<b>12</b>	2	255	55	1000
<b>13</b>	2	72	6	1000
<b>14</b>	13	698	370	1000
<b>15</b>	17	773	457	1000
<b>16</b>	24	854	572	1000
<b>17</b>	41	949	758	1000
<b>18</b>	78	996	935	1000

**Table S5.2:** Table of the 18 cost scenarios systematically derived from Keeley et al.'s (2016) and Shirk et al.'s (2010) approaches. Each scenario leads to 6 variants, corresponding to all the possible permutations of costs associated with urban, matrix and sea areas.

**S5.2.3. Table of cost scenarios resulting in the best resistance models according to the AIC criterion for the *WITHIN-ISLAND-SCALE MODELS* – Continuation**

Scenario Forest - Urban - Matrix - Sea	Model	dAIC RD - Euclidean	dAIC Current – Historical	Best support period
13 - 1000 - 698 - 370	FST ~ Current RD	-60,64	-2,71	Current
10 - 314 - 1000 - 643	FST ~ Current RD	-60,59	-1,32	Identical
5 - 154 - 442 - 1000	FST ~ Historical RD	-60,48	0,48	Identical
7 - 237 - 556 - 1000	FST ~ Current RD	-60,11	0,56	Identical
7 - 1000 - 237 - 556	FST ~ Current RD	-60,07	-1,30	Identical
2 - 55 - 255 - 1000	FST ~ Current RD	-60,00	-0,68	Identical
5 - 154 - 442 - 1000	FST ~ Current RD	-59,99	0,48	Identical
5 - 1000 - 154 - 442	FST ~ Current RD	-59,91	-1,51	Identical
10 - 314 - 643 - 1000	FST ~ Historical RD	-59,77	0,53	Identical
2 - 1000 - 55 - 255	FST ~ Current RD	-59,56	-1,10	Identical
2 - 55 - 1000 - 255	FST ~ Current RD	-59,39	-12,58	Current
13 - 370 - 1000 - 698	FST ~ Current RD	-59,32	-0,72	Identical
2 - 55 - 255 - 1000	FST ~ Historical RD	-59,31	-0,68	Identical
10 - 314 - 1000 - 643	FST ~ Historical RD	-59,26	-1,32	Identical
10 - 314 - 643 - 1000	FST ~ Current RD	-59,24	0,53	Identical
17 - 773 - 1000 - 457	FST ~ Current RD	-59,22	-2,04	Current
2 - 255 - 55 - 1000	FST ~ Historical RD	-59,16	1,76	Identical
7 - 556 - 237 - 1000	FST ~ Historical RD	-59,02	0,39	Identical
10 - 1000 - 314 - 643	FST ~ Current RD	-58,99	-1,61	Identical
17 - 1000 - 773 - 457	FST ~ Current RD	-58,94	-2,41	Current
10 - 1000 - 643 - 314	FST ~ Historical RD	-58,90	-3,02	Current
7 - 1000 - 556 - 237	FST ~ Historical RD	-58,82	-3,81	Current

**Table S5.3:** Spatial representation of Flux values of forest habitat patches for each species, derived from landscape graphs constructed from species-specific data.

Scenario Forest - Urban - Matrix - Sea	Model	dAIC		Best support period
		RD - Euclidean	Current – Historical	
10 - 643 - 1000 - 314	FST ~ Historical RD	-58,80	-3,26	Current
5 - 442 - 154 - 1000	FST ~ Historical RD	-58,78	0,69	Identical
7 - 1000 - 237 - 556	FST ~ Historical RD	-58,76	-1,30	Identical
7 - 237 - 1000 - 556	FST ~ Historical RD	-58,73	-2,67	Current
7 - 556 - 237 - 1000	FST ~ Current RD	-58,62	0,39	Identical
13 - 370 - 1000 - 698	FST ~ Historical RD	-58,59	-0,72	Identical
2 - 1000 - 55 - 255	FST ~ Historical RD	-58,46	-1,10	Identical
5 - 1000 - 154 - 442	FST ~ Historical RD	-58,40	-1,51	Identical
13 - 370 - 698 - 1000	FST ~ Historical RD	-58,39	0,39	Identical
13 - 698 - 1000 - 370	FST ~ Historical RD	-58,31	-2,57	Current
4 - 63 - 1000 - 317	FST ~ Current RD	-58,25	-7,81	Current
5 - 442 - 154 - 1000	FST ~ Current RD	-58,09	0,69	Identical
13 - 370 - 698 - 1000	FST ~ Current RD	-58,00	0,39	Identical
13 - 1000 - 698 - 370	FST ~ Historical RD	-57,92	-2,71	Current
10 - 643 - 314 - 1000	FST ~ Current RD	-57,78	-0,32	Identical
7 - 556 - 1000 - 237	FST ~ Historical RD	-57,74	-4,79	Current
17 - 457 - 1000 - 773	FST ~ Current RD	-57,73	-0,41	Identical
13 - 1000 - 370 - 698	FST ~ Current RD	-57,58	-1,98	Identical
4 - 63 - 317 - 1000	FST ~ Historical RD	-57,53	0,59	Identical
10 - 643 - 314 - 1000	FST ~ Historical RD	-57,46	-0,32	Identical
2 - 255 - 55 - 1000	FST ~ Current RD	-57,40	1,76	Identical
10 - 1000 - 314 - 643	FST ~ Historical RD	-57,38	-1,61	Identical
17 - 457 - 1000 - 773	FST ~ Historical RD	-57,32	-0,41	Identical
17 - 773 - 1000 - 457	FST ~ Historical RD	-57,18	-2,04	Current
4 - 1000 - 317 - 63	FST ~ Current RD	-57,04	-7,76	Current
4 - 63 - 317 - 1000	FST ~ Current RD	-56,94	0,59	Identical
5 - 1000 - 442 - 154	FST ~ Historical RD	-56,76	-5,24	Current
17 - 457 - 773 - 1000	FST ~ Historical RD	-56,72	0,09	Identical
2 - 1000 - 255 - 55	FST ~ Current RD	-56,67	-8,49	Current
17 - 457 - 773 - 1000	FST ~ Current RD	-56,63	0,09	Identical
13 - 698 - 370 - 1000	FST ~ Current RD	-56,53	-0,92	Identical
17 - 1000 - 773 - 457	FST ~ Historical RD	-56,53	-2,41	Current
5 - 154 - 1000 - 442	FST ~ Historical RD	-56,28	-4,90	Current
17 - 1000 - 457 - 773	FST ~ Current RD	-56,12	-2,18	Current
24 - 854 - 1000 - 572	FST ~ Current RD	-55,97	-1,89	Identical
24 - 1000 - 854 - 572	FST ~ Current RD	-55,71	-2,34	Current





## **Annexe 5 :**

### **Etude de la connectivité des habitats forestiers entre la Grande-Terre et la Basse-Terre**

Garnier, Stéphane, Daniel, Alexandrine, Khimoun, Aurélie, Faivre Bruno

Rapport final à l'attention de la DEAL de Guadeloupe

30 septembre 2021



# Etude de la connectivité des habitats forestiers entre la Grande-Terre et la Basse-Terre

## Rapport final

Stéphane Garnier, Alexandrine Daniel, Aurélie Khimoun, Bruno Faivre

Biogéosciences, UMR CNRS/uB/EPHE 6282  
Université Bourgogne Franche-Comté

30 septembre 2021



## 1/ Objet du projet

### ***1.1. Préambule***

Le projet "Etude de la connectivité des habitats forestiers entre Grande-Terre et Basse-Terre" s'inscrit dans la poursuite des activités de recherche conduites dans le cadre du projet de recherche européen FRAG&BINV (Conséquences de la fragmentation des forêts et conditions pour les invasions biologiques : le cas des oiseaux des Caraïbes). Ce projet était coordonné par le responsable scientifique du présent projet (Stéphane Garnier, laboratoire Biogéosciences, UMR CNRS/uB 6282, Dijon), et il s'est déroulé entre 2012 et 2016.

Le projet FRAG&BINV avait pour objectif d'étudier les conséquences de la fragmentation des forêts sur les populations de plusieurs espèces d'oiseaux de quatre territoires en zone caribéenne : Guadeloupe, Martinique, Montserrat, et la Guyane française. Ce programme visait à comparer les caractéristiques génétiques et phénotypiques (morphologie, immunologie, couleur) des individus/populations, ainsi que leurs interactions avec leurs parasites, entre milieux forestiers plus ou moins fragmentés, pour un cortège d'espèces montrant un gradient de spécialisation envers l'habitat forestier.

Le projet FRAG&BINV s'est officiellement terminé en 2016, et l'analyse du grand nombre d'échantillons et de données qu'il a permis de collecter est actuellement en cours de finalisation. Ce projet a également permis d'obtenir (et continue de produire) plusieurs résultats intéressants et originaux dont une partie a déjà été valorisée sous forme de publications scientifiques (dont certaines ont été jointes en annexe du présent rapport) dans des revues scientifiques internationales de premier plan. Ces résultats ont également été présentés lors de colloques nationaux (Rencontres d'écologie du paysage en 2017 à Toulouse et en 2019 à Bordeaux, Rencontre du groupe de biologie et génétique des populations à Gif-sur-Yvette en 2019, Réunion du Réseau Ecologie de Interactions Durables à Montpellier en 2019) et internationaux (2nd International Conference on Island Evolution, Ecology and Conservation aux Açores en 2016, Spatial Data Meetings for Interdisciplinary Sciences à Dijon en 2015, Regional Meeting, Society for the Conservation and Study of Caribbean Birds à la Grenade en 2013). Enfin, une partie des résultats obtenus a été présentée lors de réunions de restitution et d'échanges avec les acteurs locaux de la biodiversité, en particulier en Guadeloupe (réunion organisée à la DEAL de Guadeloupe le 3/03/2020).

Dans le cadre du projet FRAG&BINV, des résultats particulièrement intéressants (voir ci-dessous) soulèvent certaines questions en matière d'aménagement du territoire et de préservation des habitats naturels. C'est dans ce cadre que s'insère le présent projet qui a bénéficié d'un complément de financement de la DEAL de Guadeloupe (**arrêté 971-2021-04-06-00004**). Il s'agit en effet d'un projet à vocation appliquée (au sein du territoire de la Guadeloupe) destiné à compléter les résultats issus de travaux de recherche essentiellement fondamentaux.

Si le projet "Etude de la connectivité des habitats forestiers entre Grande-Terre et Basse-Terre" est focalisé sur le secteur de l'isthme qui connecte les deux parties de la Guadeloupe, il est à noter qu'un autre projet en cours (projet CANON : *Landscape Connectivity and Biodiversity: Networks of Networks*), également piloté par le laboratoire Biogéosciences (Stéphane Garnier, UMR CNRS/UB 6282, Dijon), traite également des continuités écologiques en Guadeloupe, mais à l'échelle de l'île entière. Les deux grands objectifs du projet CANON sont de déterminer comment la structure du paysage influence 1) la connectivité fonctionnelle (*i.e.* les déplacements réels des organismes) des populations et des communautés d'oiseaux et 2) la circulation des communautés de parasites au sein des communautés d'oiseaux sauvages.

### ***1.2. Thèmes de la fragmentation des habitats et des continuités écologiques***

La fragmentation des habitats est l'une des principales menaces pour la biodiversité, en particulier pour les forêts tropicales et inter-tropicales qui constituent l'habitat terrestre abritant la biodiversité la plus importante alors même qu'il subit la plus forte régression liée aux activités

anthropiques (Fahrig 2003, Laurence 2010). Malgré une riche littérature centrée sur l'impact de la fragmentation des habitats sur l'abondance des espèces et leur diversité, les conséquences de ces changements globaux sur les processus écologiques et évolutifs restent mal comprises. Se traduisant généralement par une réduction de la taille des parcelles d'habitat et un accroissement de leur isolement, la fragmentation affecte la diversité génétique des populations. D'une part, la réduction de la taille de la population locale (d'autant plus importante que le fragment est petit) entraîne une dérive génétique accrue. D'autre part, la baisse de connectivité entre fragments réduit la migration, et donc les flux de gènes entre les populations. L'action (et l'interaction) de ces deux processus conduit à un appauvrissement de la diversité génétique au sein des populations et une augmentation de la divergence entre populations (Keyghobadi 2007).

Pour faire face à une modification de son environnement, un organisme n'a que deux options, la migration ou l'adaptation. Or, les possibilités de migration d'un organisme sont liées à ses capacités intrinsèques de dispersion mais également au paysage dans lequel il évolue. Quant au potentiel adaptatif (*i.e.*, capacité d'une population ou d'une espèce à répondre à la sélection naturelle), il est directement lié à la diversité génétique au sein des populations/espèces. Dans le contexte actuel des changements globaux, comprendre comment la structure du paysage influence la dispersion des individus et les flux de gènes entre populations est donc primordial pour préserver le potentiel adaptatif des populations naturelles et limiter leur risque d'extinction (Allendorf & Luikart 2007). Ces préoccupations sont particulièrement évidentes pour des espèces insulaires et/ou endémiques qui montrent une fragilité singulière en raison de leur abondance généralement faible et de leur distribution géographique restreinte (Pimm 1991, Gaston 1994).

### ***1.3. Résultats obtenus dans le cadre du projet FRAG&BINV***

Un objectif majeur du projet FRAG&BINV était donc de déterminer **comment la fragmentation de l'habitat forestier, et plus généralement la structure du paysage, influencent la connectivité entre les populations forestières de plusieurs espèces d'oiseaux**, en adoptant les approches de la génétique des populations et de la génétique du paysage. Plusieurs résultats importants ont déjà été obtenus dans le cadre de ce projet, notamment en Guadeloupe, et ces résultats ont fondé la base de l'étude décrite dans le présent rapport.

- 1) Des analyses génétiques nous ont déjà permis de mettre en évidence des différences génétiques entre les îles de la Guadeloupe, de la Dominique et de la Martinique pour 9 espèces d'oiseaux : Viréo à moustaches, Elénie siffleuse, Colibri madère, Sucrier à ventre jaune, Saltator gros-bec, Sporophile rouge-gorge, Moqueur grivotte, Paruline caféïette et Grive à pieds jaunes (Arnoux et al. 2014, Khimoun et al 2016a). Ce type de résultat n'est pas fondamentalement nouveau dans le contexte d'un archipel, mais la délimitation des différentes entités génétiques détectées n'est parfois pas concordante avec la délimitation des sous-espèces définies sur la base de critères morphologiques principalement.
- 2) De manière plus remarquable, nous avons mis en évidence l'existence d'une différenciation génétique au sein même de la Guadeloupe entre les populations de la Grande-Terre et de la Basse-Terre, ceci pour plusieurs espèces : Elénie siffleuse, Colibri madère, Saltator gros-bec, Sporophile rouge-gorge, Moqueur grivotte, Paruline caféïette et Grive à pieds jaunes (Arnoux et al. 2014, Khimoun et al 2016a, Khimoun et al 2016b, Khimoun et al. 2017a). Ce type de résultat est plus surprenant si l'on considère l'échelle spatiale de la Guadeloupe et le fait de travailler sur des oiseaux auxquels on prête généralement d'importantes capacités de dispersion en raison du vol.
- 3) Par ailleurs, nous avons mis en évidence que plus une espèce est spécialiste de la forêt (ex Grive à pieds jaunes et Paruline caféïette), plus elle est sensible à la fragmentation de cet habitat (cette sensibilité étant quantifiée par le niveau de divergence génétique entre populations situées dans des fragments forestiers déconnectés). Autrement dit, la même

fragmentation du paysage impacte beaucoup plus durement ces espèces que les espèces généralistes. Ce résultat est conforme à ce qui était prédit, mais il n'y a à notre connaissance aucune étude similaire dans la littérature (comparaison de 8 espèces à une si petite échelle spatiale et avec des mesures de la réponse à la fragmentation et du niveau de spécialisation écologique issues de données de terrain ; Khimoun et al 2016b).

- 4) Enfin, des analyses plus approfondies conduites à ce jour seulement chez la Paruline caféïette (mais en cours chez d'autres espèces) révèlent que les populations sont également différenciées génétiquement au sein de la Basse-Terre et de la Grande-Terre, même si les niveaux de différenciation sont inférieurs à ceux observés entre les deux entités de la Guadeloupe (Khimoun et al 2017b). Les analyses montrent également que la structure du paysage est le principal déterminant du niveau de différenciation génétique entre les populations, qui rend compte du déplacement des oiseaux, et que certains habitats (notamment l'habitat urbain) constituent des freins importants à la dispersion des oiseaux.

Dans le cadre des autres objectifs du projet FRAG&BINV, nous avons également mis en évidence que la fragmentation des forêts au sens strict (c'est-à-dire configuration spatiale du manteau forestier : nombre de fragments forestiers, isolement de ces fragments, agrégation de ces fragments dans le paysage...) influence de manière bien plus importante les relations entre les oiseaux et leurs parasites que la quantité de forêt dans le paysage, en Guadeloupe et en Martinique (Perez-Rodriguez et al. 2018). Nous avons ainsi observé chez le Sporophile rouge-gorge que la proportion d'individus infectés par les parasites sanguins Haemosporidés (responsable de la malaria aviaire) augmente avec le niveau de fragmentation (au sens strict) du manteau forestier. Si ces résultats traitent des relations entre hôtes et parasites, ils ne sont pas forcément indépendants des résultats obtenus en matière de connectivité. En effet, l'appauvrissement génétique des populations d'oiseaux en réponse à la fragmentation des forêts, et donc une rupture au moins partielle de connectivité, pourrait être responsable d'une immunité des oiseaux moins performante. Par ailleurs, la manière dont la structure du paysage influence les déplacements des oiseaux dans l'espace, pas nécessairement identique chez toutes les espèces, pourrait in fine influencer la manière dont les parasites se déplacent entre fragments forestiers.

Enfin, d'autres résultats ont été obtenus sur les autres territoires ciblés par le projet FRAG&BINV. Nous avons par exemple observé des effets négatifs de la fragmentation des forêts en Guyane française sur la diversité génétique du Grimpar bec-en-coïn, une espèce forestière largement répandue en Amérique du Sud (Perrin et al. 2021). Plus la distance à traverser en dehors de la forêt est grande, pour passer d'un fragment forestier à un autre, moins la dispersion des oiseaux est importante. Comme en Guadeloupe, on observe ces effets à de très petites échelles spatiales.

L'ensemble des résultats du projet FRAG&BINV suggère donc que les effets de la fragmentation des habitats pourraient avoir été sous-estimés chez les oiseaux en milieu tropical, en raison des capacités de dispersion a priori importantes qu'on prêt à ces organismes.

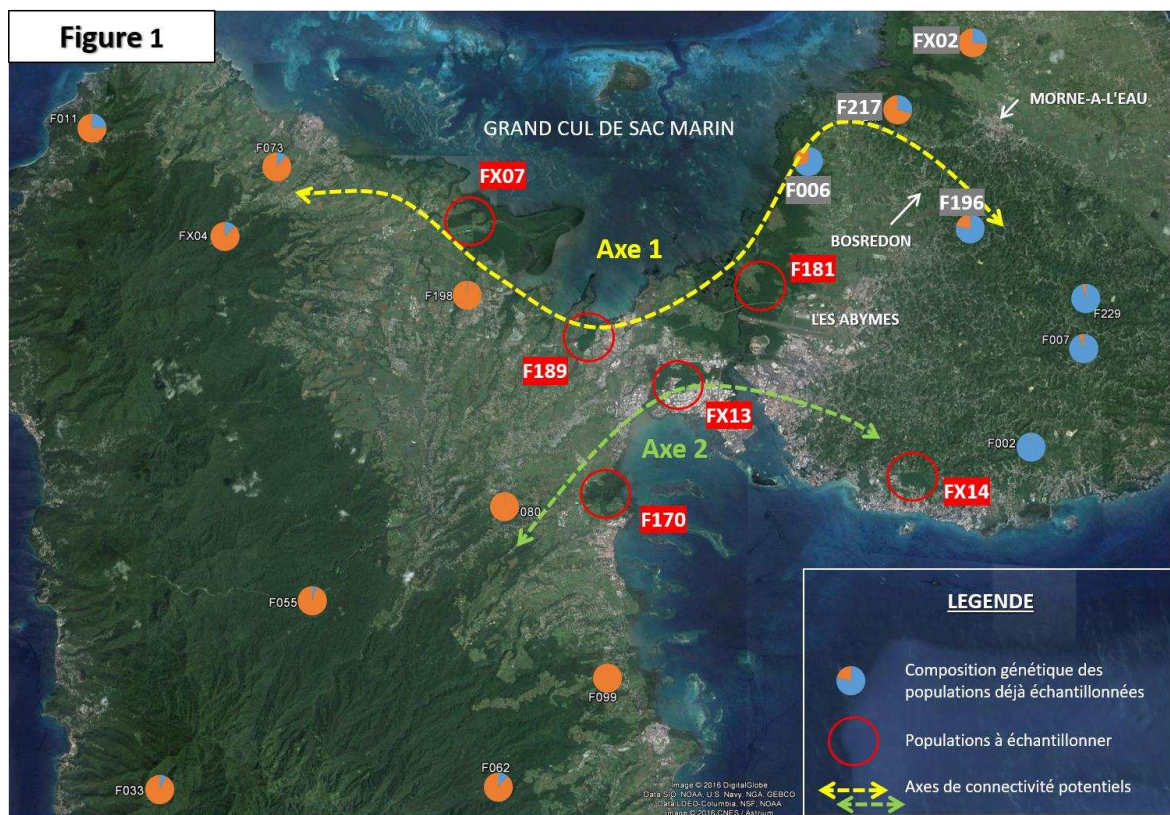
## 2/ Objectifs poursuivis

Si les résultats obtenus à travers le projet FRAG&BINV montrent clairement qu'il existe une divergence génétique entre les populations de la Basse-Terre et celles de la Grande-Terre chez plusieurs espèces d'oiseaux, il reste à déterminer précisément quelle(s) voie(s) permet(tent) le maintien de la connectivité entre ces deux parties de la Guadeloupe.

Les analyses de génétique des populations conduites chez la Paruline caféïette montrent l'existence de deux grands pools génétiques au sein de la Guadeloupe, l'un sur la Basse-Terre et l'autre sur la Grande-Terre (respectivement orange et bleu sur la figure 1). Elles semblent aussi indiquer l'existence d'un gradient de différenciation génétique entre ces deux entités. En effet, trois populations (F006, F217, FX02) échantillonnées dans les forêts marécageuses et mangroves à l'est du Grand Cul

de Sac Marin renferment une diversité génétique apparemment mixte, et la constitution génétique de la population F196 suggère une certaine connexion entre les forêts des Grands Fonds et les forêts marécageuses et mangroves (situées sur la Grande-Terre le long du Grand Cul de Sac Marin) grâce aux habitats forestiers situés entre Morne-A-L'Eau et Bosredon (voir axe 1 sur figure 1). Les forêts marécageuses et les mangroves dispersées le long du Grand Cul de Sac Marin pourraient donc constituer des relais à la dispersion (de proche en proche) pour les oiseaux entre la Basse-Terre et les forêts des Grands Fonds.

Par ailleurs, nous avons utilisé des analyses récentes de génétique du paysage chez cette même espèce (Khimoun et al. 2017b). Brièvement, la génétique du paysage résulte du mariage entre la génétique des populations et l'écologie du paysage. Il s'agit d'analyser la structure génétique des populations en prenant explicitement en compte la localisation géographique des populations et la structure du paysage entre ces populations. Ces analyses suggèrent que la connexion entre les milieux forestiers de la Grande-Terre et la Basse-Terre passe par le sud de l'isthme reliant les deux parties de l'île (voir axe 2 sur figure 1).



L'objectif ultime de la présente étude était donc de **tester l'existence de ces deux axes de connectivité entre la Grande-Terre et la Basse-Terre et de mesurer leur importance relative**. Déterminer le rôle potentiel de ces deux voies de dispersion pour une espèce forestière telle que la Paruline caféïette (qui montre les niveaux de structuration génétique les plus élevés parmi les espèces étudiées et donc la plus grande sensibilité à la fragmentation des forêts ; Khimoun et al. 2016b) permettra de mieux cibler les fragments et corridors à préserver ou éventuellement à restaurer. Les données et échantillons disponibles jusqu'alors ne permettaient pas de tester ces hypothèses parce que l'échantillonnage réalisé souffrait d'un déficit dans le secteur de l'isthme entre la Basse-Terre et la Grande-Terre (voir figure 1). **Le premier objectif du présent projet était donc de combler ce déficit par l'échantillonnage** d'une vingtaine d'individus par espèce dans chacun des 6 sites forestiers complémentaires identifiés (voir figure 1). Même si les analyses de génétique du paysage n'avaient pas été complètement réalisées pour d'autres espèces d'oiseaux que la Paruline caféïette avant le présent projet, nous avons souhaité anticiper le même complément d'échantillonnage pour quelques autres

espèces pour lesquelles nous savons déjà qu'il existe aussi deux pools génétiques distincts sur la Grande-Terre et sur la Basse-Terre. Le complément d'échantillonnage vise donc les 6 espèces suivantes : Paruline caféïette, Grive à pieds jaunes, Saltator gros-bec, Moqueur grivotte, Sporophile rouge-gorge et Elénie siffleuse. **Le second objectif était d'inférer les axes par lesquels la connectivité entre la Basse-Terre et la Grande-Terre est éventuellement maintenue** pour ces différentes espèces.

### 3/ Résultats

*Conscients du caractère un peu technique de la description de certains résultats, en particulier les résultats des analyses génétiques, nous avons fait ressortir le message principal des différents résultats en gras et dans un encadré dans le texte ci-dessous. Néanmoins, nous avons choisi de présenter l'ensemble des résultats qui alimentent chaque message, dans un souci de rigueur scientifique.*

#### **3.1. Echantillonnage de sites forestiers à proximité de l'isthme**

L'échantillonnage des 6 espèces visées s'est déroulé début 2020. Sur les six fragments forestiers retenus initialement, un site (F189, Figure 1) n'a pas pu être échantillonné du tout en raison de l'impossibilité de trouver des accès permettant d'installer les filets de captures à l'intérieur du fragment forestier. En revanche, l'échantillonnage a pu être conduit comme prévu sur les cinq autres fragments : F170, F181, FX07, FX13, FX14 (Figure 1).

En termes d'effectifs capturés par espèce et par site d'échantillonnage, les résultats sont présentés dans le tableau 1 ci-dessous.

**Tableau 1.** Nombre d'individus capturés par espèce et par site d'échantillonnage. Tous ces individus ont été inclus dans les analyses moléculaires décrites dans la suite du document.

Site de capture	Latitude	Longitude	Dates (5 jours dans la fenêtre indiquée)	Elénie siffleuse	Grive à pieds jaunes	Moqueur grivotte	Paruline caféïette	Saltator gros-bec	Sporophile rouge-gorge
F170	16,20424996	-61,59134167	26/03 - 30/03	7	0	1	23	17	26
F181	16,27298489	-61,54472549	30/03 - 04/04	5	2	1	36	8	27
FX07	16,29924887	-61,64208953	06/04 - 10/04	1	0	0	36	8	24
FX13	16,2481076	-61,57427943	22/02 - 08/03	4	0	16	21	27	27
FX14	16,21518138	-61,48948167	25/05 - 29/05	50	0	33	18	22	27

*Note : Pour un même site de capture, 4 à 5 sessions de captures ont été réalisées (une session correspondant à une matinée). Une dizaine de filets ont été posés à chaque session, mais les filets n'ont jamais été posés deux fois au même endroit. Les coordonnées géographiques données pour chaque site correspondent donc au centre de la zone échantillonnée avec les filets, sachant que les filets ont été posés dans un rayon de 500 mètres autour de ce point.*



On constate que le succès des captures est très hétérogène entre espèces et entre sites. La Paruline caféïette, le Sporophile rouge-gorge, le Saltator gros-bec ont été capturés avec des effectifs suffisants pour les analyses génétiques dans les cinq fragments (même si les effectifs sont limités pour le Saltator gros-bec pour les sites F181 et FX07). En revanche, pour les autres espèces, le nombre d'individus capturés est insuffisant pour effectuer les analyses génétiques robustes dans certains fragments. Il s'agit des fragments F170, F181 et FX07 pour le Moqueur grivotte et F181, FX07 et FX13 pour l'Elénie siffleuse. Quant à la Grive à pieds jaunes, seuls deux individus ont été capturés sur le site F181. L'échantillonnage de ce projet ne permet donc pas d'apporter un complément d'information génétique dans ce secteur pour cette espèce par rapport aux données déjà acquises il y a quelques années dans le cadre du projet FRAG&BINV. Néanmoins, nous disposons de données génétiques pour la Grive à pieds jaunes pour le site F181 datant de ce projet (ce qui n'était pas le cas pour les autres espèces, d'où l'intégration de ce site d'échantillonnage dans le projet initial).

**Nous pouvons donc considérer que l'objectif 1 de ce projet est pleinement atteint pour trois espèces (Paruline caféïette, Saltator gros-bec, Sporophile rouge-gorge), partiellement atteint pour deux espèces (Moqueur grivotte et Elénie siffleuse), et pas du tout atteint pour la Grive à pieds jaunes.**

Même si ces données de capture doivent être interprétées avec prudence, elles suggèrent néanmoins que certaines espèces pourraient être localement absentes ou très rares (notamment la Grive à pieds jaunes), ce qui pourrait avoir un impact sur la connectivité des populations de la Basse-Terre et de la Grande-Terre, compte tenu de la position géographique de ces cinq sites d'échantillonnage.

## 3.2. *Analyses génétiques*

### 3.2.1. *Analyses moléculaires et production des données génétiques*

Les analyses des données génétiques décrites par la suite intègrent les données produites dans le cadre du présent projet (correspondant aux cinq nouveaux sites échantillonnés) et des données produites dans le cadre d'autres projets (FRAG&BINV et CANON). Globalement, le détail des protocoles de biologie moléculaire utilisés est décrit dans plusieurs publications déjà produites par notre équipe. Cette partie n'a donc vocation qu'à mentionner les éventuelles différences avec ces protocoles pour les échantillons collectés dans le cadre du présent projet.

Une des six espèces n'a pour l'heure pas pu être intégrée aux analyses génétiques. Il s'agit de l'Elénie siffleuse. En effet, le succès de l'échantillonnage de cette espèce dans le cadre du présent projet est très limité dans la mesure où un seul fragment (FX14, voir Tableau 1) a pu permettre de capturer plus de 10 individus. Par ailleurs, les échantillons disponibles grâce aux projets FRAG&BINV et CANON font apparaître un déficit global d'échantillonnage sur l'ensemble de la Basse-Terre. L'espèce y avait été capturée dans plusieurs sites d'échantillonnage, mais la plupart du temps avec des effectifs ne permettant pas d'avoir la puissance statistique suffisante pour les analyses envisagées. Enfin, nous avons pu déterminer à l'aide de points d'écoute que cette espèce était une espèce faiblement spécialiste de l'habitat forestier (Khimoun et al. 2016b), comme le sont le Sporophile rouge-gorge et le Saltator gros-bec. Les échantillons et données disponibles pour ces deux autres espèces étant suffisant pour conduire les analyses envisagées initialement, nous avons fait l'hypothèse que les résultats obtenus chez ces deux espèces reflèteraient la connectivité paysagère pour les espèces ayant un niveau de spécialisation à l'habitat forestier similaire.

Une prise de sang a été effectuée sur chacun des individus capturés sur le terrain et a ensuite été conservée dans 500 µL de tampon (Queen's Lysis Buffer) puis stockée à 4°C. L'ADN de ces échantillons a été extrait en utilisant un kit d'extraction d'ADN ('Blood Genomic DNA Mini-Preps Kit', BIO BASIC INC., Markham, Canada). Une large partie des données génétiques utilisées dans les analyses décrites par la suite provient d'extraction d'ADN au phénol-chloroforme (Khimoun et al.

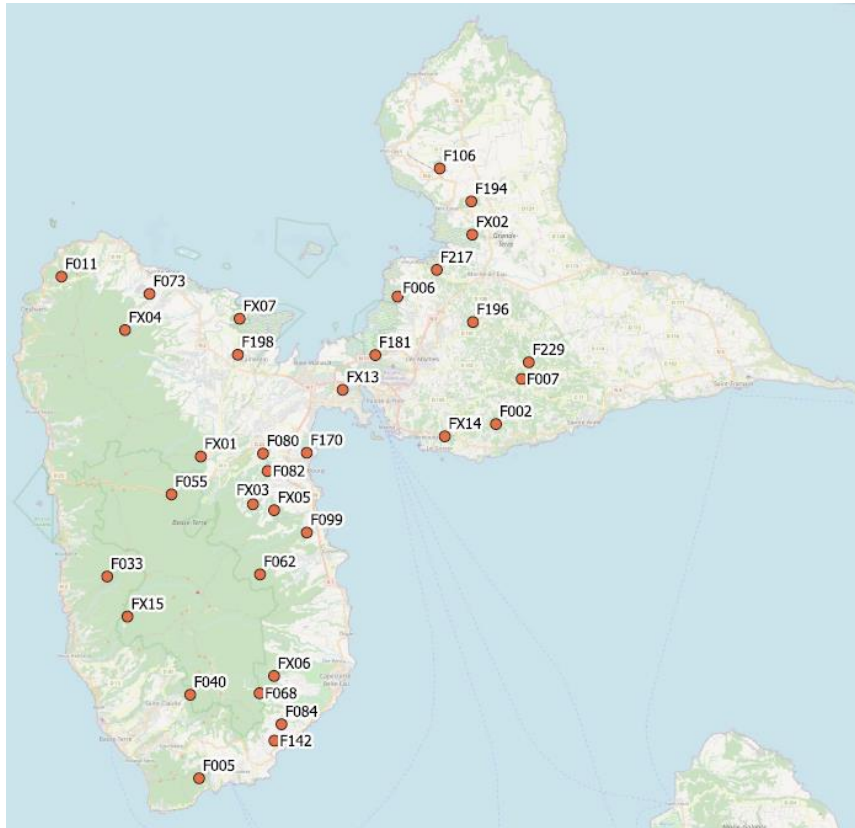
2016a). Néanmoins, si cette différence méthodologique est susceptible d'influencer la proportion de données manquantes dans le jeu de données génétiques (génotype difficile ou impossible à obtenir pour certains individus sur certains locus), elle ne peut pas affecter le génotype d'un individu, d'autant plus que des contrôles positifs dont le génotype est connu sont systématiquement inclus dans toutes les séries d'analyses moléculaires.

Les individus ont ensuite été génotypés à l'aide de locus microsatellites, dont le nombre et l'identité varient en fonction de l'espèce considérée (Tableau 2). La liste des locus utilisés par espèce ainsi que les conditions exactes de leur amplification par PCR sont décrites dans plusieurs articles précédemment publiés (Khimoun et al. 2016a, Khimoun et al. 2016b, Khimoun et al. 2017a, Khimoun et al. 2017b). Après amplification, les produits PCR ont été multiplexés avant de migrer sur un séquenceur automatique (ABI3730), et les génotypes des individus ont été obtenus après interprétation des chromatogrammes avec le logiciel Geneious (Kearse et al. 2012).

**Tableau 2.** Nombre d'individus, nombre de populations et nombre de locus microsatellites utilisés par espèce pour les analyses génétiques.

Espèces	Nombre d'individus génotypés	Nombre de populations	Nombre de locus microsatellites
Grive à pieds jaunes	503	22	10
Moqueur grivotte	322	17	10
Saltator gros-bec	563	24	14
Sporophile rouge-gorge	817	29	10
Paruline caféïette	712	27	12

Après le complément d'échantillonnage obtenu à travers le présent projet, nous disposons d'un jeu de données comptant un total de 322 à 817 individus par espèce, issus d'un nombre de fragments forestiers compris entre 17 et 29 selon l'espèce considérée (Tableau 2). C'est sur ce jeu de données que portent les analyses décrites dans la suite du document. Comme dans le tableau 2, il sera fait mention de "populations" dans la suite du document, une population désignant dans le cadre de ce projet l'ensemble des individus d'une espèce vivant dans un fragment forestier (ou un site d'échantillonnage, la totalité des sites d'échantillonnage étant forestiers). L'échantillon d'individus capturés dans chaque site et qui a ensuite été caractérisé à l'aide des locus microsatellites permet d'estimer les différents niveaux de diversité génétique de ces populations. La figure 2 décrit la localisation de l'ensemble des sites d'échantillonnage pris en compte dans ces analyses, sachant que des données génétiques ne sont pas disponibles pour toutes les espèces sur chacun de ces sites (par exemple parce qu'aucun individu (ou trop peu) n'a été capturé localement pour une espèce donnée).



**Figure 2.** Localisation des sites d'échantillonnage utilisés dans les analyses génétiques (toutes espèces confondues).

### 3.2.2. Analyses préliminaires

- Polymorphisme des locus microsatellites

Seuls les locus polymorphes à l'échelle de la zone d'étude peuvent être informatifs en termes de structure génétique des populations, autrement dit permettre d'estimer des niveaux de diversité génétique au sein des populations et des niveaux de différenciation (i.e. divergence) génétique entre populations, afin de les relier à la structure du paysage. Le nombre d'allèles observés par locus sur l'ensemble des individus a donc été calculé pour chaque espèce. En moyenne (sur l'ensemble des locus utilisés pour une même espèce), le nombre d'allèles observés varie de 4.4 pour le Moqueur grivotte à 14.9 pour la Paruline caféïette (Tableau 2). Le nombre d'allèles observés par locus varie de 2 à 25 (Tableau 3).

**Le niveau de polymorphisme de l'ensemble des locus microsatellites utilisés est satisfaisant pour les analyses de génétique des populations envisagées, même s'il semble un peu plus contraint pour le Moqueur grivotte que pour les quatre autres espèces.**

**Tableau 3.** Niveau de polymorphisme des marqueurs moléculaires utilisés

Espèces	Nombre moyen d'allèles	Nombre minimum et maximum d'allèles observés pour un locus
Grive à pieds jaunes	8,7	4 - 18
Moqueur grivotte	4,4	2 - 7
Saltator gros-bec	7,1	2 - 17
Sporophile rouge-gorge	9,9	3 - 18
Paruline caféïette	14,9	5 - 25

- Equilibre de Hardy-Weinberg

Afin de vérifier qu'il n'y a pas eu de problème d'amplification par PCR de certains allèles pour certains locus, il est d'usage de comparer la proportion observée d'individus hétérozygotes à cette même proportion attendue dans le cas de l'équilibre de Hardy-Weinberg. Un écart significatif entre ces deux proportions peut en effet provenir de la non amplification de certains allèles en raison de la présence de mutations au niveau des régions flanquantes des locus. Ces allèles bien réels mais non visibles par défaut d'amplification sont appelés des allèles nuls.

Afin de s'assurer de l'absence d'allèles nuls dans le jeu de données de chaque espèce, nous avons testé l'écart à l'équilibre d'Hardy-Weinberg pour chaque locus et dans chaque population, avec le logiciel GENEPOP (Rousset 2008). Les résultats sont synthétisés par espèce dans le Tableau 4.

**Tableau 4.** Résultats des tests de déviation par rapport à l'équilibre de Hardy-Weinberg, après correction de Bonferroni.

Espèce	Nombre de locus en déséquilibre	Sens du déséquilibre éventuel
Grive à Pieds Jaunes	0 / 220 tests	-
Moqueur Grivotte	1 / 170 tests	Déficit d'hétérozygotes
Saltator Gros Bec	0 / 336 tests	-
Sporophile Rouge Gorge	4 / 290 tests	Déficit d'hétérozygotes
Paruline Caféïette	2 / 320 tests	Déficit d'hétérozygotes

Les résultats montrent que sur l'ensemble des jeux de données, la très grande majorité des locus est à l'équilibre d'Hardy-Weinberg. Sept tests restent significatifs après correction de Bonferroni, et tous correspondent à un déficit d'hétérozygotes. Néanmoins, ces sept tests concernent un locus en déséquilibre dans une population uniquement. Dans le cas d'un problème sérieux d'amplification de certains allèles, on s'attend à ce que ce problème touche le même locus dans plusieurs populations (mais pas forcément toutes), ce qui n'est pas le cas ici. Nous considérons donc l'ensemble des locus pour chaque espèce dans la suite des analyses.

**Aucun biais d'amplification n'a été détecté pour les différents loci testés. Ils sont donc tous appropriés pour l'étude de la diversité génétique des espèces ciblées.**

- Déséquilibre de liaison

L'objectif ultime étant d'estimer les niveaux de diversité génétique et de décrire leur structuration spatiale, il est important d'utiliser plusieurs loci, c'est-à-dire plusieurs régions du génome des organismes étudiés. Néanmoins, il convient de s'assurer de l'indépendance statistique des données issues des différents locus avant de calculer les estimateurs de diversité génétique multilocus. Il s'agit de vérifier l'absence d'association préférentielle entre les allèles de deux locus, qui conduirait à des fréquences alléliques différentes de ce qui serait attendu par le jeu d'une association aléatoire. Nous avons donc testé les déséquilibres de liaisons entre toutes les paires de locus dans chaque population, en utilisant le logiciel GENEPOP (Rousset 2008). Les résultats sont synthétisés par espèce dans le Tableau 5.

**Tableau 5. Résultats des tests de déséquilibre de liaison, après correction de Bonferroni.**

Espèce	Nombre de paires de locus en déséquilibre
Grive à Pieds Jaunes	0 / 990 tests
Moqueur Grivotte	0 / 765 tests
Saltator Gros Bec	0 / 2184 tests
Sporophile Rouge Gorge	1 / 1305 tests
Paruline Caféïette	4 / 1782 tests

D'après les résultats, l'indépendance des loci est confirmée dans la quasi-totalité des tests réalisés. Le couple de locus en déséquilibre dans une population de Sporophile Rouge Gorge n'est pas inquiétant car ce couple ne ressort significativement lié dans aucun autre test. Pour la Paruline Caféïette, trois déséquilibres sur quatre surviennent dans une même population (F080). La lecture des génotypes pour cette population a été contrôlée une seconde fois, sans repérer d'anomalie. Le déséquilibre de liaison observé dans cette population pour trois paires de locus devra donc être pris en compte dans l'interprétation des résultats des analyses de structure génétique pour cette espèce.

**Pour chaque espèce, les loci microsatellites utilisés sont bien localisés dans des régions différentes du génome. Ils apportent des informations génétiques indépendantes et ils peuvent donc bien être utilisés en combinaison dans les analyses génétiques.**

- Contrôle de l'évolution temporelle de la diversité génétique

Les jeux de données totaux pour chaque espèce regroupent des données provenant de plusieurs campagnes d'échantillonnage réalisées en Guadeloupe à des dates différentes. En effet, la campagne de terrain de 2020 (présent projet) visait à réaliser un échantillonnage complémentaire à celui réalisé en 2012 et 2013, dans l'objectif d'améliorer la compréhension des patrons de structure génétique observés suite à la première campagne. Par ailleurs, le projet CANON, dont le travail de terrain a aussi eu lieu en 2020, a également permis de compléter cet échantillonnage.

Cependant, la réunion des données acquises en 2012-2013 avec celles de 2020 en un même jeu de données nécessite de vérifier en amont l'absence de différenciation temporelle au sein des populations dans cette période de temps. Dans cet objectif, entre trois et quatre populations échantillonnées en 2012-2013 ont été rééchantillonnées en 2020 pour chaque espèce. La différenciation génétique entre les individus capturés en 2012-2013 et en 2020 a ensuite été testée avec le logiciel GENEPOP (Rousset 2008) (Tableau 6).

**Tableau 6.** Résultats des tests de différenciation génétique dans le temps. Les tests significatifs sont indiqués avec un astérisque avec la probabilité du test.

Espèce	Population	Probabilité du test
Grive à Pieds Jaunes	F073	0.39
	F080	0.047 *
	F217	0.298
Moqueur Grivotte	F073	0.935
	F080	0.955
	F217	0.103
Saltator Gros Bec	F002	0.483
	F073	0.978
	F080	0.341
	F217	0.829
Sporophile Rouge Gorge	F002	0.497
	F080	0.338
	F217	0.664
Paruline Caféïette	F073	0.010 *
	F002	0.187
	F080	0.512
	F217	0.729

Les résultats montrent une absence de différenciation génétique entre 2012-2013 et 2020 pour toutes les populations de Saltator Gros Bec, de Sporophile Rouge Gorge et du Moqueur Grivotte. Les jeux de données de 2012-2013 et de 2020 peuvent donc être fusionnés pour ces espèces.

Pour les deux autres espèces, deux populations sur trois et trois populations sur quatre ne montrent aucune variation temporelle pour la Grive à pieds jaunes et la Paruline caféïette, respectivement. En revanche, on observe pour ces deux espèces qu'une population unique est significativement différenciée entre 2012-2013 et 2020, autrement dit que les fréquences alléliques semblent avoir changé au cours du temps (population F080 pour la Grive à pieds jaunes et population F073 pour la Paruline caféïette). Pour la Grive à pieds jaunes, le niveau de différenciation est extrêmement proche du seuil de significativité et ce résultat n'est pas alarmant, d'autant plus qu'il ne sera pas significatif après application d'une correction de Bonferroni pour les 3 populations de l'espèce. Néanmoins, une analyse complémentaire a été réalisée. Elle consiste en un test de corrélation visant à évaluer l'évolution du niveau de différenciation entre la population concernée et l'ensemble des autres populations entre 2012-2013 et 2020, dans le cas de ces deux espèces. Une corrélation élevée traduirait une certaine stabilité temporelle des niveaux de différenciation entre populations, ce qui permettrait de fusionner les données des différentes années.

Les résultats de cette analyse complémentaire montrent que la corrélation est significative et très élevée pour les deux espèces (Grive à pieds jaunes :  $r=0.94$ ,  $p=3.3 \times 10^{-10}$  ; Paruline caféïette :  $r=0.91$ ,  $p=6.2 \times 10^{-11}$ ). Autrement dit, la proximité génétique entre la population en cause dans le test de différenciation temporelle et les autres populations de l'espèce considérée est quasiment identique en 2012-2013 et 2020. Nous avons donc également fusionné les données des différentes années pour la Grive à pieds jaunes et la Paruline caféïette.

**Pour toutes les espèces étudiées, les jeux de données issus des échantillonnages de 2012-2013 et de 2020 peuvent être fusionnés car aucune différenciation génétique n'a été détectée à cette échelle temporaire. La combinaison de ces jeux de données permet une augmentation considérable de la puissance statistique des analyses et donc une étude plus fine des patrons de diversité génétique des espèces.**

### 3.2.3. Structure génétique des populations à l'échelle de la Guadeloupe et axes de connectivité

Pour caractériser la structure génétique des populations à l'échelle de la Guadeloupe, nous avons utilisé une analyse de partition (ou clustering) basée sur une approche bayésienne et implémentée dans le logiciel STRUCTURE (Pritchard et al. 2000). Schématiquement, cette analyse consiste à décrire la structure génétique des populations en estimant les fréquences alléliques d'un nombre de clusters (entités génétiquement différenciées) que l'on peut faire varier *a priori*, tout en assignant les individus (autrement dit les génotypes multilocus) échantillonnés à ces différents clusters de manière probabiliste (chaque individu a donc une information génétique qui a une certaine probabilité d'appartenir à tel ou tel cluster). En faisant varier le nombre K de clusters dans l'analyse, il est ainsi possible de déterminer sans *a priori* combien d'entités génétiques différentes sont représentées dans les données. Enfin, en regroupant les probabilités d'assignation des différents individus issues d'une même population (*i.e.* un même site d'échantillonnage), il est possible de caractériser le pool génétique de chaque population et ainsi d'identifier les populations qui partagent (ou pas) une partie de leur diversité génétique. Si deux populations échangent régulièrement des migrants et que ces migrants se reproduisent dans leur nouvelle population, autrement dit si les flux de gènes entre ces populations sont importants, alors ces deux populations doivent partager tout ou partie de leur information génétique. En revanche, en l'absence de flux de gènes entre populations, la dérive génétique qui se produit au sein de chacune conduit, génération après génération, à une divergence croissante des deux pools génétiques correspondant, voire *in fine* à deux pools génétiques complètement distincts (donc deux clusters tels que déterminés par le logiciel STRUCTURE).

Nous avons étudié la structure génétique des populations des cinq espèces considérées en comparant des partitions comprenant entre 1 et 10 clusters. La première étape de l'analyse est d'identifier le nombre de clusters induisant une partition optimale, c'est-à-dire celle qui confère aux données génétiques observées la "vraisemblance" (il s'agit en toute rigueur de la probabilité postérieure des données issue du modèle bayésien) la plus élevée. Dans la pratique, il s'agit de représenter l'augmentation de vraisemblance (autrement dit le gain d'information) associée à l'ajout d'un cluster supplémentaire dans le modèle, et de repérer pour quelle valeur de K (nombre de clusters) ce gain devient peu intéressant compte tenu de la complexification du modèle (Garnier et al. 2004). Par ailleurs, il est possible de moduler la valeur de K ainsi obtenue à partir des graphiques représentant la probabilité d'assignation du génotype de chaque individu aux différents clusters. Lorsqu'un cluster supplémentaire dans le modèle n'apporte aucune information, on observe que très peu (voire aucun) d'individus sont assignés à ce cluster avec une probabilité élevée. A noter que le nombre de clusters ainsi déterminé ne doit pas être interprété comme le nombre d'entités génétiques différenciées et contenant un pool génétique parfaitement homogène. Cette analyse de partition permet généralement de détecter les patrons de différenciation génétique assez tranchés. En d'autres termes, deux sites d'échantillonnage peuvent appartenir majoritairement au même cluster génétique et pourtant être significativement différenciés ; simplement leur niveau de divergence génétique est nettement inférieur au niveau de divergence entre clusters.

Une fois le nombre de clusters génétiques déterminé il est possible de représenter sur une carte les pools génétiques de chaque population échantillonnée en faisant apparaître leur composition par rapport aux différents clusters. Cette visualisation cartographique permet non seulement d'identifier les populations qui partagent (ou pas) des flux de gènes (et donc d'individus), mais aussi de repérer les éventuelles barrières à la dispersion ou au contraire les éventuels corridors de dispersion.

Il est important de noter que la manière dont les clusters supplémentaires dans le modèle se répartissent spatialement apporte une information supplémentaire, et parfois plus intéressante, dans la mesure où les clusters sont généralement établis de manière hiérarchique (Garnier et al. 2004). Par exemple, un des clusters identifiés pour une valeur de K = 2 peut être subdivisé en deux clusters pour une valeur de K = 3, de telle sorte que ces deux clusters constituent une subdivision d'un des clusters pour le modèle avec K = 2. Il est également possible que le passage d'un modèle à K = 2 vers un modèle à K = 3 génère un cluster supplémentaire à l'interface des deux clusters du modèle à K = 2, suggérant que la transition entre les pools génétiques des deux clusters initiaux est graduelle dans l'espace.

Les résultats de ces analyses sont d'abord décrits espèce par espèce. Une synthèse des résultats obtenus chez les différentes espèces est ensuite présentée par rapport aux axes potentiels de connectivité des populations de la Basse-Terre et de la Grande-Terre.

- **Grive à pieds jaunes**

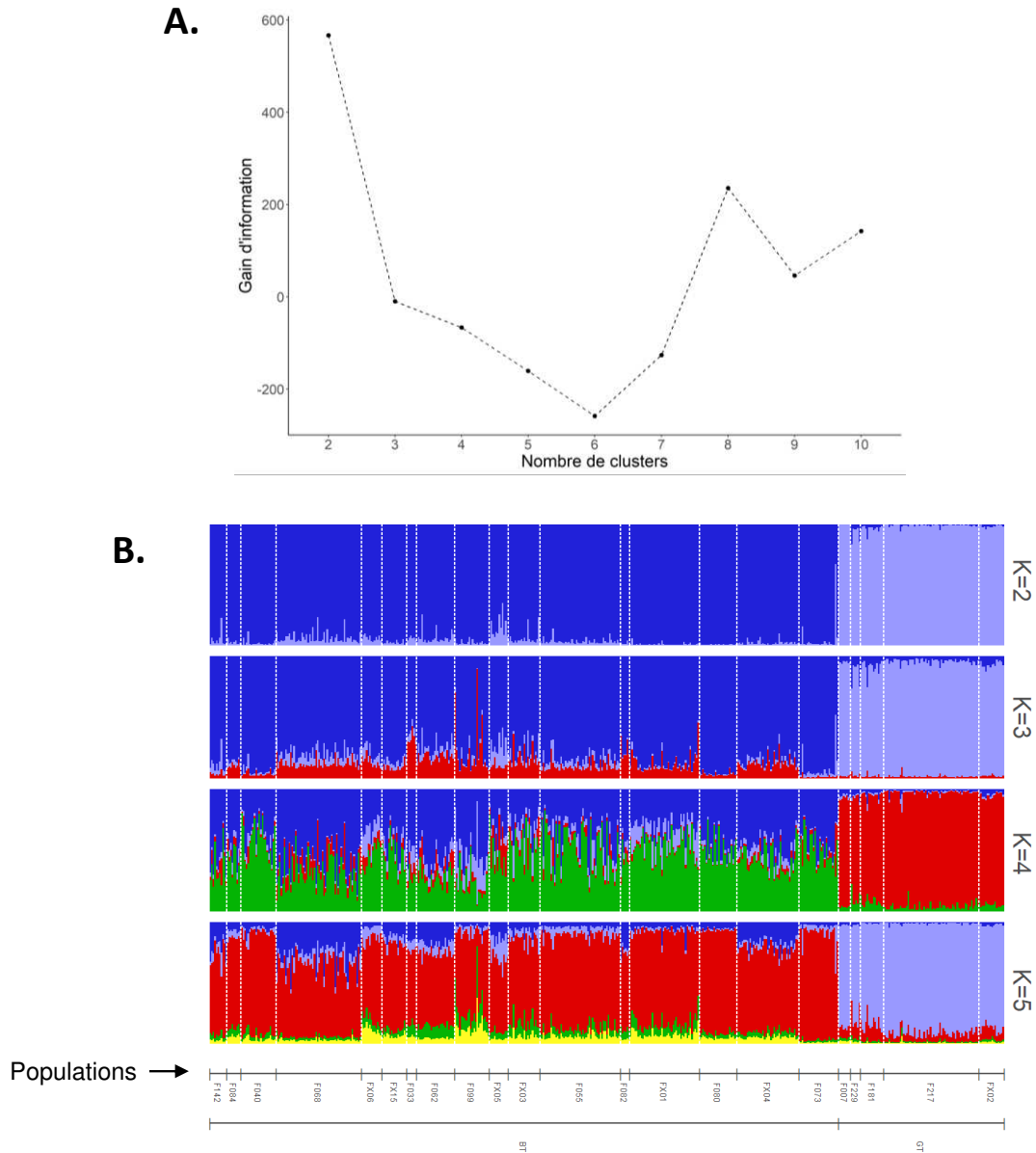
Le gain d'information dans le modèle en passant de 1 à 2 clusters est conséquent, alors que ce gain est nul voire négatif entre  $K = 2$  et  $K = 3$  (Figure 3A). Par ailleurs, on voit que l'ajout du 3<sup>e</sup> cluster dans le modèle n'apporte aucune information supplémentaire dans la mesure où quasiment aucun individu ne semble assigné à ce cluster (cluster "rouge" dans la figure 3B pour  $K = 3$ ). Enfin, lorsque l'information est synthétisée par population et spatialisée (Figure 4), les 3<sup>e</sup> et 4<sup>e</sup> clusters ne semblent correspondre à aucune logique géographique supplémentaire.

Ainsi, les populations de Grive à pieds jaunes sont très clairement isolées génétiquement entre la Basse-Terre et la Grande-Terre. Nous avons déjà mis en évidence une forte différenciation génétique entre les populations de ces deux parties de la Guadeloupe, mais avec un échantillonnage plus réduit (Arnoux et al. 2014). Nous n'avons par exemple pas d'échantillons (et donc pas de données génétiques) issus du secteur des Grands Fonds. L'échantillonnage présent permet donc de confirmer cette structuration à l'échelle de la Basse-Terre et de la Grande-Terre.

Comme mentionné plus haut, seuls deux individus de Grive à pieds jaunes ont été capturés dans le secteur de l'isthme connectant la Basse-Terre et la Grande-Terre (Tableau 1). L'absence de données génétiques dans les sites d'échantillonnage visés par le présent projet (à l'exception du fragment F181) empêche ainsi de déterminer la manière dont la transition se fait spatialement entre les deux pools génétiques. L'approche analytique utilisée ne permet donc pas d'identifier les axes de connectivité dans ce secteur. Cependant, le très faible nombre de captures réalisées dans le secteur de l'isthme et l'absence de contact sonore avec l'espèce dans certains sites (notamment le FX13) suggère que la Grive à pieds jaunes pourrait être très peu présente dans le secteur, limitant considérablement la connectivité entre les populations de la Basse-Terre et la Grande-Terre, ce qui pourrait expliquer les niveaux de différenciation génétique très élevés entre ces deux ensembles de populations (Arnoux et al. 2014).

**Les populations de Grive à pieds jaunes de la Basse-Terre sont très clairement différentes d'un point de vue génétique de celles de la Grande-Terre. Les faibles succès de capture dans les sites de la région de l'isthme suggèrent des effectifs considérablement réduits dans cette zone, limitant ainsi la connectivité entre les deux parties de l'île.**

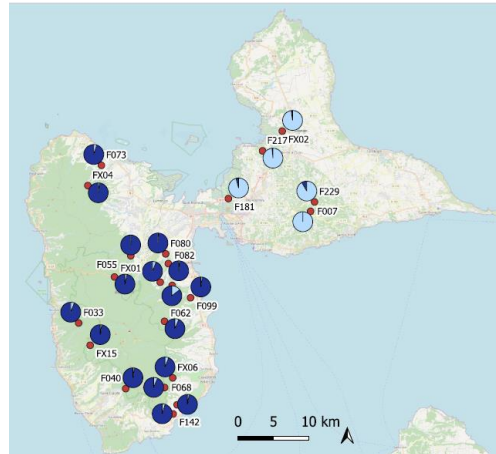




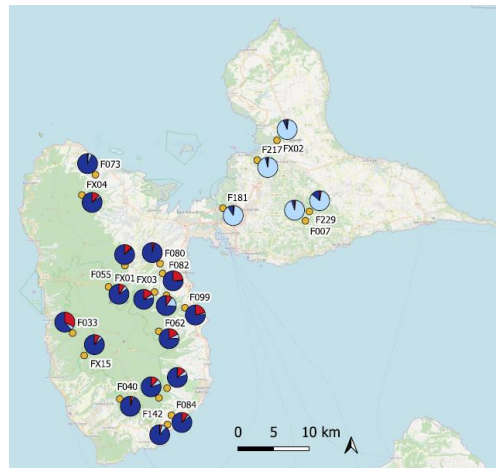
**Figure 3.** Résultats de l'analyse de partition pour la *Grive à pieds jaunes*. **A.** Gain d'information en termes de structure génétique des populations en fonction du nombre de cluster dans le modèle. Pour un modèle avec  $K$  clusters, ce gain correspond à l'augmentation de la probabilité postérieure des données observées entre les modèles à  $K-1$  et  $K$  clusters. **B.** Assignment des individus aux différents clusters du modèle. Chaque individu est représenté par une barre verticale partitionnée en  $K$  segments colorés ( $K$  étant le nombre de cluster dans le modèle), dont la longueur est proportionnelle à la probabilité d'assignation de l'individu à chaque cluster (cette probabilité pouvant être interprétée comme la proportion du génome de l'individu issue de chaque cluster). Les individus des différentes populations sont séparés par une barre en pointillés blancs. Les codes populations sont indiqués sous la figure et sont approximativement ordonnées du sud vers le nord pour la Basse-Terre (BT) et du Grand Cul-de-Sac Marin vers les Grands Fonds pour la Grande-Terre (GT). L'assignation n'est pas illustrée au-delà de  $K=5$  car non informative.

Il n'y a pas nécessairement de correspondance entre les couleurs des différents modèles (différentes valeurs de  $K$ ). Autrement dit, le cluster bleu foncé (par exemple) pour un modèle (i.e. une valeur de  $K$ ) n'est pas nécessairement le même que le cluster bleu foncé pour un autre modèle (i.e. une autre valeur de  $K$ ).

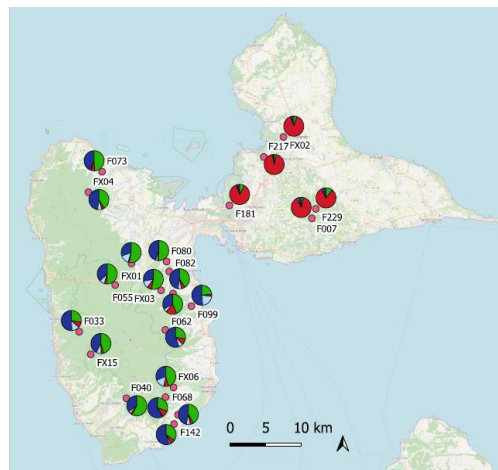
**K = 2**



**K = 3**



**K = 4**



**Figure 4.** Représentation spatiale des clusters et de la composition des pools génétiques des différentes populations de *Grive à pieds jaunes* identifiés avec l'algorithme STRUCTURE. À côté de chaque population figure un diagramme circulaire représentant la part du pool génétique de cette population issue de chacun des  $K$  clusters. Cette spatialisation des clusters n'est pas illustrée au-delà de  $K=4$  car non informative.

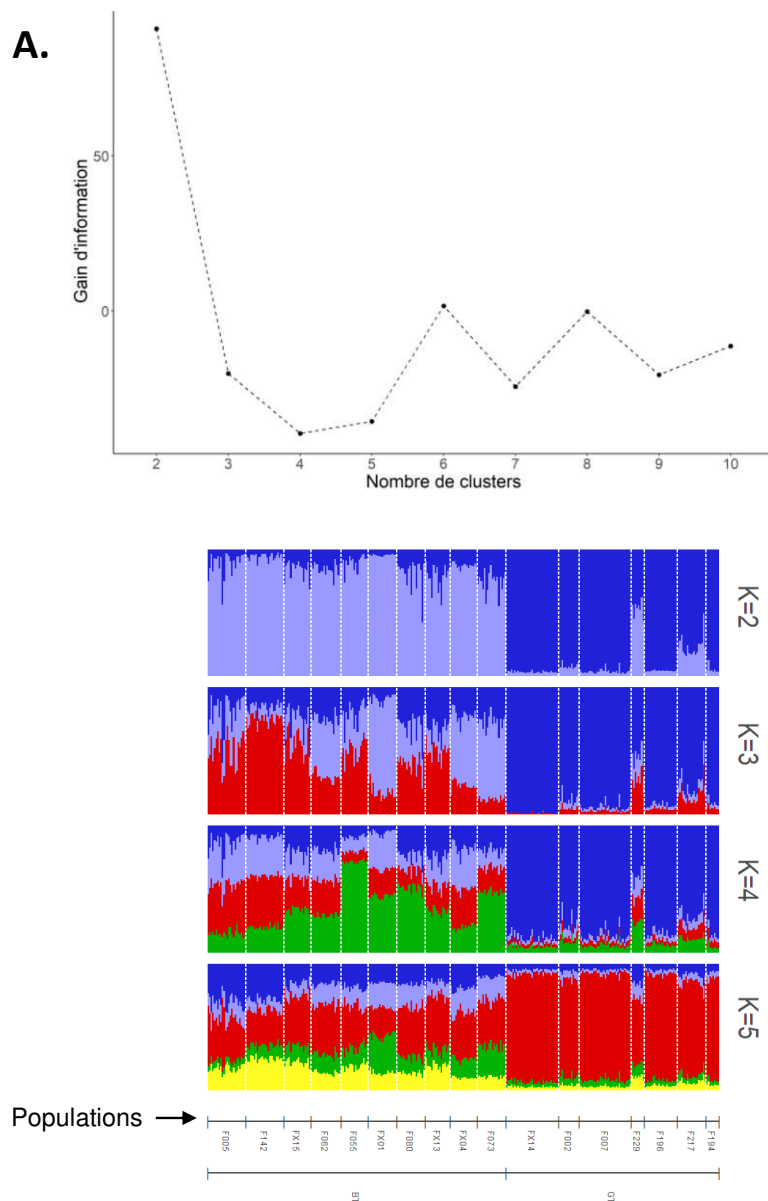
Il n'y a pas nécessairement de correspondance entre les couleurs des différents modèles (différentes valeurs de  $K$ ). Autrement dit, le cluster bleu foncé (par exemple) pour un modèle (i.e. une valeur de  $K$ ) n'est pas nécessairement le même que le cluster bleu foncé pour un autre modèle (i.e. une autre valeur de  $K$ ).

- **Moqueur grivotte**

Si l'ajout d'un second cluster dans le modèle apporte une information importante en termes de structure génétique des populations, l'ajout du troisième cluster entraîne un gain négatif, autrement dit une baisse de vraisemblance des données (Figure 5A). Les résultats d'assignation des individus aux différents clusters montrent que la plupart des individus sont assignés à un des deux clusters du modèle  $K = 2$  avec une probabilité proche de 1. En revanche, si aucun individu ne semble assigné avec une quasi-certitude au 3<sup>e</sup> cluster du modèle  $K = 3$ , la majorité des individus de certaines populations (FX01, FX04, F062, F073) semblent avoir un génotype issu principalement de ce 3<sup>e</sup> cluster (cluster "bleu clair" pour  $K=3$ , figure 5B). Il existe donc une divergence forte entre deux ensembles de populations, mais un troisième cluster semble pouvoir se dessiner. En effet, lorsque l'on regarde la manière dont ces clusters sont spatialisés (Figure 6), on constate que pour  $K = 2$ , on identifie un pool génétique sur la Basse-Terre et un sur la Grande-Terre. Lorsqu'un 3<sup>e</sup> cluster est ajouté au modèle ( $K=3$ ), on distingue les populations du nord de la Basse-Terre (dont le pool génétique est majoritairement issu du cluster "bleu clair") des populations du sud de la Basse-Terre (dont le pool génétique est majoritairement issu du cluster "rouge"). Le 3<sup>e</sup> cluster du modèle semble donc permettre de faire apparaître une certaine différenciation entre le nord et le sud de la Basse-Terre (clusters "bleu clair" et "rouge", respectivement, pour  $K=3$ ), même si la transition entre ces 2 clusters est sans doute graduelle dans l'espace et ne correspond pas à une barrière à la dispersion particulière. Le 4<sup>e</sup> cluster ne semble lui apporter qu'une information très limitée.

**En termes de connectivité, la synthèse des inférences pour le Moqueur grivotte est présentée figure 7. Il est difficile avec les données disponibles de déterminer la manière dont les flux de gènes circulent au niveau de l'isthme puisque les sites d'échantillonnage du secteur n'ont pas tous permis de collecter les échantillons nécessaires (un seul individu capturé pour les sites F170 et F181 et aucun individu capturé pour le site FX07 ; tableau 1). En particulier, les résultats obtenus ne permettent ni d'exclure ni de soutenir fortement un des deux axes de connectivité entre l'isthme et les populations de la Basse-Terre. Néanmoins, il semble que la population FX13 partage une partie plus importante de son pool génétique avec les populations du sud de la Basse-Terre qu'avec celles du nord, ce qui pourrait plutôt supporter la partie ouest de l'axe 2 que la partie ouest de l'axe 1 (voir Figure 1). Cette inférence reste néanmoins une hypothèse à confirmer pour cette espèce.**

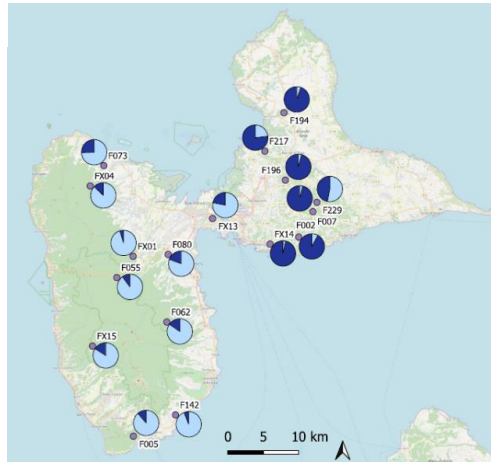
**En revanche, les analyses permettent d'exclure une connexion directe entre l'isthme et le sud des Grands Fonds. Les résultats montrent en revanche une certaine continuité génétique entre les populations des Grands Fonds et les populations situées le long du Grand Cul-de-Sac Marin entre Morne-à-l'Eau et Petit-Canal. Il est donc très probable que la connexion avec les populations de la Basse-Terre passe par les mangroves et forêts marécageuses le long du Grand Cul-de-Sac Marin.**



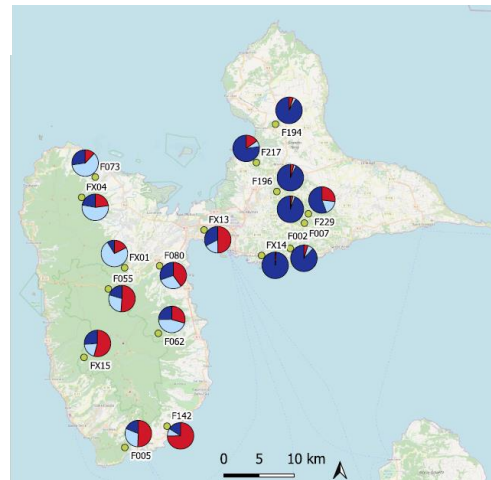
**Figure 5.** Résultats de l'analyse de partition pour le *Moqueur grivotte*. **A.** Gain d'information en termes de structure génétique des populations en fonction du nombre de cluster dans le modèle. Pour un modèle avec  $K$  clusters, ce gain correspond à l'augmentation de la probabilité postérieure des données observées entre les modèles à  $K-1$  et  $K$  clusters. **B.** Assignment des individus aux différents clusters du modèle. Chaque individu est représenté par une barre verticale partitionnée en  $K$  segments colorés ( $K$  étant le nombre de cluster dans le modèle), dont la longueur est proportionnelle à la probabilité d'assignation de l'individu à chaque cluster (cette probabilité pouvant être interprétée comme la proportion du génome de l'individu issue de chaque cluster). Les individus des différentes populations sont séparés par une barre en pointillés blancs. Les codes populations sont indiqués sous la figure et sont approximativement ordonnés du sud vers le nord pour la Basse-Terre (BT) et du Grand Cul-de-Sac Marin vers les Grands Fonds pour la Grande-Terre (GT). L'assignation n'est pas illustrée au-delà de  $K=5$  car non informative.

Il n'y a pas nécessairement de correspondance entre les couleurs des différents modèles (différentes valeurs de  $K$ ). Autrement dit, le cluster bleu foncé (par exemple) pour un modèle (i.e. une valeur de  $K$ ) n'est pas nécessairement le même que le cluster bleu foncé pour un autre modèle (i.e. une autre valeur de  $K$ ).

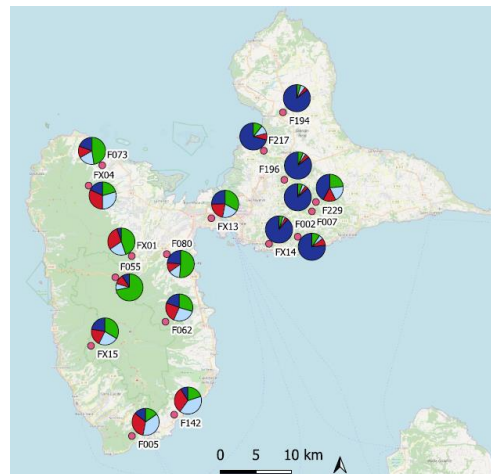
**K = 2**



**K = 3**

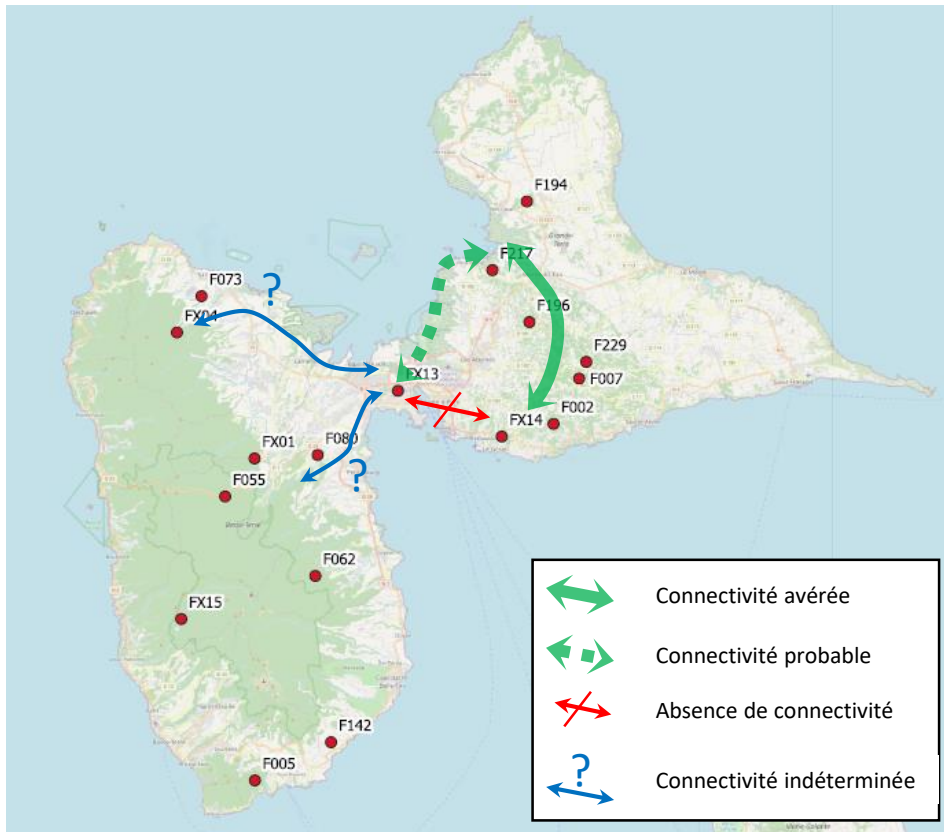


**K = 4**



**Figure 6.** Représentation spatiale des clusters et de la composition des pools génétiques des différentes populations de *Moqueur grivotte* identifiés avec l'algorithme STRUCTURE. A côté de chaque population figure un diagramme circulaire représentant la part du pool génétique de cette population issue de chacun des K clusters. Cette spatialisation des clusters n'est pas illustrée au-delà de K=4 car non informative.

Il n'y a pas nécessairement de correspondance entre les couleurs des différents modèles (différentes valeurs de K). Autrement dit, le cluster bleu foncé (par exemple) pour un modèle (i.e. une valeur de K) n'est pas nécessairement le même que le cluster bleu foncé pour un autre modèle (i.e. une autre valeur de K).

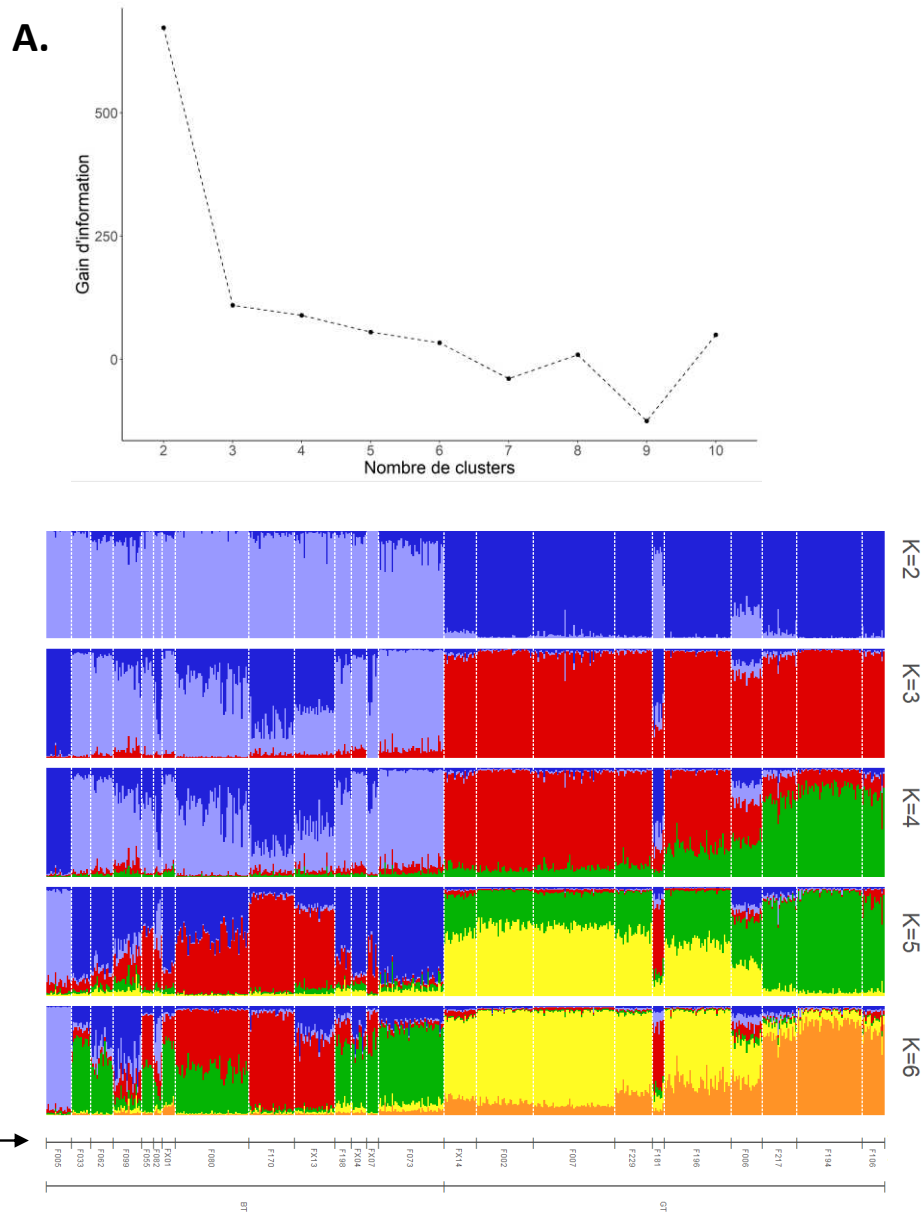


**Figure 7.** Représentation schématique de la connectivité des populations de *Moqueur grivotte* entre la Basse-Terre et la Grande-Terre.

- **Saltator gros-bec**

Le gain d'information avec l'ajout d'un second cluster dans le modèle est bien plus élevé que le gain d'information associé à l'ajout d'un 3<sup>e</sup>, 4<sup>e</sup>, ou 5<sup>e</sup> cluster dans le modèle (Figure 8A). Il y a donc un premier niveau de structuration génétique marquée avec deux pools génétiques au sein de la Guadeloupe. Néanmoins, l'ajout des 3<sup>e</sup>, 4<sup>e</sup>, 5<sup>e</sup> et même 6<sup>e</sup> clusters semble apporter un gain d'information en termes de structuration génétique. Il semble donc y avoir une structuration génétique plus fine, même si d'un niveau inférieur, au sein de la Guadeloupe. L'examen du schéma d'assignation des différents modèles (correspondant aux différentes valeurs de K) amène à la même interprétation (Figure 8B). L'examen de la distribution spatiale des clusters génétiques fait apparaître une logique géographique assez claire (Figure 9). Les deux premiers clusters identifiés traduisent une différenciation forte entre les populations de la Basse-Terre et de la Grande-Terre. Le 3<sup>e</sup> cluster (cluster "bleu foncé" pour K=3) regroupe des populations proches de l'isthme et, de manière *a priori* surprenante, la population F005 au sud de la Basse-Terre. L'ajout du 4<sup>e</sup> cluster permet de distinguer les populations des Grands Fonds (cluster "rouge" pour K=4) de celles situées au nord-ouest de la Grande-Terre, le long du Grand Cul-de-Sac Marin (cluster "vert" pour K=4). Enfin, le 5<sup>e</sup> cluster permet d'individualiser la population F005 (cluster "bleu clair" pour K=5) dans les Monts Caraïbes. Cette population semble posséder un pool génétique distinct de toutes les autres, ce qui peut expliquer qu'elle ait été "temporairement" groupée avec le 3<sup>e</sup> cluster. Le 6<sup>e</sup> cluster (cluster "bleu foncé" pour K=6) ne semble lui apporter qu'une information très limitée.

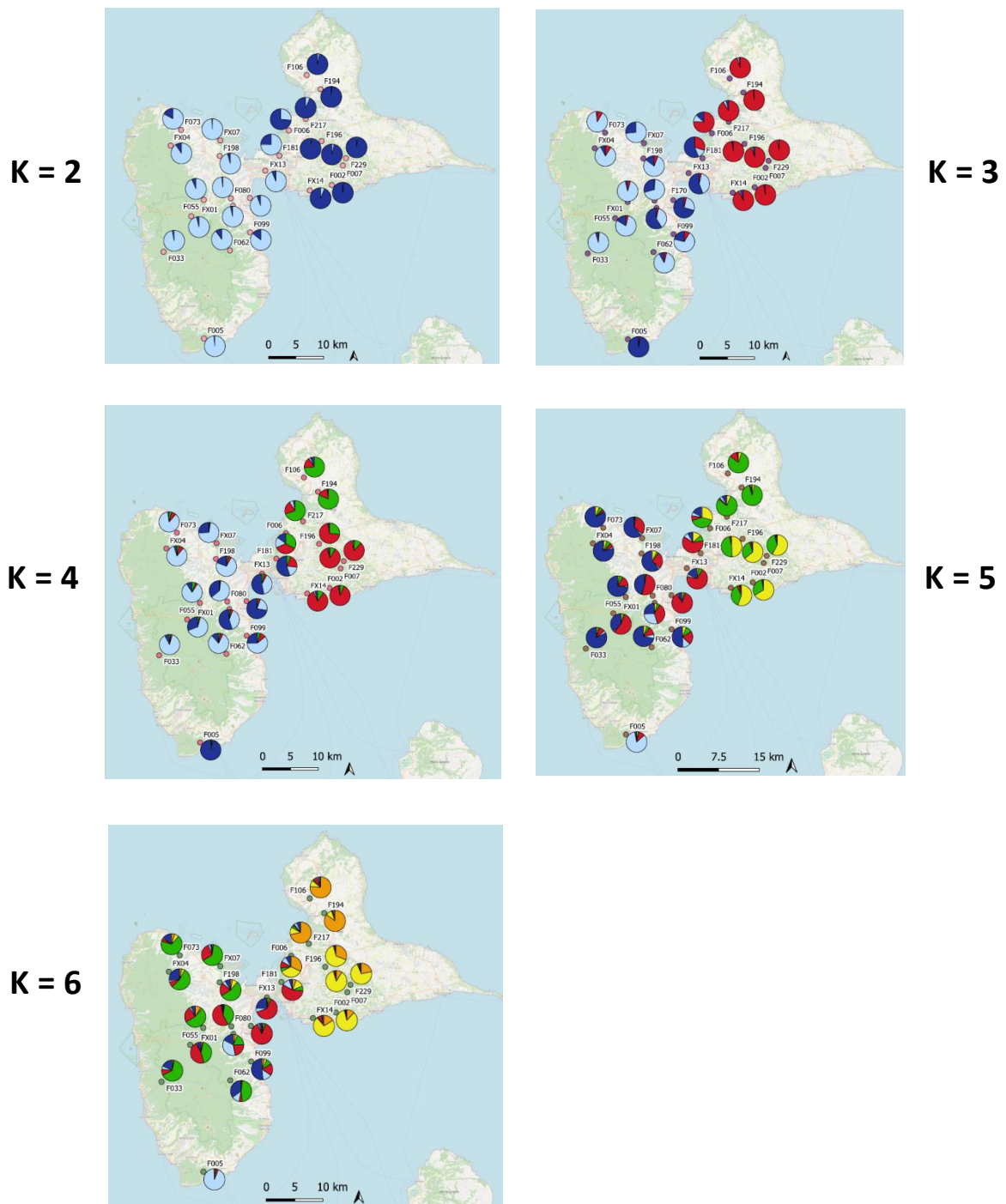
**En matière de connectivité, ces résultats permettent de faire les inférences décrites schématiquement dans la figure 10. Les analyses permettent d'exclure une connexion directe entre l'isthme et le sud des Grands Fonds. Les résultats montrent en revanche une certaine continuité génétique entre les populations des Grands Fonds et les populations situées le long du Grand Cul-de-Sac Marin jusqu'au nord de Petit-Canal (jusqu'au lac de Gaschet d'après nos échantillons, mais peut-être un peu plus loin encore). On voit également que flux de gènes et donc la dispersion des individus se font de proche en proche le long du Grand Cul-de-Sac Marin sur la Grande-Terre, depuis le nord de Petit-Canal jusqu'à l'isthme et le secteur de Jarry, puis sur la Basse-Terre en passant par le nord de la Pointe à Bacchus (à l'endroit où la Lézarde rejoint le Petit-Cul-de-Sac-Marin) avant de rejoindre le grand massif forestier de la Basse-Terre via le secteur de Fougère à l'est de Petit-Bourg. Enfin, si la connexion du secteur de Jarry avec les populations du grand massif forestier de la Basse-Terre semble se faire plutôt en direction du sud-ouest, nos résultats ne permettent ni d'exclure ni de valider l'existence d'une certaine connectivité (*a priori* plus faible) avec les mangroves et forêts marécageuses situées le long du Grand Cul-de-Sac Marin sur la Basse-Terre (vers le nord-est, donc).**



**Figure 8.** Résultats de l'analyse de partition pour le *Saltator gros-bec*. **A.** Gain d'information en termes de structure génétique des populations en fonction du nombre de cluster dans le modèle. Pour un modèle avec  $K$  clusters, ce gain correspond à l'augmentation de la probabilité postérieure des données observées entre les modèles à  $K-1$  et  $K$  clusters. **B.** Assignment des individus aux différents clusters du modèle. Chaque individu est représenté par une barre verticale partitionnée en  $K$  segments colorés ( $K$  étant le nombre de cluster dans le modèle), dont la longueur est proportionnelle à la probabilité d'assignation de l'individu à chaque cluster (cette probabilité pouvant être interprétée comme la proportion du génome de l'individu issue de chaque cluster). Les individus des différentes populations sont séparés par une barre en pointillés blancs. Les codes populations sont indiqués sous la figure et sont approximativement ordonnés du sud vers le nord pour la Basse-Terre (BT) et du Grand Cul-de-Sac Marin vers les Grands Fonds pour la Grande-Terre (GT). L'assignation n'est pas illustrée au-delà de  $K=6$  car non informative.

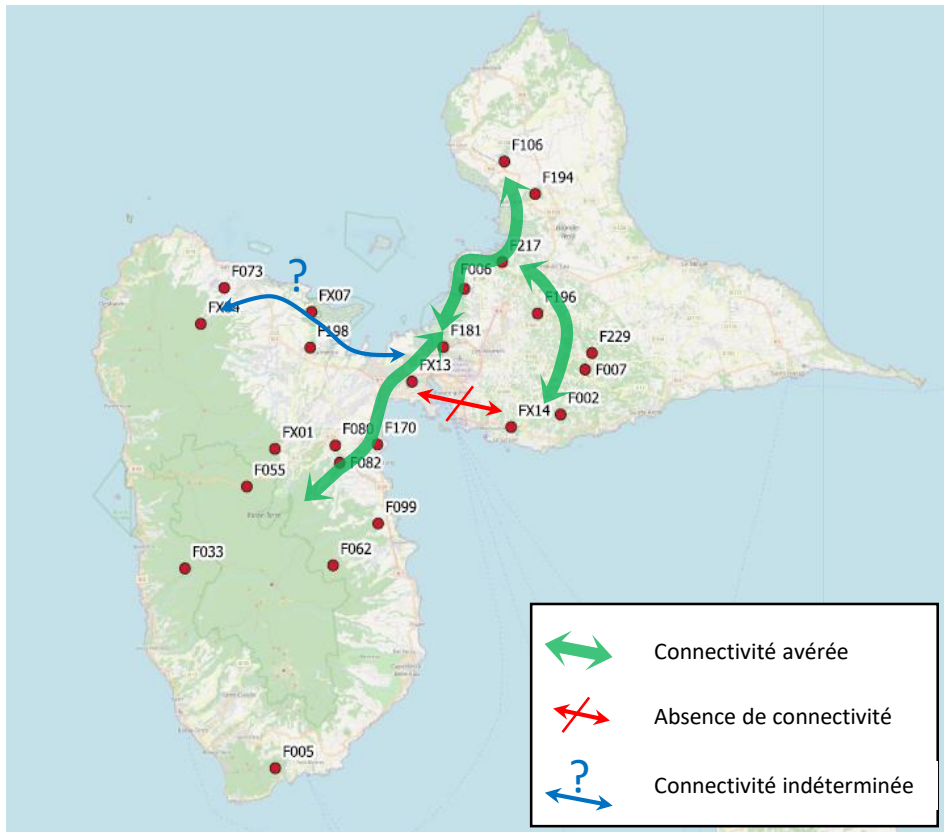
Il n'y a pas nécessairement de correspondance entre les couleurs des différents modèles (différentes valeurs de  $K$ ). Autrement dit, le cluster bleu foncé (par exemple) pour un modèle (i.e. une valeur de  $K$ ) n'est pas nécessairement le même que le cluster bleu foncé pour un autre modèle (i.e. une autre valeur de  $K$ ).





**Figure 9.** Représentation spatiale des clusters et de la composition des pools génétiques des différentes populations de *Saltator gros-bec* identifiés avec l'algorithme STRUCTURE. A côté de chaque population figure un diagramme circulaire représentant la part du pool génétique de cette population issue de chacun des K clusters. Cette spatialisation des clusters n'est pas illustrée au-delà de K=6 car non informative.

Il n'y a pas nécessairement de correspondance entre les couleurs des différents modèles (différentes valeurs de K). Autrement dit, le cluster bleu foncé (par exemple) pour un modèle (i.e. une valeur de K) n'est pas nécessairement le même que le cluster bleu foncé pour un autre modèle (i.e. une autre valeur de K).

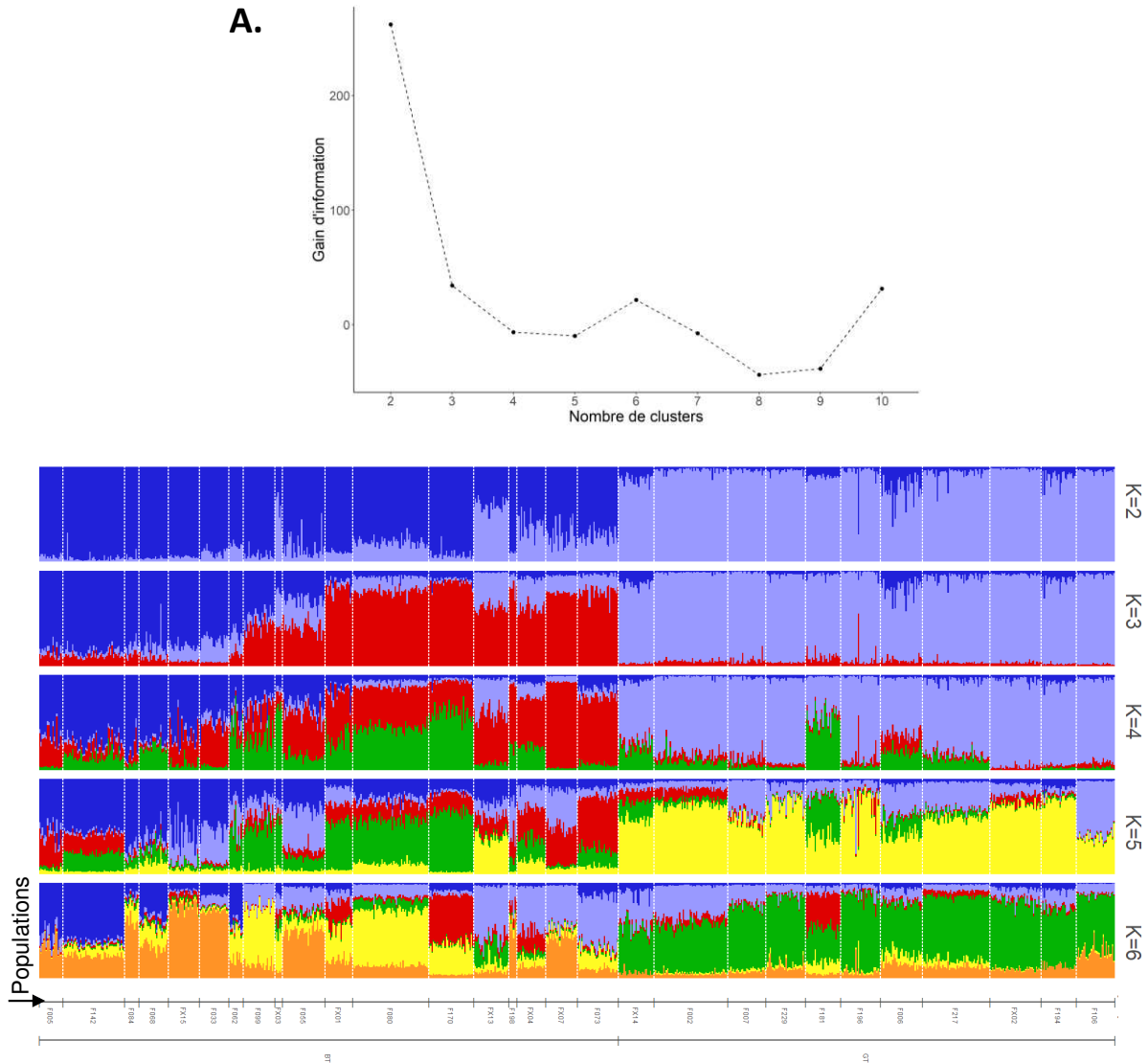


**Figure 10.** Représentation schématique de la connectivité des populations de *Saltator gros-bec* entre la Basse-Terre et la Grande-Terre.

- **Sporophile rouge-gorge**

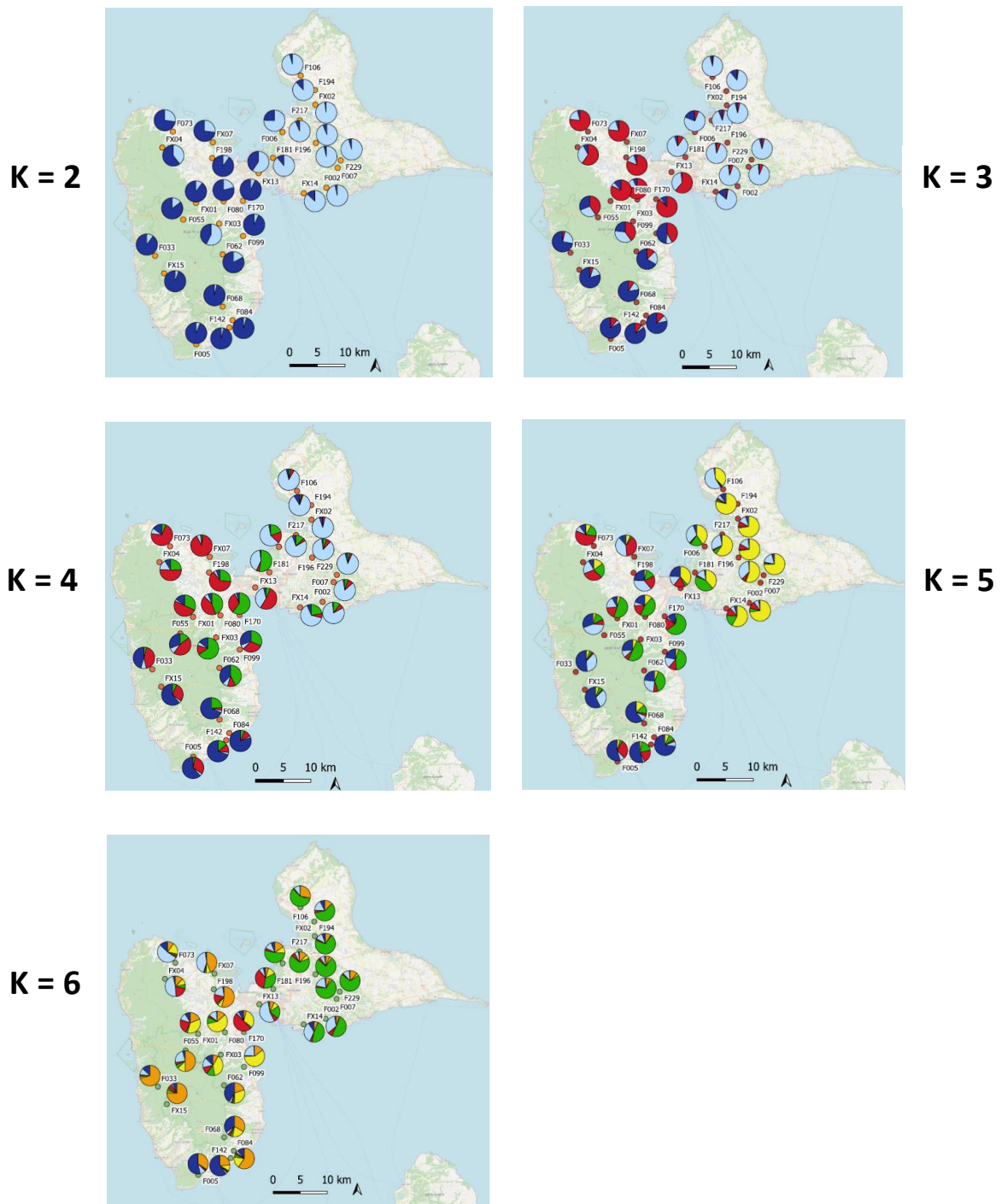
Si l'ajout d'un second cluster dans le modèle apporte une information importante en termes de structure génétique des populations, l'ajout du troisième cluster entraîne un gain beaucoup plus faible, bien que positif (Figure 11A). Il y a donc un premier niveau de structuration génétique marquée avec deux pools génétiques au sein de la Guadeloupe. Néanmoins, l'ajout du 3<sup>e</sup> cluster semble apporter une information complémentaire, bien que correspondant à un niveau de différenciation génétique bien plus faible. Les diagrammes d'assignation des individus dans les différents modèles (*i.e.* pour les différentes valeurs de K) montrent sensiblement la même chose (Figure 11B). En effet, dans le modèle à 3 clusters, les individus de plusieurs populations sont assignés très majoritairement à un même cluster, et ceci est observé pour chacun des 3 clusters de ce modèle. Dans le modèle à 4 clusters, seules deux populations (FX03 et F170) semblent posséder un pool génétique principalement issu du 4<sup>e</sup> cluster (cluster "vert", figure 11B), même si les probabilités d'assignation à ce cluster sont voisines de 60%. Lorsque l'on examine la répartition spatiale des clusters (Figure 12), on remarque que le premier niveau de structuration correspond à une différenciation marquée entre la Basse-Terre et la Grande-Terre. L'ajout d'un troisième cluster dans le modèle permet de faire apparaître une différenciation entre les populations du nord (cluster "rouge" pour K=3) et celles du sud de la Basse-Terre (cluster "bleu foncé" pour K=3). Le quatrième cluster ajouté au modèle (cluster "vert" pour K=4) regroupe les populations de l'est de la Basse-Terre et intègre également la population F181 située au niveau du Pont de l'Alliance sur la Grande-Terre. Le 5<sup>e</sup> cluster (cluster "bleu clair" pour K=5) ne semble correspondre à aucune logique géographique particulière.

**Concernant les axes de connectivité des populations (Figure 13), les résultats montrent une certaine continuité génétique, et donc une certaine connectivité entre les populations des Grands Fonds et celles situées le long du Grand Cul-de-Sac Marin sur la Grande-Terre, depuis le nord de Petit-Canal (jusqu'au lac de Gaschet d'après nos échantillons, mais peut-être un peu plus loin encore) jusqu'à l'isthme et le secteur de Jarry (population FX13). Par ailleurs, il ne semble pas y avoir de connexion directe importante entre l'isthme (population FX13) et le sud des grands Fonds (population FX14), car le pool génétique de la population FX13 ne semble pas contenir d'information qui serait partagée avec la population FX14 mais ne serait pas partagée avec les autres populations des grands Fonds ou celles situées le long du Grand Cul-de-Sac Marin sur la Grande-Terre. Néanmoins, cette connexion ne peut pas être définitivement exclue sur la base des résultats présentés, car lorsque le modèle contient 6 clusters (Figure 12), une partie du pool génétique des populations FX13 et FX14 semble provenir du cluster "bleu clair" (même si nous avons vu précédemment qu'au-delà du 4<sup>e</sup> cluster, les clusters ajoutés au modèle n'apportaient que très peu d'information). D'autres analyses devront être conduites pour pouvoir trancher définitivement. Enfin, les résultats obtenus ne permettent pas de distinguer la partie ouest des deux axes 1 et 2 (Figure 1), autrement dit la manière dont les flux de gènes se font entre les populations de la Basse-Terre et l'isthme connectant les deux parties de la Guadeloupe. Ces résultats montrent en effet un certain partage du pool génétique des populations FX13 et F181 (secteur de l'isthme) et des populations situées sur la partie ouest des deux axes 1 (e.g. F198 et FX07, avec notamment le clusters "bleu clair" pour K=5) et 2 (e.g. F170 et F080, avec notamment le clusters "vert" pour K=5) de la Figure 1. Il paraît donc pertinent de considérer ces deux axes comme probables à la lumière des résultats présentés ici.**



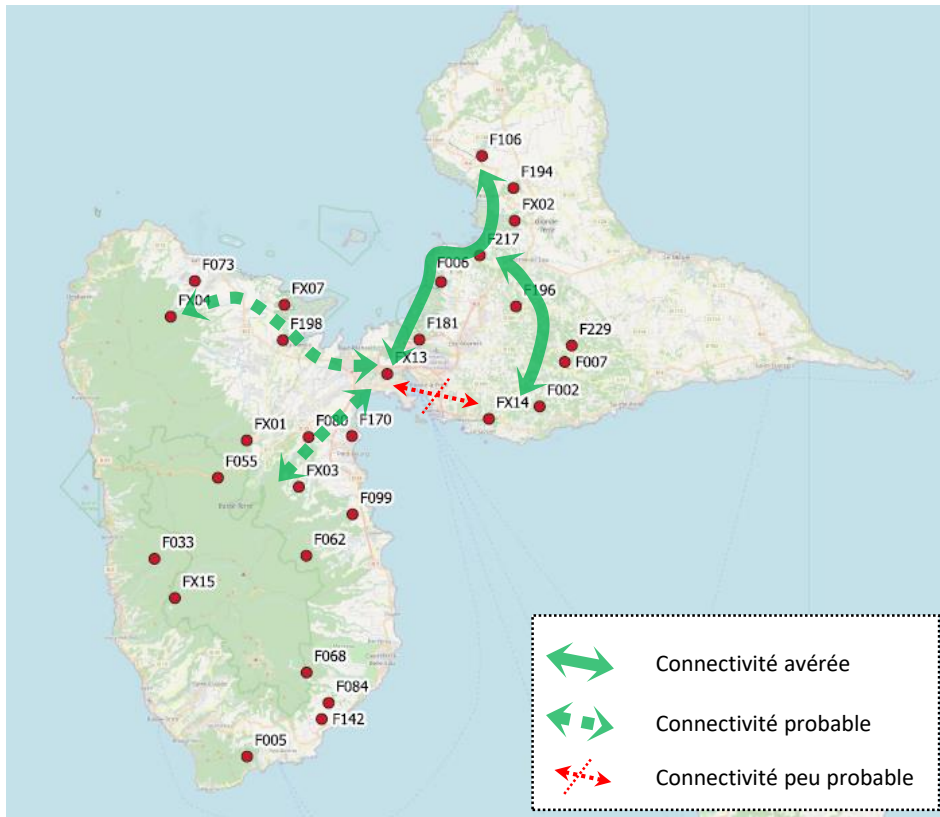
**Figure 11.** Résultats de l'analyse de partition pour le *Sporophile rouge-gorge*. **A.** Gain d'information en termes de structure génétique des populations en fonction du nombre de cluster dans le modèle. Pour un modèle avec  $K$  clusters, ce gain correspond à l'augmentation de la probabilité postérieure des données observées entre les modèles à  $K-1$  et  $K$  clusters. **B.** Assignment des individus aux différents clusters du modèle. Chaque individu est représenté par une barre verticale partitionnée en  $K$  segments colorés ( $K$  étant le nombre de cluster dans le modèle), dont la longueur est proportionnelle à la probabilité d'assignation de l'individu à chaque cluster (cette probabilité pouvant être interprétée comme la proportion du génome de l'individu issue de chaque cluster). Les individus des différentes populations sont séparés par une barre en pointillés blancs. Les codes populations sont indiqués sous la figure et sont approximativement ordonnés du sud vers le nord pour la Basse-Terre (BT) et du Grand Cul-de-Sac Marin vers les Grands Fonds pour la Grande-Terre (GT). L'assignation n'est pas illustrée au-delà de  $K=6$  car non informative.

Il n'y a pas nécessairement de correspondance entre les couleurs des différents modèles (différentes valeurs de  $K$ ). Autrement dit, le cluster bleu foncé (par exemple) pour un modèle (i.e. une valeur de  $K$ ) n'est pas nécessairement le même que le cluster bleu foncé pour un autre modèle (i.e. une autre valeur de  $K$ ).



**Figure 12.** Représentation spatiale des clusters et de la composition des pools génétiques des différentes populations de *Sporophile rouge-gorge* identifiés avec l'algorithme STRUCTURE. A côté de chaque population figure un diagramme circulaire représentant la part du pool génétique de cette population issue de chacun des  $K$  clusters. Cette spatialisation des clusters n'est pas illustrée au-delà de  $K=6$  car non informative.

Il n'y a pas nécessairement de correspondance entre les couleurs des différents modèles (différentes valeurs de  $K$ ). Autrement dit, le cluster bleu foncé (par exemple) pour un modèle (i.e. une valeur de  $K$ ) n'est pas nécessairement le même que le cluster bleu foncé pour un autre modèle (i.e. une autre valeur de  $K$ ).



**Figure 13.** Représentation schématique de la connectivité des populations de *Sporophila rouge-gorge* entre la Basse-Terre et la Grande-Terre.

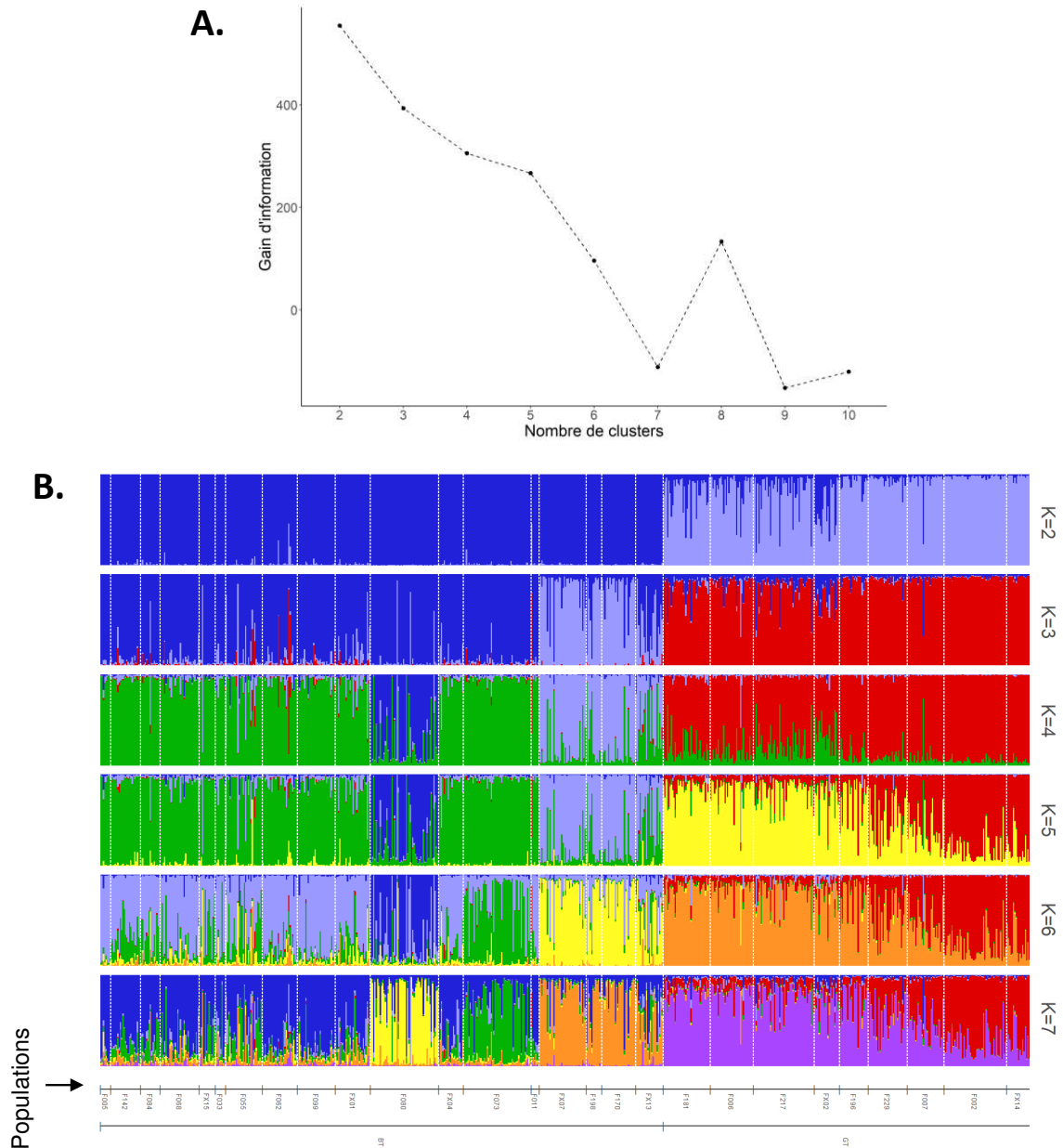
- **Paruline caféïette**

Dans le cas de la Paruline caféïette, le gain d'information associé à l'ajout d'un cluster dans le modèle reste positif jusqu'à  $K=6$  (Figure 14A), même si ce gain est de moins en moins important à mesure que  $K$  augmente. La décroissance graduelle de ce gain traduit l'existence d'une structuration génétique des populations assez fine. Contrairement à certaines autres espèces, il ne semble pas y avoir d'abord un niveau de divergence fort entre la Grande-Terre et la Basse-Terre puis un niveau de divergence plus faible au sein de ces deux parties de la Guadeloupe. Les résultats d'assignation des individus aux différents clusters montrent que la plupart des individus d'une même population sont tous assignés à un des clusters du modèle avec une probabilité proche de 1, et que chaque cluster semble correspondre au pool génétique majoritaire d'au moins une population jusqu'au modèle contenant six clusters (Figure 14B). Lorsqu'un septième cluster est ajouté au modèle, quasiment aucun individu d'aucune population n'est fortement assigné à ce cluster (cluster "bleu clair" pour  $K=7$ ). Le fait d'identifier six clusters au sein de la Guadeloupe témoigne de l'existence d'une forte structuration génétique des populations de Paruline caféïette à l'échelle de la Guadeloupe.

L'examen de la répartition spatiale des clusters génétiques fait apparaître une logique géographique très claire (Figure 15). Les deux premiers clusters séparent les populations de la Basse-Terre et celles de la Grande-Terre. Le troisième cluster (cluster "bleu clair" pour  $K=3$ ) regroupe trois populations sur la Basse-Terre proches de l'isthme (FX07, F198, F170) et la population du secteur de Jarry (FX13). Le quatrième cluster (cluster "bleu foncé" pour  $K=4$ ) permet d'isoler la population F080 des autres, un résultat déjà observé dans des analyses antérieures (Khimoun et al. 2017b). L'ajout d'un cinquième cluster permet de séparer les populations des grands Fonds (cluster "rouge" pour  $K=5$ ) des populations situées le long du Grand Cul-de-Sac Marin sur la Grande-Terre (cluster "jaune" pour  $K=5$ ). Enfin, le sixième cluster (cluster "vert" pour  $K=6$ ) ne semble pas très localisé géographiquement et permet surtout de distinguer la population F073 (près de Sainte-Rose) des populations voisines.

**Ces résultats permettent de dresser un schéma relativement clair de la connectivité entre les populations de Paruline caféïette (Figure 16). On observe une transition graduelle entre les pools génétiques des populations du sud des Grands Fonds et celles situées le long du Grand Cul-de-Sac Marin sur la Grande-Terre jusqu'au pont de l'Alliance (population F181). Il existe donc un axe de connectivité entre ces populations (FX14 et F002 d'un côté, et F181 de l'autre) qui passe par le nord des Grands Fonds entre Morne-à-l'Eau et Bosrédon, puis redescend le long du Grand Cul-de-Sac Marin. Il n'y a en revanche aucune connectivité directe entre l'isthme (population FX13) et le sud des Grands Fonds (population FX14). En ce qui concerne la connectivité entre l'isthme et les populations de la Basse-Terre, les résultats présentés dans ce rapport ne permettent de distinguer la partie ouest des deux axes hypothétiques de la figure 1. En effet, la population FX13 possède un pool génétique en grande partie commun avec celui des populations F198 et FX07 situées sur l'axe 1 et celui de la population F170 située sur l'axe 2. D'autres analyses devront être effectuées afin de pouvoir éventuellement trancher, mais pour l'heure, ces deux axes de connectivité apparaissent probables.**

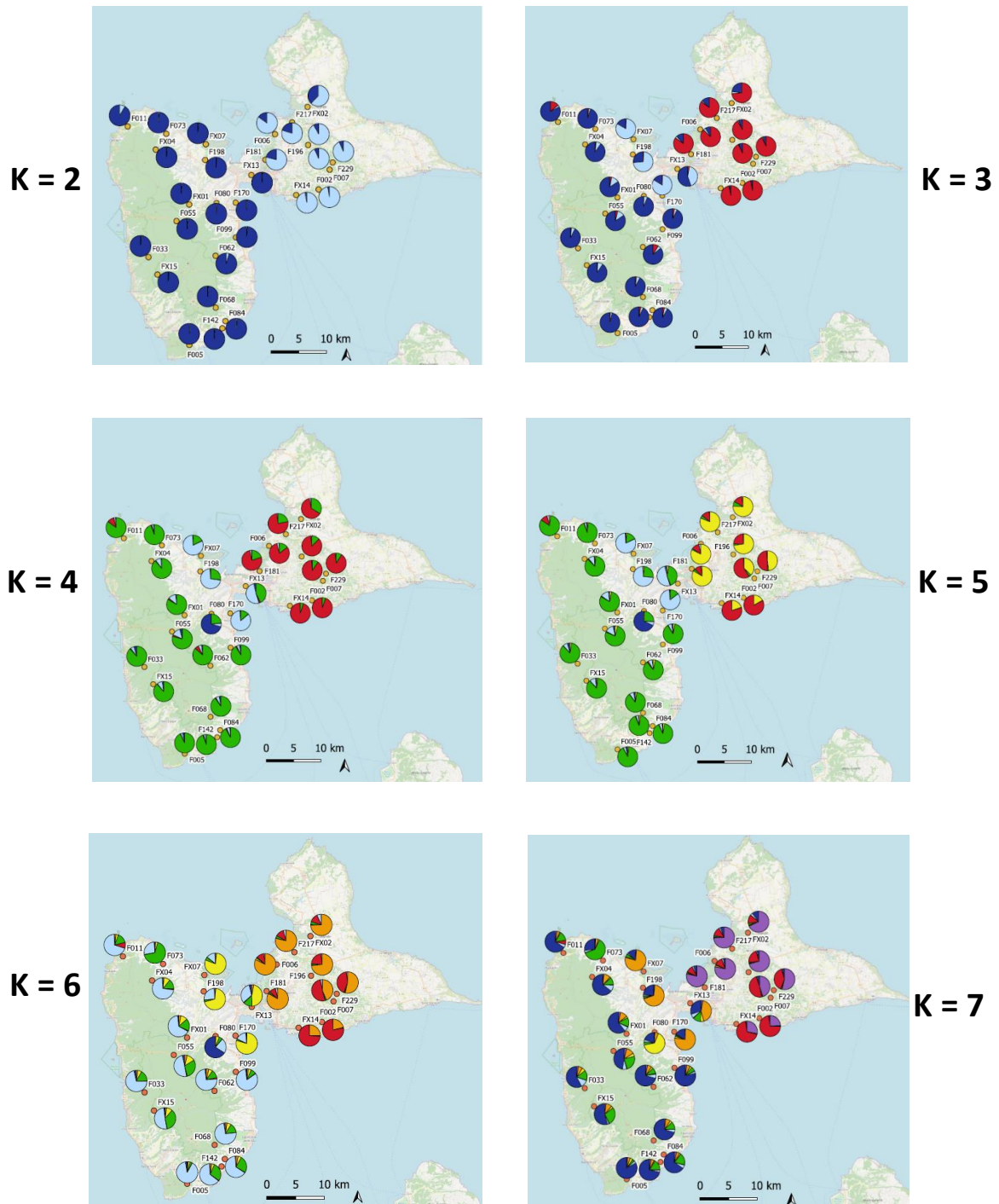
**Il est intéressant de noter que les populations FX13 (forêt marécageuse de Jarry) et F181 (située juste au nord-est du Pont de l'Alliance) semblent avoir deux pools génétiques nettement distincts, ce qui pourrait traduire des flux de gènes très réduits voire absents entre ces deux localités. L'échantillonnage de la population FX13 s'est déroulé à l'intérieur de la forêt marécageuse mais à proximité de la lisière (pour des contraintes d'accès au cœur du massif). Six matinées de capture ont été réalisées en des points différents autour de la partie ouest de la forêt marécageuse de Jarry (entre la N1, la N10 et la D32 (voie verte). Une matinée de capture a eu lieu le long de la voie verte (D32) et une matinée le long de la N10. C'est au cours de ces matinées d'échantillonnage que le nombre de Paruline caféïette capturées est le plus faible, et aucun individu de cette espèce n'a été capturé le long de la voie verte. Même s'il ne s'agit que de chiffres de capture sur deux matinées de terrain seulement, ces données suggèrent que les voies de communication dans le secteur de Jarry pourraient constituer des freins importants à la dispersion de cette espèce, ce qui pourrait expliquer la différenciation génétique observée entre les populations FX13 et F181 alors qu'elles sont éloignées d'à peine plus de 3 kilomètres.**



**Figure 14.** Résultats de l'analyse de partition pour la *Paruline caféiette*. **A.** Gain d'information en termes de structure génétique des populations en fonction du nombre de cluster dans le modèle. Pour un modèle avec  $K$  clusters, ce gain correspond à l'augmentation de la probabilité postérieure des données observées entre les modèles à  $K-1$  et  $K$  clusters. **B.** Assignment des individus aux différents clusters du modèle. Chaque individu est représenté par une barre verticale partitionnée en  $K$  segments colorés ( $K$  étant le nombre de cluster dans le modèle), dont la longueur est proportionnelle à la probabilité d'assignation de l'individu à chaque cluster (cette probabilité pouvant être interprétée comme la proportion du génome de l'individu issue de chaque cluster). Les individus des différentes populations sont séparés par une barre en pointillés blancs. Les codes populations sont indiqués sous la figure et sont approximativement ordonnés du sud vers le nord pour la Basse-Terre (BT) et du Grand Cul-de-Sac Marin vers les Grands Fonds pour la Grande-Terre (GT). L'assignation n'est pas illustrée au-delà de  $K=7$  car non informative.

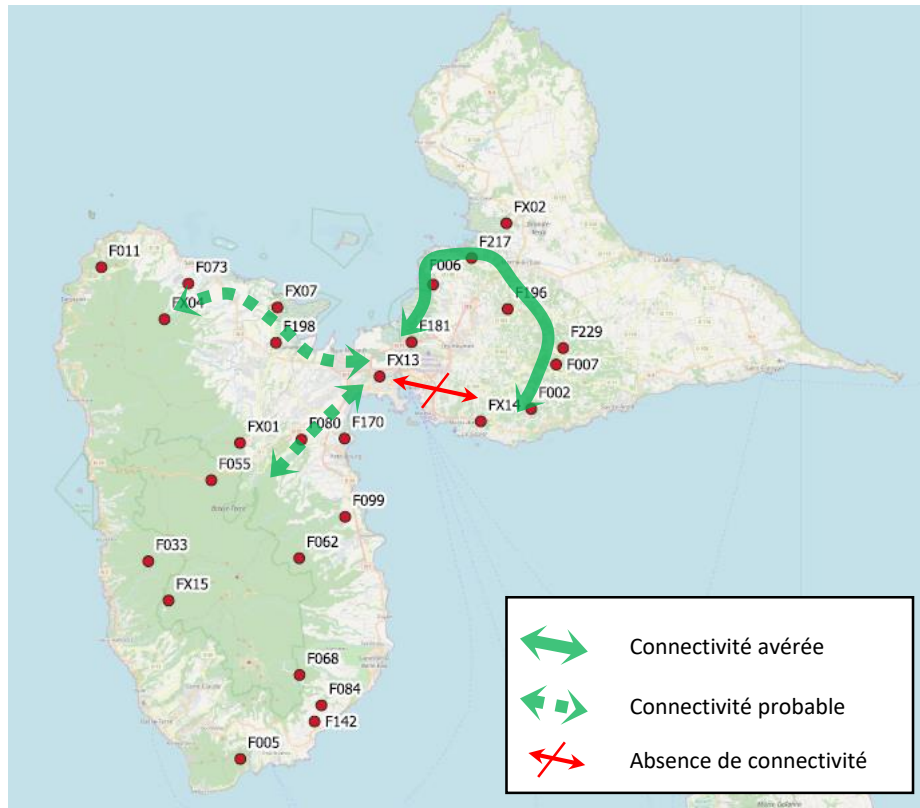
Il n'y a pas nécessairement de correspondance entre les couleurs des différents modèles (différentes valeurs de  $K$ ). Autrement dit, le cluster bleu foncé (par exemple) pour un modèle (i.e. une valeur de  $K$ ) n'est pas nécessairement le même que le cluster bleu foncé pour un autre modèle (i.e. une autre valeur de  $K$ ).





**Figure 15.** Représentation spatiale des clusters et de la composition des pools génétiques des différentes populations de *Paruline caféïette* identifiés avec l'algorithme STRUCTURE. A côté de chaque population figure un diagramme circulaire représentant la part du pool génétique de cette population issue de chacun des K clusters. Cette spatialisation des clusters n'est pas illustrée au-delà de K=7 car non informative.

Il n'y a pas nécessairement de correspondance entre les couleurs des différents modèles (différentes valeurs de K). Autrement dit, le cluster bleu foncé (par exemple) pour un modèle (i.e. une valeur de K) n'est pas nécessairement le même que le cluster bleu foncé pour un autre modèle (i.e. une autre valeur de K).



**Figure 16.** Représentation schématique de la connectivité des populations de *Paruline caféiette* entre la Basse-Terre et la Grande-Terre.

## 4/ Conclusions et perspectives

Le premier objectif du présent projet était de combler un déficit d'échantillonnage dans le secteur de l'isthme connectant la Basse-Terre et la Grande-Terre. Même s'il n'a pas été possible de collecter des échantillons en nombre suffisant dans tous les sites et pour toutes les espèces ciblées, cet objectif a été pleinement atteint pour trois espèces (Paruline caféïette, Saltator gros-bec, Sporophile rouge-gorge), partiellement atteint pour deux espèces (Moqueur grivotte et Elénie siffleuse), et pas du tout atteint pour la Grive à pieds jaunes. Il est important de noter que la non capture ou le nombre de captures très faible d'une espèce dans certains sites peut elle-même traduire une information en termes de connectivité, même si ces données (qui peuvent signifier l'absence de l'espèce) doivent être interprétées dans ce cadre avec la prudence nécessaire.

L'analyse des données génétiques issues des échantillons collectés à travers le présent projet et groupées avec celles des deux projets FRAG&BINV et CANON a permis de mettre en évidence un certain nombre de résultats qui semblent être partagés par plusieurs espèces d'oiseaux forestiers.

Nous avons observé une différenciation génétique nette entre les populations de la Basse-Terre et celles de la Grande-Terre chez les cinq espèces étudiées dans ce projet. Si nous avons déjà décrit un tel patron de différenciation pour la Grive à pieds jaunes (Arnoux et al. 2014) et la Paruline caféïette (Khimoun et al. 2017b), les résultats présentés dans ce rapport incluent un nombre de populations supérieur et donc une meilleure couverture spatiale de la Guadeloupe, ce qui vient renforcer ces résultats. Pour les trois autres espèces, la description de ce patron avec ce niveau de résolution spatiale est nouvelle. Il est donc raisonnable de penser qu'un nombre plus important d'espèces d'oiseaux, et peut-être d'autres taxons, présentent également ce patron de divergence génétique entre les deux parties de la Guadeloupe. Il existe donc des enjeux réels et majeurs en termes de connectivité écologique dans le secteur de l'isthme connectant la Basse-Terre et la Grande-Terre. Néanmoins, toutes les espèces d'oiseaux ne sont pas nécessairement affectées. En effet, une analyse préliminaire de la structure génétique des populations de Sucrier à ventre jaune conduite sur une douzaine de populations réparties sur la Basse-Terre et la Grande-Terre n'avait révélé aucun patron de différenciation entre ces deux parties de la Guadeloupe (Kreus 2017). Si ce résultat est a priori peu surprenant compte tenu du caractère généraliste de cette espèce, les patrons de différenciation observés ici sont plus inattendus chez certaines espèces qui ne paraissent pas très spécialistes de l'habitat forestier telles que le Sporophile rouge-gorge (Khimoun et al. 2016b).

Par ailleurs, les patrons de différenciation génétique décrits dans ce rapport sont d'autant plus intéressants qu'ils ne se bornent pas à l'existence de deux entités génétiques au sein de chaque espèce, l'une sur la Basse-Terre et l'autre sur la Grande-Terre. En effet, un nombre de clusters génétiques supérieur à deux a été identifié chez toutes les espèces à l'exception de la Grive à pieds jaunes. Il est également à noter que l'analyse de partition utilisée ici est généralement utilisée pour détecter et caractériser les niveaux de structuration majeure, ce qui n'empêche pas la possibilité d'avoir une structuration génétique plus fine au sein des clusters génétiques identifiés (Garnier et al. 2004; Pritchard et al. 2000). Ces patrons de structuration génétique laissent donc entrevoir la possibilité de décrire plus finement les effets de la structure du paysage sur les patrons de dispersion des oiseaux sur le territoire de la Guadeloupe.

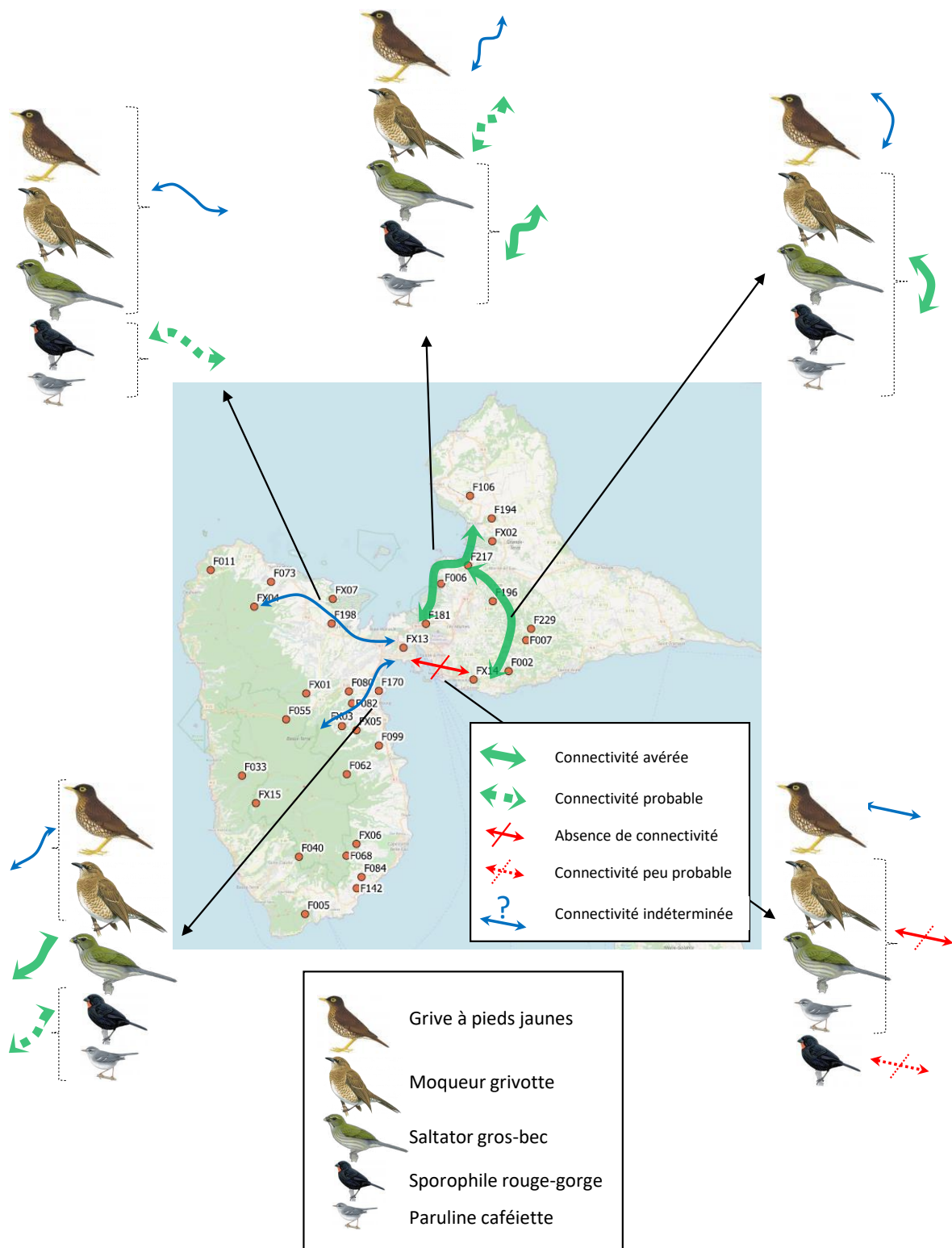
Avant de décrire le schéma de connectivité synthétique issu de l'étude de cinq espèces sur le même territoire, il convient de rappeler le cadre dans lequel un patron de structuration génétique peut être traduit en termes de dispersion des organismes. Le patron de différenciation génétique entre populations observé à un instant  $t$  traduit le niveau de flux de gènes entre ces populations en intégrant les multiples générations qui ont précédé cet instant (leur nombre étant très difficile à estimer, de quelques dizaines à plusieurs centaines voire milliers ; Keyghobadi 2007). Il y a en effet un délai dans la réponse génétique par rapport au facteur qui modifie les patrons de dispersion. En d'autres termes, l'absence de différence génétique entre deux populations ne signifie pas nécessairement que ces populations échangent actuellement des gènes (à travers la dispersion d'individus), et peut correspondre à une situation où la dispersion aurait été récemment stoppée ou très restreinte entre ces populations, sans avoir eu le temps de se révéler au niveau génétique. Les données génétiques intègrent donc en quelque sorte les patrons de dispersion d'un passé plus ou moins récent.

Les patrons de structuration génétique observés ont permis de faire des inférences en termes de connectivité écologique qui ont été décrites indépendamment pour chaque espèce. Il s'agit désormais de décrire le patron synthétique qui émerge de la confrontation des schémas de connectivité spécifiques. Même si un schéma de connectivité unique n'est pas nécessairement pertinent pour toutes les espèces (et qu'il n'y a donc pas de nécessité absolue à établir un tel schéma), cette confrontation laisse entrevoir certains éléments qui sont *a priori* partagés par plusieurs espèces, sans doute au-delà des quelques-unes étudiées dans le présent travail. La synthèse des inférences de connectivité est présentée schématiquement dans la figure 17.

Au sein de la Grande-Terre, il ressort très clairement que les populations des Grands Fonds ne sont pas déconnectées entre elles malgré la structure particulière du manteau forestier dans ce secteur. Par ailleurs les populations situées le long du Grand Cul-de-Sac Marin, du nord de Petit-Canal (ou un peu plus au sud pour les espèces *a priori* absentes ou très rares à cette latitude telles que la Grive à pieds jaunes et la Paruline caféïette) à l'isthme connectant la Basse-Terre et la Grande-Terre, semblent également connectées entre elles et permettre les flux de gènes de proche en proche. De manière plus intéressante, des échanges génétiques semblent exister entre ces deux ensembles de populations, la connexion se faisant *a priori* par le nord des Grands Fonds entre Morne-à-l'Eau et Bosrédon au niveau de la route nationale 5. Le couloir entre Morne-à-l'Eau et Bosrédon présente donc un intérêt majeur en termes de connectivité écologique sur la Grande-Terre. Ces connexions sont soutenues par les données génétiques de quatre espèces sur cinq, la cinquième (Grive à pieds jaunes) ne contredisant pas ce schéma mais ne pouvant simplement apporter aucune information à ce niveau.

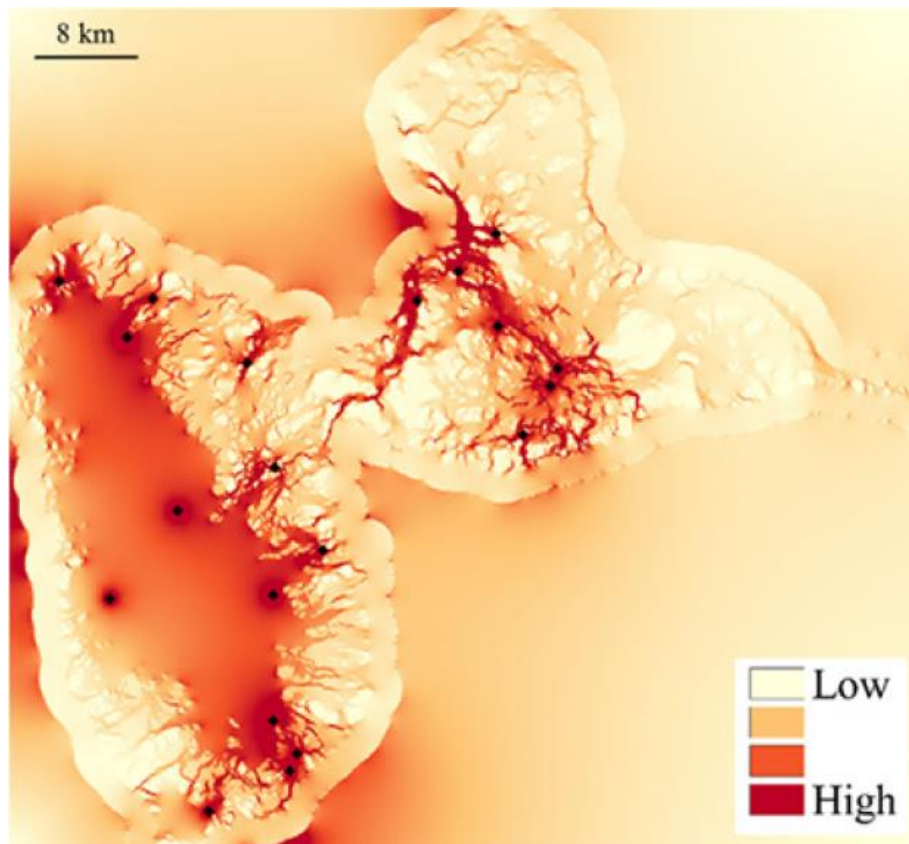
Par ailleurs, nous pouvons écarter l'hypothèse d'une connexion directe entre les populations de l'isthme et celles du sud des Grands Fonds malgré une distance géographique relativement faible (4-5 kilomètres). Ce résultat est soutenu par les données de trois espèces (Moqueur grivotte, Saltator grosbec et Paruline caféïette), et probablement soutenu par celles du Sporophile rouge-gorge, alors que les données de la Grive à pieds jaunes ne permettent pas d'établir un diagnostic à ce niveau. Cette étude apporte donc des résultats assez clairs en matière de connectivité entre les populations de la Grande-Terre et l'isthme assurant la connexion avec la Basse-Terre. Le second objectif est donc atteint, au moins en partie.

En revanche, les données génétiques et l'approche analytique utilisée ici ne permettent pas complètement de déterminer par quelle(s) voie(s) les flux de gènes s'opèrent entre l'isthme et les populations de la Basse-Terre. L'axe de connectivité entre l'isthme et passant par le nord de la Pointe à Bacchus (à l'endroit où la Lézarde rejoint le Petit-Cul-de-Sac-Marin) pour rejoindre le grand massif forestier de la Basse-Terre via le secteur de Fougère à l'est de Petit-Bourg est validé au moins chez une espèce (le Saltator grosbec) et probable chez deux espèces (Sporophile rouge-gorge et Paruline caféïette). Les données des deux autres espèces (Moqueur grivotte, Grive à pieds jaunes) ne permettent ni de confirmer, ni d'invalider cet axe de connectivité. Enfin, en ce qui concerne l'axe de connectivité entre l'isthme et le grand massif forestier de la Basse-Terre en passant par le nord de la Basse-Terre le long du Grand Cul-de-Sac Marin, les données d'aucune espèce ne permettent de le valider ou de l'invalider complètement. Néanmoins, même si la manière dont la connexion se fait entre l'isthme et le grand massif forestier de la Basse-Terre reste encore à être déterminée, la forêt marécageuse de Jarry et les mangroves et forêts marécageuses de part et d'autre de la Rivière Salée semblent jouer un rôle majeur dans la connectivité écologique entre la Basse-Terre et la Grande-Terre.



**Figure 17.** Représentation schématique et synthétique de la connectivité des populations d'oiseaux forestiers entre la Basse-Terre et la Grande-Terre. Les axes représentés sur la carte résultent de la synthèse des résultats obtenus pour les cinq espèces (voir texte). Pour chaque axe indiqué, les résultats obtenus pour chaque espèce sont rappelés autour de la carte.

Si l'approche analytique présentée dans ce rapport a déjà permis d'apporter des réponses extrêmement intéressantes en termes de connectivité écologique et parfois directement mobilisables par les gestionnaires de l'environnement, nous poursuivons actuellement un effort d'analyse mobilisant des méthodes beaucoup plus puissantes en génétique du paysage. Il s'agira notamment de faire des inférences, à partir des mêmes données génétiques, sur les coûts de déplacements des organismes associés aux différents éléments composant le paysage, qu'il s'agisse du type d'occupation des sols, du réseau routier ou encore du relief. Ces analyses nous permettront à terme de dresser une carte relativement fine des flux de gènes pour chaque espèce, comme nous l'avions fait pour la Paruline caféïette il y a quelques années. La figure 18 (tirée de l'article de Khimoun et al. 2017b) présente ce type de carte, sachant que les couleurs les plus chaudes correspondent aux flux de gènes les plus importants et que les couleurs les plus froides correspondent à des flux de gènes réduits voire absents. Même si ces analyses ont déjà été effectuées chez la Paruline caféïette, il est nécessaire de les reconduire avec le complément d'échantillonnage issu du présent travail, car ces analyses souffraient d'un déficit de données génétiques dans le secteur de l'isthme (qui constitue le nœud de la connectivité entre la Basse-Terre et la Grande-Terre). Ces analyses devraient permettre de retrouver l'essentiel des résultats décrits dans ce rapport, mais aussi et surtout de les dépasser en permettant de trancher sur les axes de connectivité là où cela n'a pas été possible avec l'approche présentée dans ce rapport. En parallèle, nous travaillons actuellement dans le cadre du projet CANON sur la validation des graphes paysagers à partir de données génétiques. Les graphes paysagers sont en effet des modèles très utilisés (par les gestionnaires et les chercheurs) en aménagement du territoire par rapport aux problématiques de trames vertes et bleues. L'objectif ultime est de pouvoir décrire le schéma de connectivité global à l'échelle de la Guadeloupe et de déterminer les fragments forestiers et/ou corridors à préserver (ou restaurer) en priorité sans qu'il soit nécessaire de disposer de données génétiques sur l'ensemble des fragments forestiers (ce qui est naturellement impossible).



**Figure 18.** Intensité des flux de gènes entre les populations échantillonnées de Paruline caféïette (Figure tirée de l'article Khimoun et al. 2017b, joint en annexe du rapport).

Ces nouvelles analyses viennent de débiter et vont nécessiter des calculs intensifs sur un centre de calcul informatique pendant a priori plusieurs mois. Mais lorsqu'ils seront connus, les résultats seront naturellement communiqués à la DEAL de Guadeloupe et aux acteurs locaux de la biodiversité et de l'environnement. Bien conscients de la difficulté d'accès de ces acteurs à la littérature académique, nous envisageons d'assurer un retour de ces résultats localement et directement, comme nous l'avions fait pour le projet FRAG&BINV lors d'une réunion de restitution hébergée par la DEAL le 3/03/2020.

## 5/ Remerciements

Le présent projet a pu être réalisé grâce à l'obtention de nombreuses autorisations, qu'il s'agisse de la capture d'oiseaux et du transport des échantillons collectés (DEAL de Guadeloupe, CRBPO) ou de l'accès à certains sites d'échantillonnage (Parc National de Guadeloupe, Conservatoire du Littoral, Office National des Forêts, plusieurs propriétaires privés). En plus des auteurs du présent rapport, plusieurs personnes ont participé aux opérations de terrain : Jeanne Abbou, Pierre Bulens, Camille Fagot, Luc Guillaume, Gilles Leblond, Gilles Mourgaud, Thomas Pagnon, Laura Pinto, Célia Sineau, Alexandre Vong. Enfin, ce projet a été financé par la DEAL de Guadeloupe et le programme ISITE-BFC.

## 6/ Références bibliographiques

*Les références indiquées ci-dessous par un astérisque sont fournies en annexe du rapport.*

- Allendorf, F.W. & Luikart, G. (2007) Conservation and the genetics of populations. Blackwell Publishing, Malden.
- (\*) Arnoux E, Eraud C, Navarro N, Tougard C, Thomas A, Cavallo F, Vetter N, Faivre B & Garnier S (2014) Morphology and genetics reveal an intriguing pattern of differentiation at a very small geographic scale in a bird species, the forest thrush *Turdus lherminieri*. *Heredity*, 113: 514-525.
- Fahrig L (2003) Effects of habitat fragmentation on biodiversity. *Ann. Rev. Ecol. Evol. Syst.* 34: 487-515.
- Garnier S, Alibert P, Audiot P, Prieur B & Rasplus J-Y (2004) Isolation by distance and sharp discontinuities in gene frequencies: implications for the phylogeography of an alpine insect species, *Carabus solieri*. *Molecular Ecology*, 13: 1883-1897.
- Gaston K (1994) *Rarity*. Chapman & Hall, London.
- Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, et al. 2012. « Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data ». *Bioinformatics* 28 (12): 1647-49.
- Keyghobadi N (2007) The genetic implications of habitat fragmentation for animals. *Can. J. Zool.* 85: 1049-1064.
- (\*) Khimoun A, Arnoux E, Martel G, Pot A, Eraud C, Condé B, Loubon M, Théron F, Covas R, Faivre B & Garnier S (2016a) Contrasted patterns of genetic differentiation across eight bird species in the Lesser Antilles. *Genetica*, 144: 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 (2016b) Habitat specialisation predicts genetic response to fragmentation in tropical birds. *Molecular Ecology*, 25: 3831-3844.

- (\*) Khimoun A, Ollivier A, Faivre B, Garnier S (2017a) Level of genetic differentiation affects relative performances of EST- and genomic SSRs. *Molecular Ecology Resources*, 17: 893-903.
- (\*) Khimoun A, Peterman W, Eraud C, Faivre B, Navarro N & Garnier S (2017b) Landscape genetic analyses reveal fine-scale effects of forest fragmentation in an insular tropical bird. *Molecular Ecology*, 26: 4906-4919.
- Kreus R (2017) Effets de la fragmentation forestière sur la structure génétique d'une espèce généraliste des Petites Antilles : le Sucrier à ventre jaune. Mémoire de stage de master 1<sup>ère</sup> année encadré par S Garnier. Université des Antilles – Master ECOTROP.
- Laurence WF (2010) Habitat destruction: death by a thousand cuts. In *Conservation Biology for all* (eds NS Sodhi & PR Ehrlich). Oxford Univ. Press.
- (\*) Pérez-Rodríguez A, Khimoun A, Ollivier A, Eraud C, Faivre B, Garnier S (2018) Habitat fragmentation, but not habitat loss, drives the prevalence of blood parasites in a Caribbean passerine. *Ecography*, 41: 1835-1849
- (\*) Perrin A, Khimoun A, Faivre B, Ollivier A, de Pracontal N, Théron F, Loubon M, Leblond G, Duron O, Garnier S (2021) Habitat fragmentation differentially shapes neutral and immune gene variation in a tropical bird species. *Heredity*, 126: 148-162.
- Pimm S (1991) *The balance of nature?* University of Chicago Press, Chicago.
- Pritchard, Jonathan K, Matthew Stephens, et Peter Donnelly. 2000. « Inference of Population Structure Using Multilocus Genotype Data ». *Genetics* 155 (2): 945-59.
- Rousset, François. 2008. « Genepop'007: A Complete Re-Implementation of the Genepop Software for Windows and Linux ». *Molecular Ecology Resources* 8 (1): 103-6.





