
Article:
Beninde, J. orcid.org/0000-0002-1677-1809, Wittische, J. and Frantz, A.C. (2023) Quantifying uncertainty in inferences of landscape genetic resistance due to choice of individual-based genetic distance metric. Molecular Ecology Resources. ISSN 1755-098X

https://doi.org/10.1111/1755-0998.13831

Received: 15 March 2021 | Revised: 12 May 2023 | Accepted: 22 June 2023

DOI: 10.1111/1755-0998.13831

MOLECULAR ECOLOGY RESOURCES WILEY

FROM THE COVER

Quantifying uncertainty in inferences of landscape genetic resistance due to choice of individual-based genetic distance metric

Joscha Beninde[1,2] | Julian Wittische[3,4] | Alain C. Frantz[3,4,5]

[1]LA Kretz Center for California Conservation Science, Institute of the Environment and Sustainability, University of California, Los Angeles, California, USA

[2]IUCN WCPA Connectivity Conservation Specialist Group, Gland, Switzerland

[3]Musée National d'Histoire Naturelle, Luxembourg City, Luxembourg

[4]The Fondation Faune-Flore, Luxembourg City, Luxembourg

[5]The University of Sheffield, Sheffield, UK

Correspondence
Joscha Beninde, LA Kretz Center for California Conservation Science, Institute of the Environment and Sustainability, University of California, Los Angeles, CA, USA.
Email: joscha.research@gmail.com

Alain C. Frantz, Musée National d'Histoire Naturelle, Luxembourg City, Luxembourg.
Email: afrantz@mnhn.lu

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: BE 6887/1-1; Fonds National de la Recherche Luxembourg, Grant/Award Number: C20/SR/14748041; Musée National d'Histoire Naturelle, Luxembourg

Handling Editor: Jeremy B. Yoder

Abstract

Estimates of gene flow resulting from landscape resistance inferences frequently inform conservation management decision-making processes. Therefore, results must be robust across approaches and reflect real-world gene flow instead of methodological artefacts. Here, we tested the impact of 32 individual-based genetic distance metrics on the robustness and accuracy of landscape resistance modelling results. We analysed three empirical microsatellite datasets and 36 simulated datasets that varied in landscape resistance and genetic spatial autocorrelation. We used ResistanceGA to generate optimised multi-feature resistance surfaces for each of these datasets using 32 different genetic distance metrics. Results of the empirical dataset demonstrated that the choice of genetic distance metric can have strong impacts on inferred optimised resistance surfaces. Simulations showed accurate parametrisation of resistance surfaces across most genetic distance metrics only when a small number of environmental features was impacting gene flow. Landscape scenarios with many features impacting gene flow led to a generally poor recovery of true resistance surfaces. Simulation results also emphasise that choosing a genetic distance metric should not be based on marginal R^2-based model fit. Until more robust methods are available, resistance surfaces can be optimised with different genetic distance metrics and the convergence of results needs to be assessed via pairwise matrix correlations. Based on the results presented here, high correlation coefficients across different genetic distance categories likely indicate accurate inference of true landscape resistance. Most importantly, empirical results should be interpreted with great caution, especially when they appear counter-intuitive in light of the ecology of a species.

KEYWORDS
comparative landscape genetics, connectivity, corridor, current flow, fragmentation, multiple-path analysis

1 | INTRODUCTION

Spatial genetic patterns form the basis for many ecological and evolutionary research avenues and are frequently used to guide decision-making in conservation management. With the appropriate tools of analysis, patterns of isolation by distance (Wright, 1943), by resistance (McRae, 2006), by environment (Wang & Bradburd, 2014) or by barriers (Cushman et al., 2006) can be detected and reveal the extent to which a species' environment can impact its gene flow pattern. Results of spatial genetic analyses are used in many different ways, for example, to identify common drivers of patterns in the life history of species (Medina et al., 2018), or to identify corridors and generate scientific guidance for conservation management, as is the goal, specifically, of many landscape genetic studies (Keller et al., 2015). Underlying these efforts is the assumption that models of a landscape's impact on spatial genetic patterns are ecologically meaningful and robust to different methods of inference.

Landscape genetic approaches that statistically relate the distribution of genetic similarities among individuals to landscape characteristics (Cushman et al., 2006; Schwartz et al., 2009) hold great promise as solutions to several methodological problems have recently been proposed. For example, linear-mixed effects models (LME) can account for the non-independence of pairwise comparisons by the use of maximum-likelihood population effect (MLPE) parameterisation (Shirk et al., 2017b). Also, simulations have shown that LMEs with corrected Akaike information criteria (AICc) model selection consistently outperform other regression methods, or Mantel-based methods, in correctly identifying true models of landscape impact (Shirk et al., 2017b). Furthermore, the recently developed R-package ResistanceGA (Peterman, 2018), which combines LME models with MLPE parameterisation, makes use of a machine-learning algorithm to infer resistance values of landscape factors that maximise fit to user-specified genetic distances and has proven to be a robust tool in inferring true landscape resistance (Winiarski et al., 2020). This approach circumvents the inherent subjectivity and biases associated with user-specified resistance values, which are frequently based solely on expert opinion (Peterman et al., 2014, 2019; Richardson et al., 2016).

However, approaches of landscape genetic inferences are still evolving (Richardson et al., 2016) and some methodological aspects remain underexplored. For instance, the genetic distance metrics used in (individual-based) landscape resistance modelling seem to be frequently chosen with little or no justification provided. The proportion of shared alleles, D_{PS}, (Draheim et al., 2018; Landguth et al., 2010; Trumbo et al., 2013) and kinship or relatedness coefficients (Dellicour et al., 2019a; Renner et al., 2016) are common choices and continue to be implemented in new applications developed for landscape genetic analyses (e.g. Savary et al., 2020). However, many of these estimators have a large sample variance, which has recently been shown to negatively impact landscape genetic inferences (Winiarski et al., 2020). In addition, Kimmig et al. (2020a) have shown that AICc-based support and rank of the same single-surface resistance model can vary, depending on the genetic distance metric used for inference. These authors suggested that the use of a genetic distance based on 10 axes of a factorial correspondence analysis (FCA; an eigenvector-based multivariate approach similar to principal component analysis, PCA) leads to model selection results with the highest support. In contrast, results from a simulation study suggested that genetic distances derived from 64 PCA axes allowed to correctly identify the underlying resistance surfaces most frequently, particularly in simulation scenarios that were more challenging for landscape genetic inferences (Shirk et al., 2017a).

The uncertainty regarding the choice of genetic distance metric and how it influences landscape genetic inferences led us to explore this issue more thoroughly. We first quantified the variation in landscape genetic inferences given 32 different genetic distance metrics for three empirical datasets of different taxa, based on single- and multi-feature optimisations using ResistanceGA. In the next step, we simulated 36 population-genetic datasets, varying the number of features resistant to gene flow, maximum resistance values and the degree of spatial autocorrelation. We then applied the same ResistanceGA optimisation framework to the simulated datasets as applied to the empirical datasets. This generated 1152 optimised multi-feature resistance surfaces (OMFRS; 36 population-genetic datasets, each analysed with 32 different genetic distance metrics), which we correlated with the true resistance surfaces, providing a measure of accuracy for the recovery of true landscape resistance. We then used recursive partitioning approaches to identify the simulation settings, for example, resistance or isolation-by-distance scenario and genetic distance metrics that were most strongly associated with accurate resistance recovery. In a final step, and with the aim of directly comparing empirical datasets with the simulated results, we test if resistance recovery can be predicted based on the convergence of OMFRS, and if resistance recovery is associated with marginal R^2 (mR^2) of optimised models across genetic distance metrics.

2 | MATERIALS AND METHODS

2.1 | Basis of the landscape genetic analyses

Analyses of landscape resistance were based on ResistanceGA. This R package evaluates how landscape features influence genetic connectivity by statistically relating the distribution of inter-individual genetic distances to the current flow across alternative landscape resistance models. This method uses a machine-learning algorithm (a genetic algorithm in R package GA; Scrucca, 2013) to optimise resistance surfaces that best fit genetic data, thereby avoiding the need for users to specify resistance values. ResistanceGA can optimise single categorical and continuous resistance surfaces, as well as multiple resistance surfaces simultaneously. The optimisation process uses log-likelihood as the objective function; a statistic obtained from LME models fit with pairwise genetic distance as the response variable and pairwise current flow as the explanatory variable, the latter being calculated from resistance models. In order to account for non-independence among pairwise genetic

and environmental distances, mixed effects models are fitted using the MLPE parameterisation implemented in R package LME4 (Bates et al., 2015). Environmental variables are included as fixed effects in the model and identity for each pair of individuals as random effects. Model support for optimised resistance surfaces is assessed based on Akaike information criteria corrected for small sample size (AICc).

In order to gain an understanding of the resistance to movement of the entire landscape (Peterman & Pope, 2021), environmental variables need to be combined into a composite multi-feature resistance surface. Given a large number of environmental features and ensuing analyses (see below), we followed a stepwise optimisation procedure (Kimmig et al., 2020a) to generate OMFRS. At first, single-feature analyses optimised resistance surfaces separately for each environmental feature. Multi-feature models were then created by sequentially adding environmental features to a multi-feature resistance surface, based on model support in the prior single-feature analysis (starting with the highest-ranking environmental feature and only including those that explained genetic differences better than the *distance-only* model). However, only those environmental feature models were retained in the multi-feature model when support was $\Delta AICc > 2$.

2.2 | Genotype datasets

The empirical analyses were based on three previously published datasets: (1) The BOARS dataset consisted of 790 samples of a large omnivorous mammal, the wild boar (*Sus scrofa* L.) that were collected (2005–2009) randomly across the southern Walloon part of Belgium by the local Services of the Nature and Forest Department of the Public Service of Wallonia and genotyped at 14 microsatellite loci (Dellicour et al., 2019a); (2) the FOXES dataset contained 184 samples of a medium-sized mammalian meso-predator, the red fox (*Vulpes vulpes* L.), collected as road-kills (2010–2015) across the metropolitan area of Berlin and genotyped at 15 microsatellite loci (Kimmig et al., 2020a); (3) the LIZARDS dataset contained 223 samples of a small insectivorous reptile, the common wall lizard (*Podarcis muralis* Laurenti), that were collected across the German city of Trier (2011–2012) and genotyped at 17 microsatellite loci (Beninde et al., 2016a). The geographical distribution of the samples of all three datasets is shown in Figure 1. Wild boars and foxes are continuously distributed across the Walloon and Berlin study areas, respectively, and samples covered this space comprehensively. The distribution of lizards in Trier is more restricted and samples exhaustively cover all known occurrence locations. The three datasets thus follow random sampling schemes that either comprehensively cover continuously distributed areas or all known, spatially discontinuous occurrences. Such schemes increase the robustness of spatial genetic inferences and are preferred over clustered or transect sampling (Landguth & Schwartz, 2014; Schwartz & McKelvey, 2009).

The isolation-by-distance (IBD) pattern underlying each empirical dataset was described by regressing pairwise estimates of Loiselle's kinship coefficient F_{ij} (Loiselle et al., 1995; Vekemans & Hardy, 2004) against the natural logarithm of inter-individual straight-line geographical distances using SPAGEDI v.1.5 (Hardy & Vekemans, 2002). The slope of this regression was estimated to be $b = -0.009$ (SE = 0.002) in the BOARS dataset, $b = -0.013$ (SE = 0.002) in the FOXES dataset and $b = -0.013$ (SE = 0.003) in LIZARDS. Previous analyses of genetic discontinuities using clustering algorithms identified varying levels of population structure in the three datasets. Five geographically distinct and coherent clusters were detected for BOARS (Dellicour et al., 2019a), but the authors concluded that clusters were an artefact of the underlying isolation-by-distance pattern as cluster boundaries did not correspond to any environmental barriers (see also Frantz et al., 2009). Spatially explicit clustering could not confirm a genetic discontinuity, identified by conventional clustering, induced by rivers in FOXES (Kimmig et al., 2020a). Both spatially explicit and conventional clustering identified strong genetic discontinuity associated with a major river in LIZARDS (Beninde et al., 2016a).

2.3 | Genetic distance metrics

For each of these three datasets, we computed a total of 32 different genetic distance metrics. Altogether, 22 of these were based on two eigenvector-based multivariate analyses, the principal component analysis (PCA) and the closely related factorial correspondence analysis (FCA). Both approaches cluster variance between samples into composite gradients, maximising differences between samples. PCA assumes continuous, normally distributed data (Dytham, 2011), whereas FCA assumes data consisting of multistate categorical variables (She et al., 1987) suitable for the analysis of microsatellite data. We used GENETIX v.4.05.2 (Belkhir et al., 1996–2004) to generate a contingency table of allele count (0, 1 or 2) by individual for all alleles in the population. We then used the R package ADE 4 1.7.13 (Dray & Dufour, 2007) to perform FCAs and PCAs (with rescaling of allele counts) on the contingency table and obtain the position of every individual on every axis ranging from 1 to 64. We then used the R package ECODIST 2.0.1 (Goslee & Urban, 2007) to obtain genetic distance matrices based on the Euclidean distance between positions of individuals on an incremental number of axes (first distance matrix based on the position on axis 1, second distance matrix based on positions on axes 1 & 2, etc.) up to the first 10 axes. As recommended by Shirk et al. (2017a) we also calculated a genetic distance metric based on the first 64 PCA and FCA axes, thus generating 11 distance matrices each for PCA- and FCA-based genetics distance metrics.

The remaining 10 genetic distance metrics represent other commonly used metrics investigated by Shirk et al. (2017b) that fall into the following categories, whose definitions were taken from the manual of program SPAGEDI v.1.5 (Hardy & Vekemans, 2002). **Kinship coefficients** (the probability that two homologous alleles drawn randomly from each of two individuals are identical by descent): (i) Kc.Lo (Loiselle et al., 1995); (ii) Kc.R (Ritland, 1996); **relatedness coefficients** (the proportion of gene copies in one individual with alleles identical to these of another individual): (iii) Rc.L&R (Lynch & Ritland, 1999); (iv) Rc.Li (Li et al., 1993); (v) Rc.Q&G (Queller & Goodnight, 1989); (vi) Rc.W (Wang, 2002); **fraternity coefficients** (the probability that two

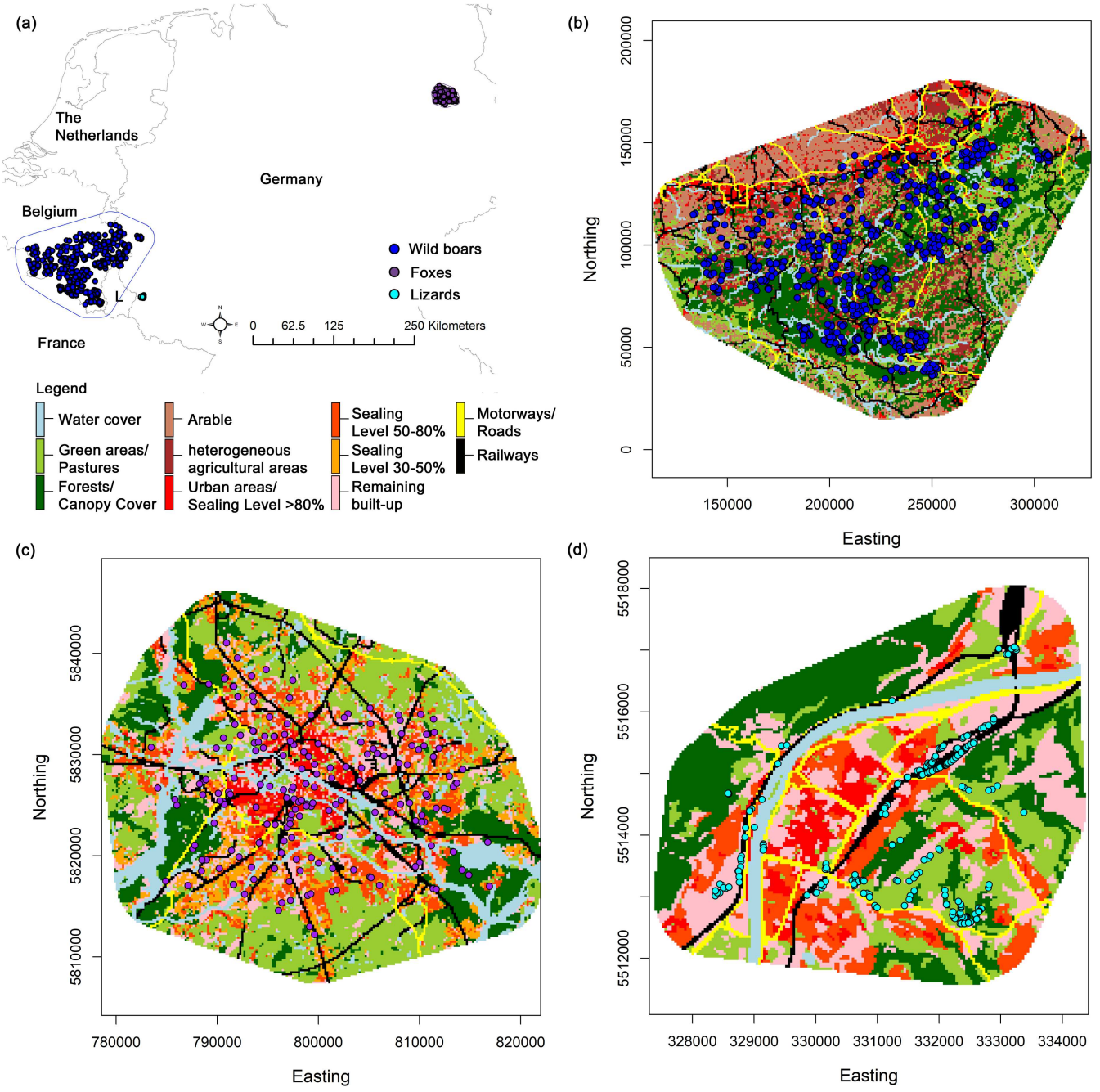

FIGURE 1 Geographical origin of the three empirical datasets. (a) Location of the study regions in Central Europe (L = Luxemburg). Distribution of sampling sites (coloured dots) for BOARS (b), FOXES (c) and LIZARDS (d). The background colours of the maps are indicative of land use types.

individuals share both of their alleles identical by descent): (vii) Fc.L&R (Lynch & Ritland, 1999); (viii) Fc.W (Wang, 2002); **other metrics**: (ix) Rousset's $\hat{a}$ (Rousset, 2000); (x) the proportion of shared alleles D_{PS} (Bowcock et al., 1994). The bias of the kinship, relatedness and fraternity estimators tends to be small, but they are generally characterised by a large sample variance (Lynch & Ritland, 1999; Van de Casteele et al., 2001; Vekemans & Hardy, 2004; Wang, 2002). Rousset's $\hat{a}$ estimates genetic distance between individuals in a continuous population taking isolation-by-distance into account, while D_{PS} is a dissimilarity measure that considers the number of direct differences between genotypes. With the exception of D_{PS}, which was estimated using the R package ADEGENET 2.1.1 (Jombart, 2008), all metrics were calculated using SPAGEDI v.1.5. For ease of reference, we will refer to all of these 10 metrics jointly as 'Other Metrics'.

2.4 | Spatial datasets

We analysed nine environmental features per dataset (Table S1) that fell into three categories (see original publications for ecological justifications regarding choice of the features; abbreviations correspond to datasets: B = BOARS; F = FOXES; L = LIZARDS). **Biotic factors**: (i)

green areas: all types of arable land and grassland, fallow land, allotments, airports, public parks, cemeteries and bare soils [F,L]; (ii) *forests*: all forest irrespective of their composition [B,F]; (iii) *canopy cover*: all forested areas plus all street trees [L]; (iv) *hetero. agri.*: heterogeneous agricultural areas [B] (v) *arable*: arable land and permanent crops [B]; (vi) *pastures*: pastures [B]; **Anthropogenic infrastructure**: (vii) *motorways* [B,F]; (viii) *roads*: all motorways, primary and secondary roads [L]; (ix) *railways*: all railway lines [B,L], including major stations [F]; (x) *urban areas*: all artificial surfaces [B]; (xi) *sealing level (S.L.)* > 80%: sealing level > 80%: continuous urban fabric [F,L]; (xii) *S.L.* 50%–80%: sealing level 50%–80%: discontinuous dense urban fabric [F,L]; (xiii) *S.L.* 30%–50%: sealing level 30%–50% discontinuous medium dense urban fabric [F]; (xiv) *remaining built-up*: the remaining Urban Atlas categories not covered by the previous categories [F,L]; **Abiotic factors**: (xv) *water cover*: all water bodies [B,F,L]; (xiv) *slope*: calculated from aggregated digital elevation models [B,L] (Figures S1–S3).

The proximity of a sampling location to the boundary of the study area can erroneously constrain the predicted current flow during optimisation (Koen et al., 2010). The extent of each study area was therefore increased to include a buffer around a minimum convex polygon of the sampling locations (BOARS: 20 km; FOXES: 5 km; LIZARDS: 1 km). The buffer distance reflects the upper margin of known, regular dispersal distances of the focal species in the specific area or in the type of habitat considered (Prévot & Licoppe, 2013; Schulte, 2008; Trewhella & Harris, 1990). Using ArcMap v.10.3 (ERSI Inc.), the buffered study extents were converted into rasters, covering an area of 23,932 km^2 with a resolution of 750 × 750 m in the case of the BOARS, of 1206 km^2 at 200 × 200 m resolution for FOXES and of 34 km^2 at 40 × 40 m resolution for LIZARDS. We increased grid cell size for FOXES and LIZARDS compared to the original publications to reduce computational load.

2.5 | Creating landscape resistance models: preparation of single-feature surfaces

We first assessed the effects of single environmental features. To generate the relevant input rasters for each study area, we converted a shapefile containing all the environmental features of interest (except *slope*) into a raster by categorically classifying each cell as the single feature with the largest area within the cell. However, we ensured that, when a grid cell overlapped with a 'linear' feature (*motorways*, *roads*, *railways*, *water cover*), it was categorically assigned to the linear feature, irrespective of the proportion of the cell this feature occupied. When two or more linear features overlapped with a grid cell, the cell was classified as the linear feature that occupied the greatest proportion of the cell. Following this approach, each raster cell was only assigned a single categorical value. The ensuing raster with all the categorical features was re-classified into single-feature resistance surfaces where grid cells had a value of one or zero, depending on whether it contained the feature of interest. For BOARS and LIZARDS only, a continuous raster of *slope* was generated separately (Figures S1–S3).

2.6 | Landscape genetic analyses: single-feature optimisations

Optimisation of resistance values was conducted using ResistanceGA 4.0.14 (for single-feature analyses of empirical datasets) and ResistanceGA 4.1-0.46 (for multi-feature analyses of empirical datasets and all analyses of the simulated datasets). We used the *SS_optim*() function to optimise the resistances of all single-feature surfaces. We used the *commuteDistance*() function of the R-package gdistance v.1.2-1 (Van Etten, 2017) to calculate pairwise current flow between locations on resistance surfaces, which is equivalent to circuit-theory-based current flow (Kivimäki et al., 2014). Using the *GA.PREP*() function, we defined a range of 1–500 for the resistance values of categorical surfaces to be assessed during optimisation. We refrained from expanding the explored parameter space above 500 when optimised resistance values were at or near this limit, as we did not aim to identify precise resistance values, but rather aimed to quantify the variation of landscape genetic inferences across genetic distance metrics. Limiting the explored parameter space also reduced computational load.

To account for stochastic parameter estimation of the ResistanceGA algorithm, each optimisation was run twice. We chose the higher-ranking model of these paired optimisation runs for a (pseudo-) bootstrap procedure with the *resist. boot*() command. This function sub-samples a user-specified proportion of the locations without replacement, refits the MLPE model for each optimised resistance surface and re-calculates AICc values. The objective of the procedure is to identify the top-ranking surface across all bootstrap iterations and thereby increase the robustness of model selection, independently of the combinations of samples in the input file. We performed the bootstrap procedure simultaneously for all single-feature resistance surfaces of a specific combination of dataset and genetic distance metric, sampling 75% of the observations at each of 1000 iterations. Model support was evaluated based on AICc: if the difference in AICc (ΔAICc) between two models was >2 AICc units, the model with the smallest AICc value was considered to be better supported. Hence, if the difference in AICc between the *distance-only* model (i.e. the model based on Euclidean distances between all pairs of individuals) and an individual predictor was <2 AICc units, the single-feature model was not better supported than the *distance-only* model.

2.7 | Creating landscape resistance models: preparation and analysis of multi-feature surfaces

Next, we generated multi-feature resistance surfaces by sequentially adding environmental features to a resistance surface. The principle underlying this stepwise approach is to add predictors based on model support (AIC_c values) in the single-feature analysis but to only retain a new predictor in the multi-feature model if its

addition improved model support ($\Delta AIC_c > 2$ after bootstrap analysis) relative to the previous multi-feature model. We started with the highest-ranking single-feature models but only considered environmental features whose model support was $\Delta AICc > 2$ with the *distance-only* model in the single-feature analysis. We performed multi-feature optimisations by applying the single-surface optimisation *SS_optim*() procedure to multi-categorical resistance surfaces. Each grid consisted of *N* categorical predictors and each cell in the grid was assigned a value from zero to *N*, depending on whether it was classified as one of the predictors or as matrix, that is, the remaining study area.

Kimmig et al. (2020a) have shown that the statistical support of multi-feature categorical models was sensitive to the starting resistance values of the input surface. It would have been impractical to perform optimisations of grids containing all possible combinations of starting values to identify the combination with the highest support. We thus followed the approach by Kimmig et al. (2020a) where the initial value assigned to a categorical predictor in a multi-feature surface depended on its resistance/permeability value inferred in the initial single-feature analysis. For example, when manually combining two different categorical predictors in a single grid, a predictor inferred to be permeable in the initial analysis was given a value of zero in the multi-feature grid, a predictor resisting gene flow a grid value of two and the remaining cells a value of one. If multiple predictors were shown in the initial analysis to, say, resist gene flow, the feature with the highest initially inferred resistance value was given the highest value in the multi-feature grid, followed by the predictor with the second-highest initial resistance value and so on.

We performed each optimisation twice but only included the run with the lowest AICc value in bootstrap analysis. In the case of the BOARS and LIZARDS, we also included a continuous resistance surface (*slope*) in the analyses. When it was the turn to consider *slope* in the stepwise optimisation procedure, we used the multi-surface *MS_optim*() command to combine the previous multi-feature resistance surface with the continuous *slope* raster. If the addition of *slope* improved model support, we added the next categorical environmental feature to the multi-feature resistance surface and continued with the *MS_optim*() command to combine both rasters. Otherwise, we only optimised the new multi-categorical surface using the *SS_optim*() command. In some instances, the best-supported single-feature resistance surface remained the best model after the stepwise optimisation procedure. For ease of reference, we will nevertheless refer to all final models as 'multi-feature' models (OMFRS).

We assessed the similarities of the final OMFRS obtained for different genetics distance metrics based on correlation analyses. We used the *layerStats*() command in R package RASTER v.3.6.11 (Hijmans, 2022) to estimate Pearson correlation coefficients ρ for pairwise comparisons of all OMFRS within each dataset and plotted matrices of correlation coefficients using R package CORRPLOT v.0.92 (Wei & Simko, 2021). Cases where no single-features model was better than the *distance-only* model were omitted from this analysis. Plots were generated using the R-packages GGPLOT2 v.3.1.1 (Wickham, 2016) and TREEMAP v.2.4.3 (Tennekes, 2021). We used a Scheirer-Rare-Hare extension of the Kruskal–Wallis test (Dytham, 2011) in R package RCOMPANION v.2.1.7 (Mangiafico, 2019) to test for the effect of the dataset and the type of genetic distance metric (FCA- or PCA-based metric, Other Metric) on the number of single-feature models that explained the distribution of genetic distances better than the *distance-only* model as well as on the number of features that were included in the final OMFRS.

2.8 | Generation of simulated data

We conducted spatially explicit demo-genetic simulations using CDPOP v.1.3.12 (Landguth & Cushman, 2010; Figure 2) in order to objectively quantify the ability of the different genetic metrics to re-create true landscape resistance. In order to investigate realistic simulation scenarios, we based simulation settings on the LIZARDS dataset. We simulated 100 generations of reproduction and dispersal among 3500 individuals in a grid based on the original LIZARDS study area and some of the surrounding landscape (Figure S3). We simulated 16 loci with 12 alleles each to mimic the original dataset (16 loci and an average of 11.8 alleles per locus). In order to generate the cost-distance matrices required by CDPOP, we created resistance surfaces consisting of five different categorical environmental features (derived from the empirical lizard dataset): (a) *green areas*: obtained by merging the *canopy cover* and *forests* features; (b) *transport infrastructure*: obtained by merging the *railways* and *roads* features; (c) *remain. built-up*; (d) *urban dense*: obtained by merging the *S.L.* >80% and *S.L.* 50%–80%: features; (e) *water cover*. To decrease computational load for simulations and for subsequent optimisations, we increased the cell size of the resistance grid by a factor of three (120 × 120 m resolution), while preserving continuity of linear features in the landscape (Figure S3, Appendix S1). It has been shown previously that spatial grain has limited effect on landscape genetic inference (Cushman & Landguth, 2010; Winiarski et al., 2020), and we therefore expect minimal effects on results.

We then ran simulations where an individual's dispersal probability was determined by a dispersal kernel (10^{-x}; see CDPOP manual) and cumulative resistance between source and target grid cells (Landguth & Cushman, 2010). Following Beninde et al. (2016a), we excluded the major river and very urbanised areas from carrying lizards, while dispersal through these areas was permitted. We set the mean clutch size (reproductive ability) to four (Ji & Braña, 2000). At the end of a simulation, we only considered the cells where lizards had been caught in the field and subsampled the same number of individuals per cell as in the empirical dataset (223 individuals) to generate the genetic datasets used in the downstream landscape genetic analyses.

In order to quantify the ability of the genetic distance metrics to re-create true landscape resistance in different ecological and demographic contexts, we varied three hierarchical factors in simulations, *fragmentation*, *resistance* and *isolation-by-distance*, resulting in a total of 12 different modelling scenarios (Figure 2). At the

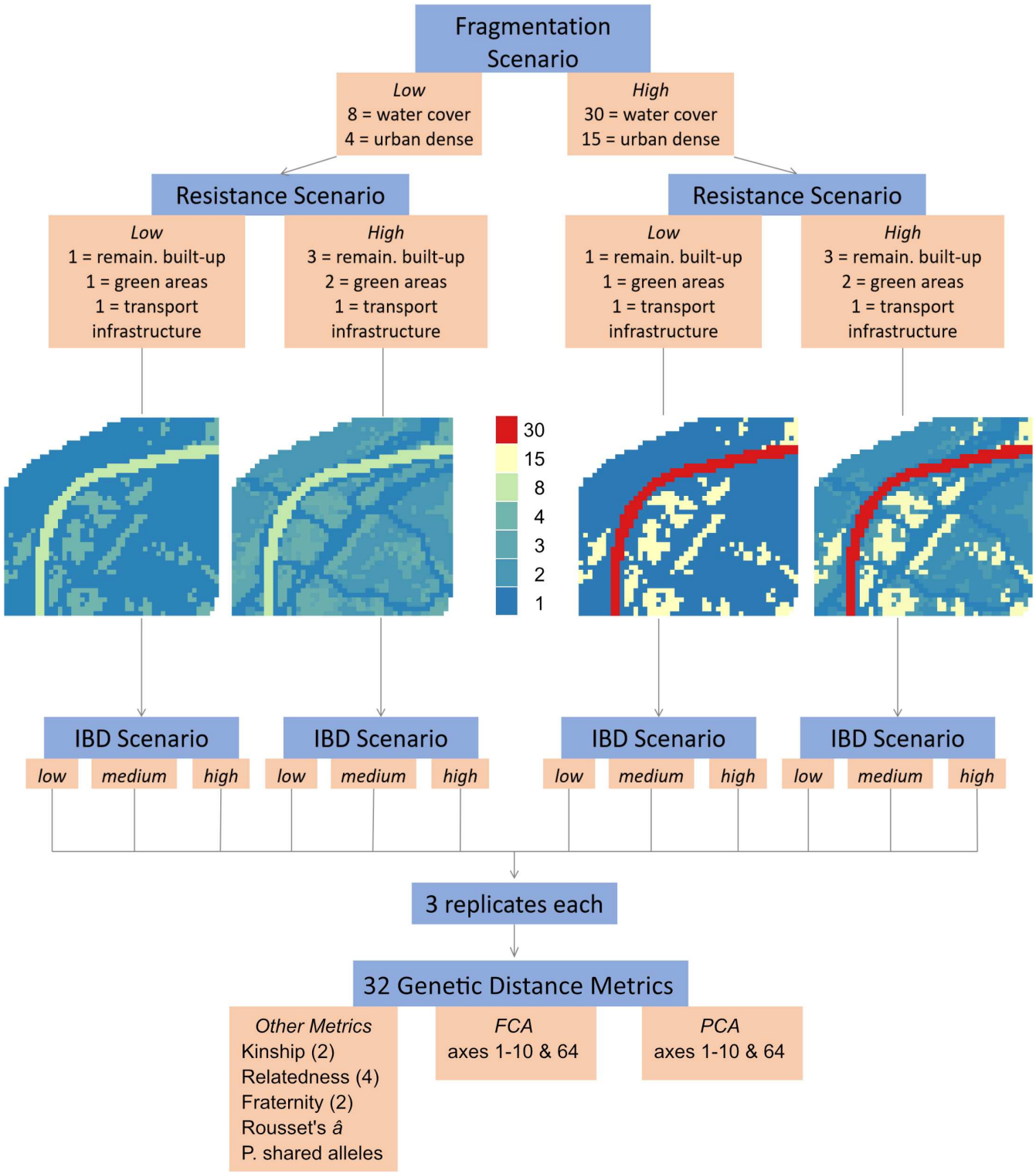

FIGURE 2 Overview of the simulation datasets. Fragmentation and Resistance Scenarios describe the different ways the landscapes were specified. IBD scenario and replicates refer to the settings used in CDPOP to generate the simulated datasets. The 32 Genetic distance metrics refer to the analysis framework applied to each of the simulated (and empirical) datasets, conducted using ResistanceGA.

uppermost hierarchical level, the *fragmentation scenario*, the two features with the highest resistance were either weakly (resistance values: *water cover* = 8, *urban dense* = 4; *low-fragmentation* scenario) or strongly (*water cover* = 30, *urban dense* = 15; *high-fragmentation* scenario) resisting gene flow. Within each of these two scenarios, we further simulated a *low-resistance scenario* in which the remaining three features did not impact gene flow (all resistance values = 1) as well as a *high-resistance* scenario where two of the three remaining features weakly impacted gene flow (*remain. built-up* = 3, *green areas* = 2, *transport infrastructure* = 1). For each combination of

fragmentation and *resistance* scenario, we further simulated three different strengths of IBD by changing dispersal function parameters, as this was previously shown to influence accurate landscape genetic inferences (Shirk et al., 2017a). The slope of this regression in LIZARDS was estimated as $b = -0.013$ (SE = 0.003; see above). In the *high-IBD* scenarios, IBD slopes were $b > -0.013$, in the *medium-IBD* scenario slopes ranged from $-0.010 < b \leq -0.013$ and slopes in the *low-IBD* scenario were $b < -0.010$. We simulated three replicates for each of these 12 modelling scenarios, resulting in 36 simulated datasets.

2.9 | Analysis of simulation results

2.9.1 | ResistanceGA optimisations

The 36 simulated genetic datasets were analysed using ResistanceGA, in the same way as the empirical datasets for single- (see Section 2.6) and multi-feature optimisations (see Section 2.7), and using the same 32 genetic distance metrics as with the empirical datasets (see Section 2.3). In combination, this yielded 1152 analyses. The result of each of these analyses was a final optimised landscape resistance surface (OMFRS), identified following the model-selection framework laid out in sections 2.6 and 2.7. We interpret and refer to this as the 'optimised resistance surface', which we compared to the simulation-underlying resistance model, the 'true resistance surface', using raster correlations. The correlation coefficient between the true and the optimised resistance surfaces is a direct measure of the ability to re-create true resistances following ResistanceGA optimisations, and we refer to this correlation as 'resistance recovery'. Resistance recovery values close to one indicate the optimised resistance was close to the true resistance, while negative values, or those close to zero, indicate no, or little, resistance recovery respectively. In cases when the best-supported model overall was the *distance-only* model, that is, no optimised resistance surface had $\Delta AICc > 2$, correlations to the true raster cannot be performed. In these cases, we assigned a resistance recovery of 0, accounting for the fact that the distance model was a bad fit to the true resistance while allowing us to include the results for further analyses.

2.9.2 | Identifying factors most important for accurate resistance recovery

The next set of analyses quantified the importance of the genetic distance metrics and the simulation scenarios' ecological and demographic contexts for resistance recovery. We first used a Kruskal–Wallis test to check for differences in resistance recovery between the three replicate runs performed for each set of simulation scenarios. In the next step, we used two different approaches of recursive partitioning to identify the most important variables determining resistance recovery. Both approaches used resistance recovery as the response variable and the genetic distance metrics and all simulation scenarios (*replicate*, *fragmentation*, *resistance* and *IBD*) as predictor variables, to create homogenous outcome groups.

The first approach used Random Forest models (Breiman, 2001) to quantify the overall amount of variation in resistance recovery explained by predictor variables, the relative importance of the predictor variables (based on the increase in mean-squared error when variables are omitted, %IncMSE) and the estimated association between feature levels of predictor variables and resistance recovery (using partial dependence plots). The machine-learning regression tree-based algorithm implemented in Random Forest models is particularly well-suited for large datasets and has high predictive ability (Buri et al., 2022). Models were fitted with ntrees = 10,000 and the default value of mtry using the *randomForest*() function in the R package by the same name (v4.7-1; Liaw & Wiener, 2002).

The second approach was to create more readily interpretable results to use for further analyses and to provide guidance for empirical approaches. For this, we made use of the conditional inference framework implemented by *ctree*() in the R package partykit v.1.2-16 (Hothorn & Zeileis, 2015), to generate a single decision-tree. These models are comparable in predictive ability of ensemble-based models of single trees, as produced by Random Forest, but allow straightforward interpretation (Buri et al., 2022). We generated Random Forest models for the full dataset, and ctree models for a subset of the genetic distance metrics only (excluding the two lowest performing metrics, see Section 3), both jointly and separately for the three simulation replicates.

2.9.3 | Association between true resistance recovery and the convergence of OMFRS

The ctree model predicts the response variable for each terminal node, given the predictor variable levels at splits of the decision tree (Hothorn & Zeileis, 2015). In our models, the response variable is the value of resistance recovery. To test if resistance recovery is linked to the convergence of all optimised rasters falling within terminal nodes, we calculated the median pairwise correlation coefficient of all OMFRS within terminal nodes and correlated it to the predicted resistance recovery. Such an association would allow to directly compare simulation results to empirical data, by calculating the median of pairwise correlation coefficient of all OMFRS for a given empirical dataset.

2.9.4 | Association between resistance recovery and marginal R^2

In order to understand if the mR^2 of models is associated with the resistance recovery achieved by models, and thus a suitable metric to identify the best genetic distance metric to use for a given dataset, we used generalised linear models and model selection based on Akaike Information Criterion (AIC). In all models, we specified the resistance recovery of the OMFRS as the response variable and

the following as predictor variables: only the mR^2 of models (model 1), both the mR^2 and the genetic distance metrics (model 2), and both the mR^2, the genetic distance metrics and their interactions (model 3). Models were run with the glm() function and ranked using the AIC() in stats package in R, using a gamma distribution that best fitted the negative exponential transformation of the response variable.

3 | RESULTS

3.1 | Single-feature optimisations

There was substantial variation in the number of single-feature models that explained the distribution of genetic distances better than the *distance-only* model ('better-than-distance models') across all genetic metrics (Figure 3a). The Scheirer-Ray-Hare extension of the Kruskal–Wallis test showed that there was an effect of the genetic metric category on the number of better-than-distance models (Table 1). The FCA- and PCA-based metrics generally inferred a higher number of better-than-distance models than the Other Metrics and the Other Metrics were less consistent in the number of inferred better-than-distance models (Table 1, Figure 3a).

Independently of the dataset, the AICc-based model rank of single-feature models (only considering better-than-distance landscape factors), varied substantially across the different genetic distance metrics (Figure S4). Seven different landscape features received highest AICc-based model support at least once in BOARS, six in the case of the FOXES and five in the case of the LIZARDS (Figure 4).

3.2 | Multi-feature optimisations

Across all genetic metrics, there was substantial variation in the number of better-than-distance environmental features that were included in the final multi-feature models (Figure 3b). The Scheirer–Ray–Hare extension of the Kruskal–Wallis test showed that there was an effect of the genetic metric category on the number of features included in the final multi-feature models (Table 2). Using an FCA- or PCA-based metric generally resulted in a higher number of features included in the final model than when using one of the Other Metrics (Figure 3b).

The final OMFRS did not converge across many genetic metrics and the degree of convergence differed between datasets. The highest correlation coefficients were obtained for pairwise comparisons of the final OMFRS in the LIZARDS dataset (median ± IQR: 0.875 ± 0.967; Figure 5e). Pairwise correlation coefficients were lower in BOARS (median ± IQR: 0.408 ± 0.600; Figure 5a) and FOXES (median ± IQR: 0.241 ± 0.734; Figure 5c). LIZARDS was also the only dataset where we systematically obtained high correlation coefficients across the three categories of

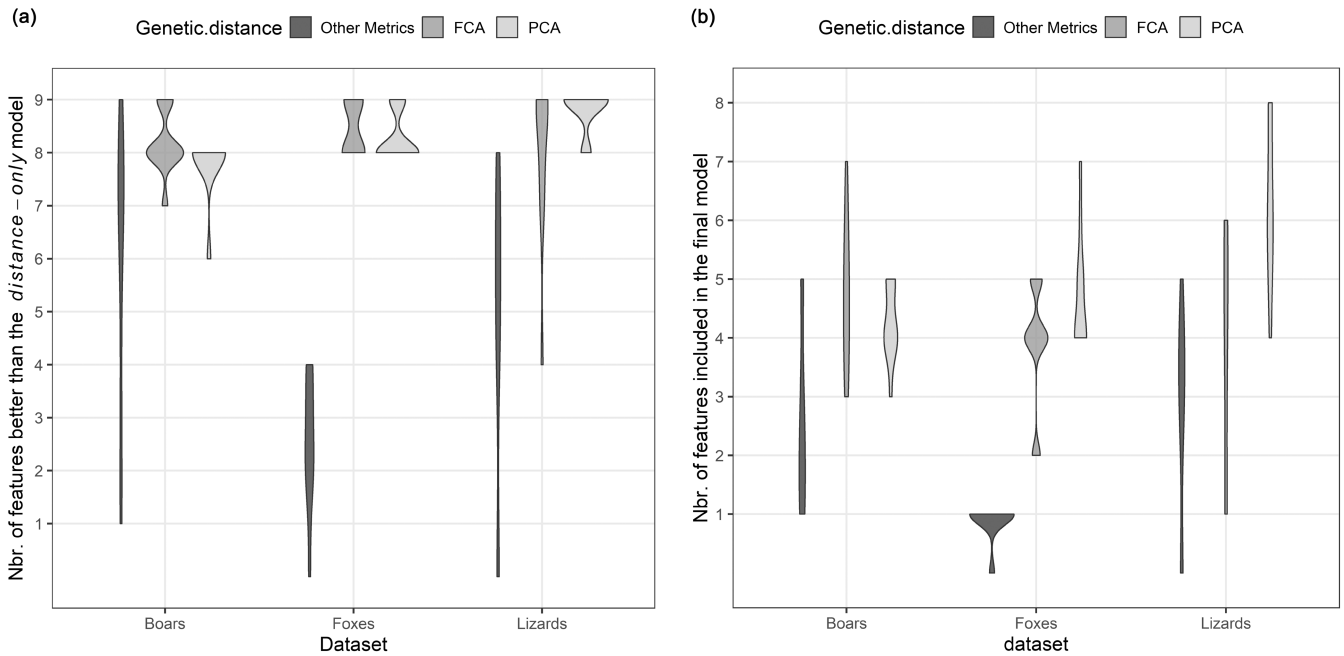

FIGURE 3 Number of features in models of empirical datasets. The number of environmental features are shown that impact the distribution of the genetic variation in the three empirical datasets after bootstrapping. (a) Number of models with a single environmental feature with higher AICc-based model support than the *distance-only* model. (b) Number of environmental features that are included in the final multi-feature model. FCA: Eleven metrics based on position of individuals on differing number of axes of a factorial correspondence analysis. PCA: Eleven metrics based on position of individuals on differing number of axes of a principal component analysis. Other Metrics: Eight metrics based on fraternity, kinship and relatedness coefficients in addition to Rousset's $\hat{a}$ and the proportion of shared alleles D_{PS}. For further information, please refer to the Section 2.

Source	df	SS	SS/MS_{total}	*p*-Value
Dataset	2	1323	1.865	.393
Genetic distance category	2	34,173	48.168	<.001
Dataset*Genetic distance	4	6265	8.830	.065
Residuals	87	25,637		
Total				

TABLE 1 Number of single-features better than the *distance-only* model in empirical datasets.

Note: Results from a Scheirer-Ray-Hare extension of the Kruskal–Wallis test for the number of environmental features that explain the distribution of genetic distances better than the *distance-only* model. We investigated the effects of dataset (BOARS, FOXES & LIZARDS; see Section 2) and type of genetic distance (Other metrics, FCA-based & PCA-based; see Section 2).

Abbreviations: df, degrees of freedom; MS_{total}, total of the sum of squares values divided by the total degrees of freedom; SS, sum of squares.

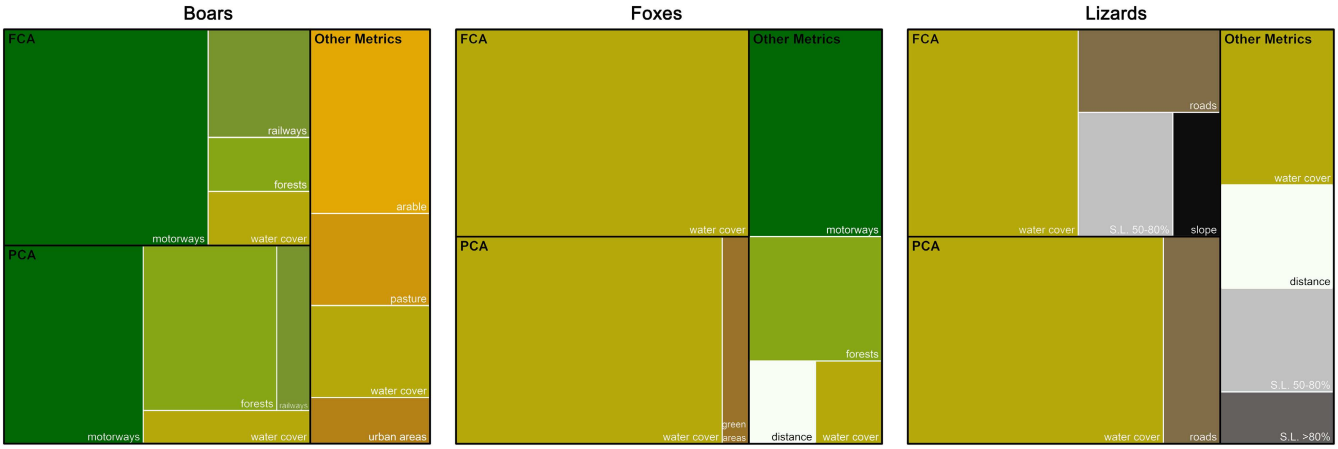

FIGURE 4 Frequency of highest-ranking models in empirical datasets. The frequency of single-feature models being identified as the AICc-based best-supported model in each of the three empirical datasets, ordered by the three different categories of genetic distance metrics. FCA: 11 metrics based on position of individuals on differing number of axes of a factorial correspondence analysis. PCA: 11 metrics based on position of individuals on differing number of axes of a principal component analysis. Other Metrics: Eight metrics based on fraternity, kinship and relatedness coefficients in addition to Rousset's $\hat{a}$ and the proportion of shared alleles D_{PS}. The size of each rectangle is proportional to the number of times a model was identified as the best-supported model. For further information, please refer to the Section 2.

Source	df	SS	SS/MS_{total}	*p*-Value
Dataset	2	3546	4.746	.093
Genetic distance category	2	28,602	38.286	<.001
Dataset*Genetic distance	4	5285	7.074	.132
Residuals	87	33,538		
Total				

TABLE 2 Number of features included in final, multi-feature models of empirical datasets.

Note: Results from a Scheirer-Ray-Hare extension of the Kruskal–Wallis test for the number of features that are included in the final multi-feature model. We investigated the effects of dataset (BOARS, FOXES & LIZARDS; see Section 2) and type of genetic distance (Other Metrics, FCA-based & PCA-based; see Section 2).

Abbreviations: df, degrees of freedom; MS_{total}, total of the sum of squares values divided by the total degrees of freedom; SS, sum of squares.

distance metrics (Figure 5f). In the FOXES dataset, we only obtained highly correlated resistance surfaces within the same category of distance metric (Figure 5d). While there was some good convergence between a few individual metrics, it was difficult to discern an overall pattern in the correlation coefficients in the case of the BOARS (Figure 5f).

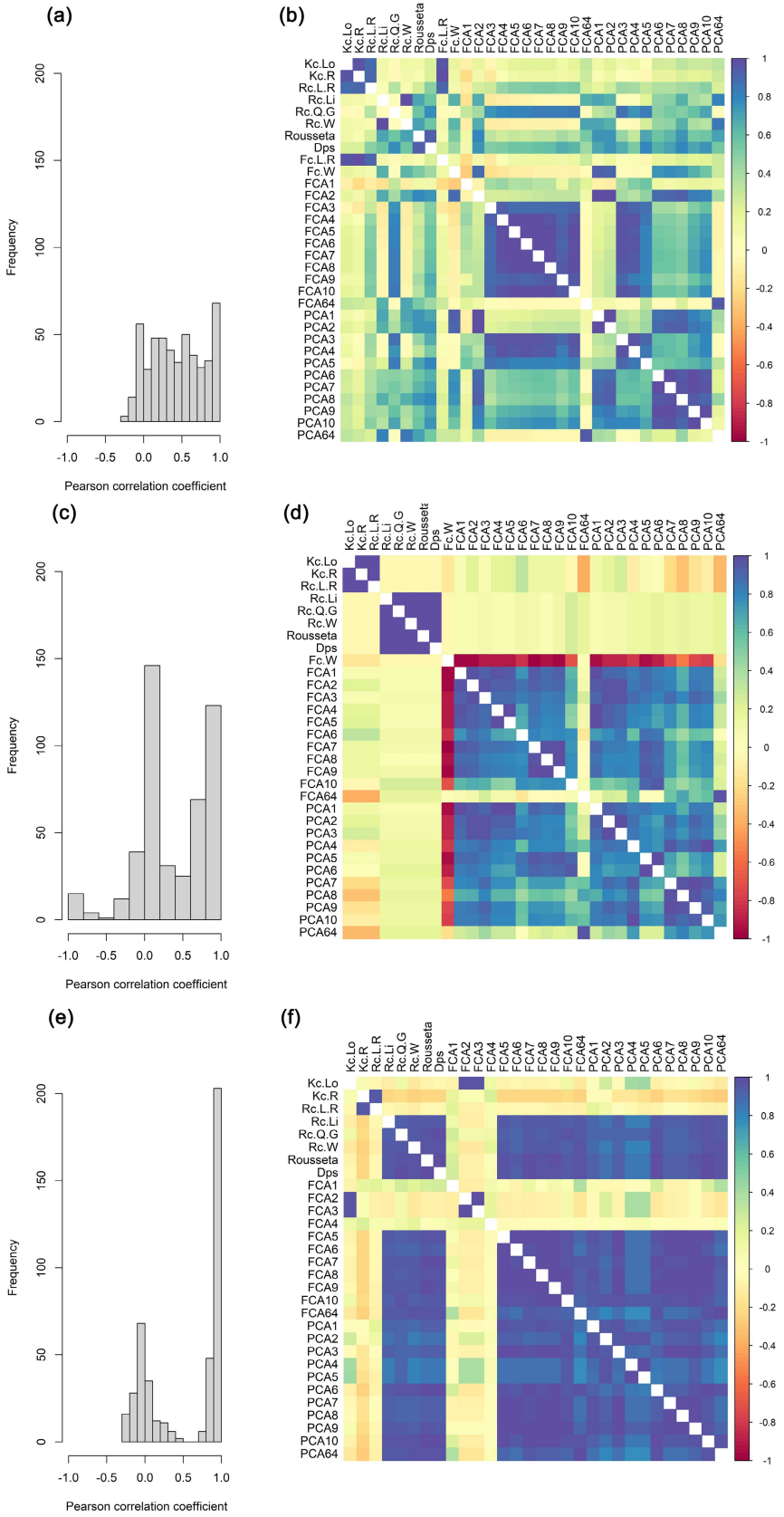

FIGURE 5 Similarity of multi-feature models of empirical datasets. Distribution and matrix analysis of the Pearson correlation coefficients (ρ) obtained for pairwise comparisons of the final optimised resistance surfaces generated using different genetic distances of the empirical datasets. Histogram showing the distribution of ρ in the (a) BOARS, (c) FOXES and (e) LIZARDS datasets and correlation matrix of ρ in the (b) BOARS (d) FOXES and (f) LIZARDS datasets. The scale colours in the correlation matrix denote whether the correlation is positive (closer to 1, dark blue) or negative (closer to −1, dark red). Explanations on the precise nature of the different genetic distance metric can be found in the Section 2.

3.3 | Simulation results

3.3.1 | Differences in replicate runs

Resistance recovery was not significantly different between the three replicate runs (Kruskal–Wallis: $\chi^2 = 3.294$, df = 2, $p = .193$).

3.3.2 | Basic summary statistics of resistance recovery for genetic distance metrics

Grouping the genetic distance metrics into Other, FCA and PCA and averaging the minimum, median and maximum resistance recovery of all models within these categories showed that Other Metrics have both the highest maximum and median resistance recovery (0.999 and 0.867 respectively). FCA- and PCA-derived metrics performed similarly, in general, and achieved comparable maximum resistance recovery to Other Metrics (0.988 and 0.989 respectively) and slightly lower median resistance recovery (0.813 for both). However, Other Metrics were more variable and obtained strikingly lower minimum resistance recovery (−0.826) than both FCA- and PCA-derived metrics (−0.182 and −0.212 respectively).

3.3.3 | Recursive partitioning of resistance recovery

Random Forest models of the full dataset explained 52.2% of the variation and identified *resistance* scenario (56.2% IncMSE), genetic distance metric (41.2%) and *fragmentation* scenario (35.6%) as the most important predictor variables, in descending order, while *replicate* (17.5%) and *IBD* scenario (12.9%) were less important. Partial Dependence of genetic distance metrics further identified particularly poor resistance recovery by the fraternity coefficients Fc.L&R and Fc.W (Figure S5), which we subsequently dropped from all further analyses. For this subset of the data, *resistance* scenario remained the most important factor, followed by *fragmentation* scenario, based on %IncMSE-rank of Random Forest models and the highest hierarchy-level of ctree models (Figure S6). The importance of the remaining predictors differed and both *IBD* scenario and *replicate* became more important, while *genetic distance metric* ranked lowest in %IncMSE and only appeared in one node in the third hierarchy level. We therefore also generated Random Forest and ctree models for each of the three simulation replicates separately.

Random Forest models across the three replicates varied in the amount of variation explained (replicate 1 = 48.0%, replicate 2 = 61.9%, replicate 3 = 44.6%). Models of all replicates identified *resistance* scenario as the most important predictor variable, that is, with highest %IncMSE, and genetic distance metric as the least important, while *fragmentation* and *IBD* scenario varied in %IncMSE rank. Ctree models similarly identified resistance scenario as most important (1st hierarchy level) in models for all replicates, genetic distance metric as least important (only featured in one node across models for all replicates), while fragmentation and IBD scenarios varied in importance (Figure 6; Figures S6–S8).

Based on the dataset of all replicates, resistance recovery was among the lowest in all *high-resistance* & *low-fragmentation* scenarios (Figure S6). Within the *high-resistance* & *high-fragmentation* scenario, resistance recovery was highest when using most of the Other Metrics, that is, Kc.Lo, Kc.R, Rc.L&R, Rc.Li, Rc.Q&G, Rc.W and D_{PS}, and lower when using Rousset's $\hat{a}$ or any FCA- or PCA-derived metric. Based on the data for replicate 1 (Figure S7), genetic distance metrics were significant in one sub-node of the *low-resistance* scenario, and showed higher resistance recovery for all FCA- and PCA-based metrics (except those based on 64 axes) than for the remaining metrics.

3.3.4 | Association between true resistance recovery and the convergence of OMFRS

Correlation analyses of the median pairwise correlation coefficient of all OMFRSs within terminal nodes of ctree models and the predicted resistance recovery of terminal nodes were performed for the full dataset, separately for all three replicates and jointly for all three replicates. Results for the full dataset were insignificant ($\rho = .342$, df = 14, p-value = .195; Figure 7a). Correlation results for the replicates analysed separately varied considerably (replicate 1: $\rho = -.094$, df = 11, p-value = .759; replicate 2: $\rho = .863$, df = 3, p-value = .059; replicate 3: $\rho = .866$, df = 4, p-value = .026). Correlation results for the combined data of all three replicates were significant ($\rho = .47$, df = 22 and p-value = .021; Figure 7b).

3.3.5 | Association between resistance recovery and marginal R^2

There was a positive and significant association of resistance recovery and mR^2 of models (Figure S9). The model with the highest AIC support (Table 3), also showed significant interactions of mR^2 and genetic distance metrics, rendering mR^2 an unsuitable measure by which to choose the best model among those generated with different genetic distance metrics. As shown in Figure S9, Other metrics had lower mR^2 than PCA- or FCA-derived models, while having higher median resistance recovery values (see above).

4 | DISCUSSION

Landscape genetic approaches need to deliver ecologically meaningful and methodologically robust results in order to provide scientific guidance for conservation applications (Keller et al., 2015). A large body of simulation studies shows that conclusions of landscape genetic studies can be impacted by a variety of methodological and statistical choices (Peterman et al., 2014, 2019; Richardson et al., 2016; Shirk et al., 2017a, 2017b). Surprisingly, the effect of genetic distance

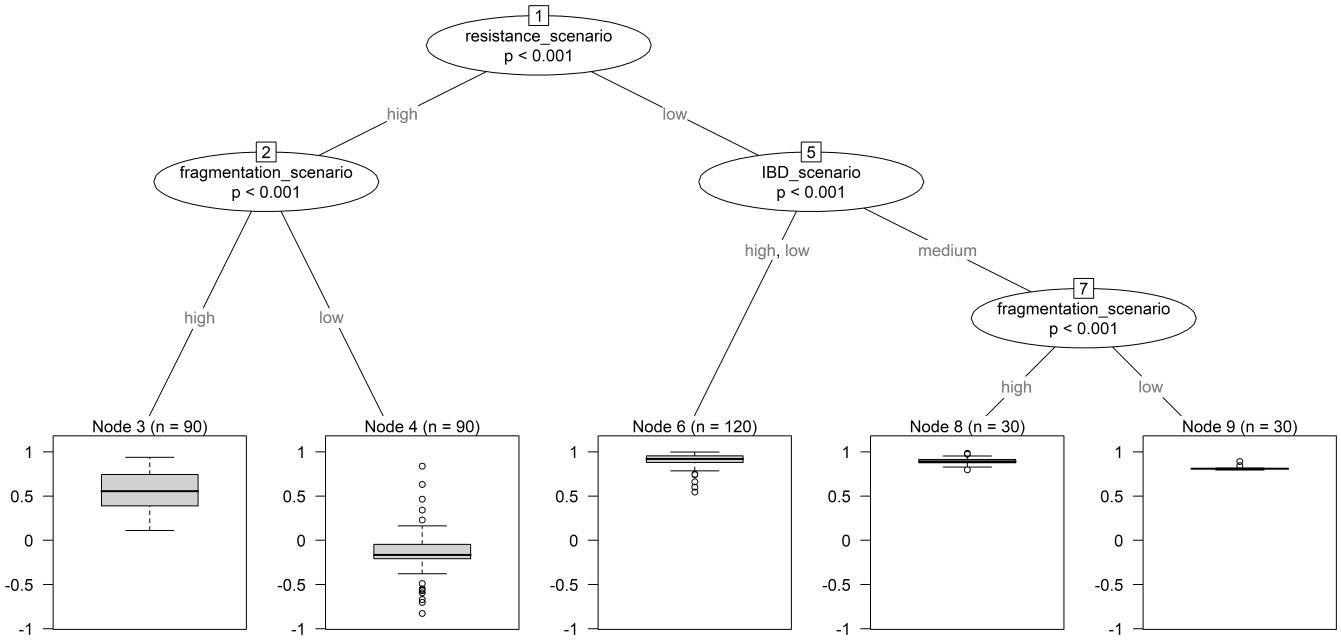

FIGURE 6 Recursive partitioning results of the simulated datasets. This ctree analysis shows the overarching importance of the *resistance-*, *fragmentation-* and *IBD-scenarios* for the accurate parameterization of landscape resistance surfaces. The analysis shown here is based only on the data of replicate 2 of the simulation datasets and excludes data generated using the fraternity coefficients Fc.L&R and Fc.W (see Section 3). This ctree analysis is representative of the main simulation results and emphasises that the choice of genetic distance metric may only be marginally important across landscape scenarios (see Figures S6–S8 for ctree analyses of the full simulation dataset, replicate 1 and replicate 3). Terminal nodes show resistance recovery in boxplots, the response variable, with a value of one indicating perfect recovery of the true landscape resistance surface. In summary, high resistance recovery can be expected in low resistance scenarios, and low resistance recovery in high resistance scenarios.

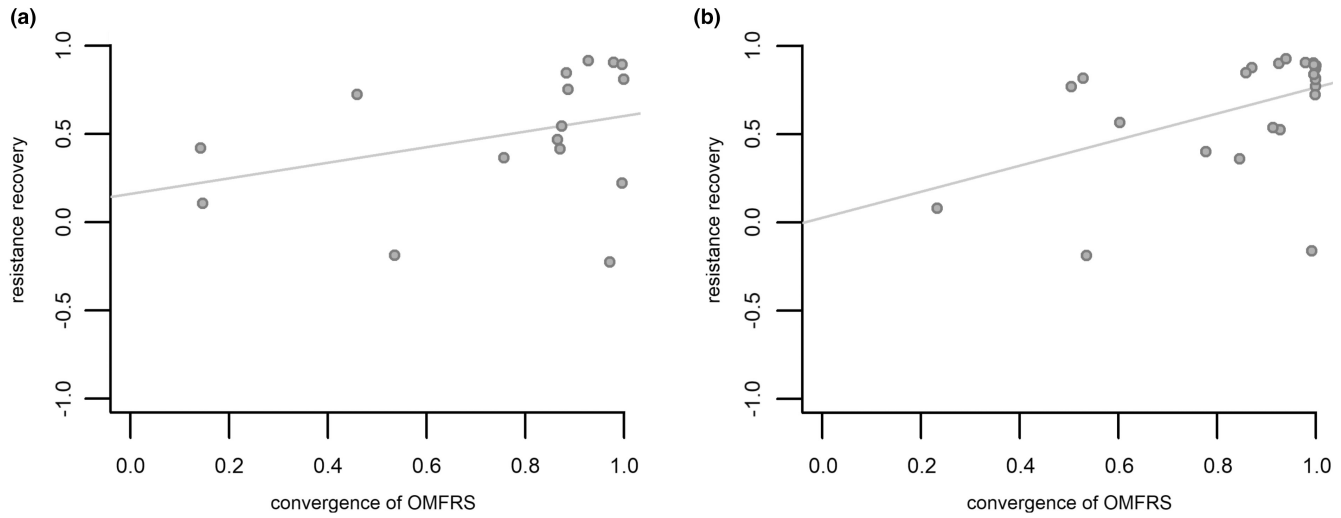

FIGURE 7 Resistance recovery and convergence of OMFRS in simulated datasets. The association of resistance recovery and the convergence of optimised multi-feature resistance surfaces of the simulated dataset, calculated separately for (a) terminal nodes of ctree models for the full dataset (see Figure S6) and (b) the combined terminal nodes of all ctree models run separately for replicates 1–3 (Figure 6; Figure S7 and S8).

metrics on landscape genetic inferences has received relatively little attention (but see Shirk et al., 2017a; Winiarski et al., 2020). We here provide evidence for a strong impact of the choice of genetic distance metric in empirical datasets and demonstrate via simulations that the ability to correctly identify true landscape resistance, which we refer to as resistance recovery, strongly depends on the characteristics of the landscape under consideration. Our simulation scenarios demonstrate further that resistance recovery can be low, irrespective of the utilised genetic distance metric. Interpreting the empirical findings in light of simulation results leads us to conclude that accurate resistance recovery can be limited, even when study designs are well-suited for landscape genetic inference.

TABLE 3 Association of resistance recovery and marginal R^2 in simulation datasets.

GLM	df	AIC
Resistance recovery ~ mR^2 + genetic distance metric + mR^2 * genetic distance metric	3	1639.9
Resistance recovery ~ mR^2 + genetic distance metric	34	1665.9
Resistance recovery ~ mR^2	65	1814.7

Note: Results of GLMs testing the association of resistance recovery (i.e. the accuracy of re-creating the true landscape resistance surface), marginal R^2 of models and the genetic distance metric used to generate models (11 PCA- and FCA-based metrics and 10 Other metrics), ranked by AIC.

Comprehensive single-feature analyses of three individual-based empirical datasets demonstrated a strong impact of the choice of genetic distance metric on the results of landscape genetic inferences. Given the large number of environmental features and ensuing analyses, we followed a stepwise optimisation procedure, which first optimised single-feature resistance surfaces and then created multi-feature surfaces based on model support in the single-feature analysis, dropping all models that were no better than the *distance-only* model. This was a more viable approach than testing model support for all possible combinations of all single-feature models, as it greatly reduced computational load. We found significant differences between genetic distance metrics in the number of single-feature models that were better than the *distance-only* model (Figure 3a), and thus, the number of single-feature models considered in multi-feature analyses. The choice of genetic distance metric was similarly important in identifying the best-supported single-feature model, which varied substantially across all metrics (Figure 4). Users, therefore, need to be cautious in the interpretation of results that appear counter-intuitive in light of the biology of the species, especially when the best-supported single-feature model(s) did not correspond to environmental features causing major genetic discontinuities, for example, those identified by clustering methods (e.g. Kimmig et al., 2020a, 2020b).

It is recommended to extend landscape genetic inferences beyond single-feature models, more generally, to capture landscape resistance more comprehensively (Peterman & Pope, 2021). However, multi-feature results for the three empirical datasets also demonstrated a strong effect of genetic distance metrics on the outcome of multi-feature modelling approaches. Similar to single-feature analyses, there were large differences in the number of environmental features included in the final optimised models (Figure 3b), and pairwise correlation coefficients of all final OMFRS varied greatly across genetic distance metrics (Figure 5). Therefore, the usage of some genetic distance metrics must have led to erroneous OMFRS. Furthermore, there was no combination of genetic distance metrics across datasets that consistently gave rise to highly correlated resistance surfaces. For the most part, FCA- and PCA- derived metrics provided more similar models, which were very different from models generated using Other metrics (except in LIZARDS, see Figure 5f). Our empirical results are thus rather sobering and imply limited robustness of inferences of landscape genetic resistance across individual-based genetic distance metrics.

A more nuanced picture emerges, however, when considering simulation results. Recursive partitioning of results showed particularly poor resistance recovery by the fraternity coefficients Fc.L&R and Fc.W. (Figure S5), corroborating findings by Shirk et al. (2017a), and we recommend they find no further application in landscape genetic studies. Beyond that, the choice of genetic distance metric is of lesser importance for high-resistance recovery and only becomes relevant within specific simulation settings. Overarchingly, results of the *low-resistance* scenario, that is, when only two of the five environmental features were simulated to restrict gene flow (Figure 2), produced high resistance recovery, robustly across genetic distance metrics. In contrast, in the *high-resistance* scenarios, that is, when all five environmental features were simulated to restrict gene flow, resistance recovery was generally low. Further exploring results specifically with respect to genetic distance metrics showed that all FCA- and PCA-based metrics (except those based on 64 axes) can lead to slightly better resistance recovery in the *low-resistance* scenario, for example, when resistance recovery was already high. Further, in the *high-resistance & high-fragmentation* scenario, resistance recovery was markedly higher when using most of the Other metrics, that is, Kc.Lo, Kc.R, Rc.L&R, Rc.Li, Rc.Q&G, Rc.W and D_{PS}, as opposed to Rousset's *â* or any FCA- or PCA-derived metric. Thus, when working with empirical data and expecting *low-resistance* levels, resistance recovery may be increased when using FCA- or PCA-derived genetic distances based on axes 1–10. While, when *high resistance* and *high fragmentation* are expected, any of the Other metrics (except Rousset's *â*) may increase resistance recovery. We need to emphasise that these conclusions need to be further substantiated by expanding on the three simulation replicates presented here.

Other simulation studies have shown that the optimisation of resistance models followed by AICc ranking of LME models is methodologically sound and suited for pairwise distance matrices (Shirk et al., 2017b; Winiarski et al., 2020). The simulation study by Shirk et al. (2017a) came to different conclusions from those presented here, despite using a very similar set of distance metrics. These authors showed similarly high model selection accuracy of most genetic distance metrics across simulation scenarios (with the exception of Fc.L&R and Fc.W.), while a PCA metric derived from 64 axes led to recovering true resistance models most frequently, particularly in simulation scenarios that are more challenging for landscape genetic inferences. This is in contrast to the results presented here, which also show similar performance of most genetic distance metrics, but emphasise that performance varies greatly in different landscape scenarios. The fact that many metrics performed relatively well in Shirk et al. (2017a) might be an artefact of the results being based on a single simulated landscape, although this needs to be substantiated. Winiarski et al. (2020) found that, in principle, ResistanceGA had a high ability to recover true resistance surfaces under a variety

of simulation scenarios and using single- and multi-surface analyses. However, the authors used a synthetic genetic distance metric in order to better control levels of random noise introduced by variance in genetic distances. Similar to our study, the authors found that ResistanceGA had difficulty identifying the true resistance surfaces when simulating highly stochastic individual-based data.

Our simulation results create an obvious dilemma. Overarchingly, recovering accurate resistance surfaces appears to depend on the number of environmental features influencing gene flow or, in the words of the present study, on the *resistance scenario*. Knowledge of this, however, is usually limited, and more often than not, it is the explicit goal of empirical landscape genetic studies to generate such knowledge on patterns of landscape resistance in the first place. Moreover, our results clearly show that researchers should refrain from choosing a genetic distance metric based on the mR^2 of models (e.g. Kimmig et al., 2020a; Mignotte et al., 2021), as the GLM model with the highest AIC support (Table 3) showed significant interactions of mR^2 and genetic distance metrics.

The results from the simulation study broadly supported that one possible way of assessing the accuracy of landscape genetic inference in a given dataset might be the convergence of OMFRS obtained using genetic distance metrics. When there are high levels of convergence between OMFRS obtained from using different genetic distance metrics, resistance recovery is likely high, whereas resistance recovery is likely low when convergence is low (although we need to acknowledge that it is also possible to have high convergence on a poorly performing model, corresponding to low resistance recovery). The correlation between the median values of the pairwise correlation coefficients of all OMFRSs within terminal nodes and the mean resistance recovery predicted for terminal nodes showed positive associations with clear trends in recursive partitioning analyses, although some remained insignificant (Figure 7).

Applying this approach to the three empirical datasets, the highest value of convergence between OMFRSs was obtained for the LIZARDS dataset and indicated fair expected resistance recovery, while the level of convergence in BOARS and FOXES indicated low expected resistance recovery. LIZARDS was the only empirical dataset which systematically obtained high correlation coefficients of OMFRSs across the three categories of distance metrics, while in BOARS and FOXES FCA- and PCA-derived models had a generally low correlation with models generated with Other metrics. LIZARDS was also the only dataset where Bayesian clustering methods identified a genetic discontinuity that was clearly associated with a landscape feature (see Section 2.2 Genotype datasets). It thus appears plausible that the highly converging resistance surfaces accurately represent the landscape features acting as drivers of gene flow in LIZARDS. Given the lack of convergence between optimised resistance surfaces in the BOARS and FOXES datasets, we cannot confidently identify accurate resistance surfaces for them. One limitation of the present study is the single spatial scale used in analyses, as spatial scale can influence the inferred effects of landscape variables in landscape genetic studies (Angelone et al., 2011; Bauder et al., 2021). However, as OMFRSs did not converge at the spatial scale analysed here, we assume a similar influence of genetic distance metrics on results also at other scales.

Future work needs to disentangle the effect of genetic distance metrics on true landscape resistance more generally, especially to provide better guidance on which individual-based genetic distance metrics to use in different empirical settings. In line with Kimmig et al. (2020a), we chose mostly categorical surfaces for this study, however, it has been postulated that continuous surfaces are better suited for accurate inference of landscape resistance (Peterman et al., 2019). Furthermore, the modelling framework implemented here can be expanded to test all possible pairwise combinations of single features, using the multi-surface function in ResistanceGA. Simulations have further suggested an increase in model selection accuracy with lower variance in genetic distances and with population-based approaches (Winiarski et al., 2020). Given we here presented microsatellite-based data, single-nucleotide polymorphisms (SNPs) approaches may allow for more accurate resistance recovery, due to the expected lower variance estimates of SNP-based genetic distances (e.g. Galla et al., 2020; Lemopoulos et al., 2019). In addition, the question of whether inference based on population-based approaches is less affected by the choice of genetic distance metric needs to be investigated. Furthermore, a ResistanceGA-based empirical study has recently suggested that convergence of the results obtained with different population-based genetic distance metrics and AICc-based model support improved when using a least-cost path approach, rather than the current flow method implemented here, to calculate resistance distances (Reyne et al., 2023). The study was performed on natterjack toads (*Bufo camalita*) and further research will show whether this conclusion is generally valid or a result of the limited mobility of the study species.

Comparisons of ResistanceGA and more recently implemented recursive partitioning approaches to parameterise resistance surfaces will be particularly valuable to explore. One of these approaches utilises iterative Random forest models paired with least-cost transect analysis (LCTA) to parameterise resistance surfaces (Pless et al., 2021). Statistically, this is a very different approach from the genetic algorithm and MLPE models utilised in ResistanceGA but it also requires users to choose a genetic distance metric, with the resulting values used as the response variable. A Gradient forest approach, on the other hand, uses allele frequencies directly as the response variable, alleviating users from choosing genetic distance metrics (Vanhove & Launey, 2023). A first comparison of this method looks promising, showing Gradient forest approaches (as implemented in resGF), to outperform both ResistanceGA and Random forest modelling paired with LCTA in landscape genetic inference (Vanhove & Launey, 2023).

5 | CONCLUSION

The extensive optimisations presented here, using ResistanceGA on 36 simulated and three empirical datasets, could not identify a single genetic distance metric, or a group of metrics, that

systematically recovered accurate resistance surfaces across different landscape scenarios. Furthermore, simulation results indicate that mR^2 of models is not a suitable indicator for the best-performing genetic distance metric. In summary, our results call into question the ability to recover complex spatial resistance surfaces using simplified summary statistics of genetic distances, highlighting the need for new approaches to infer landscape resistance robustly across landscape scenarios and taxonomic groups. When using ResistanceGA, we emphasise the need to optimise resistance surfaces with different genetic distance metrics and assess the convergence of results via pairwise correlation coefficients of all optimised multi-feature resistance surfaces. These need to be checked for high coefficients across different genetic distance categories (Other, FCA and PCA) and can be compared to the simulated results presented here for expected true resistance recovery values. Most importantly, we want to call for a cautious interpretation of empirical data, especially when results appear counter-intuitive in light of the ecology of species, (e.g. Peterman et al., 2014), or are based on a single genetic distance metric only (e.g. Beninde et al., 2018), and, particularly, when inferences are translated into conservation recommendations.

AUTHOR CONTRIBUTIONS

ACF and JB designed the study and analysed the data. JW generated the simulated datasets. ACF and JB wrote the manuscript with contributions from JW.

ACKNOWLEDGEMENTS

We thank Alain Licoppe, Sabine Bertouille and the local Services of the Nature and Forest Department (General Directorate for Agriculture, Natural Resources and Environment of the Public Service of Wallonia) for sample collection and Marie-Christine Flamand (University of Louvain) for genotyping of the wild boar samples. We thank Paul Wilmes, Cédric Laczny and Maharshi Vyas for granting us access to the computing cluster of the Systems Ecology research group at the University of Luxembourg. Funding for JB was provided by the UCLA La Kretz Center for California Conservation Science and by the German Science Foundation (DFG: BE 6887/1-1). JW was funded by the Luxembourg research fund FNR (C20/SR/14748041). The work was further supported by an internal grant from the Musée National d'Histoire Naturelle, Luxembourg.

CONFLICT OF INTEREST STATEMENT

No conflict of interest to declare.

DATA AVAILABILITY STATEMENT

The empirical datasets are available under Beninde et al. (2016b) (https://doi.org/10.5061/dryad.31qg7), Dellicour et al. (2019b) (https://doi.org/10.5061/dryad.c6t0470) and Kimmig et al. (2020b) (https://doi.org/10.5061/dryad.dv41ns1ts). The script used to generate the simulated data are currently available under https://github.com/julian-wittische/BestDistance.

BENEFIT-SHARING STATEMENT

Benefits from this research accrue from the sharing of our data and results on public databases as described above.

ORCID

Joscha Beninde https://orcid.org/0000-0002-1677-1809

REFERENCES

Angelone, S., Kienast, F., & Holderegger, R. (2011). Where movement happens: Scale-dependent landscape effects on genetic differentiation in the European tree frog. *Ecography*, *34*(5), 714–722. https://doi.org/10.1111/j.1600-0587.2010.06494.x

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

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*(14), 3422–3438. https://doi.org/10.1111/mec.15979

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., & Bonhomme, F. (1996–2004). *GENETIX 4.05*. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171: Université de Montpellier II.

Beninde, J., Feldmeier, S., Veith, M., & Hochkirch, A. (2018). Admixture of hybrid swarms of native and introduced lizards in cities is determined by the cityscape structure and invasion history. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1883), 20180143. https://doi.org/10.1098/rspb.2018.0143

Beninde, J., Feldmeier, S., Werner, M., Peroverde, D., Schulte, U., Hochkirch, A., & Veith, M. (2016a). Cityscape genetics: Structural vs. functional connectivity of an urban lizard population. *Molecular Ecology*, *25*(20), 4984–5000. https://doi.org/10.1111/mec.13810

Beninde, J., Feldmeier, S., Werner, M., Peroverde, D., Schulte, U., Hochkirch, A., & Veith, M. (2016b). Data from: Cityscape genetics: Structural vs. functional connectivity of an urban lizard population. *Dryad Dataset*. https://doi.org/10.5061/dryad.31qg7

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*(6470), 455–457. https://doi.org/10.1038/368455a0

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Buri, M., Tanadini, L. G., Hothorn, T., & Curt, A. (2022). Unbiased recursive partitioning enables robust and reliable outcome prediction in acute spinal cord injury. *Journal of Neurotrauma*, *39*(3–4), 266–276. https://doi.org/10.1089/neu.2020.7407

Cushman, S. A., & Landguth, E. L. (2010). Scale dependent inference in landscape genetics. *Landscape Ecology*, *25*(6), 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*(4), 486–499. https://doi.org/10.1086/506976

Dellicour, S., Prunier, J. G., Piry, S., Eloy, M.-C., Bertouille, S., Licoppe, A., Frantz, A. C., & Flamand, M.-C. (2019a). Landscape genetic analyses of *Cervus elaphus* and *Sus scrofa*: Comparative study and analytical developments. *Heredity*, *123*(2), 228–241. https://doi.org/10.1038/s41437-019-0183-5

Dellicour, S., Prunier, J. G., Piry, S., Eloy, M.-C., Bertouille, S., Licoppe, A., Frantz, A. C., & Flamand, M.-C. (2019b). Data from: Landscape genetic analyses of *Cervus elaphus* and *Sus scrofa*: Comparative study and analytical developments. *Dryad Dataset*. https://doi.org/10.5061/dryad.c6t0470

Draheim, H. M., Moore, J. A., Fortin, M.-J., & 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*(8), 1219–1230. https://doi.org/10.1111/eva.12617

Dray, S., & Dufour, A.-B. (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, *22*(4), 1–20. https://doi.org/10.18637/jss.v022.i04

Dytham, C. (2011). *Choosing and using statistics: A biologist's guide* (3rd ed.). Wiley-Blackwell.

Frantz, A. C., Cellina, S., Krier, A., Schley, L., & Burke, T. (2009). Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: Clusters or isolation by distance? *Journal of Applied Ecology*, *46*(2), 493–505. https://doi.org/10.1111/j.1365-2664.2008.01606.x

Galla, S. J., Moraga, R., Brown, L., Cleland, S., Hoeppner, M. P., Maloney, R. F., Richardson, A., Slater, L., Santure, A. W., & Steeves, T. E. (2020). A comparison of pedigree, genetic and genomic estimates of relatedness for informing pairing decisions in two critically endangered birds: Implications for conservation breeding programmes worldwide. *Evolutionary Applications*, *13*(5), 991–1008. https://doi.org/10.1111/eva.12916

Goslee, S. C., & Urban, D. L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, *22*(7), 1–19. https://doi.org/10.18637/jss.v022.i07

Hardy, O. J., & Vekemans, X. (2002). Spagedi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, *2*(4), 618–620. https://doi.org/10.1046/j.1471-8286.2002.00305.x

Hijmans, R. (2022). *Raster: Geographic data analysis and modeling*. R Package Version 3.6–11. https://CRAN.R-project.org/package=raster

Hothorn, T., & Zeileis, A. (2015). Partykit: A modular toolkit for recursive partytioning in R. *Journal of Machine Learning Research*, *16*(118), 3905–3909.

Ji, X., & Braña, F. (2000). Among clutch variation in reproductive output and egg size in the wall lizard (*Podarcis muralis*) from a lowland population in of northern Spain. *Journal of Herpetology*, *34*(1), 54–60. https://doi.org/10.2307/1565238

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

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. https://doi.org/10.1007/s10592-014-0684-y

Kimmig, S., 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. (2020a). Beyond the landscape: Resistance modelling infers physical and behavioural gene flow barriers to a mobile carnivore across a metropolitan area. *Molecular Ecology*, *29*(3), 466–484. https://doi.org/10.1111/mec.15345

Kimmig, S., 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. (2020b). Data from: Beyond the landscape: Resistance modelling infers physical and behavioural gene flow barriers to a mobile carnivore across a metropolitan area. *Dryad Dataset*. https://doi.org/10.5061/dryad.dv41ns1ts

Kivimäki, I., Shimbo, M., & Saerens, M. (2014). Developments in the theory of randomized shortest paths with a comparison of graph node distances. *Physica A: Statistical Mechanics and its Applications*, *393*, 600–616. https://doi.org/10.1016/j.physa.2013.09.016

Koen, E. L., Garroway, C. J., Wilson, P. J., Bowman, J., & Bersier, L.-F. (2010). The effect of map boundary on estimates of landscape resistance to animal movement. *PLoS One*, *5*(7), e11785. https://doi.org/10.1371/journal.pone.0011785

Landguth, E. L., & Cushman, S. A. (2010). Cdpop: A spatially explicit cost distance population genetics program. *Molecular Ecology Resources*, *10*(1), 156–161. https://doi.org/10.1111/j.1755-0998.2009.02719.x

Landguth, E. L., Cushman, S. A., Murphy, M. A., & Luikart, G. (2010). Relationships between migration rates and landscape resistance assessed using individual-based simulations. *Molecular Ecology Resources*, *10*(5), 854–862. https://doi.org/10.1111/j.1755-0998.2010.02867.x

Landguth, E. L., & Schwartz, M. K. (2014). Evaluating sample allocation and effort in detecting population differentiation for discrete and continuously distributed individuals. *Conservation Genetics*, *15*(4), 981–992. https://doi.org/10.1007/s10592-014-0593-0

Lemopoulos, A., Prokkola, J. M., Uusi-Heikkilä, S., Vasemägi, A., Huusko, A., Hyvärinen, P., Koljonen, M.-L., Koskiniemi, J., & Vainikka, A. (2019). Comparing RADseq and microsatellites for estimating genetic diversity and relatedness—Implications for brown trout conservation. *Ecology and Evolution*, *9*(4), 2106–2120.

Li, C. C., Weeks, D. E., & Chakravarti, A. (1993). Similarity of DNA fingerprints due to chance and relatedness. *Human Heredity*, *43*(1), 45–52. https://doi.org/10.1159/000154113

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22.

Loiselle, B. A., Sork, V. L., Nason, J., & Graham, C. (1995). Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany*, *82*(11), 1420–1425. https://doi.org/10.2307/2445869

Lynch, M., & Ritland, K. (1999). Estimation of pairwise relatedness with molecular markers. *Genetics*, *152*(4), 1753–1766. https://doi.org/10.1093/genetics/152.4.1753

Mangiafico, S. (2019). *Rcompanion: Functions to support extension education program evaluation*. R Package Version 2.1.7. https://CRAN.R-project.org/package=rcompanion

McRae, B. H. (2006). Isolation by resistance. *Evolution*, *60*(8), 1551–1561. https://doi.org/10.1554/05-321.1

Medina, I., Cooke, G. M., & Ord, T. J. (2018). Walk, swim or fly? Locomotor mode predicts genetic differentiation in vertebrates. *Ecology Letters*, *21*(5), 638–645. https://doi.org/10.1111/ele.12930

Mignotte, A., Garros, C., Dellicour, S., Jacquot, M., Gilbert, M., Gardès, L., Balenghien, T., Duhayon, M., Rakotoarivony, I., de Wavrechin, M., & Huber, K. (2021). High dispersal capacity of *Culicoides obsoletus* (Diptera: Ceratopogonidae), vector of bluetongue and Schmallenberg viruses, revealed by landscape genetic analyses. *Parasites & Vectors*, *14*, 1–14. https://doi.org/10.1186/s13071-020-04522-3

Peterman, W. E. (2018). ResistanceGA: An R package for the optimisation of resistance surfaces using genetic algorithms. *Methods in Ecology and Evolution*, *9*(6), 1638–1647. https://doi.org/10.1111/2041-210X.12984

Peterman, W. E., Connette, G. M., Semlitsch, R. D., & Eggert, L. S. (2014). Ecological resistance surfaces predict fine-scale genetic differentiation in a terrestrial woodland salamander. *Molecular Ecology*, *23*(10), 2402–2413. https://doi.org/10.1111/mec.12747

Peterman, W. E., & Pope, N. S. (2021). The use and misuse of regression models in landscape genetic analyses. *Molecular Ecology*, *30*(1), 37–47. https://doi.org/10.1111/mec.15716

Peterman, W. E., Winiarski, K. J., Moore, C. E., Carvalho, C. S., Gilbert, A. L., & Spear, S. F. (2019). A comparison of popular approaches to optimise landscape resistance surfaces. *Landscape Ecology*, *34*(7), 2197–2208. https://doi.org/10.1007/s10980-019-00870-3

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 of the United States of America*, *118*, e2003201118. https://doi.org/10.1073/pnas.2003201118

Prévot, C., & Licoppe, A. (2013). Comparing red deer (*Cervus elaphus* L.) and wild boar (*Sus scrofa* L.) dispersal patterns in southern Belgium. *European Journal of Wildlife Research*, *59*(6), 795–803. https://doi.org/10.1007/s10344-013-0732-9

Queller, D. C., & Goodnight, K. F. (1989). Estimating relatedness using genetic markers. *Evolution*, *43*(2), 258–275. https://doi.org/10.1111/j.1558-5646.1989.tb04226.x

Renner, S. C., Suarez-Rubio, M., Wiesner, K. R., Drögemüller, C., Gockel, S., Kalko, E. K., Ayasse, M., & Frantz, A. C. (2016). Using multiple landscape genetic approaches to test the validity of genetic clusters in a species characterized by an isolation-by-distance pattern. *Biological Journal of the Linnean Society*, *118*(2), 292–303.

Reyne, M., 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., Brady, S. P., Wang, I. J., & Spear, S. F. (2016). Navigating the pitfalls and promise of landscape genetics. *Molecular Ecology*, *25*(4), 849–863. https://doi.org/10.1111/mec.13527

Ritland, K. (1996). Estimators for pairwise relatedness and individual inbreeding coefficients. *Genetical Research*, *67*(2), 175–185. https://doi.org/10.1017/S0016672300033620

Rousset, F. (2000). Genetic differentiation between individuals. *Journal of Evolutionary Biology*, *13*(1), 58–62. https://doi.org/10.1046/j.1420-9101.2000.00137.x

Savary, P., Foltête, J.-C., Moal, H., Vuidel, G., & Garnier, S. (2020). 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

Schulte, U. (2008). *Die Mauereidechse: Erfolgreich im Schlepptau des Menschen [The Wall Lizard: Success in Tow Behind Man]*. Laurenti-Verlag.

Schwartz, M. K., Copeland, J. P., Anderson, N. J., Squires, J. R., Inman, R. M., McKelvey, K. S., Pilgrim, K. L., Waits, L. P., & Cushman, S. A. (2009). Wolverine gene flow across a narrow climatic niche. *Ecology*, *90*(11), 3222–3232. https://doi.org/10.1890/08-1287.1

Schwartz, M. K., & McKelvey, K. S. (2009). Why sampling scheme matters: The effect of sampling scheme on landscape genetic results. *Conservation Genetics*, *10*(2), 441–452. https://doi.org/10.1007/s10592-008-9622-1

Scrucca, L. (2013). GA: A package for genetic algorithms in R. *Journal of Statistical Software*, *53*(4), 1–37. https://doi.org/10.18637/jss.v053.i04

She, J. X., Autem, M., Kotulas, G., Pasteur, N., & Bonhomme, F. (1987). Multivariate analysis of genetic exchanges between *Solea aegyptiaca* and *Solea senegalensis* (Teleosts, Soleidae). *Biological Journal of the Linnean Society*, *32*(4), 357–371. https://doi.org/10.1111/j.1095-8312.1987.tb00437.x

Shirk, A. J., Landguth, E. L., & Cushman, S. A. (2017a). A comparison of individual-based genetic distance metrics for landscape genetics. *Molecular Ecology Resources*, *17*(6), 1308–1317. https://doi.org/10.1111/1755-0998.12684

Shirk, A. J., Landguth, E. L., & Cushman, S. A. (2017b). A comparison of regression methods for model selection in individual-based landscape genetic analysis. *Molecular Ecology Resources*, *18*(1), 55–67. https://doi.org/10.1111/1755-0998.12709

Tennekes, M. (2021). *Treemap*: *Treemap visualization*. R Package Version 2.4–3. https://CRAN.R-project.org/package=treemap

Trewhella, W. J., & Harris, S. (1990). The effect of railway lines on urban fox (*Vulpes vulpes*) numbers and dispersal movements. *Journal of Zoology*, *221*(2), 321–326. https://doi.org/10.1111/j.1469-7998.1990.tb04004.x

Trumbo, D. R., Spear, S. F., Baumsteiger, J., & Storfer, A. (2013). Rangewide landscape genetics of an endemic Pacific northwestern salamander. *Molecular Ecology*, *22*(5), 1250–1266. https://doi.org/10.1111/mec.12168

Van de Casteele, T., Galbusera, P., & Matthysen, E. (2001). A comparison of microsatellite-based pairwise relatedness estimators. *Molecular Ecology*, *10*(6), 1539–1549. https://doi.org/10.1046/j.1365-294x.2001.01288.x

Van Etten, J. (2017). R package gdistance: Distances and routes on geographical grids. *Journal of Statistical Software*, *76*(13), 1–21. https://doi.org/10.18637/jss.v076.i13

Vanhove, M., & Launey, S. (2023). Estimating resistance surfaces using gradient forest and allelic frequencies. *Molecular Ecology Resources*. https://doi.org/10.1111/1755-0998.13778

Vekemans, X., & Hardy, O. J. (2004). New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, *13*(4), 921–935. https://doi.org/10.1046/j.1365-294x.2004.02076.x

Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, *23*(23), 5649–5662. https://doi.org/10.1111/mec.12938

Wang, J. (2002). An estimator for pairwise relatedness using molecular markers. *Genetics*, *160*(3), 1203–1215.

Wei, T., & Simko, V. (2021). *Corrplot*: *Visualization of a correlation matrix*. R Package Version 0.92. https://github.com/taiyun/corrplot

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.

Winiarski, K. J., Peterman, W. E., & McGarigal, K. (2020). Evaluation of the R package 'resistancega': A promising approach towards the accurate optimisation of landscape resistance surfaces. *Molecular Ecology Resources*, *20*(6), 1583–1596. https://doi.org/10.1111/1755-0998.13217

Wright, S. (1943). Isolation by distance. *Genetics*, *28*(2), 114–138.

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: 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*, *00*, 1–18. https://doi.org/10.1111/1755-0998.13831