

ARTICLE



Species distribution models predict genome-wide polymorphism and gene flow in an endangered amphibian

Benjamin Monod-Broca¹✉, Hugo Cayuela^{2,3}, Jérôme G. Prunier⁴, Lelio Belloir¹, Igor Boyer⁵, Francis Isselin-Nondedeu^{5,6,7}, Glenn Yannic⁸, Heike Pröhl⁹, Quentin Rougemont^{10,11} and Jean-Paul Léna^{1,11}

© The Author(s), under exclusive licence to The Genetics Society 2026

Species distribution models (SDMs) are widely used to predict habitat suitability but their usefulness and accuracy for inferring population health is still debated. Here, we evaluate whether SDM-derived relative habitat suitability (RHS) predicts genome-wide genetic diversity and connectivity—which are key proxies for population health and the functional integrity of landscapes. We addressed this issue in the Yellow-bellied toad (*Bombina variegata*), an endangered amphibian species with limited dispersal. We combined hierarchical SDMs, integrating both continental-level bioclimatic data and regional-level landscape variables, with genome-wide SNP data from 404 individuals sampled across 92 sites in southeastern France. We then used a multi-scale modelling framework to test the effect of bioclimatic (BRHS) and landscape (LRHS) habitat suitability on observed heterozygosity and pairwise genetic differentiation, accounting for heterogeneous genetic drift using gravity models. Our results show that both BRHS and LRHS are significant predictors of heterozygosity, with their effects expressed at different spatial scales—11 km and 3 km for BRHS and LRHS, respectively. Connectivity patterns also widely varied depending on scale and were best explained by gravity models integrating LRHS, BRHS, and local heterozygosity, underscoring the combined role of landscape resistance and population size in shaping patterns of genetic differentiation. These findings show that SDMs, when carefully calibrated and interpreted, can provide proxies for genetic diversity and landscape resistance in species with limited dispersal.

Heredity; <https://doi.org/10.1038/s41437-026-00885-y>

INTRODUCTION

The rapid growth of citizen science platforms over the past two decades has led to a remarkable increase in biodiversity monitoring data, particularly species presence records (Schmeller et al. 2009; Chandler et al. 2017; Feldman et al. 2021). In parallel, significant advances have been made in the development of species distribution models (SDMs), which combine presence data with environmental predictors to estimate current or future species distributions or relative habitat suitability (RHS thereafter; Araújo et al. 2019; Bowler et al. 2022; Valavi et al. 2022; Della Rocca et al. 2024). Together, the SDM framework and citizen science-based data have become essential tools for assessing the impacts of global change on biodiversity and guiding conservation strategies for threatened species (Razgour et al. 2018). However, the extent to which SDM-derived RHS values can reliably predict the health and long-term viability of wild populations remains debated (Guisan et al. 2013; Lee-Yaw et al. 2022). Understanding how species distributions and habitat suitability relate to the demographic and genetic processes determining population health, dynamics, and persistence is

therefore critical for evaluating the potential of these tools in wildlife management and conservation planning.

The link between habitat suitability and population health seems relatively intuitive: environmental conditions strongly influence population dynamics and structure (Moran 1953), which in turn shape species distributions (Lawton 1994; Holt et al. 2005). Species are expected to persist in areas where environmental conditions support stable or positive population growth rates (Hutchinson 1957; Holt 2009). Favourable environments promote recruitment and reduce mortality (Southwood 1977), thereby increasing population growth and ultimately resulting in larger population sizes and higher local abundances (e.g., Denoël and Lehmann 2006). Furthermore, unsuitable environmental conditions in the surrounding landscape matrix are expected to limit both the intensity and distance of dispersal (Berry et al. 2005; Wang et al. 2008). These demographic processes can have cascading effects on the genetic characteristics of populations. Larger populations tend to maintain higher genetic diversity due to reduced genetic drift, whereas smaller populations are more vulnerable to genetic erosion and inbreeding (Frankham 1996). In

¹Université Claude Bernard Lyon 1, LEHNA UMR 5023, CNRS, ENTPE, Villeurbanne F-69622, France. ²Department of Biology, University of Oxford, Oxford OX1 3RB, UK. ³ECOBIO, CNRS, UMR6553, Université de Rennes, Rennes, France. ⁴ADENEKO, Saint-Girons, France. ⁵UMR 7324, CNRS CITERES, Tours, France. ⁶Institut Méditerranéen de Biodiversité et d'Écologie (IMBE), UMR Université Aix-Marseille Avignon, 7263-CNRS, 237-IRD IRPNC (Ingénierie de la Restauration des Patrimoines Naturels et Culturels), Avignon, France. ⁷Département d'Aménagement et d'Environnement, Ecole Polytechnique de l'Université de Tours, Tours, France. ⁸Université Grenoble Alpes, Université Savoie Mont Blanc, CNRS, LECA, Grenoble 38000, France. ⁹Institute of Zoology, University of Veterinary Medicine of Hannover, Hannover, Lower Saxony, Germany. ¹⁰CNRS, AgroParisTech, Laboratoire Ecologie Systématique et Evolution, UMR 8079, Bâtiment 680, Université Paris-Saclay, Gif-sur-Yvette 91190, France. ¹¹These authors contributed equally: Quentin Rougemont, Jean-Paul Léna. Associate editor: Giorgio Bertorelle; benjamin.monod-broca@neuf.fr

Received: 28 November 2025 Revised: 31 August 2026 Accepted: 31 August 2026

Published online: 17 September 2026

parallel, reduced dispersal leads to decreased gene flow, which contributes to greater genetic differentiation between populations (Ellstrand and Elam 1993; Cayuela et al. 2018). Together, the erosion of genetic diversity and gene flow can undermine a population's adaptive potential and capacity in the face of environmental change and may ultimately increase the risk of inbreeding depression and extinction (Hedrick and Kalinowski 2000; Seaborn et al. 2021).

Although based on clear theoretical predictions, relationships between SDM-derived relative habitat suitability (RHS) and the demographic and genetic properties of populations still lack strong empirical support, partly due to the scarcity of suitable datasets (Holt 2009; Pagel and Schurr 2012). Despite numerous studies demonstrating the key role of environmental factors in shaping genetic diversity (e.g., Carnaval et al. 2009; De Kort et al. 2021; Yannic et al. 2014), studies explicitly testing links between contemporary SDM-derived RHS and population genetic features have largely failed to detect consistent patterns. Among the 18 studies reviewed by Lee-Yaw et al. (2022), only two reported a clear positive relationship between RHS and genetic diversity, and none found evidence that RHS predicts genetic differentiation. As proposed by these authors, this limited empirical support may stem from key factors—such as historical colonization and related founder effects, past climatic conditions, and environmental changes—that can obscure relationships between genetic patterns and static RHS derived from SDMs based on contemporary conditions.

Beyond these factors, we argue that additional limitations may further reduce the detectability of relationships between RHS and population genetic structure. First, mismatches in spatial scale between the processes underlying genetic variation (e.g., drift and gene flow) and the resolution at which RHS is estimated may obscure associations. Second, RHS may capture only a subset of the environmental drivers shaping genetic structure, particularly when SDMs incompletely represent the species' ecological niche (Araújo and Guisan 2006). Third, commonly used genetic markers (e.g., microsatellites, AFLPs) may lack the resolution required to detect such relationships (Sunde et al. 2020; Skey et al. 2023). Fourth, the strength of RHS–genetics relationships is likely contingent on species-specific dispersal ability, with stronger signals expected in low-vagility species for which landscape features more strongly constrain gene flow (Bohonak 1999). Finally, disentangling the relative contributions of genetic drift and gene flow remains technically challenging, further complicating inference of relationships between RHS and genetic structure (Prunier et al. 2017; Savary et al. 2023).

In this study, we sought to test whether explicitly addressing these limitations influenced the ability of SDM-derived RHS to predict genetic diversity and genetic differentiation in a threatened vertebrate. We leveraged the yellow-bellied toad (*Bombina variegata*) as a model system, as its ecological characteristics—particularly its low dispersal ability and strong sensitivity to environmental conditions—combined with the availability of high-resolution occurrence and genomic datasets, make it especially well-suited for a rigorous assessment of RHS–genetics relationships. To overcome the limitations outlined above, we (1) implemented a multiscale framework to identify the spatial scales at which RHS best explains variation in genetic diversity and differentiation; (2) used a large, high-resolution occurrence dataset (15,344 georeferenced records) to model both climatic and landscape components of habitat suitability, generating bioclimatic RHS (BRHS) and landscape RHS (LRHS); (3) generated genome-wide data (404 individuals from 92 sites genotyped at 37,169 SNPs) to maximize analytical resolution; (4) focused on a low-dispersal species in which landscape effects on gene flow are expected to be strong; and (5) applied gravity models to quantify the effects of BRHS and LRHS on genetic differentiation while explicitly accounting for spatial variation in genetic drift via heterozygosity.

We first characterized the genetic structure of *Bombina variegata* within the study area in southeastern France. We then specifically tested three hypotheses: (H1) genetic diversity should increase with both bioclimatic habitat suitability (BRHS), estimated at the continental scale, and landscape habitat suitability (LRHS), estimated at the local scale; (H2) genetic differentiation should be better predicted by BRHS and LRHS than isolation-by-distance; and (H3) BRHS should influence genetic diversity and differentiation at broader spatial scales than LRHS. Finally, we discuss the generality of our findings and their implications for interpreting RHS in other systems.

MATERIAL AND METHODS

Model species and study area

Model species – The yellow-bellied toad is a small pond-breeding amphibian distributed in central and western Europe (Speybroeck et al. 2016), that has suffered a sharp decline over the last century, particularly in the western margin of its range (Lescure et al. 2011). The conservation of this species is therefore a major priority in these countries, which has led to its inclusion in Annexes II and IV of the EU Habitat Directive. This pioneer species generally breeds in small, shallow and sunny ponds in their early ecological succession stages, either newly created or regularly reset to those stages by frequent disturbances. It occurs in forested or extensive grazing areas with rugged terrain, often in foothills or hilly regions, closely associated with springs and dense hydrological networks. The proximity of water sources has been hypothesized to play a crucial role in the subsistence of animals by providing them with refuge during periods of drought (Pichenot 2008).

Study area – We conducted this study at the south-western margin of the species' global range (Fig. 1A) in a geographic area covering ~70,000 km² centred around Lyon and Grenoble, in southeastern France (Fig. 1B). Due to its peripheral position, the study area spans a gradient from a core area of high occurrence to the north, subjected to a continental climate (Jura, Saône-et-Loire) to peripheral, relict populations to the south, in subalpine or Mediterranean climates (Hautes-Alpes). A broader study area encompassing the species' full distribution range (see next section) was used to calibrate the bioclimatic SDMs, with projections subsequently analysed within the regional area. In contrast, landscape SDMs were both calibrated and analysed directly at the regional scale.

General overview of the analytical approach

Our study proceeded as follows: we first calibrated a bioclimatic SDM at the scale of the studied *Bombina variegata* lineage's range (to obtain a continental-level BRHS), and a landscape SDM at a regional level using fine-resolution data (to obtain LRHS). After describing the genetic structure within the sampled sites, we assessed the relative influence of BRHS and LRHS on local genetic diversity, considering the spatial scale at which each RHS was inferred. To further assess the explanatory power of our models and evaluate whether LRHS adequately summarized the environmental information relevant to genetic patterns, we replaced LRHS with the raw variables used in the landscape SDM and compared their performance. Next, we examined the influence of each RHS on gene flow using gravity models fitted at multiple spatial scales, and compared their performance against that of distance-based models. To evaluate the contribution of each RHS, we compared these models to isolation-by-distance (IBD) null models. Following the approach used for genetic diversity, we finally compared the performance of models built on raw landscape variables to those using LRHS.

Modelling habitat suitability

Hierarchical modelling, spatial scales, and environmental data. We modelled RHS at two spatially nested levels, considering two sets of separately-built SDMs. At the 'continental' level, we modelled the influence of bioclimatic variables on *B. variegata* occurrence. As the bioclimatic niche differs between the different lineages of *B. variegata* (Boyer et al. 2021), we focused on the Carpathian-Western *B. v. variegata* clade occurring in our study area (Dufresnes et al. 2025) to improve model prediction (Boyer et al. 2021; Preuk et al. 2025). We used monthly temperatures and precipitations between 2000 and 2020 from the CHELSA database (Karger et al. 2017; Karger et al. 2018) to build 19 bioclimatic variables using the R package *dismo* (Hijmans et al. 2017). These climatic data were downloaded at the

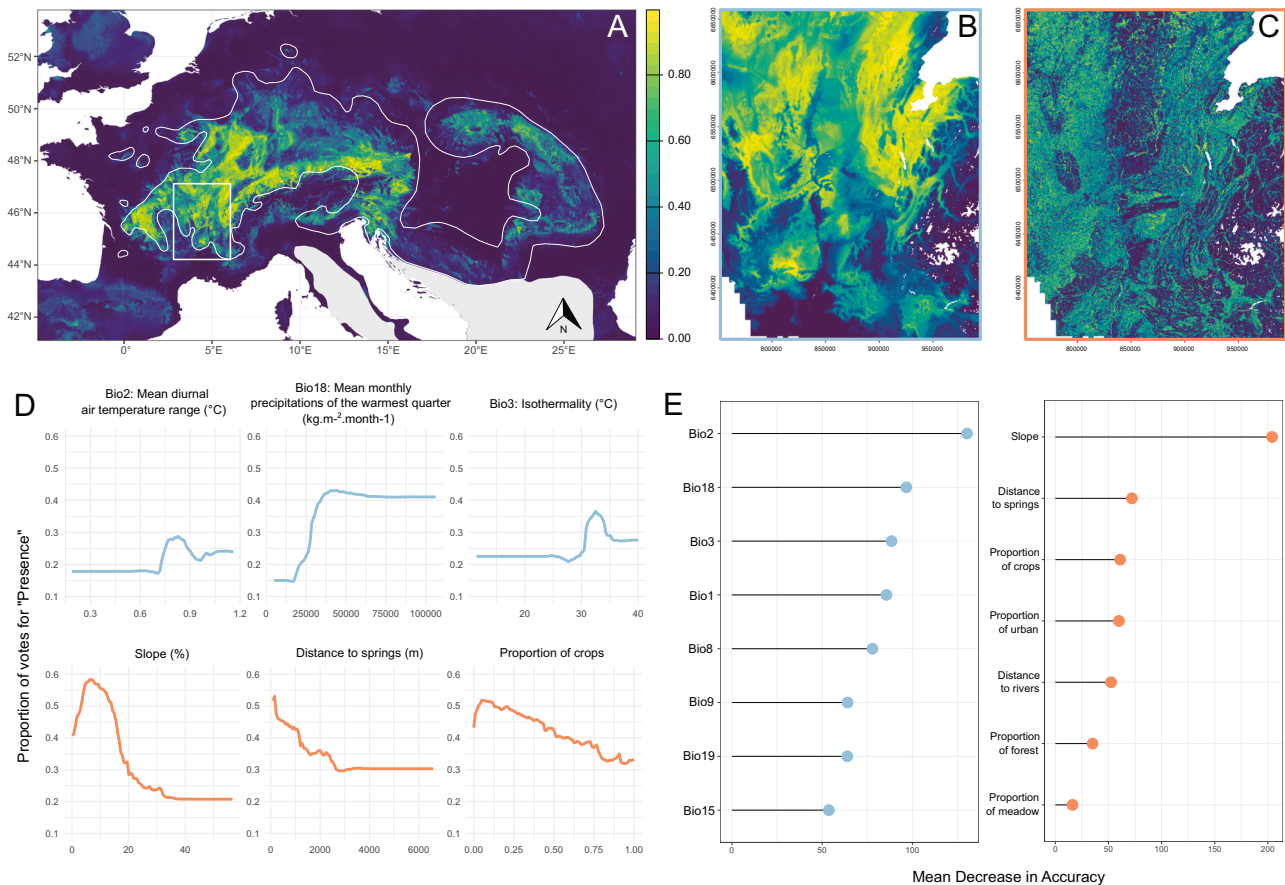


Fig. 1 **SDM predictions.** Projection of the bioclimatic SDM over (A) its calibration area, *i.e.*, the extent of occurrence for the Carpathian-Western focal lineage excluding other lineages (in grey), overlaid with its range map modified from IUCN (Ogrodowczyk et al. 2024), and (B) over the regional study area displayed by the white rectangle in (A); C projection of the landscape SDM; D response curves for the three most contributing variables in the bioclimatic (above) and landscape SDMs (below); E variable importance measured by the Mean Accuracy Decrease (MDA) for the bioclimatic SDM (left) and the landscape SDM (right). Note the core-to-periphery gradient in the regional study area (B–C), especially visible with the bioclimatic SDM (B). Coordinates in B and C are in Lambert 93.

30'' arc resolution to fit with our high-resolution occurrence data. To reduce multicollinearity, bioclimatic variables were selected using the function *vifstep* from the R library *usdm* (Naimi et al. 2014) with a Variable Inflation Factor (VIF) threshold of 5, leading to the selection of 8 variables including mean annual air temperature (*bio1*), mean diurnal temperature range (*bio2*), isothermality (*bio3*), mean daily mean air temperatures of the wettest quarter (*bio8*), mean daily mean air temperatures of the driest quarter (*bio9*), precipitation seasonality (*bio15*), precipitation of the warmest quarter (*bio18*), and mean monthly precipitation amount of the coldest quarter (*bio19*).

At the 'regional' level, we modelled the influence of landscape variables describing geomorphology and land cover on *B. variegata* occurrence. This spatially hierarchical approach helps identifying fine-grain drivers of RHS and improves our ability to predict habitat suitability compared to coarse, bioclimatic-only modelling (Gardner et al. 2019; Goicolea et al. 2025). This modelling design has been specifically shown to produce better SDMs for our model species (Arntzen et al. 2025), albeit based on AUC scores only. The selection of landscape variables was based on previous studies that modelled the ecological niche (Pichenot 2008) or examined occurrence-environment relationships using occupancy models (Cayuela et al. 2015) for this species. We considered four main categories of land cover – forest, meadow, crop and urban areas – which were obtained after reclassification from the high-resolution land cover map (10 m) of the CES-OSO database (<https://www.theia-land.fr/>). These 10-metres resolution rasters were then aggregated to a 250 m resolution to obtain the proportion of each land cover type in each pixel. This resolution was chosen to reduce the noise due to spatial uncertainty associated to land cover and occurrence data. Moreover, geomorphologic variables particularly relevant to the species (Pichenot 2008; Cayuela et al. 2015) – slope, distance to springs and

distance to rivers – were derived at the 250 m resolution from the RGE Alti® (<https://geoservices.ign.fr/rgealti>) and European catchments and rivers network system (Ecrins) databases, respectively. Applying the same procedure as for the bioclimatic variables (VIF threshold of 5, *vifstep* function from the *usdm* package) would have resulted in the removal of one variable (forest cover). However, as forest cover has been repeatedly highlighted as a biologically significant predictor for this species (Pichenot 2008; Scheele et al. 2014; Cayuela et al. 2015), and given that all pairwise Pearson correlation coefficients between landscape variables were below the $|r| < 0.7$ threshold commonly recommended for variable selection in SDMs (Dormann et al. 2013; maximum $|r| = 0.55$, see Fig. S1), all landscape variables were retained in the analysis.

Occurrence data. For the continental-level SDMs, recent occurrence data (2000–2024) with a location accuracy higher than 5000 m were downloaded from the GBIF database (GBIF.org 2024). We kept the occurrence data of the focal Carpathian-Western *B. v. variegata* clade, dropping the records from the other clades. The remaining 46,365 occurrence points were then spatially filtered to eliminate pseudo-replications (mainly due to sampling biases) as recommended by Beck et al. (2014) and Boria et al. (2014). For this purpose, we selected only one occurrence by 30'' arc bioclimatic raster cell to obtain the final dataset used for the bioclimatic SDM consisting of 5689 occurrence records (Figure S1), with a median precision of 1209 m (range 1–5000 m, mean = 1966 m). The 5 km accuracy threshold was chosen as a compromise between minimizing geographic sampling bias and maintaining spatial precision. Although this threshold exceeds the resolution of the 30'' arc climatic grid (approximately 0.6–0.9 km in our study area), potentially creating point-to-pixel mismatches, it reduced the risk of severe geographic bias. In several countries (e.g., France, Switzerland) records are

overwhelmingly reported to GBIF with a 5 km accuracy, and stricter filtering would discard most occurrences from these regions (see comparison in Fig. S2), ultimately leading to severely underestimating the suitability and presence of the species in our study area (Fig. S3). In the end, the actual location where each observation was made is expected to be no more than 5 km from the recorded coordinates. Consequently, at least half of the retained occurrences are expected to fall within approximately one to two grid cells of their recorded position on the bioclimatic layers, and all within 5 km. Given the continental scope of our analyses and the spatial thinning to the predictor resolution, any residual mismatch should add noise rather than induce directional bias, a trade-off commonly accepted in continental-level SDMs (Guisan et al. 2007). Furthermore, the bioclimatic data show a strong spatial autocorrelation at small scales (Fig. S4), including up to 5,000 m, so the climatic conditions associated with occurrence records remain highly similar within the range of their spatial uncertainty.

For the regional-level SDMs, recent occurrence data (2000–2022) provided by wildlife managers were gathered and combined (see acknowledgements), which ensured a more comprehensive coverage than cropping data from the GBIF database over the regional extent. The initial regional dataset included 15,344 georeferenced records, which were then filtered to keep only one record by 250 m land cover raster cell, to prevent pseudo-replication. The final regional dataset included 3,210 occurrence records, which represented a substantial increase compared to the number of occurrence records (only 158) obtained by applying the same procedure to GBIF data (Fig. S2). This high-resolution occurrence dataset of *B. variegata* results from the survey intensification in this region – and in France in general – following the national action plans implemented for its conservation since 2011 (PNA Sonneur, ECOTER 2011). The resulting exhaustive coverage is expected to leave few populations undetected and hence to overcome one of the major weaknesses of SDMs, namely the spatial biases caused by incomplete or uneven sampling (Jarnevich et al. 2015).

Species distribution models (SDMs). SDMs were built using random forest (RF) algorithms implemented in the R package *randomForest* (Liaw and Wiener 2002). This approach was selected for its strong predictive performance in the context of species distribution modelling, in which it shows similar performances to ensemble models while offering faster computation and a more flexible implementation (Valavi et al. 2022). Pseudo-absences, set at 10 times the number of records, were randomly generated in disks around occurrence data. Disks were defined between a minimal distance of 1 km (the maximum of 95% of annual dispersal distances, expected to include the home range of the observed individuals; Cayuela et al. 2020) and a maximal distance of 200 km away from presence points (for bioclimatic SDMs; still excluding areas where other lineages occur) and between 1 and 10 km away (for landscape SDMs). Although occurrence uncertainty may overlap with the lower bound of pseudo-absence generation in continental SDMs (1–5 km), additional analyses using pseudo-absences generated only beyond 5 km produced nearly identical predictions (Pearson's $r = 0.992$; Fig. S5), indicating negligible effects on model outcomes. The higher thresholds were set to avoid sampling environmental variables in places where the species might not have had the opportunity to disperse historically. Several random forest sampling strategies were tested, including default RF, weighted sampling (increasing misclassification cost), equal sampling (averaging predictions across subsets with an equal number of presence and pseudo-absence), and down-sampling (same as equal sampling but at the tree level instead of model level), following Valavi et al. (2021) and Valavi et al. (2022). For each model, an internal cross-validation procedure was applied with 70% of the dataset used for training the model and 30% for testing. Random forest parameters – number of variables selected by split (*mtry*), number of trees by RF (*ntree*), and size of each node (*nodesize*) – were first tuned using the R library *caret* (Kuhn 2008) before building each model with optimized parameter values and k-folds cross-validation ($k = 5$) with 10 replicates.

The best-performing models were then selected based on the area under the receiver operating characteristic curve (hereafter AUC) and true skill statistic (TSS). To check the most contributing variables within each model, the Mean Decrease in Accuracy (MDA) was computed by removing each variable and counting the number of misclassifications it introduced (Han et al., 2016). The response curve for each variable was then visualized by computing model predictions across its range of values with all other variables set to their mean.

Genomic analyses

Sampling protocol and molecular analyses. A total of 404 *B. variegata* individuals were sampled from 92 sites within the regional study zone where landscape SDMs were built (Fig. 2A). The median sample size was 4

individuals per locality (range: 1–20). Toads were captured by hand or using nets at breeding sites during the activity period (April–August). Individuals were sexed (when possible) and DNA was collected using buccal swab following the protocol presented in a previous study (Cayuela et al. 2020).

DNA was extracted with a salt-based method and an RNase A (Qiagen Blood and Tissue) treatment was applied following the manufacturer's recommendation. DNA quality was assessed using gel electrophoresis and DNA quantity was evaluated using a NanoDrop spectrophotometer (Thermo Scientific). Concentration of DNA was normalized to 20 ng/μl. Libraries were constructed following a double-digest RAD (restriction site-associated DNA sequencing; Peterson et al. 2012) protocol modified from Mascher et al. (2013). Genomic DNA was digested with two restriction enzymes (*Pst*I and *Msp*I, Poland et al. 2012) by incubating at 37 °C for two hours followed by enzyme inactivation by incubation at 65 °C for 20 min. Sequencing adaptors and a unique individual barcode (provided in Supplementary Data S4) were ligated to each sample using a ligation master mix including T4 ligase. The ligation reaction was completed at 22 °C for 2 h followed by 65 °C for 20 min to deactivate the enzymes. Samples were pooled in multiplexes of 48 individuals, ensuring that individuals from each sampling site were sequenced as part of at least six different multiplexes to avoid pool effects. Libraries were size selected (target size: 250–500 bp) using a BluePippin prep (Sage Science), amplified by PCR and sequenced with Illumina NovaSeq at the Genome Québec sequencing platform. Sequencing was performed to obtain at least 20 million reads per sample, ensuring a minimal depth of approximately 30X per sample (the mean number of reads per sample was 24 million and the sequencing depth was ~34X).

RAD-sequencing and genotyping. Barcodes were removed using *cuta-dapt* (Martin 2011) and trimmed to 130 bp, allowing for an error rate of 0.1. They were then demultiplexed using the “process_radtags” module of *stacks* version 2.64 (Catchen et al. 2013) and aligned to the genome assembly of *Bombina bombina* (a sister species of *B. variegata*; GenBank assembly accession: CA_027579735.1) using *bwa-mem2* (Li 2013; Vasi-muddin et al. 2019) with default parameters and sorted with *samtools* version 1.16 (Daneczek et al. 2021). The median mapping rate excluding secondary reads or supplementary alignment was 74.2% (mean = 74.14%, min = 70.67, max = 76.00). Then, aligned reads were processed with *Sstacks* version 2.64 for SNP calling and genotyping using *gstacks* with default parameters. We then ran the “populations” module to produce a *vcf* file that was further filtered using *bcftools* and bash scripts. SNPs were kept if they displayed a read depth greater than 4 and less than 120 (based on the mean distribution of read depth computed with *vcftools*), the latter being set to control for paralogy and highly repetitive regions. The data were filtered with a Minor/Alternate Allele Count (MAC) of 2, a genotype quality score of 20, and a maximum rate of missing data of 20% per SNP and per individual. The data were also filtered for DNA repetitive elements using the *HDplot* approach (McKinney et al. 2017) modified by Dorant et al. (2020), leading to the conservation of 73,825 SNPs located among 37,169 singleton loci. This *vcf* file was then filtered to keep only one randomly selected SNP per RAD locus to control for linkage disequilibrium, resulting in a final *vcf* file of 37,169 high quality SNPs.

Population structure, heterozygosity and genetic differentiation. We started by describing population genetic structure using two approaches. First, a principal component analysis (PCA) was performed to investigate the molecular variation among individuals using the R package *ade4* (Dray and Dufour 2007). We allowed 5% of missing data and a minor allele count of 4 keeping a total of 15,529 SNPs and imputed missing data using the mean genotype. The significance of the PCA axes was assessed using the *bstick* function implemented in the R package *vegan* (Oksanen et al. 2022) as recommended elsewhere (Jackson 1993). We quantified ancestry and admixture proportion using the *snmf* function included in the R package *LEA* (Frichot and François 2015), with K-values ranging from 1 to 30. Maps of ancestry coefficients were plotted in R following recommendations available elsewhere (Caye et al. 2016) and using the R package *ggplot2*.

We quantified the level of genetic differentiation among the genetic groups inferred by LEA using different approaches. First, we calculated F_{ST} (Weir & Cockerham 1984) implemented in the *hierfstat* R package along with 95% confidence intervals using 1000 bootstraps (Goudet 2005). For the needs of landscape genomics analyses, we computed Bray-Curtis distance among individuals on the unlinked genotype data using the *vegdist* function in the R package *vegan* (Oksanen et al. 2022), as this metric is suitable for individual-based landscape genetic analyses (Shirk et al.

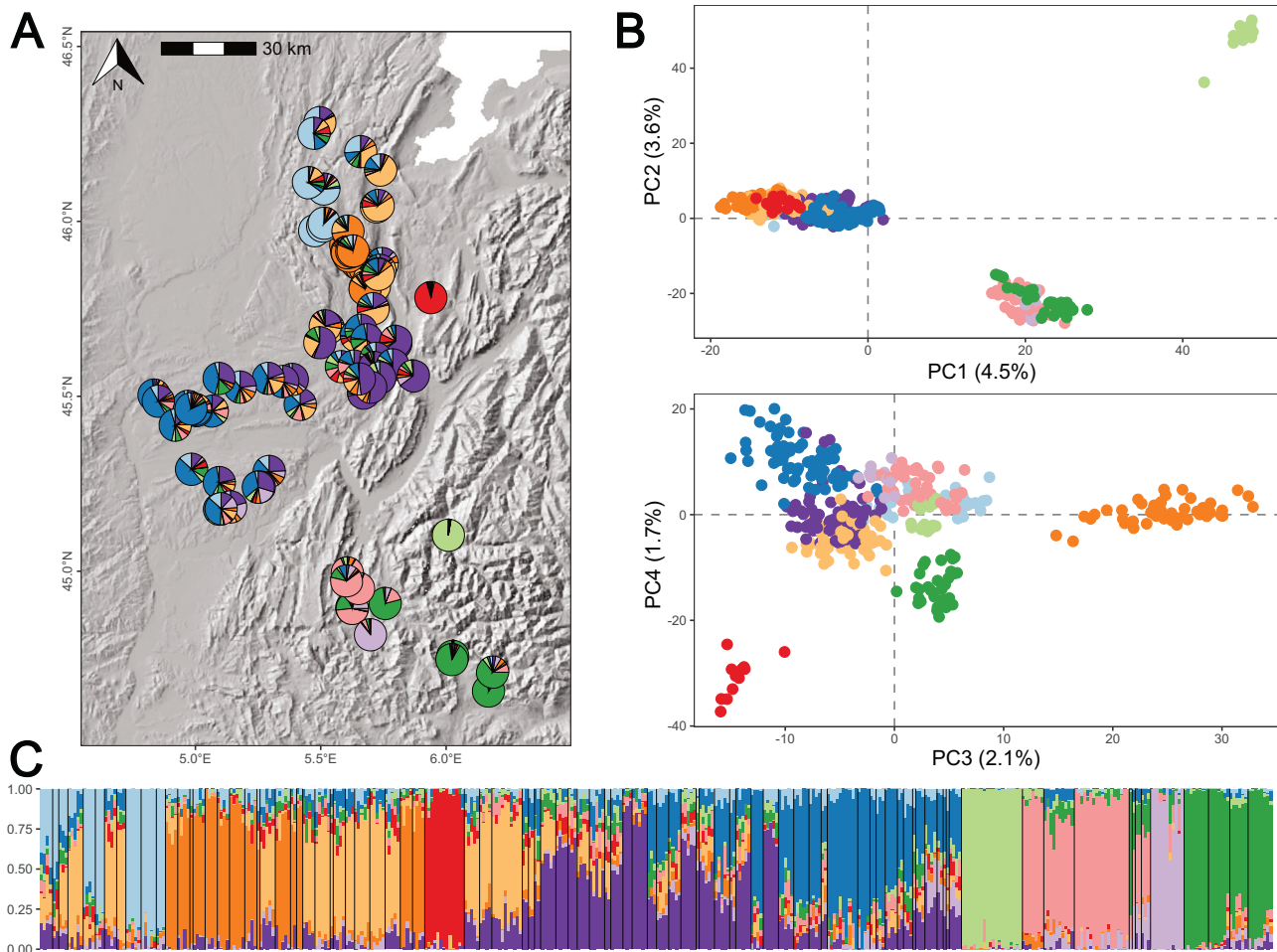


Fig. 2 Population genomic structure in the study area. **A** Map of the sampled populations showing the distribution of genetic groups; **B** PCA showing the axes of variation in allele frequencies matching geographic clusters; and **C** ancestry from LEA ($K = 10$). Colours among the three panels match, and reflect ancestry identified in LEA.

2017). We then computed individual observed heterozygosity (H_0) and inbreeding for each chromosome separately as well as genome-wide using the *vcftools* package (Danecek et al. 2011). The distribution of heterozygosity among genetic clusters was visualised using the R-package *ggplot2* (Wickham 2011).

Landscape genetics analyses

We investigated the degree to which bioclimatic RHS (BRHS) and landscape RHS (LRHS) from SDMs predict individual genome-wide heterozygosity (*i.e.*, observed heterozygosity, H_0) and genetic differentiation. In addition, we evaluated how the landscape variables used to compute LRHS at the regional level explains heterozygosity and genetic differentiation. Importantly, all analyses were conducted at the individual level (see *e.g.*, Balkenhol et al. 2015) using mixed models to take into account the non-independence of individuals belonging to a same sampled site. This approach is especially suited in the case of the yellow-bellied toad since defining biologically meaningful populations is challenging in this species and ponds are usually occupied by few adults. Indeed, *Bombina variegata* breeds in dynamic networks of temporary water bodies with highly variable spatial extent, characterized by high annual turnover (~50% creation–destruction) and short persistence times (typically ~3 years), particularly in managed forest systems (Cayuela, Boualit et al. 2016). Furthermore, we did not perform landscape genomics analyses using expected heterozygosity as this metric is best computed at the population level, thus requiring the definition of biological populations. This could have been done using the clusters from LEA, but landscape metrics are of little relevance at the scale of these major—often geographically overlapping—clusters.

Relationship between RHS and heterozygosity. To examine the scale effect of BRHS and LRHS on observed heterozygosity across a biologically relevant range of values, we extracted these indices around each sampled site by averaging the predicted RHS values within buffer zones ranging from 0.5 km to 21 km of radius, with increments of 500 m. Additionally, we extracted the raw variables used to build the land use SDM to investigate their direct contributions, following the same procedure.

We then investigated the effect of both BRHS and LRHS on the observed heterozygosity using mixed models implemented in SAS® *Proc MIXED* (SAS institute Inc. 2023). More specifically, we treated the observed heterozygosity – calculated at the individual level—as the dependent variable and each scale combination between BRHS and LRHS as well as their interactive effect were successively introduced as fixed terms in the model. For all these combinations, all reduced model sets were performed—including the models with both or only one RHS metric, and the null model, *i.e.*, with only the intercept term—resulting in a whole set of 3,445 models. For each model, an anisotropic power covariance structure including a nugget effect was used to model spatial autocorrelation in model residuals. In this structure, ρ_{lat} and ρ_{long} describe the rate of decrease of residual spatial correlation with distance along the latitudinal and longitudinal axes, respectively. Values closer to 1 indicate stronger spatial persistence of residual similarity. This anisotropic spatial autocorrelation may thus capture the effect of historical colonization from north to south through repeated founder effects on heterozygosity patterns, in which case we expect to find ρ_{long} greater than ρ_{lat} . Although we initially introduced the sampling site as a random effect, we further discarded it from the analyses since it was always estimated at its lower bound (*i.e.*, 0). This was predictable because this effect is fully accounted

for in the estimation of the spatial autocorrelation parameters (given the small number of individuals per sampling site). All models were optimised using the maximum likelihood method and ranked according to their normalized AIC (wAIC) to investigate the effect of BRHS and LRHS—as well as their interactive effect—and the scale at which these effects are the most pronounced. For each model, the F statistic was then computed using the method of Kenward-Roger to approximate the denominator degree of freedom (SAS Institute Inc. 2023) and then used to both compute the marginal pseudo- R^2 (hereafter R^2_{marg}) of the mixed model according to the method proposed by Edwards et al. (2008) and to perform commonality analyses (see below). Lastly, we investigated the influence of the original raw landscape variables used to build the LRHS using the best-supported model. For this, we replaced the LRHS variable with all raw landscape variables and retained only those that had a significant effect on H_O . We then compared the AIC performance to the model including the LRHS.

We also conducted a commonality analysis (CA), an approach used for partitioning the explained variance between unique effects and common effects (i.e., shared variance) associated with the predictors introduced into the model. This method allowed us to investigate the respective contribution of BRHS and LRHS in the presence of multicollinearity—a method that is particularly valuable in the landscape genetic framework where environmental predictors are usually collinear (Prunier et al. 2015). We applied Mood's procedure (1971) for partitioning variance and used the method of Edwards et al. (2008) and Lai et al. (2022) to compute the semi-partial (pseudo) R^2 for each predictor.

Lastly, we repeated the same procedure using sampling-site-level data to assess whether the conclusions differed from those obtained with the individual-level analyses.

Relationship between RHS and genetic differentiation. We first computed effective pairwise distances between each pair of sampling sites using *Circuitscape* (McRae 2006; McRae et al. 2009) based on each RHS layer specified as a conductance map. We also computed topographic pairwise distances, calculated as the shortest topographic path between sites using the R package *topoDistance* (Wang 2020) to infer the landscape distance matrix in the absence of both BRHS and LRHS effects (i.e., the isolation-by-distance hypothesis, Wright 1943). Lastly, we computed landscape pairwise distances based on layers used to build LRHS using *Circuitscape*. Individual genetic distances were inferred from the Bray-Curtis dissimilarities (Bray and Curtis 1957). As LRHS and BRHS become highly collinear at large scales ($r > 0.7$ above 40 km, see Fig. S6), essentially carrying the same information, we restricted the analyses to all pairwise Euclidean distances under 44 km. This led to the separation of two distinct geographic sectors, one in the Alps and the other in the lowlands, corresponding to the main geographic divide highlighted in population structure analyses. This separation was taken into account in the connectivity models by partitioning random effects according to the origin of individuals (and site), thus capturing the variance explained by both site and individuals separately for each sector (see equation in Supplementary Text S1).

Gravity models are particularly useful in landscape genetics because they explicitly consider the local effect of each node of origin (often corresponding to a site or a population) on the pairwise genetic distance along with the effect of the landscape between nodes (Murphy et al. 2010; Savary et al. 2023). Such a feature may be particularly valuable in landscape genetics since pairwise genetic distances depend not only on the gene flow between nodes but also on the respective population effective size (N_e) at each node. This model framework could thus be particularly relevant when N_e is heterogeneous among sampled sites (Prunier et al. 2017; Savary et al. 2023). Because estimating N_e is challenging, studies often rely on proxies such as the census size or the habitat patch area (Prunier et al. 2017; Savary et al. 2023). For species like *B. variegata*, such proxies are not practical, the first one requiring capture-mark-recapture data and the second requiring arbitrary buffer delimitation around ponds. In the present study, we therefore grouped all individuals located within the same raster pixel (corresponding to a single sampling site for all but two pairs of sites) to form nodes and used the mean observed heterozygosity (H_O) at each node as a proxy for N_e . These nodes should be regarded as analytical rather than biological units. They represent the minimum spatial units required to estimate local genetic drift while avoiding pseudo-replication among individuals sharing the same local environment, and should not be interpreted as distinct demographic units (i.e., biological populations).

Six models were thus fitted using the general gravity model structure presented in Supplementary Text S1: (i) the null model including only an

intercept as explanatory term, (ii) the model with the topographic distance matrix only, thus corresponding to the simple effect of the isolation by distance (IBD), the models with either (iii) BRHS or (iv) LRHS distance matrices, (v) the model with both BRHS and LRHS distance matrices and (vi) the model including the interactive effect between both RHS distance matrices. The relative support of each model was evaluated based on AIC. Based on observed patterns of genetic diversity in relation to RHS, we assumed that individuals sampled within 3 km could be considered as belonging to genetically homogeneous units. This distance was therefore selected as a minimum threshold for including inter-individual distances in the models, although a sensitivity analysis was also performed for values ranging from 1 to 5 km (Supplementary Data S2). We then constructed models by progressively increasing the maximum distance threshold from 4 km to 44 km, thereby mechanically pruning the network by retaining only arcs corresponding to inter-individual distances between 3 km and the maximum threshold. To evaluate the contribution of individual variables (H_O , LRHS, BRHS, LRHS*BRHS), we focused on the full interaction model and computed partial R^2 values across the full range of maximum distances. Additionally, we calculated pairwise correlations among explanatory variables at each spatial scale to examine how collinearity might influence variance partitioning across distances. Lastly, we examined models where LRHS distances were replaced by pairwise landscape distances based on layers used to build LRHS in the SDMs.

RESULTS

Species distribution models

Species distribution models reflected accurately the known range of the lineage (Fig. 1). Among the four random forest sampling strategies used for the continental-level (bioclimatic) and regional-level (landscape) SDMs, the best performance was achieved with the down-sampling approach in both cases, which provided good predictions of occurrence (AUC = 0.9957 and 0.9331, respectively, and TSS = 0.9427 and 0.7378). In bioclimatic SDM, Mean Decrease in Accuracy (MDA; Fig. 1E) – which quantifies variable importance – showed that the mean diurnal temperature range (*bio2*), the mean monthly precipitations of the warmest quarter (*bio18*) and the isothermality (*bio3*) were the three main contributing variables. For all these variables, higher values correspond to higher predicted BRHS (Fig. 1D), which indicates that they have positive effects on *B. variegata* occurrence. In landscape SDM, MDA showed that slope was the most influential factor, followed by the distance to springs and the proportion of crops. The slope predicted occurrence with an optimal value around 10%, highlighting that slope values on either side of this optimum negatively impact habitat suitability. On the other hand, the distance to springs and proportion of crops negatively predicted occurrence: habitat suitability is greater closer to springs and with low crop proportions.

Population structure, heterozygosity and genetic differentiation. Our analyses revealed some genetic structure among the sampled *B. variegata* populations (Fig. 2A). First, the PCA (Fig. 2B) showed a clear genetic subdivision between individuals from two highland sites in the Alps (Bourg d'Oisans valley, mean elevation: 720 m; light green cluster) and those from the lowland sites (mean elevation: 416 m) along the first PCA axis with a proportion of variance explained of 4.5%—a typical value for high-dimensional data with diffuse variance, caused here by the large number of SNPs. The second PCA axis separated the individuals from the remaining highland sites in the Alps (mean elevation = 823 m; in pink and green) from those from the lowlands, with a variance explained of 3.6%. The third and fourth PCA axes were significant and captured 2.1% and 1.7% of the variance, respectively, also reflecting local geography (Fig. 2A).

Clustering analyses in LEA suggest the presence of 15–20 genetic groups based on the cross-entropy criterion with a plateau around 15 and 18. For illustration purposes, we only present results for $K = 10$ (Fig. 2C) but the outcomes for additional K values can be

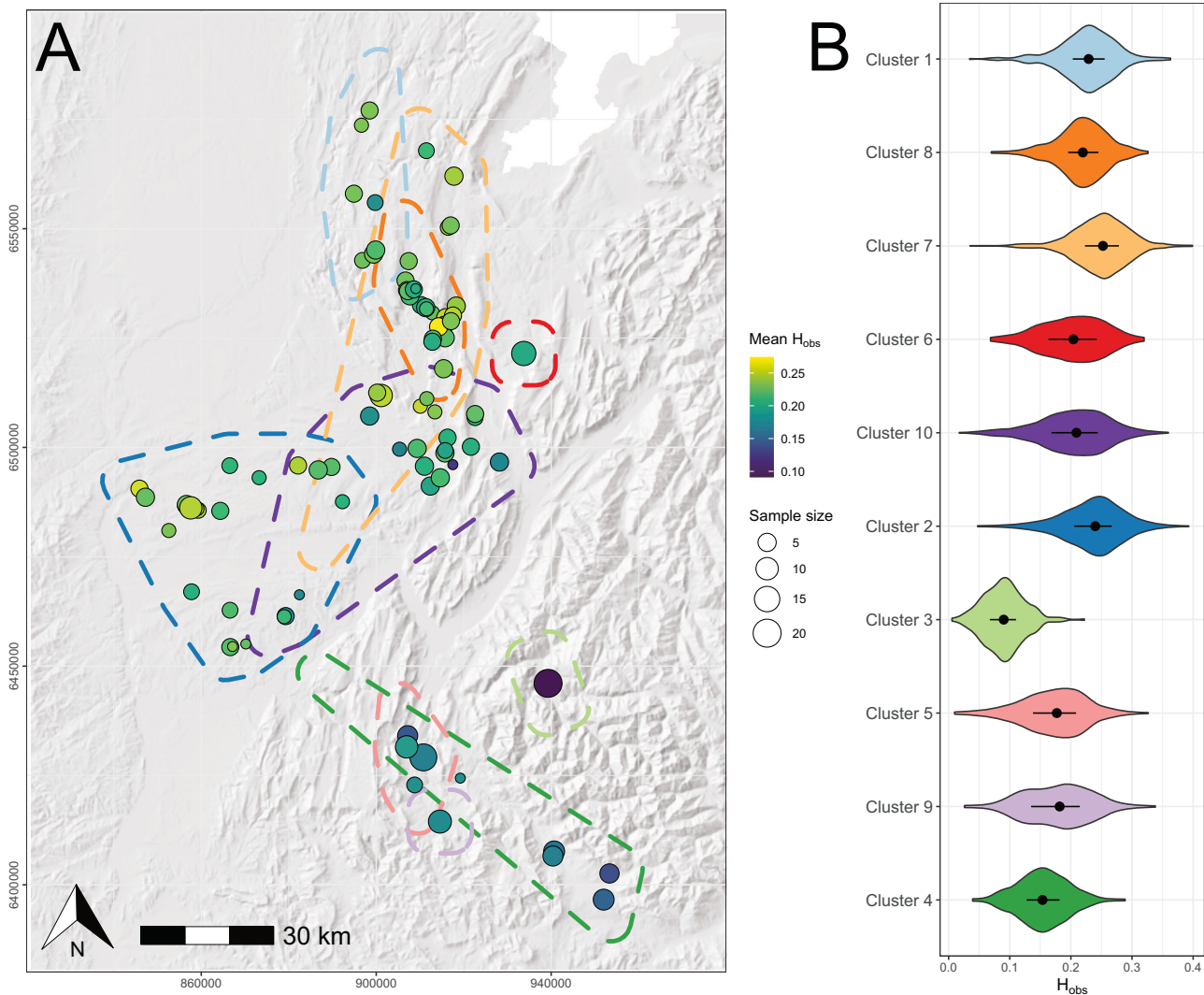


Fig. 3 Genetic diversity in the study area. **A** observed heterozygosity (H_0) averaged for each population (coordinates in Lambert 93); and **B** distribution of individual H_0 within each LEA cluster sorted by latitude ($K = 10$ for illustration purposes, with colours matching those from Fig. 2), showing mean and SD values.

found in Supplementary materials (Fig. S7). Clusters broadly match geographic features and are consistent with PCA axes, clearly segregating areas previously identified. Even finer-scale structure is visible, separating sites from various watersheds in the Bugey region (consistently with the third PCA axis). Genetic differentiation measured by F_{ST} among LEA clusters revealed high genetic structure (mean $F_{ST} = 0.163$, range 0.021–0.374) and the same pattern of three differentiated groups with some substructure (Fig. S8).

Our data also highlight a large range of values among individual observed heterozygosity (mean = 0.202, range 0.006–0.399) and among averaged values at sampling sites (mean = 0.216, range 0.091–0.275), generally following a geographic gradient from the core to the margin of the range (Fig. 3A). This core-to-margin pattern is also apparent among clusters. However, because cluster membership was assigned at the individual level, a single sampling site could include individuals assigned to different clusters. As a result, the geographic gradient is less distinct among clusters than at the sampling-site level. Inbreeding revealed by F_{IS} similarly shows large disparities (Fig. S9), both when considered at the individual level (mean = 0.244, range –0.648–0.974) and when averaged for each sampling site (mean = 0.192, range –0.023–0.661), translating a general deficit of heterozygotes across the study region, compared to neutral expectations.

Relationship between RHS and heterozygosity. LRHS and BRHS were strong predictors of the spatial variation in heterozygosity. The model selection procedure (among models presented in Supplementary data S1) indicated that models including the effect of both BRHS and LRHS (across all scale combinations) were better supported than the alternative models including only one or the other. The best supported model that included both BRHS and LRHS was indeed 51 times more supported than the best supported model without BRHS, 631 times more supported than the one without LRHS, and 250,000 times more supported than the null model. We did not find evidence for an interactive effect of BRHS and LRHS: excluding the interactive effect resulted in a model that was 2.6 times better supported by the data compared to the model that included the interaction. Residual spatial autocorrelation was stronger along the longitudinal axis ($\rho_{long} = 0.96$) than along the latitudinal axis ($\rho_{lat} = 0.90$). This indicates a slower decay of residual similarity east-west than north-south after accounting for fixed effects.

We found that the scale effect of LRHS ranges from 3 km to 3.5 km since models at a larger or lower scale were at least 3 times less supported than the best model performed at a scale of 3 km (Fig. 4A, Supplementary Data S1). The scale effect of BRHS was more difficult to define since wAIC peaks at 11 km, but other

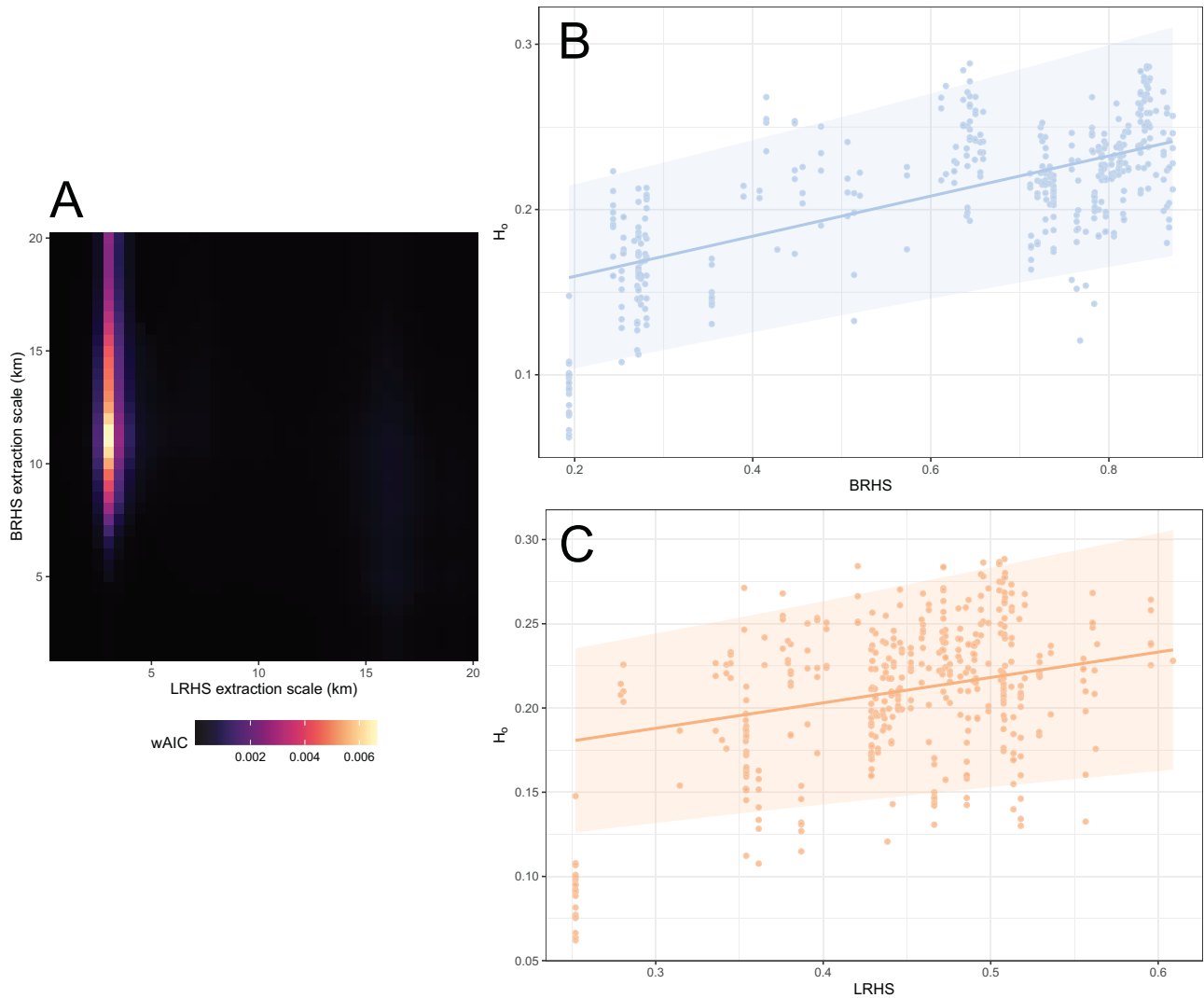


Fig. 4 Modelling the effect of LRHS and BRHS on genetic diversity. **A** Heatmap of wAIC for selecting the extraction scale of LRHS and BRHS; marginal effect of BRHS (**B**) and LRHS (**C**) on observed heterozygosity.

models at similar scales had comparable support: all models within a range of 9.5 km and 13.5 km were at most 1.5 times less supported than this best supported model (Fig. 4A, Supplementary data S1).

Our analyses showed that observed heterozygosity increases with RHS. The best supported model, with a marginal R^2 of 0.69, indicated that both BRHS and LRHS had a positive effect on observed heterozygosity (Fig. 4B, Table S1). In addition, the commonality analysis (Table S1) showed that the unique variance of BRHS (0.694) was 12 times higher than that of LRHS (0.057), the contribution of which was exclusively related to negative shared variance (-0.057), suggesting a 'suppressor effect' from LRHS.

Lastly, we found that the models including LRHS had better or comparable support to those incorporating the original set of landscape variables. Specifically, the model in which LRHS was replaced by all original landscape variables had lower support ($AIC = -1924.7$) than the model including LRHS ($AIC = -1928.8$). However, when non-significant landscape variables were removed and only slope and distance to springs were kept, the resulting model showed a better support ($AIC = -1945.3$; Table S2).

Relationship between RHS and genetic differentiation. Across all structures and all scales, gravity models—including site-specific heterozygosity – were better supported (Fig. 5, Supplementary

data S2) and at most scales H_0 explained most of the variance in genetic differentiation based on partial R^2 (Fig. 6B). For a minimal distance of 3 km—complete results for other minimal distances are available in Supplementary data S2—the best supported structure based on wAIC was the gravity model including only the LRHS at 4 km. The additive model including LRHS and BRHS was the most supported at 5 km and 10 km and the full model including the interactive effect between LRHS and BRHS was better supported for the remaining scales, peaking in wAIC at 6 and above 12 km (Fig. 6A).

When focusing on the full gravity model including the interaction, the partial R^2 values revealed that the explanatory power of each variable is strongly scale-dependent (Fig. 6B). Across most scales, the observed heterozygosity H_0 – 'node' effect in the gravity model—explained more variance than all the other variables. The LRHS distances variable showed two partial R^2 maxima, explaining the most variance at a scale of 10 km (partial $R^2 = 0.332$) and then again at 37 km (partial $R^2 = 0.26$). At this second partial maximum of 37 km, the contribution of the BRHS (partial $R^2 = 0.40$) and of the interaction between LRHS and BRHS (partial $R^2 = 0.54$, 0.6 at 36 km) were also maximized. This also corresponded to the scale at which the correlation between LRHS and BRHS rises significantly (Fig. S6). The random effects capturing the variance associated to the sampling site in each sector were often

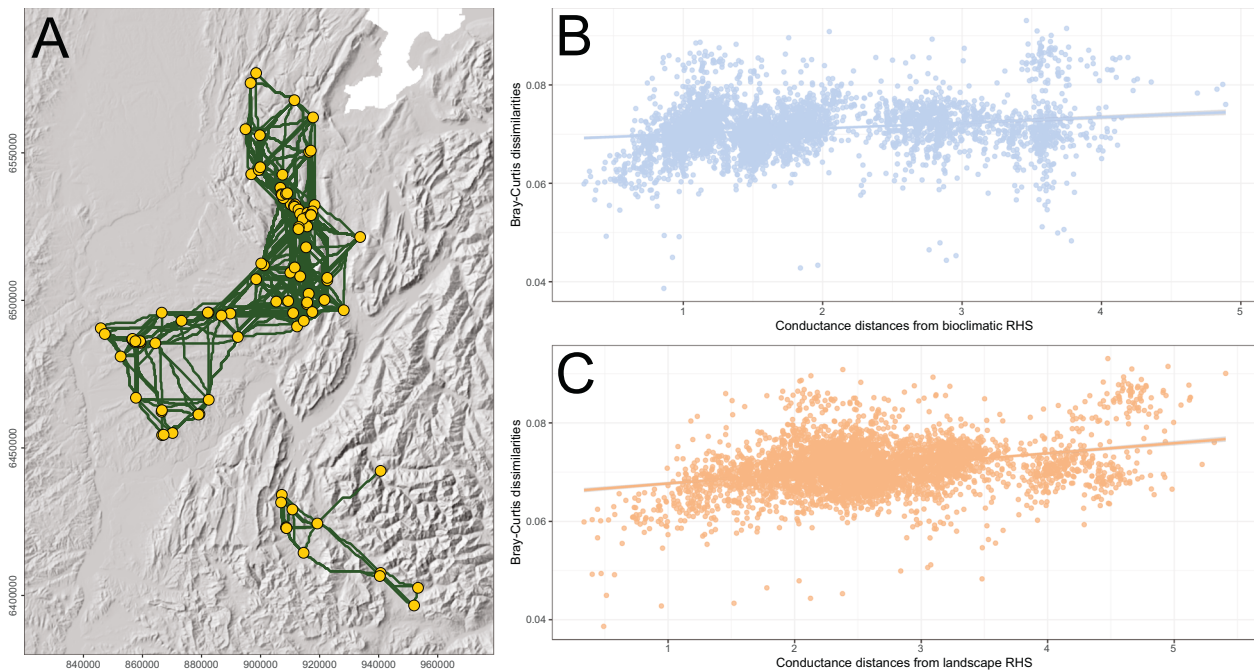


Fig. 5 Modelling the effect of LRHS and BRHS on genetic connectivity. **A** Map of the sampling sites and the topographic distances of less than 30 km computed with *topoDist* (Wang, 2020); and relationship between genetic differentiation measured by Bray–Curtis dissimilarities and conductance distances computed in *Circuitscape* (McRae 2006) for BRHS (**B**) and LRHS (**C**).

significant and revealed a contrasting pattern, with a consistently smaller value for the lowlands (0.001 in the full model between 3 km and 12 km) than for the Alps (0.079, Supplementary data S2). Similarly, the random effects for the individuals of each sector were significant at most scales and showed the same contrasting pattern (0.001 for the lowlands sector and 0.002 for the Alps).

Lastly, replacing LRHS by pairwise distances based on raw landscape variables led to similar model performance, although the support was generally higher for these models. For instance, the interactive model at a scale of 12 km—when the wAIC of this model structure reaches a plateau in Fig. 6B—shows an AIC of -19027.7 , when the corresponding model with LRHS replaced by raw variables has an AIC of -19294.96 (Supplementary data S3). At scales at which the additive gravity model including LRHS is the best supported, replacing it with raw variables always leads to a model with interaction being better supported (Supplementary data S3).

DISCUSSION

Our study provides one of the few empirical demonstrations of clear relationships between SDM-derived relative habitat suitability (RHS) and population genetic structure in animals. Specifically, we show that RHS derived from bioclimatic variables at the continental scale (BRHS) and from landscape variables at the regional scale (LRHS) reliably predicts spatial variation in both heterozygosity and genetic differentiation. Consistent with hypothesis H1, heterozygosity increased with both BRHS and LRHS. Likewise, in line with hypothesis H2, genetic differentiation was better predicted by BRHS and LRHS than isolation by distance in models accounting for site-specific variation in observed heterozygosity. Finally, supporting hypothesis H3, BRHS influenced both heterozygosity and genetic differentiation at broader spatial scales than LRHS.

Species distribution models

SDMs closely reflected the known distribution and biology of the focal Carpathian-Western lineage of *B. variegata* (Dufresnes et al.

2025). At the continental scale, climate continentality emerged as a critical driver, with occurrence positively influenced by mean diurnal temperature range (*bio2*) and isothermality (*bio3*) – two variables capturing the high amplitude of temperatures under continental climate (Snow 2005). These conditions match with those currently prevailing in the Carpathian glacial refugium from which the lineage spread across central and western Europe (Dufresnes et al. 2025). Precipitation of the warmest quarter also increased predicted relative occurrence, consistent with the specialised breeding ecology of *B. variegata* in shallow pools, where summer rainfall reduces the risk of drying—one of the primary causes of egg and larval mortality (Sinsch et al. 2020; Kijanović et al. 2023). Overall, these findings are congruent with the outcomes of a previous study modelling lineage-specific bioclimatic niches in this species (Boyer et al. 2021).

Regional landscape models highlighted slope and distance to springs as major predictors, consistent with earlier work (Pichenot 2008). This last study emphasized the same relationship between *B. variegata* occurrence and distance to spring and slope using a different niche modelling method (ENFA) in a different region (northeastern France). This strongly supports the species' close association with foothill regions where springs likely play a crucial role by providing moist refugia during summer droughts (Pichenot 2008), and echoes results from bioclimatic models regarding water availability during the warmest quarter. Our models also revealed that land use exerts a moderate, yet detectable, influence on the distribution of *B. variegata* in Western Europe. We identified a negative effect of crop cover on the species' occurrence, consistent with previous findings based on different SDM approaches (Pichenot 2008) and occupancy models (Cayuela et al. 2015) in northeastern France. While some *B. variegata* populations can occasionally persist in intensive agricultural landscapes, agricultural expansion has resulted in the destruction of habitats essential for the species' presence (Cayuela et al. 2015). Specifically, the intensification of agricultural practices during the 20th century has led to the degradation and destruction of crucial habitats for the species, including springs and low-order rivers. The lower TSS observed in the regional

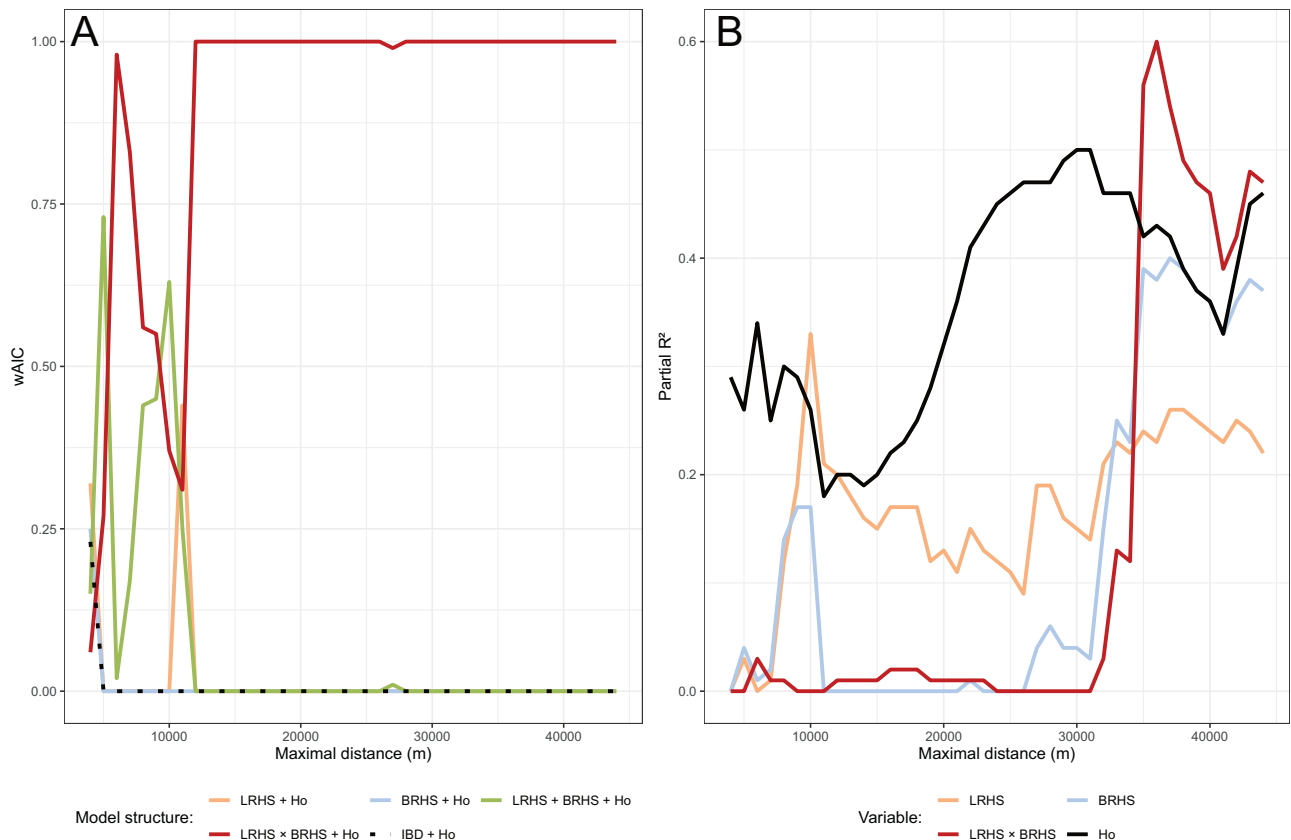


Fig. 6 Scale-dependence of model support and variable contribution in genetic connectivity modelling. **A** wAIC by model structure as a function of the maximum individual pairwise distances considered (minimum distance = 3 km); **B** partial R^2 associated with each variable in the full model with interaction, as a function of the maximum individual pairwise distances considered (minimum distance = 3 km).

landscape SDMs likely reflects a noisier and more stochastic occupancy process than in the continental climatic SDMs, resulting in greater overlap between presence and absence predictions and, consequently, lower TSS values despite relatively high AUC, owing to the threshold dependence of TSS (Allouche et al. 2006).

Our regional-level models nevertheless failed to detect effects of land-use variables that have been shown to play a key role in the species' distribution in previous studies (Pichenot 2008; Cayuela et al. 2015; Dufresnes et al. 2021). We found no substantial effect of forest cover on the occurrence of *B. variegata* within our study area, despite this species often being described as a primarily forest-dwelling amphibian (AmphibiaWeb 2023; Arntzen et al. 2025). Similarly, we found no effect of roads and urbanization, in contrast to previous research (Cayuela et al. 2015). These discrepancies are likely explained by the regional context: the study area features small and fragmented lowland forest patches, and the effect of urbanization and roads may appear only at finer scales (Cayuela et al. 2015).

Overall, our results thus agree with previous studies on climatic and landscape drivers of *B. variegata* distribution – although we also report discrepancies that could be due to the heterogeneity of the spatial scale at which the studies were conducted, and the landscape composition and structure of the study areas. The hierarchical modelling approach combining both bioclimatic and landscape predictors described both large-scale and fine-grain patterns, improving accuracy as in other studies (Gardner et al. 2019; Goicolea et al. 2025), including results specific to this species (Arntzen et al. 2025), or sister species *B. pachypus* (Bernabo et al. 2022), and in the context of climate change (Préau et al. 2019; Boyer et al. 2025).

Population structure, heterozygosity, and genetic differentiation. Our analyses revealed the presence of genetic structure, reflecting geographic differentiation between regions, sometimes with contrasted environmental conditions (e.g., lowland vs. mountain areas). Such structure is typical of pond-breeding amphibians at this scale (Beebee 2005) and well documented in *B. variegata* (Cayuela et al. 2020; Hantzschmann et al. 2021; Kleißen et al. 2021; Pröhl et al. 2021). It likely results from both low gene flow caused by limited dispersal—95% of between-patch movement being less than 1 km (Cayuela et al. 2020)—and strong drift in these small marginal populations.

The wide range of observed heterozygosity and inbreeding values is consistent with previous studies based on other markers (microsatellites: Hantzschmann et al. 2021; Kleißen et al. 2021; Pröhl et al. 2021) and reflects the strong variability in population sizes and degrees of geographic isolation across our study area. Mean F_{IS} is high, with locally very elevated individual values (up to 0.974), particularly in isolated lowland populations and in mountain sites. Such patterns point to strong local inbreeding, which may expose these populations to inbreeding depression. Combined with their likely limited adaptive potential, this increases the probability of local extinction, especially at the range margins where demographic fluctuations are pronounced. These findings also emphasize that the effects of genetic drift are spatially heterogeneous in our system, and this heterogeneity should thus be incorporated in the landscape connectivity models (Prunier et al. 2017; Savary et al. 2023).

Relationship between RHS and heterozygosity. Both BRHS and LRHS predicted heterozygosity, showing a positive relationship consistent with our expectations. This association likely reflects the

positive effects of habitat quality on demographic rates—such as survival and/or fertility—which, in turn, drive higher population growth rates and larger census sizes. Larger census sizes are typically associated with increased effective population sizes and reduced genetic drift, thereby facilitating the maintenance of higher genome-wide polymorphism.

Interestingly, we showed that most of the variance was explained by BRHS alone, with LRHS providing only a modest increase in explanatory power. At broader spatial scales, BRHS and LRHS exhibited a strong positive shared variance, likely due to large-scale collinearity between landscape and climate. However, at finer scales, the shared variance between these variables was negative, indicating a potential ‘suppressor effect’ (Friedman and Wall 2005), whereby the inclusion of LRHS increases the unique explanatory power of BRHS by accounting for variance that is not directly related to the response variable. This highlights the value of hierarchical modelling in refining predictions from broad-scale bioclimatic models by incorporating finer-scale landscape variables, as previously reported (Gardner et al. 2019; Goicolea et al. 2025).

The RHS indices operated at distinct scales: BRHS peaked at a larger spatial scale (11 km radius), whereas LRHS was most influential at a finer scale (3 km radius). This likely reflects underlying demographic and evolutionary processes: landscape quality tends to influence local population dynamics by affecting survival and reproduction within or around breeding sites, as well as dispersal between nearby patches—especially considering that the maximum dispersal distances reported in this species are around ~3.8–4.5 km and that 95% of annual dispersal is under 1 km (Cayuela et al. 2020). In contrast, bioclimatic conditions exert more gradual effects on population size and demography along altitudinal and latitudinal gradients, which operate at broader spatial scales. Notably, accounting for spatial autocorrelation had only a marginal effect on the estimated effect of BRHS, likely due to altitudinal variation in BRHS within mountainous regions. Furthermore, the observed anisotropy shows that residuals are less correlated latitudinally than longitudinally, which is consistent with the expected direction of the colonization (from north to south in our study area, Dufresnes et al. 2025). The autocorrelation structure used could thus be interpreted as capturing the effect of the historical colonization process on heterozygosity through successive founder effects. However, these parameters describe residual spatial dependence only and do not provide direct evidence for specific historical mechanisms. Results from population-level models are fully consistent with these conclusions, as they show the same exact patterns, albeit with slightly different values (Supplementary Text S2, Table S4, Supplementary Data S4).

Lastly, we found that selected raw landscape variables explained genetic diversity a little better than LRHS, but only when non-significant predictors were removed. In particular, the two retained landscape variables—slope and distance to springs—emerged as the most important predictors, and were also the top contributors in the landscape SDM. This further confirms that these variables are key determinants of this species’ biology (Pichenot 2008). Overall, the similar performance of LRHS and raw landscape variables suggests that the SDM effectively captures most of the environmental variation relevant to genetic diversity while providing a single integrative predictor.

Relationship between RHS and genetic differentiation. Both LRHS and BRHS predicted genetic differentiation. Strikingly, the relationship between genetic differentiation and RHS followed the same general pattern as those from observed heterozygosity: in the full model, the effect of LRHS is predominant over BRHS at lower scales and BRHS starts having an interactive effect at higher scales. At these larger scales, the LRHS and the BRHS might also carry the same information, which would explain why all three variables—

BRHS, LRHS and their interaction—reach a peak in their partial R^2 together at 36 km. Again, we expect landscape variables to be strongly correlated with climate at larger spatial scales. This is substantiated by the scale-dependent covariation between these variables, as the rising contributions of BRHS and of the interaction is associated with a substantial increase in Pearson correlation with LRHS.

Importantly, we aimed to isolate the effect of genetic drift while modelling the relationship between gene flow and the environment, by using gravity models that incorporate observed heterozygosity as a proxy for node size (i.e., subpopulation size). Our results show that heterozygosity within nodes is the strongest predictor of genetic differentiation—explaining 30% of the variance, compared to 15% for BRHS and 5% for LRHS. This finding is consistent with population genetic theory, which posits that genetic differentiation is shaped both by dispersal and resulting gene flow, and by genetic drift, whose effects increase as population size decreases (Hartl et al. 1997; Prunier et al. 2017; Savary et al. 2023). We thus emphasize the importance of accounting for genetic drift when assessing the influence of landscape features on gene flow. Importantly, our study area is located at the periphery of the species distribution range and the effect of drift might be locally intensified by founder effects. Future studies including samples from the whole distribution range would be welcome to further disentangle the relative contributions of historical versus contemporary factors on gene flow and genetic drift (Duncan et al. 2015). Interestingly, our analyses also revealed that the impact of drift on genetic differentiation varies across regions and is stronger in mountainous areas, where population sizes are likely smaller due to lower habitat suitability. Our analyses were based on Bray–Curtis dissimilarity as the measure of pairwise genetic differentiation. Although the choice of individual-based genetic distance metric has been shown to influence landscape genetic inference in some cases (Beninde et al. 2023), Bray–Curtis was highly correlated with both the proportion of shared alleles (Dps; $r = 0.993$) and Euclidean genetic distance ($r = 0.931$). More generally, all three pairwise distance metrics were strongly correlated with one another (Pearson’s $r = 0.923$ – 0.993 ; Fig. S10). These complementary results suggest that the patterns reported here are unlikely to depend on the particular genetic distance metric used, at least among the commonly used individual-based metrics evaluated.

As for genetic diversity, models based on pairwise distances derived from raw landscape variables performed comparably to those based on LRHS, although they generally received slightly stronger statistical support. This slightly better performance of the raw-variable models may reflect a loss of information when using LRHS compared to modelling the contribution of individual environmental variables separately. However, our results suggest that LRHS retains most of the environmental information influencing connectivity, while simplifying the multidimensional environmental space into a single integrative variable.

Using SDMs to predict genetic diversity and connectivity. Our results provide rare empirical evidence that SDM-derived relative habitat suitability (RHS) can reliably predict both genetic diversity and connectivity. This finding contrasts with several previous studies that failed to detect significant relationships between RHS and population genetic characteristics across a wide range of plant and animal species (Lee-Yaw et al. 2022). By showing that both bioclimatic (BRHS) and landscape (LRHS) suitability indices explain variation in heterozygosity and genetic differentiation, we demonstrate that SDMs can serve as effective proxies for population genetic health—at least under certain ecological contexts.

This success likely reflects, at least in part, several methodological advantages of our approach. First, we explicitly addressed the issue of spatial scale, showing that BRHS and LRHS capture

distinct yet complementary processes—broad-scale climatic gradients versus local landscape constraints. Second, the use of high-resolution genomic data (tens of thousands of SNPs) provided robust estimates of genetic metrics. Finally, by applying gravity models, we were able to quantify the effects of RHS on genetic differentiation while simultaneously accounting for the influence of genetic drift, thereby refining the interpretation of gene flow–habitat suitability relationships.

We also focused on a species whose demography is strongly shaped by environmental conditions (Cayuela, Arsovski et al. 2016; Cayuela et al. 2020, 2022), with highly variable population sizes and marked declines toward the edges of its range (Lescure et al. 2011). Moreover, the species exhibits limited dispersal ability and follows a diffusive dispersal mode, characterized by random and local movements (Cayuela et al. 2020). These features likely foster stronger links between habitat suitability, demography, and genetic processes than in species with greater movement capacity (e.g., birds, Grafius et al. 2017; large mammals, Zeller et al. 2018), long-distance or jump dispersal (e.g., Suarez et al. 2001), or large census and effective population sizes (N_e ; e.g., invertebrates, Clarke et al. 2024), which tend to be less dependent on local environmental conditions.

Overall, our results help clarify the ongoing debate on the appropriate conditions for using SDMs. Despite widespread application to assess population health and connectivity, empirical support for RHS–demography–genetics relationships remains weak (Lee-Yaw et al. 2022), and SDMs have faced substantial conceptual criticism (Keeley et al. 2017; Zeller et al. 2018; Sartor et al. 2022). Consistent with previous work (Cushman 2006; Keeley et al. 2017), our results show that RHS can predict genetic connectivity and may serve as a proxy for genetic diversity in species with limited dispersal and strong environmental sensitivity. Moving forward, we argue that uncovering general RHS–genetics relationships will depend on explicitly accounting for key limitations—particularly mismatches in spatial scale, incomplete niche representation in SDMs, limited resolution of genetic markers, and the difficulty of disentangling drift from gene flow—to establish a robust analytical framework capable of reliably detecting these links across systems.

DATA AVAILABILITY

Genomic data, associated metadata, and SDM projections used in this study are available from Zenodo: <https://doi.org/10.5281/zenodo.22210263>.

REFERENCES

- Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *J Appl Ecol* 43(6):1223–1232.
- AmphibiaWeb (2023) *Bombina variegata*: Yellow-Bellied Toad <https://amphibiaweb.org/species/2046> University of California, Berkeley, CA, USA. Accessed Jul 16, 2025.
- Araújo MB, Guisan A (2006) Five (or so) challenges for species distribution modelling. *J Biogeogr* 33(10):1677–1688.
- Araújo MB, Anderson RP, Márcia Barbosa A, Beale CM, Dormann CF, Early R et al. (2019) Standards for distribution models in biodiversity assessments. *Sci Adv* 5(1):eaat4858.
- Arntzen JW, Harmos K, Vörös J (2025) Assessment of environmental variables for species distribution modelling: insight from the mosaic distribution of red and yellow-bellied toads. *Basic Appl Ecol* 87:120–127.
- Balkenhol N, Cushman SA, Storfer A, Waits LP (2015) Introduction to landscape genetics—concepts, methods, applications. In: *Landscape Genetics* (eds N. Balkenhol, S.A. Cushman, A.T. Storfer and L.P. Waits), John Wiley & Sons, Ltd 1:1–8 <https://doi.org/10.1002/9781118525258.ch01>
- Beck J, Böller M, Erhardt A, Schwanghart W (2014) Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecol Inform* 19:10–15.
- Beebe TJ (2005) Conservation genetics of amphibians. *Heredity* 95(6):423–427.
- Beninde J, Wittische J, Frantz AC (2023) Quantifying uncertainty in inferences of landscape genetic resistance due to choice of individual-based genetic distance metric. *Mol Ecol Resour* 24(1):e13831.
- Bernabo I, Biondi M, Cittadino V, Sperone E, Iannella M (2022) Addressing conservation measures through fine-tuned species distribution models for an Italian endangered endemic anuran. *Glob Ecol Conserv* 39:e02302.
- Berry O, Tocher MD, Gleeson DM, Sarre SD (2005) Effect of vegetation matrix on animal dispersal: genetic evidence from a study of endangered skinks. *Conserv Biol* 19(3):855–864.
- Bohonak AJ (1999) Dispersal, gene flow, and population structure. *Q Rev Biol* 74(1):21–45.
- Boria RA, Olson LE, Goodman SM, Anderson RP (2014) Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecol Model* 275:73–77.
- Bowler DE, Bhandari N, Repke L, Beuthner C, Callaghan CT, Eichenberg D et al. (2022) Decision-making of citizen scientists when recording species observations. *Sci Rep* 12(1):11069.
- Boyer I, Cayuela H, Bertrand R, Isselin-Nondedeu F (2021) Improving biological relevance of model projections in response to climate change by considering dispersal amongst lineages in an amphibian. *J Biogeogr* 48(3):561–576.
- Boyer I, Pagano A, Isselin-Nondedeu F, Bertrand R (2025) Elevational range shifts among *Bombina* toad populations in the Alps in response to future climate change. *Regional Environ Change* 25(1):23.
- Bray JR, Curtis JT (1957) An ordination of the upland forest communities of southern Wisconsin. *Ecol Monogr* 27(4):326–349.
- Carnaval AC, Hickerson MJ, Haddad CF, Rodrigues MT, Moritz C (2009) Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science* 323(5915):785–789.
- Caye K, Deist TM, Martins H, Michel O, François O (2016) TESS3: fast inference of spatial population structure and genome scans for selection. *Mol Ecol Resour* 16(2):540–548.
- Cayuela H, Lambrey J, Vacher JP, Miaud C (2015) Highlighting the effects of land-use change on a threatened amphibian in a human-dominated landscape. *Popul Ecol* 57(2):433–443.
- Cayuela H, Arsovski D, Thirion JM, Bonnaire E, Pichenot J, Boitard S et al. (2016) Contrasting patterns of environmental fluctuation contribute to divergent life histories among amphibian populations. *Ecology* 97(4):980–991.
- Cayuela H, Boualit L, Arsovski D, Bonnaire E, Pichenot J, Bellec A et al. (2016) Does habitat unpredictability promote the evolution of a colonizer syndrome in amphibian metapopulations? *Ecology* 97(10):2658–2670.
- Cayuela H, Rougemont Q, Prunier JG, Moore JS, Clobert J, Besnard A et al. (2018) Demographic and genetic approaches to study dispersal in wild animal populations: A methodological review. *Mol Ecol* 27(20):3976–4010.
- Cayuela H, Besnard A, Cote J, Laporte M, Bonnaire E, Pichenot J et al. (2020) Anthropogenic disturbance drives dispersal syndromes, demography, and gene flow in amphibian populations. *Ecol Monogr* 90(2):e01406.
- Cayuela H, Monod-Broca B, Lemaître JF, Besnard A, Gippet JM, Schmidt BR et al. (2022) Compensatory recruitment allows amphibian population persistence in anthropogenic habitats. *Proc Natl Acad Sci* 119(38):e2206805119.
- Chandler M, See L, Copas K, Bonde AM, López BC, Danielsen F et al. (2017) Contribution of citizen science towards international biodiversity monitoring. *Biol Conserv* 213:280–294.
- Clarke SH, Lawrence ER, Matte JM, Gallagher BK, Salisbury SJ, Michaelides SN et al. (2024) Global assessment of effective population sizes: Consistent taxonomic differences in meeting the 50/500 rule. *Mol Ecol* 33(11):e17353.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA et al. (2011) The variant call format and VCFtools. *Bioinformatics* 27(15):2156–2158.
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO et al. (2021) Twelve years of SAMtools and BCFtools. *Gigascience* 10(2):giab008.
- De Kort H, Prunier JG, Ducatez S, Honnay O, Baguette M, Stevens VM et al. (2021) Life history, climate and biogeography interactively affect worldwide genetic diversity of plant and animal populations. *Nat Commun* 12(1):516.
- Della Rocca F, Musiani M, Galaverni M, Milanese P (2024) Improving online citizen science platforms for biodiversity monitoring. *J Biogeogr* 51(12):2412–2423.
- Denoël M, Lehmann A (2006) Multi-scale effect of landscape processes and habitat quality on newt abundance: implications for conservation. *Biol Conserv* 130(4):495–504.
- Dorant Y, Cayuela H, Wellband K, Laporte M, Rougemont Q, Mérot C et al. (2020) Copy number variants outperform SNPs to reveal genotype–temperature association in a marine species. *Mol Ecol* 29(24):4765–4782.
- Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G et al. (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36(1):27–46.
- Dray S, Dufour AB (2007) The ade4 package: implementing the duality diagram for ecologists. *J Stat Softw* 22:1–20.
- Dufresnes C, Suchan T, Smirnov NA, Denoël M, Rosanov JM, Litvinchuk SN (2021) Revisiting a speciation classic: comparative analyses support sharp but leaky transitions between *Bombina* toads. *J Biogeogr* 48(3):548–560.

- Dufresnes C, Lukanov S, Gippner S, Ambu J, Strachinis I, Arsovski D et al. (2025) Historical biogeography and systematics of yellow-bellied toads (*Bombina variegata*), with the description of a new subspecies from the Balkans. *Vertebrate Zool* 75:1–30.
- Duncan SI, Crespi EJ, Mattheus NM, Rissler LJ (2015) History matters more when explaining genetic diversity within the context of the core–periphery hypothesis. *Mol Ecol* 24(16):4323–4336.
- ECOTER (2011) Plan national d'actions en faveur du Sonneur à ventre jaune (*Bombina variegata*). Ministère de l'Ecologie, de l'Energie, du Développement durable et de la Mer / DREAL Lorraine. 195 pp.
- Edwards LJ, Muller KE, Wolfinger RD, Qaqish BF, Schabenberger O (2008) An R2 statistic for fixed effects in the linear mixed model. *Stat Med* 27(29):6137–6157.
- Ellstrand NC, Elam DR (1993) Population genetic consequences of small population size: implications for plant conservation. *Annu Rev Ecol Syst* 24(1):217–242.
- Feldman MJ, Imbeau L, Marchand P, Mazerolle MJ, Darveau M, Fenton NJ (2021) Trends and gaps in the use of citizen science derived data as input for species distribution models: A quantitative review. *PLoS One* 16(3):e0234587.
- Frankham R (1996) Relationship of genetic variation to population size in wildlife. *Conserv Biol* 10(6):1500–1508.
- Frichot E, François O (2015) LEA: An R package for landscape and ecological association studies. *Methods Ecol Evolut* 6(8):925–929.
- Friedman L, Wall M (2005) Graphical views of suppression and multicollinearity in multiple linear regression. *Am Statistician* 59(2):127–136.
- Gardner AS, Maclean IM, Gaston KJ (2019) Climatic predictors of species distributions neglect biophysiological meaningful variables. *Divers Distrib* 25(8):1318–1333.
- GBIF.org (2024) GBIF Occurrence Download <https://doi.org/10.15468/dl.kaucy8>.
- Goicolea T, Adde A, Broennimann O, García-Viñas JJ, Gastón A, José Aroca-Fernández M et al. (2025) Spatially-nested hierarchical species distribution models to overcome niche truncation in national-scale studies. *Ecography* 2025(3):e07328.
- Goudet J (2005) Hierfstat, a package for R to compute and test hierarchical F-statistics. *Mol Ecol notes* 5(1):184–186.
- Grafius DR, Corstanje R, Siriwardena GM, Plummer KE, Harris JA (2017) A bird's eye view: using circuit theory to study urban landscape connectivity for birds. *Landscape Ecol* 32(9):1771–1787.
- Guisan A, Graham CH, Elith J, Huettmann F, NCEAS Species Distribution Modelling Group (2007) Sensitivity of predictive species distribution models to change in grain size. *Divers Distrib* 13(3): 332–340.
- Guisan A, Tingley R, Baumgartner JB, Naujokaitis-Lewis I, Sutcliffe PR, Tulloch AI et al. (2013) Predicting species distributions for conservation decisions. *Ecol Lett* 16(12):1424–1435.
- Han H, Guo X, Yu H (2016) Variable selection using mean decrease accuracy and mean decrease gini based on random forest. In 2016, the 7th IEEE International Conference on Software Engineering and Service Science (ICSESS) (pp. 219–224). IEEE
- Hantzschmann AM, Sinsch U, Göttlicher C, Pröhl H (2021) Conservation genetics of yellow-bellied toads (*Bombina variegata*): a matter of geographical scale and isolation. *Conserv Genet* 22(1):83–96.
- Hartl DL, Clark AG, Clark AG (1997) Principles of population genetics (Vol. 116, p. 542). Sunderland, MA: Sinauer Associates.
- Hedrick PW, Kalinowski ST (2000) Inbreeding depression in conservation biology. *Annu Rev Ecol Syst* 31(1):139–162.
- Hijmans RJ, Phillips S, Leathwick J, Elith J, Hijmans MRJ (2017) Package 'dismo'. *Circles* 9(1):1–68.
- Holt RD (2009) Bringing the Hutchinsonian niche into the 21st century: ecological and evolutionary perspectives. *Proc Natl Acad Sci USA* 106(suppl.2):19659–19665.
- Holt RD, Keitt TH, Lewis MA, Maurer BA, Taper ML (2005) Theoretical models of species' borders: single species approaches. *Oikos* 108(1):18–27.
- Hutchinson GE (1957) Concluding remarks. In Cold Spring Harbor symposia on quantitative biology (Vol. 22, pp. 415–427). Cold Spring Harbor Laboratory Press.
- Jackson DA (1993) Stopping rules in principal components analysis: a comparison of heuristic and statistical approaches. *Ecology* 74(8):2204–2214.
- Jarnevich CS, Stohlgren TJ, Kumar S, Morissette JT, Holcombe TR (2015) Caveats for correlative species distribution modeling. *Ecol Inform* 29:6–15.
- Karger DN, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza RW et al. (2017) Climatologies at high resolution for the Earth land surface areas. *Sci Data* 4: 170122.
- Karger DN, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza RW et al. (2018) Data from: Climatologies at high resolution for the earth's land surface areas. Dryad Digit Reposit. <https://doi.org/10.5061/dryad.kd1d4>.
- Keeley AT, Beier P, Keeley BW, Fagan ME (2017) Habitat suitability is a poor proxy for landscape connectivity during dispersal and mating movements. *Landscape Urban Plan* 161:90–102.
- Kleiben J, Balkenhol N, Pröhl H (2021) Landscape genetics of the yellow-bellied toad (*Bombina variegata*) in the Northern Weser Hills of Germany. *Diversity* 13(12):623.
- Kuhn M (2008) Building predictive models in R using the caret package. *J Stat Softw* 28:1–26.
- Lai J, Zou Y, Zhang S, Zhang X, Mao L (2022) glmm.hp: an R package for computing individual effect of predictors in generalized linear mixed models. *J Plant Ecol* 15(6):1302–1307.
- Lawton JH (1994) Population dynamic principles. *Philos Trans R Soc Lond Ser B: Biol Sci* 344(1307):61–68.
- Lee-Yaw JA, McCune JL, Pironon S, Sheth SN (2022) Species distribution models rarely predict the biology of real populations. *Ecography* 2022(6):e05877.
- Lescure J, Pichenot J, Cochard PO (2011) Régression de *Bombina variegata* (Linné, 1758) en France par l'analyse de sa répartition passée et présente. *Bull de la Société Herpétol de Fr* 137:5–41.
- Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.
- Liaw A, Wiener M (2002) Classification and regression by randomForest. *R N* 2(3):18–22.
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17(1):10–12.
- Mascher M, Muehlbauer GJ, Rokhsar DS, Chapman J, Schmutz J, Barry K et al. (2013) Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). *Plant J* 76(4):718–727.
- McKinney GJ, Waples RK, Seeb LW, Seeb JE (2017) Paralogues are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. *Mol Ecol Resour* 17(4):656–669.
- McRae BH (2006) Isolation by resistance. *Evolution* 60(8):1551–1561.
- McRae BH, Shah VB, Mohapatra TK (2009) Circuitscape user's guide. The University of California, Santa Barbara.
- Mood AM (1971) Partitioning variance in multiple regression analyses as a tool for developing learning models. *Am Educ Res J* 8(2):191–202.
- Moran PA (1953) The statistical analysis of the Canadian lynx cycle: II synchronisation and meteorology. *Aust J Zool* 1(3):291–298.
- Murphy MA, Dezzani R, Pilliod DS, Storer A (2010) Landscape genetics of high mountain frog metapopulations. *Mol Ecol* 19(17):3634–3649.
- Naimi B, Hamm NA, Groen TA, Skidmore AK, Toxopeus AG (2014) Where is positional uncertainty a problem for species distribution modelling? *Ecography* 37(2):191–203.
- Oksanen J, Simpson GL, Blanchet FG, Kindt R, Legendre P, Minchin PR, et al. (2022) vegan: community ecology package. R package version 2.6-2. See <https://CRAN.R-project.org/package=vegan>.
- Ogrodowczyk A, Krása A, Schmidt BR, Naumov B, Anthony BP, Cogălniceanu D, et al. (2024) *Bombina variegata*. The IUCN Red List of Threatened Species 2024: e.T77970289A228179967. <https://doi.org/10.2305/IUCN.UK.2024-2.RLTS.T77970289A228179967.en> Accessed on 15 2026.
- Pagel J, Schurr FM (2012) Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. *Glob Ecol Biogeogr* 21(2):293–304.
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* 7(5):e37135.
- Pichenot J (2008) Contribution à la Biologie de la Conservation du Sonneur à ventre jaune (*Bombina variegata* L.) Ecologie Spatiale et Approche Multi-Echelles de la Sélection de L'habitat en Limite Septentrionale de Son Aire de Répartition. Université de Reims Champagne-Ardenne.
- Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One* 7(2):e32253.
- Préau C, Isselin-Nondedeu F, Sellier Y, Bertrand R, Grandjean F (2019) Predicting suitable habitats of four range margin amphibians under climate and land-use changes in southwestern France. *Regional Environ Change* 19(1):27–38.
- Preuk N, Romero-Mujalli D, Zurell D, Steinbauer M, Juergen Kreyling A (2025) Randomising spatial patterns supports the integration of intraspecific variation in ecological niche models. *Ecography* 2025(2), e07289.
- Pröhl H, Auffarth J, Bergmann T, Buschmann H, Balkenhol N (2021) Conservation genetics of the yellow-bellied toad (*Bombina variegata*): population structure, genetic diversity and landscape effects in an endangered amphibian. *Conserv Genet* 22(3):513–529.
- Prunier JG, Colyn M, Legendre X, Nimon KF, Flamand MC (2015) Multicollinearity in spatial genetics: separating the wheat from the chaff using commonality analyses. *Mol Ecol* 24(2):263–283.
- Prunier JG, 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 Ecol Evolut* 8(12):1866–1877.
- Razgour O, Taggart JB, Manel S, Juste J, Ibanez C, Rebelo H et al. (2018) An integrated framework to identify wildlife populations under threat from climate change. *Mol Ecol Resour* 18(1):18–31.

- Savary P, Foltête JC, Moal H, Vuidel G, Garnier S (2023) Inferring landscape resistance to gene flow when genetic drift is spatially heterogeneous. *Mol Ecol Resour* 23(7):1574–1588.
- Sartor CC, Wan HY, Pereira JA, Eizirik E, Trigo TC, de Freitas TRO et al. (2022) Landscape genetics outperforms habitat suitability in predicting landscape resistance for congeneric cat species. *J Biogeogr* 49(12):2206–2217.
- Scheele BC, Boyd CE, Fischer J, Fletcher AW, Hanspach J, Hartel T (2014) Identifying core habitat before it's too late: the case of *Bombina variegata*, an inter-nationally endangered amphibian. *Biodivers Conserv* 23(3):775–780.
- Schmeller DS, Henry PY, Julliard R, Gruber B, Clobert J, Dziock F et al. (2009) Advantages of volunteer-based biodiversity monitoring in Europe. *Conserv Biol* 23(2):307–316.
- Seaborn T, Griffith D, Kliskey A, Caudill CC (2021) Building a bridge between adaptive capacity and adaptive potential to understand responses to environmental change. *Glob Change Biol* 27(12):2656–2668.
- Shirk AJ, Landguth EL, Cushman SA (2017) A comparison of individual-based genetic distance metrics for landscape genetics. *Mol Ecol Resour* 17(6):1308–1317.
- Skey ED, Ottewell KM, Spencer PB, Shaw RE (2023) Empirical landscape genetic comparison of single nucleotide polymorphisms and microsatellites in three arid-zone mammals with high dispersal capacity. *Ecol Evolut* 13(5):e10037.
- Snow R (2005) Continental climate and continentality. In *Encyclopedia of World Climatology* (pp. 303–305). Springer, Dordrecht.
- Southwood TR (1977) Habitat, the templet for ecological strategies? *J Anim Ecol* 46(2):337–365.
- Speybroeck J, Beukema W, Bok B, Van Der Voort J (2016) Field guide to the amphibians and reptiles of Britain and Europe. Bloomsbury publishing.
- Suarez AV, Holway DA, Case TJ (2001) Patterns of spread in biological invasions dominated by long-distance jump dispersal: insights from Argentine ants. *Proc Natl Acad Sci USA* 98(3):1095–1100.
- Sunde J, Yildirim Y, Tibblin P, Forsman A (2020) Comparing the performance of microsatellites and RADseq in population genetic studies: Analysis of data for pike (*Esox lucius*) and a synthesis of previous studies. *Front Genet* 11:218.
- Valavi R, Elith J, Lahoz-Monfort JJ, Guillera-Aroita G (2021) Modelling species presence-only data with random forests. *Ecography* 44(12):1731–1742.
- Valavi R, Guillera-Aroita G, Lahoz-Monfort JJ, Elith J (2022) Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. *Ecol Monogr* 92(1):e01486.
- Vasimuddin M, Misra S, Li H, Aluru S (2019) Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)* (pp. 314–324). IEEE.
- Wang JJ (2020) Topographic path analysis for modelling dispersal and functional connectivity: Calculating topographic distances using the topoDistance R package. *Methods Ecol Evolut* 11(2):265–272.
- Wang YH, Yang KC, Bridgman CL, Lin LK (2008) Habitat suitability modelling to correlate gene flow with landscape connectivity. *Landsc Ecol* 23(8):989–1000.
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 1358–1370.
- Wickham H (2011) ggplot2. *Wiley Interdiscip Rev Comput Stat* 3(2):180–185.
- Wright S (1943) Isolation by distance. *Genetics* 28(2):114.
- Yannic G, Pellissier L, Ortego J, Lecomte N, Couturier S, Cuyler C et al. (2014) Genetic diversity in caribou linked to past and future climate change. *Nat Clim Change* 4(2):132–137.
- Zeller KA, Jennings MK, Vickers TW, Ernest HB, Cushman SA, Boyce WM (2018) Are all data types and connectivity models created equal? Validating common connectivity approaches with dispersal data. *Divers Distrib* 24(7):868–879.
- Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics *Mol Ecol* 22(11):3124–3140.
- Cushman SA, McKelvey KS, Hayden J, Schwartz MK (2006) Gene flow in complex landscapes: testing multiple hypotheses with causal modeling *The American Naturalist* 168(4):486–499.
- Kijanović A, Vukov T, Mirč M, Krizmanić I, Tomašević Kolarov N (2023) Inability of yellow-bellied toad to accelerate metamorphosis in desiccation conditions *Journal of Zoology* 320(2):96–107.
- Sinsch U (2020) Carry-over effects of the larval environment on the post-metamorphic performance of *Bombina variegata* (Amphibia, Anura) *Herpetological Journal* 30(3):126–134.

ACKNOWLEDGEMENTS

This research was supported by the Conseil Général d'Isère (CG38), the PNA Sonneur à ventre jaune and the Union Nationale des Producteurs de Granulats (UNPG), as well as the Institut Rhônealpin des systèmes complexes (IXXI). We thank Rémi Bogey, Rémi Fonters, Fabrice Darinot, Léo Déjeux, and Jérémie Bruset for their help with the field sampling, as well as Vanessa Gardette and Adeline Dumet for their help with the DNA extraction. We are also much indebted to the following associations for providing access to their exhaustive occurrence database of *Bombina variegata*: CEN PACA, LPO Ain, LPO AuRA, LPO BFC, PNA Sonneur AURA, SHNA-OFAB. Tissue sampling was carried out under permits DDPP01-21-115, 38202103121110, DDT/SEEF2021-0181 and 26-2021-04-06-00005. We thank the two anonymous reviewers for their constructive feedback, which greatly contributed to improving the manuscript.

AUTHOR CONTRIBUTIONS

BMB, HC, QR, and JPL designed the research project; BMB, HC, LB, QR, and JPL performed research; BMB, QR, and JPL analysed data; all authors wrote and revised the article.

COMPETING INTERESTS

The authors declare no competing interests.

RESEARCH ETHICS STATEMENT

All research was performed in accordance with relevant guidelines and regulations.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41437-026-00885-y>.

Correspondence and requests for materials should be addressed to Benjamin Monod-Broca.

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

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.