

Text S1: Detailed methodology for gravity models

Gravity models are particularly useful in landscape genetics because they explicitly consider the local effect of each node (or site or local population) of origin 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 the N_e is heterogeneous among sampled sites (Prunier et al. 2017, Savary et al. 2023). However, gravity models, which are a kind of social relations regression models that were developed to analyse dyadic interactions within a network (reviewed in Hoff 2021), are usually intended for directed flows between the origin and the destination while the genetic differentiation between sites (or individuals) is undirected, and properties of site as origin and destination are undistinguishable (and would make no sense for individual genetic distances). In landscape genetics, it follows that the dataset (i.e., pairwise genetic distances) reduces to $\frac{N*(N-1)}{2}$ if N is the number of nodes and after linearization by log transformation, the equation of the gravity model should reduce to:

$$\ln(DG_{ij}) = \beta_0 + \beta_1 * \ln(c_i) + \beta_1 * \ln(c_j) + \beta_2 * \ln(DL_{ij}) + \tau_i + \tau_j + e_{ij},$$

with DG_{ij} the genetic distance between node i and j , DL_{ij} the landscape distance (including landscape resistance) between i and j , c_i and c_j the specific characteristic of node i and j , each beta term the estimated parameter associated to each effect, the ‘node’ random effect $\tau \sim \text{Dist } N(0, \sigma_\tau^2)$ and the residual dyadic error $e_{ij} \sim \text{Dist } N(0, \sigma_e^2)$. Since $DG_{ij} = DG_{ji}$, a symmetry constraint should be applied on the effect of node characteristic, so that the estimated effect is the same whether the node is i or j . This can be easily applied using only one explanatory term as the sum of both node characteristics. In this simple form, site random effects are implemented using a Toeplitz(1) covariance structure to explicitly account for non-independence of pairwise genetic distances according to the nodes or origin (Clarke et al. 2002, Selkoe et al. 2010).

In the present work, we used the method presented in Prunier et al. (2013) and Cayuela et al. (2020) for application on individual genetic distances in spatially hierarchized sampling schemes as is the case in our study. Moreover, as a preliminary analysis revealed substantial variance heterogeneity between pairwise distances belonging to sites sampled in the southern margin of the species range relative to others, we also allowed a separate estimation of the random effects for the two sets of sampling sites. This resulted in two covariance parameters for each set of sampling sites, one for the node random effect and the other for the individual random effect. Since obtaining an estimation of N_e is a difficult task, authors usually replace it by a proxy 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 CMR data and the second requiring arbitrary buffer delimitation around ponds. In the present study, we therefore used the mean observed heterozygosity at each node. The resulting model was thus of the form:

$$\ln(DG_{ij}) = \beta_0 + \beta_1 * (\ln(Ho_i) + \ln(Ho_j)) + \beta^T X_{ij} + \tau_i + \tau_j + \tau_k + \tau_l + e_{ij},$$

where X_{ij} is a p dimensional vector of log transformed landscape distance regressors, β^T the vector of corresponding parameters to be estimated, Ho the observed heterozygosity, τ_i and τ_j the node random effect for each sector, τ_k and τ_l the individual random effect for each sector ($\tau \sim \text{Dist } N(0, \sigma_\tau^2)$ for all τ), and $e_{ij} \sim \text{Dist } N(0, \sigma_e^2)$ the residual dyadic error.

References

- Cayuela, H., Besnard, A., Cote, J., Laporte, M., Bonnaire, E., Pichenot, J., ... & Léna, J. P. (2020). Anthropogenic disturbance drives dispersal syndromes, demography, and gene flow in amphibian populations. *Ecological Monographs*, 90(2), e01406.
- Clarke, R. T., Rothery, P., & Raybould, A. F. (2002). Confidence limits for regression relationships between distance matrices: estimating gene flow with distance. *Journal of Agricultural, Biological, and Environmental Statistics*, 7(3), 361-372.
- Hoff, P. (2021). Additive and multiplicative effects network models. *Statistical Science*, 36(1), 34-50.
- Murphy, M. A., Dezzani, R., Pilliod, D. S., & Storfer, A. (2010). Landscape genetics of high mountain frog metapopulations. *Molecular ecology*, 19(17), 3634-3649.
- Prunier, J. G., Dubut, V., Chikhi, L., & Blanchet, S. (2017). Contribution of spatial heterogeneity in effective population sizes to the variance in pairwise measures of genetic differentiation. *Methods in Ecology and Evolution*, 8(12), 1866-1877.
- Prunier, J. G., Kaufmann, B., Fenet, S., Picard, D., Pompanon, F., Joly, P., & Lena, J. P. (2013). Optimizing the trade-off between spatial and genetic sampling efforts in patchy populations: Towards a better assessment of functional connectivity using an individual-based sampling scheme. *Molecular ecology*, 22(22), 5516-5530.
- Savary, P., Foltête, J. C., Moal, H., Vuidel, G., & Garnier, S. (2023). Inferring landscape resistance to gene flow when genetic drift is spatially heterogeneous. *Molecular Ecology Resources*, 23(7), 1574-1588.
- Selkoe, K. A., Watson, J. R., White, C., Horin, T. B., Iacchei, M., Mitarai, S., ... & Toonen, R. J. (2010). Taking the chaos out of genetic patchiness: seascape genetics reveals ecological and oceanographic drivers of genetic patterns in three temperate reef species. *Molecular ecology*, 19(17), 3708-3726.

Text S2: Full details on the comparison between individual-based *versus* sampling-site-based models linking BRHS and LRHS to genetic diversity.

Since the random effect of sampling sites was estimated at the lower bound (*i.e.*, 0) in our models linking both RHS with genetic diversity while correcting for spatial autocorrelation, we constructed sampling-site-level equivalents using site-specific heterozygosity means instead of individual heterozygosity. These models produced nearly identical results at both scales (Table S4). Minor quantitative differences arise in scale selection (10.5 km vs. 11 km), explained variance, and coefficient estimates. The most notable discrepancy is in the shared variance between BRHS and LRHS (-0.131 vs. -0.057), which produces a negative total variance for LRHS at the sampling site level. This does not affect our interpretation, however: LRHS still acts as a suppressor variable correcting the BRHS estimate, albeit to a slightly greater degree. Overall, the consistency between individual- and sampling-site-level models tends to strengthen our genetic diversity analysis.

In summary, sampling-site-level analyses confirm our conclusions drawn from individual-level analyses: (1) SDM-derived LRHS and BRHS predict observed heterozygosity; (2) LRHS has an effect at smaller scales, while BRHS is expressed at larger ones; and (3) LRHS seems to act as a suppressor variable, correcting the prediction from BRHS.

Figure S1: Pearson correlation coefficients between landscape variables used to build the landscape SDM.

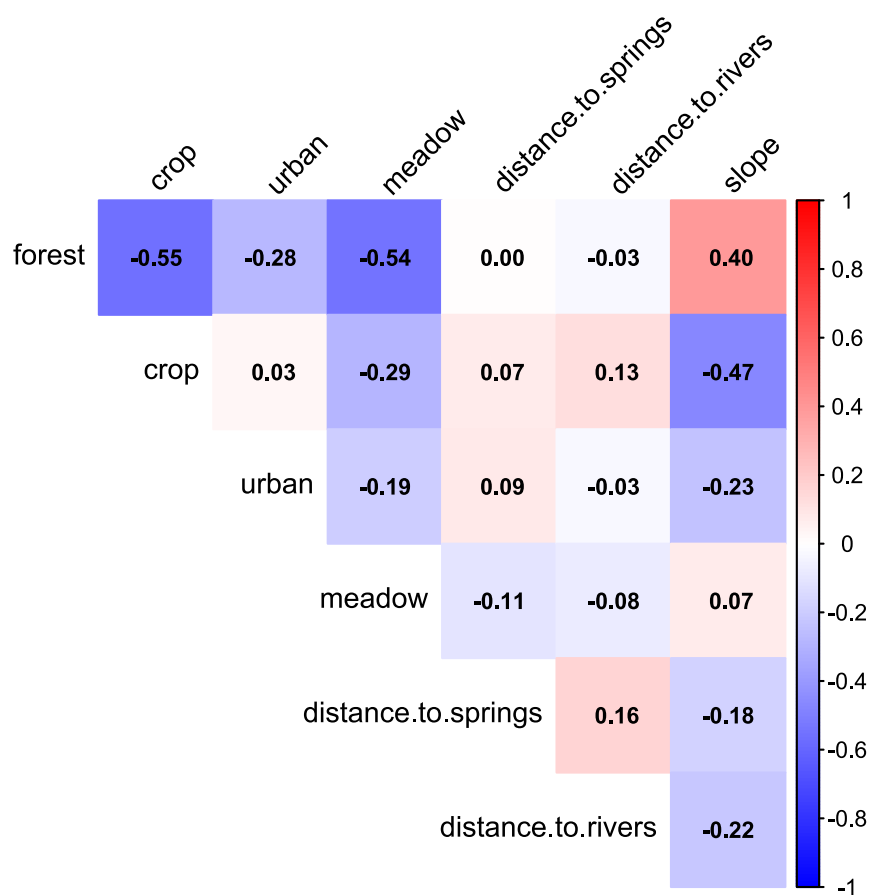


Figure S2: Maps of occurrence data for the Carpathian-Western *Bombina variegata* lineage. Continental-level records from GBIF with (A) an accuracy of 5000 m, used in the SDMs; and (B) with an accuracy of 1000 m. Regional-level records from (C) regional associations (used in the SDMs, see acknowledgements) and (D) GBIF with an accuracy of 250 m, overlaid with the bioclimatic SDM.

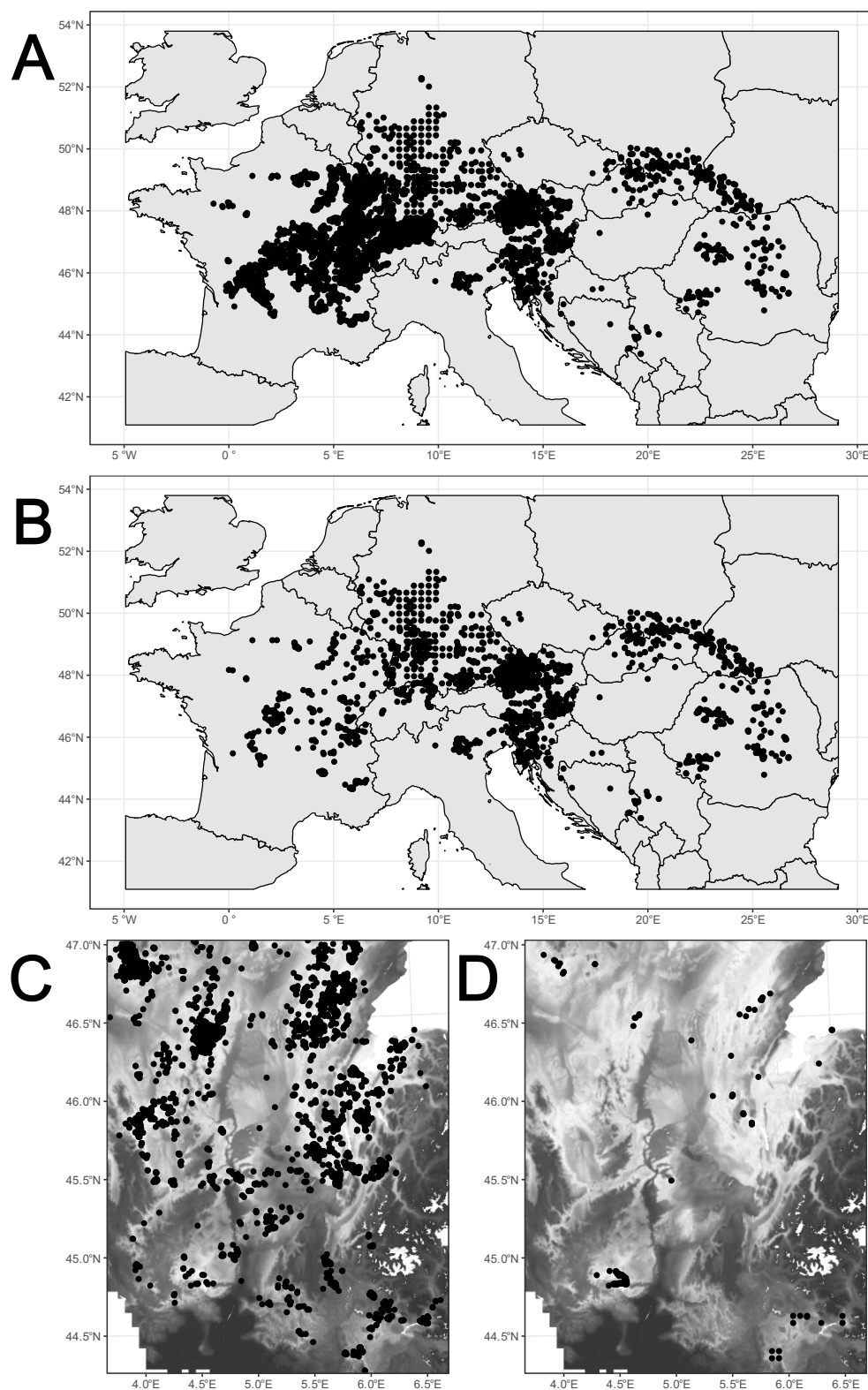


Figure S3: Comparison of SDMs built with all records with uncertainty <5,000m (used in our study, BRHS) and only records matching the resolution of bioclimatic rasters from CHELSA (uncertainty <1,000m), and residuals of the linear model (SDM 1,000m ~ SDM 5,000m). Removing occurrences with a precision between 1,000m and 5,000m leads to severely underestimating the species' presence in France, and especially in our study area.

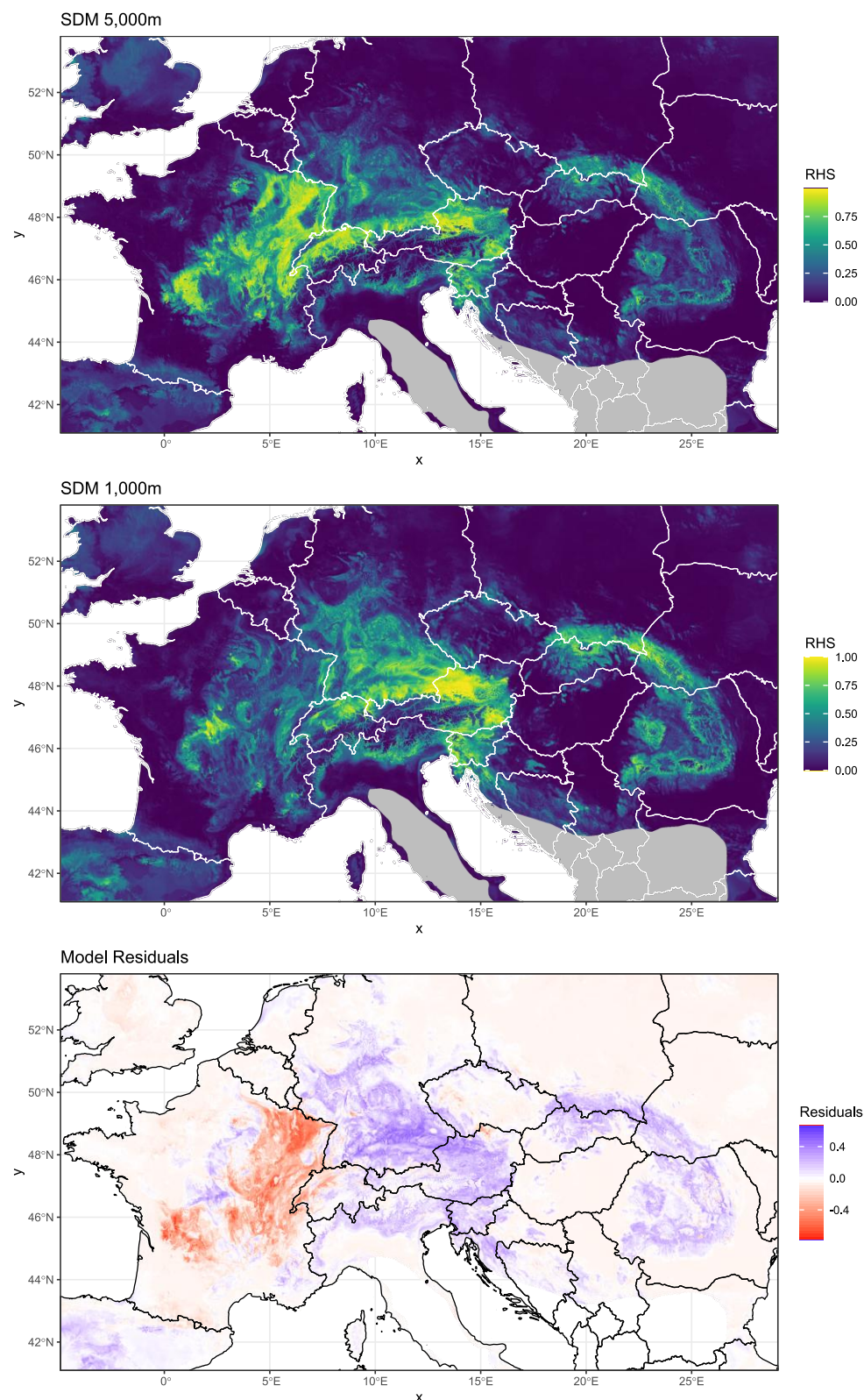


Figure S4: Semi-variograms of each bioclimatic raster used to build the continental SDMs over the whole range of pseudo-absence generation (up to 200,000m), with red areas showing the range of coordinate uncertainty from 0 to the 5,000 m upper limit for occurrence data from GBIF.

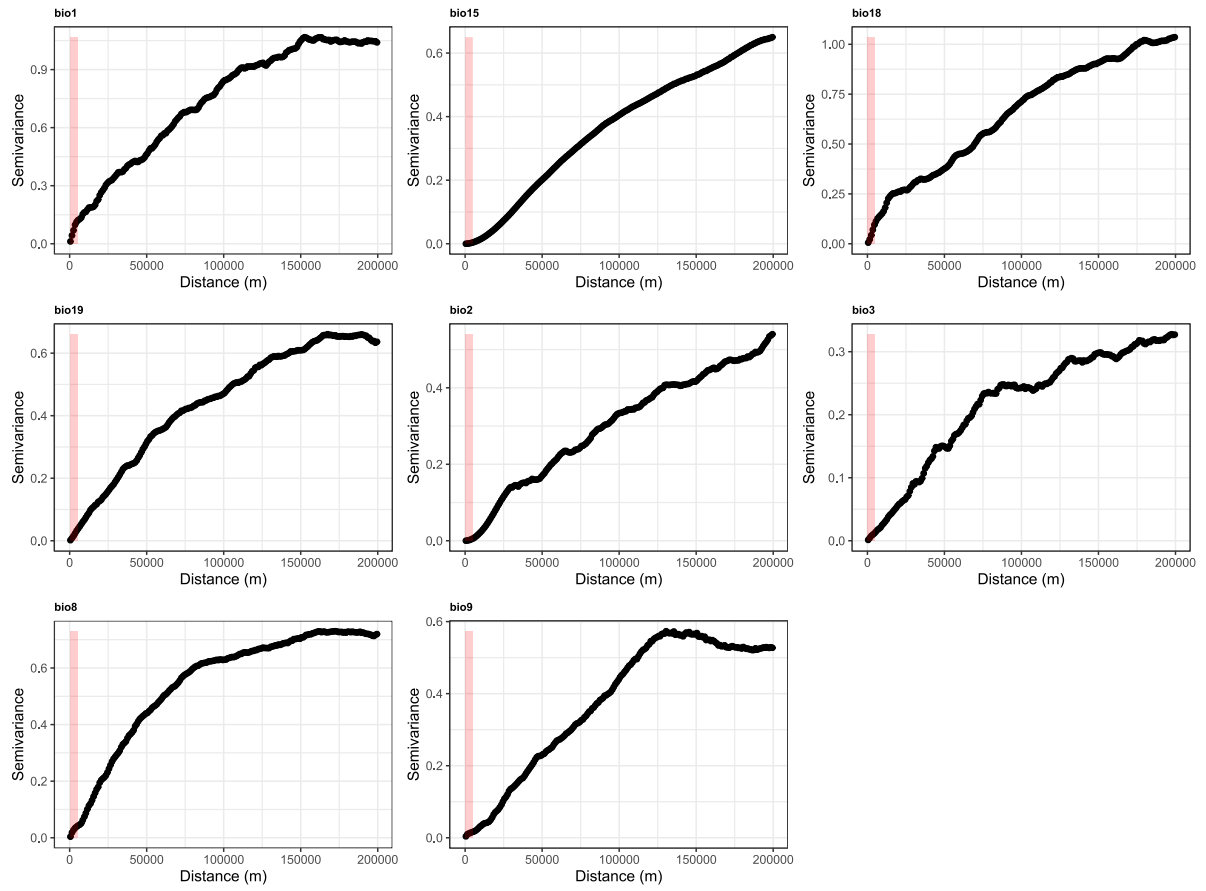


Figure S5: Comparison of SDMs built with pseudo-absences (PA) generated within 1km (used in our study, BRHS) and within 5 km, and the residuals of a simple linear model (SDM 1km ~ SDM 5km). These two models are highly correlated (Pearson's $r = 0.992$).

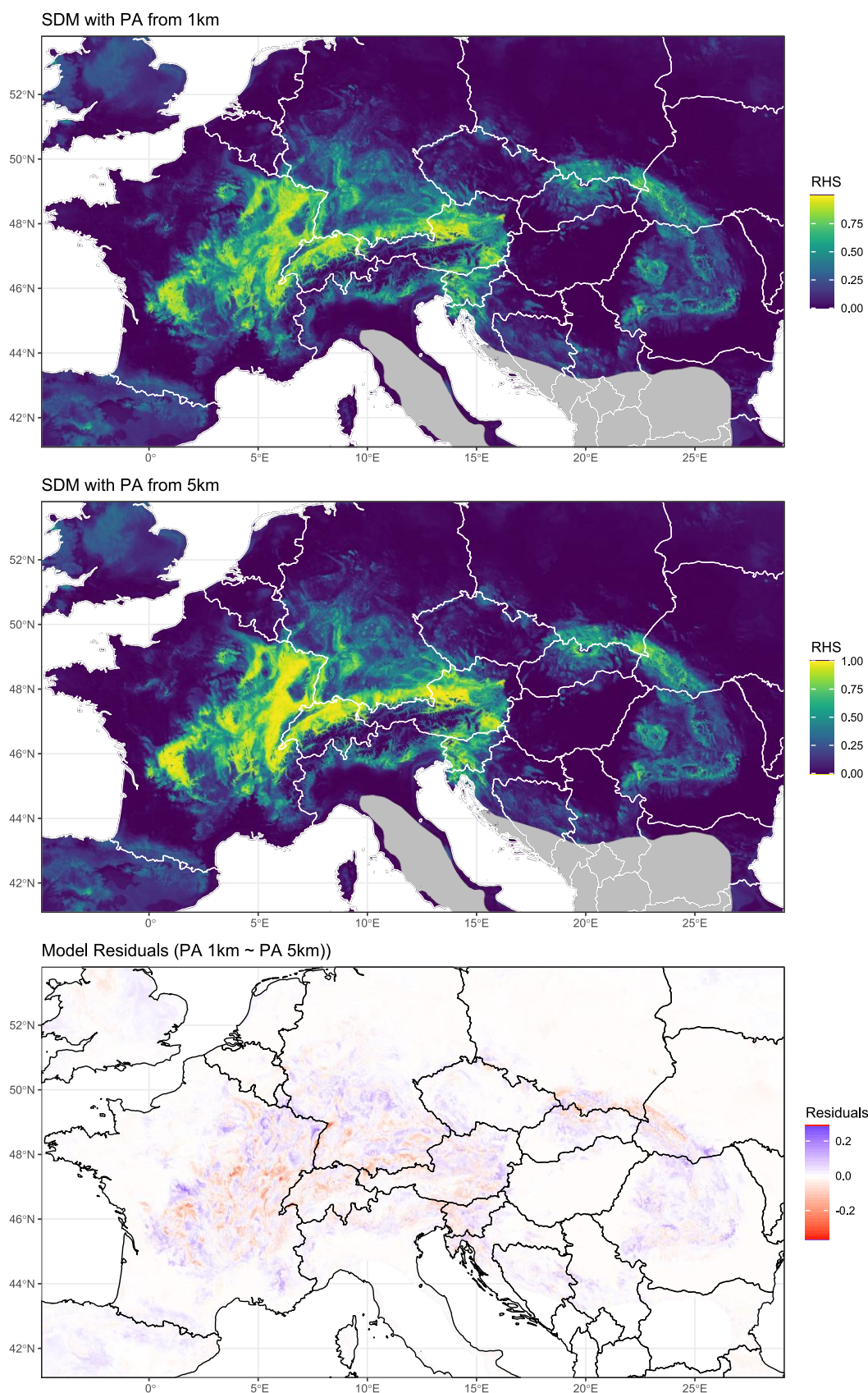


Figure S6: Correlation between LRHS and BRHS effective distances as a function of maximal distance considered (minimal distance = 3 km).

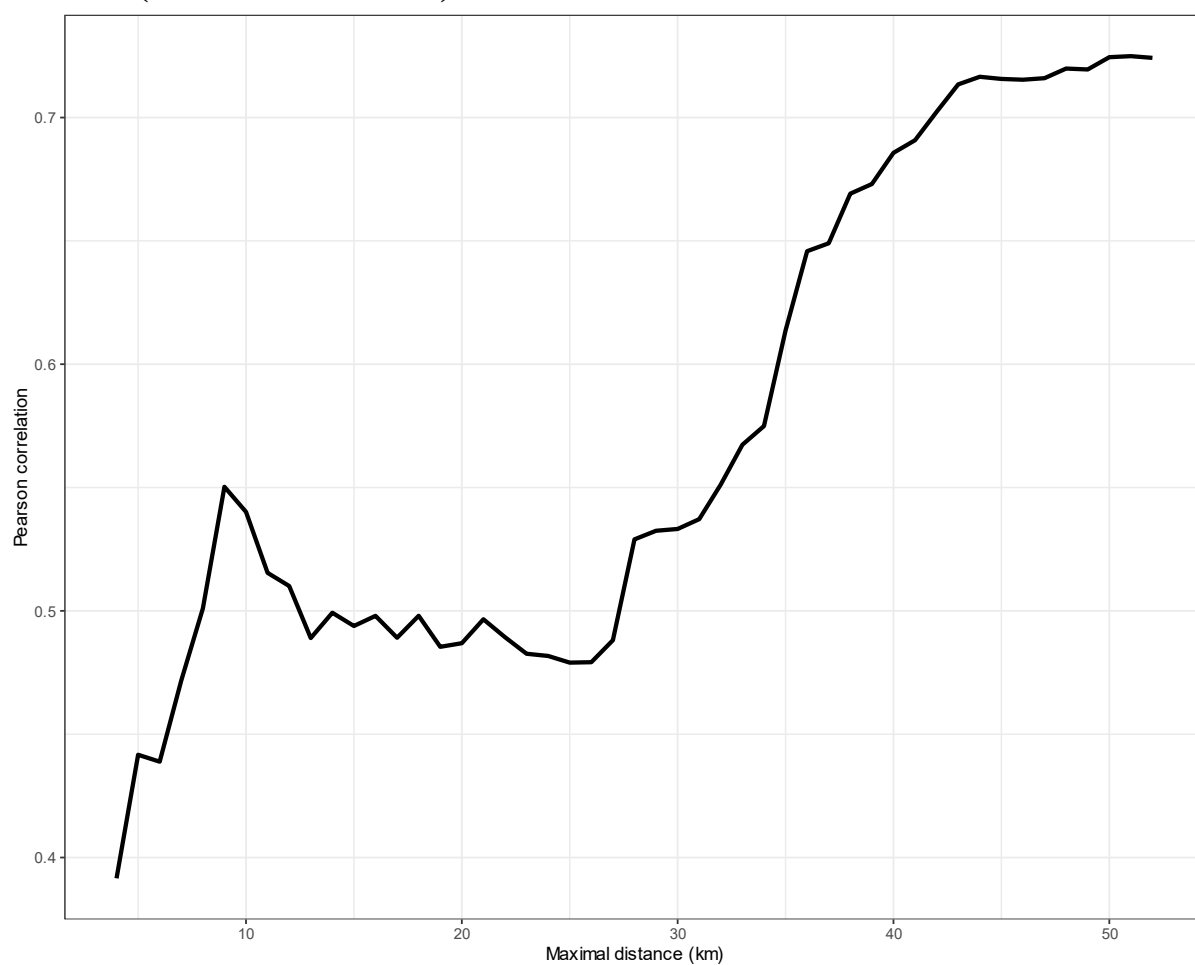


Figure S7: Structure plots for $K=2$ to $K=20$, with individuals sorted by latitude (from north to the left to south to the right).

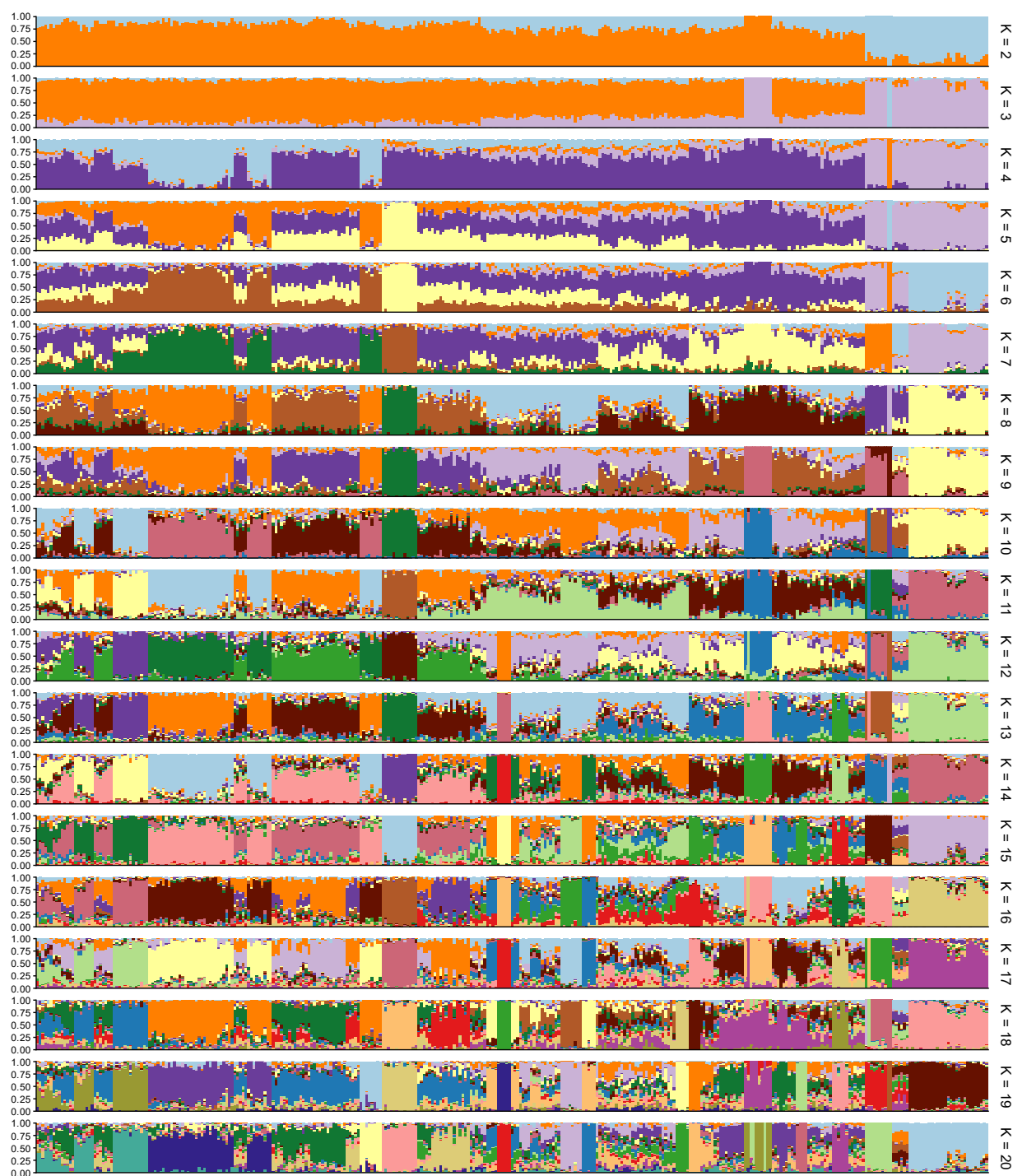


Figure S8: Heatmap of F_{ST} values among genetic clusters identified by LEA.

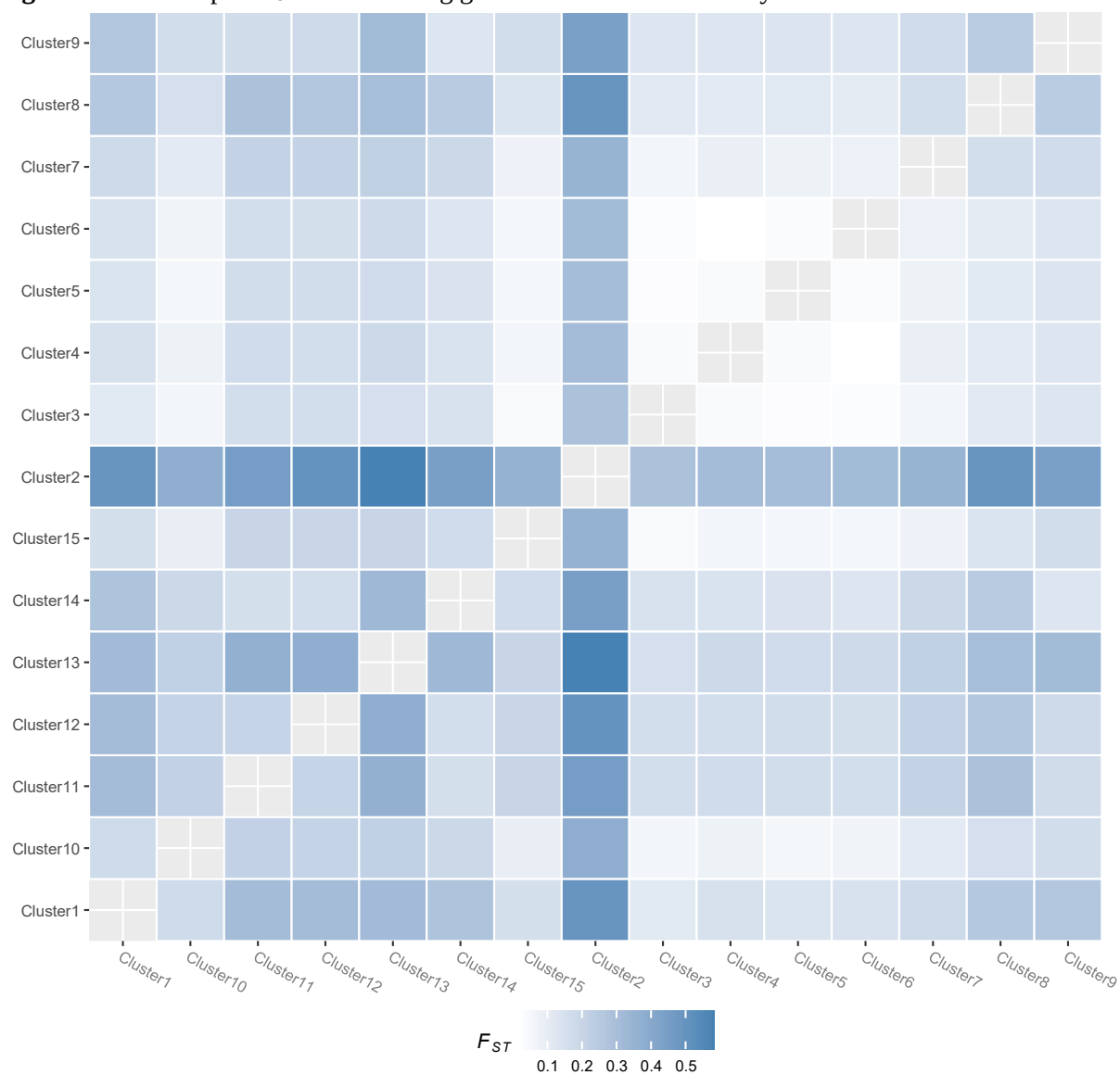


Figure S9: Mean F_{IS} across sites (A) and among clusters identified by LEA (B). Colors as in Figures 2 and 3; coordinates in Lambert 93.

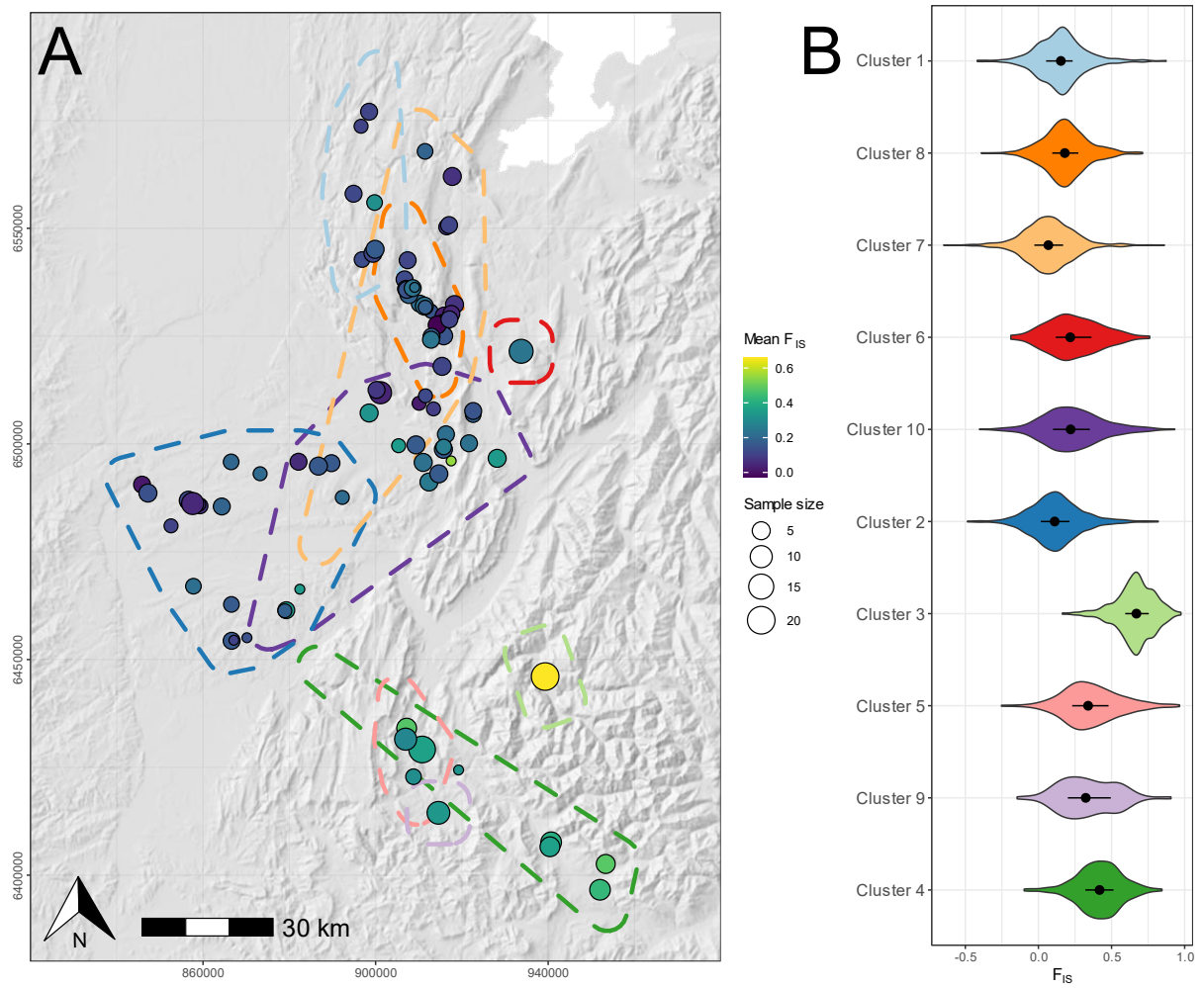


Figure S10: Correlation between Bray-Curtis dissimilarities used to measure individual genetic differentiation and commonly used genetic distance metrics in landscape genetics (proportion of alleles shared, Dps; Euclidean distance).

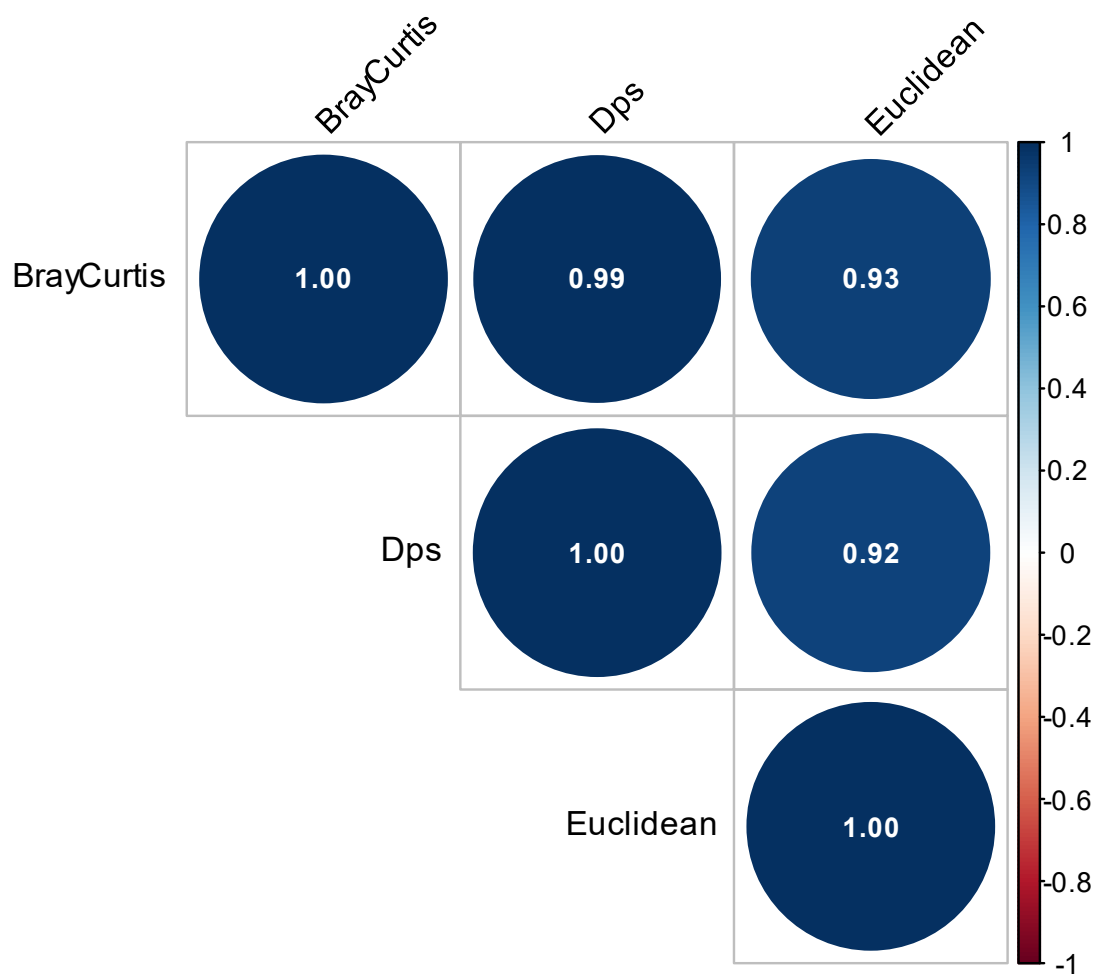


Table S1: Summary of the best supported model exploring the relationship between both RHS (BRHS and LRHS) and genetic diversity.

Covariance parameters	<i>estimate</i> $\pm se$	z	pval				
nugget	0.0001 ± 0.0001	3.593	0.00016				
Exp long	0.9633 ± 0.0160	60.25	<.0001				
Exp lat	0.9045 ± 0.0467	19.37	<.0001				
σ^2	0.0003 ± 0.00003	12.74	<.0001				
Fixed effect	<i>estimate</i> $\pm se$	$F_{df,ddf}$	pval	Unique variance	Shared variance	Total variance	R^2_{marg}
BRHS 11km	0.117 ± 0.0231	$F_{1,10.5}$ = 5.06	0.0004	0.694	-0.057	0.637	0.694
LRHS 3km	0.1418 ± 0.0424	$F_{1,70.7}$ = 3.34	0.0012	0.057	-0.057	0	

Table S2: Summary of the best supported model based on BRHS and the two significant raw landscape variables (slope and distance to springs) to explain genetic diversity.

Covariance parameters	<i>estimate</i> $\pm se$	z	pval				
nugget	0.0005 ± 0.00015	3.43	0.0003				
Exp long	0.9475 ± 0.0256	37.03	<.0001				
Exp lat	0.9111 ± 0.0525	17.35	<.0001				
σ^2	0.0003 ± 0.00003	12.66	<.0001				
Fixed effect	<i>estimate</i> $\pm se$	$F_{df,ddf}$	pval	Unique variance	Shared variance	Total variance	R^2_{marg}
BRHS 11km	0.1371 ± 0.0205	$F_{1,9.77}$ $= 5.83$	0.0002	1.059	-0.437	0.622	0.701
Slope 3km	-0.00217 ± 0.0008	$F_{1,61.2}$ $= -2.80$	0.0069	0.0374	0.075	0.112	
Distance to springs 3km	-0.00004 ± 0.00002	$F_{1,48.4}$ $= -2.48$	0.0168	0.0381	-0.258	-0.220 [†]	

[†] Note that the total variance explained by an explanatory term (here, the distance to springs) can be negative in the case of commonality analyses applied to mixed linear models, which results from a near-zero unique variance and a strong negative shared variance, revealing a strong suppressor effect.

Table S3: Commonality analysis for the model with BRHS and the two significant raw landscape variables (slope and distance to springs averaged in a 3 km radius buffer) to explain genetic diversity. Total variance explained for single variables is obtained by summing commonalities weighted by the order of the relationship as in Edwards et al. (2008) and Lai et al. (2022) (e.g., for BRHS: $1.05915 + (1/2) \times (-0.17567) + (1/2) \times (-0.50788) + (1/3) \times 0.24666 = 0.800$).

BRHS	Slope	Dist. springs	Partial R ²	Commonality	Sum of commonalities	Total variance explained
x			1.05915 [†]	1.05915	1.05915	0.800
	x		0.03739	0.03739	0.03739	0.034
		x	0.03810	0.03810	0.03810	-0.132 [‡]
x	x		0.92087	-0.17567	1.09654	0
x		x	0.58936	-0.50788	1.09724	0
	x	x	0.07910	0.00361	0.07549	0
x	x	x	0.70136	0.24666	0.45469	0

[†] For the same reason that the total variance associated to an explanatory term can be negative in commonalities analyses when applied to LMM, the (pseudo) partial R² associated to an explanatory term can exceed 1 (here, this is the case of BRHS), also revealing a strong collinearity with at least one other explanatory term.

[‡] See footnote of Table S2.

Table S4: Comparison between individual-level models linking relative habitat suitability with observed heterozygosity used in this study with the best sampling-site-level alternative.

	Individual-level model	Sampling-site-level model
Scale BRHS	11 km	10.5 km
Scale LRHS	3 km	3 km
BRHS estimate $\pm$ se	0.117 $\pm$ 0.0231	0.1222 $\pm$ 0.0244
LRHS estimate $\pm$ se	0.1418 $\pm$ 0.0424	0.1364 $\pm$ 0.0446
BRHS unique variance	0.694	0.757
LRHS unique variance	0.057	0.047
Shared variance	-0.057	-0.131
BRHS total variance	0.637	0.626
LRHS total variance	0.000	-0.084
Marginal R ²	0.694	0.674

Supplementary Data S1:

Table containing the summary of the 3345 models fit to explore the scale effect on the relationship between relative habitat suitability derived from species distribution models and observed heterozygosity, and model selection procedure (based on wAIC), at the individual level (OUT_HOBSINDIV) and at the site level (OUT_HOBSIDSITE).

Code	Meaning
clim	Bioclimatic relative habitat suitability (BRHS)
land	Landscape relative habitat suitability (LRHS)
INTER	BRHS*LRHS interaction

Supplementary Data S2:

Table containing the summary of the gravity models (grav) fit at the sampling site level (POP) or at the individual level (IND) to explain Bray-Curtis genetic differentiation with conductance distances computed with Circuitscape from bioclimatic and landscape relative habitat suitability (BRHS and LRHS, respectively) derived from species distribution models.

Code	Variable
v0	log(Wang's topographic distance)
v1	log(BRHS distance)
v2	log(LRHS distance)
v3	log(BRHS distance)*log(LRHS distance)
v4	(log(observed heterozygosity of sampling site or individual 1)+log(observed heterozygosity of sampling site or individual 2))/2

Supplementary Data S3:

Table containing the summary of the 242 gravity models fit to explain inter-individual Bray-Curtis genetic differentiation with Circuitscape conductance distances from bioclimatic relative habitat suitability (BRHS) and Circuitscape distances from raw landscape variables – used to build the SDMs – to replace landscape relative habitat suitability (LRHS).

Code	Variable
v0	$(\log(\text{observed heterozygosity of sampling site or individual 1}) + \log(\text{observed heterozygosity of sampling site or individual 2})) / 2$
v1	$\log(\text{BRHS distance})$
v2	$\log(\text{slope distance})$
v3	$\log(\text{distance to rivers distance})$
v4	$\log(\text{distance to springs distance})$
v5	$\log(\text{forest distance})$
v6	$\log(\text{crop distance})$
i2	$\log(\text{slope distance}) * \log(\text{BRHS distance})$
i3	$\log(\text{distance to rivers distance}) * \log(\text{BRHS distance})$
i4	$\log(\text{distance to springs}) * \log(\text{BRHS distance})$
i5	$\log(\text{forest distance}) * \log(\text{BRHS distance})$
i6	$\log(\text{crop distance}) * \log(\text{BRHS distance})$

Supplementary Data S4:

Table containing the barcode and plate information for all individuals included in the sampling.

Name in the table	Meaning
PlatePosition	Position of the sample on the plate
Barcode	Individual barcode used
Site	Sampling site code
Ind	Individual