Supplementary information to Beninde, J. *et al.* "Admixture of hybrid swarms of native and introduced lizards in cities is determined by the cityscape structure and invasion history"

SECTION 1: Sampling design and field work

To adequately cover the distribution of the common wall lizard, *Podarcis muralis*, within the four cities we produced city-wide distribution maps. These maps were based on field surveys conducted prior to sampling by ourselves, except for Mannheim, where unpublished data from a citywide census was made available to us (performed by the German conservation NGO NABU). These surveys captured the distribution of lizards comprehensively in all four cities. We then estimated patch specific lizard densities in a standardized procedure following the protocol of the German Federal Agency for Nature Conservation. These were issued for assessment of conservation status of this species [1], which is protected under the Habitats Directive of the European Union. Where areas were not accessible, potential distribution and abundance was estimated based on habitat availability and the presence and frequency of lizards in the vicinity. We used these fine-scale distribution maps to produce 200 random sampling points across each city within the species' respective local distribution, weighted by the relative abundance of lizards within patches (stratified random sampling).

SECTION 2. Molecular analyses

As recommended in the supplementary protocol of the Qiagen DNEasy blood and tissue kit for isolation of DNA from buccal swabs, ATL buffer was replaced by 400 µl PBS buffer. For cytochrome *b* sequences, we used 50 µl PCR reactions, containing 0.0625 µmol/µL of the primers LGlulk (5'-AACCGCCTGTTGTCTTCAACTA-3') and Hpod (3'-GGTGGAATGGGATTTTGTCTG-5'), added 2-10 ng of genomic DNA, 20 µl 5 Prime Mastermix and 25 µl purified water [2]. PCR settings were 15 min at 95°C, 35 cycles of 30s at 94°C, 30s at 43°C, 90s at 72°C and 10 min at 72°C.

GenBank accession numbers used for alignement and fitting of tree are as follows: Western France lineage AY234155 [3]; Calabria lineage DQ001023, Tuscany lineage DQ001028, Eastern France (native lineage) DQ001029, Venetia DQ001032 [4]; Central Balkans lineage HQ652887, Romagna lineage HQ652921, Southern Alps lineage HQ652963 [2]. Lineage names follow 5 [5], which can extend beyond the geographic extant the name implies, e.g. the Western France lineage also occurs in Spain. A phylogenetic tree was fitted using *P. siculus* and *P. melisellensis* as outgroups (HQ154646, AY185097, [6]).

For genotypic analyses, we used the following microsatellite loci: twelve had been developed for *Podarcis muralis* (B3, B4, C9: 7; PmurC150, PmurC168, PmurC275_278, PmurC164, PmurC038, PmurC028, PmurC356, PmurC109, PmurC103: [8]), two for *Zootoca vivipara* (Lv-319 and Lv-472: [9]) and three for *Podarcis bocagei* (Pb10, Pb50, Pb73: [10]). Primers were labelled with FAM, TAMRA or HEX. Multiplex PCR protocols for two or three loci were used with the following annealing temperatures: 57°C for C9, B4, Pb73 and all PmurC-primers; 56°C for B3, Pb10 and Lv319; 53°C for Lv472 and Pb50. Using the HotMasterMix by 5PRIME or Multiplex Mastermix by Qiagen and Multigene Gradient Thermal Cyclers (Labnet) amplifications were conducted as recommended by manufacturers. Multiplex PCRs were performed in 10 µl reaction mix containing: 2–10 ng genomic DNA, 5.0 µl MasterMix, 2.0 µl water and 0.1 µM of each primer. Fragment lengths of PCR products was determined on a MEGABACE 1000 sequencer using the software Fragment Profiler 1.2 (Amersham Biosciences) or on Applied Biosystems 3500 and 3750 Genetic Analysers using the softwares GeneMapper 5 (Applied Biosystems) or PeakScanner 2.0 (Applied Biosystems). We ran ca. 28 samples per multiplex on all sequencers to quantify differences in fragment lengths and calibrate the final dataset.

SECTION 3: Landscape models of environmental factors

Environmental factors utilized in this study:

<u>Aspect</u>

Southerly exposed slopes warm up most intensely and thus represent suitable habitat for reptiles in general and for *Podarcis muralis* in Germany, at its Northern range margin, specifically [11].

> *northern, north-western, north-eastern*
> *western, eastern*
> *flat*
> *south-western, south-eastern*
> *southern*

<u>Slope</u>

Although included in a previous publication covering the city of Trier, we could not acquire this data for the other cities. In Trier, slope did not contribute to the best model of gene flow. We had included this factor initially, as steeper slopes capture more solar irradiance, representing more suitable habitat.

Substrate

Substrate types differ in their physical properties, thermal radiation as well as absorption and storage capacities of moisture. As summarized by Schulte [11], 55% of populations occur on rocky or gravel substrates, 30% of occurrences are located on open ground substrate. The remaining occurrences were detected on sealed surfaces. Accordingly, we created three category-levels.

sealed surfaces

open ground

rocky/gravel

none of these

rock/gravel: mostly railroads, but also dry stone walls in vineyards as well as rock formations were included here.

open ground: this includes unsealed parking lots, sandy areas and sports fields. All of these were areas were not vegetated.

sealed surfaces: includes roads, houses and sealed parking lots

Canopy cover

Here, all trees were digitized due to negative effects on *Podarcis muralis* populations by shadowing [11].

trees

no trees

Vegetation type

The previous category 'vegetation height' was also evaluated in its naturalness, as more natural, unmanaged vegetation was shown to harbor a higher density of arthropods. We distinguished semi-natural, cultivated and planted vegetation.

planted vegetation

cultivated vegetation

semi-natural vegetation

no vegetation

Structural diversity

In this layer, the structural variety of different surfaces was evaluated to capture potential hiding places, such as crevices, fugues joints or hollow spaces. These are used by *Podarcis muralis* as retreats, overwintering sites and also as nesting sites [11]. We distinguished low, intermediate and high levels of structural diversity.

> *low structural diversity*
> *medium structural diversity*
> *high structural diversity*
> *no structural diversity*

low structural diversity: highly managed, uniform areas, as well as sandy areas
medium structural diversity: natural meadows, gardens in residential areas, walls and cemeteries.
high structural diversity: areas bordering railway tracks, scrapyards, allotment gardens, ruderal areas, and non-utilized areas.

Buildings

All houses and buildings were digitized individually or as larger polygons, if spatially arranged as near-contiguous blocks or rows. Additionally, a buffer of ten meters was then added to capture negative effects of the shaded area of buildings and houses.

> *buildings*
> *no buildings*

South-facing walls

To create this layer the unbuffered building layer was copied and superimposed on the other. Then, the upper layer was shifted by two meters to the north and cut out using the "Erase" function in ArcGIS. This leaves only south-facing walls. Additionally walls in vineyards were included.

> *south-facing walls*
> *no south-facing walls*

Roads

All tarred roads were digitized here, pedestrian and bicycle paths were not included. Subsequently, this line feature class was buffered by a meter to capture the negative impacts by traffic.

> *roads*
> *no roads*

Traffic

In addition to the roads layer, all major roads were scaled according to traffic volume (only those roads with an average daily traffic, ADT, greater than 2900 vehicles 24 h). Data was received from the traffic bureaus of local authorities. All roads were buffered by four meters to capture the width of roads. Roads were scaled proportionate to the maximum ADT of each city.

<u>Water</u>

All areas covered by water were digitized, representing unsuitable habitat.

water surfaces

no water surfaces

Grid development

The area of grids expanded beyond the extent digitized for environmental factors to prevent false inferences of landscape resistances: if individuals are close to a grids' boundary, movement of such individuals can be artificially constrained in simulations [12]. For landscape genetic analyses, we therefore expanded the extent of the grids where necessary, by buffering sample locations according to dispersal distances of this species. We chose a 1km buffer distance, equaling the maximum distance a wall lizard has been recorded to disperse and strongly exceeding average dispersal distances of <200m [13]. Following Koen *et al.* [12], we assigned random values to buffered areas extending beyond those areas with known data. Random data was created proportional to known data values of environmental factors in areas with known data (figure 2). This procedure has demonstrated to prevent overestimation of resistances, when compared with true environmental data, and thus alleviates potential effects of artificial boundaries [12].

Section 4: Distribution modelling

Species Distribution Modelling using Maxent

We used a presence-only (PO) method (Maxent) to build SDMs for *Podarcis muralis*. Maxent is a machine learning method following a maximum entropy approach [14] implemented in the Maxent software v3.3.3k [15,16]. Maxent is widely applied for PO data in species distribution modeling and also used to explore and interpret the environmental drivers shaping a species' distribution [17]. Maxent uses presence locations, background points and a set of predictor variables to estimate the probability of presence (logistic output) for each grid cell of the landscape. Circumventing the criticism of interpreting the logistic output in this way [18,19], it is commonly viewed as a habitat suitability [20]. The background points are taken from the landscape and are used to contrast the conditions at presence sites, therefore it is also sometimes called a presence-background method. We refrained from also testing generalized linear models for PA data as we showed previously that PO data delivers better results in the context of landscape genetic optimization in common wall lizards [21].

To avoid data collinearity and model overfitting [22,23] we applied the following procedure to reduce the number of predictors and determine the optimal model complexity: In a first step we checked the pairwise correlations between all twelve environmental factors using SDMtoolbox [24] and removed factors with a Pearson correlation coefficient larger than 0.7 [23]. We used all presence points of the sampled individuals and the remaining environmental predictors to run Maxent with the following changes to the default settings. We enabled the creation of response curves and jackknifing to measure variable performance. Only linear and quadratic features were selected to avoid the creation of overly complex response curves for the two continuous variables. Because the field survey was conducted throughout the entire city, i.e. without spatial sampling bias, we selected the entire city as the modelling background (using 10000 background points). We did not let Maxent remove duplicate records, i.e. presence points in the same grid cell, as in our study, duplicates represent higher local abundances and are not a result of varying spatial sampling intensity. In cases of variable spatial sampling intensity the removal would be advisable (see e.g. Kramer-Schadt *et al.* [25]).

We also changed the default prevalence τ. Elith *et al.* [20] state that a τ value of 0.5 corresponds to a temporal and spatial scale of sampling that results in a 50% chance of the species being present in suitable areas and leads to values of the logistic output near 0.5 at "typical" presence sites. In the case of the common wall lizard and when sampling during sunny weather conditions, the probability of encountering individuals is larger than 50% at suitable sites. Although we did not formally test the probability of encountering individuals at such sites, we set τ to the value of 0.75, a value that seems more reasonable to us than the arbitrary default value of τ=0.5. The arbitrariness of the default value of τ is one of the major points of criticism concerning the interpretation of the logistic output as a probability of presence [17,19]. For further justification see supplementary information for Beninde *et al.* [21].

In a stepwise procedure we eliminated the predictor contributing least to the model, using Maxent's own analysis of variable contribution and re-ran Maxent with the reduced predictor set. AICc values were calculated for all models using NicheAnalyst (v3.0; [26]) (Maxent's raw output is necessary for this) to determine the best model based on the minimum ΔAICc values [22]. When there was more than one equally good model we chose the one with the smallest number of environmental factors as the final predictor set. Although, to our knowledge, the use of almost only categorical predictors is uncommon in this application, Elith & Graham [27] state that categories are modeled well with Maxent.

In a last step we ran the final PO-model with 10-fold cross-validation (CV), so AUC (area under the curve of receiver-operator characteristic) values could be calculated on independent test data as a measure of model fit. Although AUC scores for PO data as a measure of performance can be misleading [28], Merow *et al.* [17] note that AUC is appropriate for high sampling intensities, which

is the case in our study. As an additional performance measure we show the omission of test localities (or extrinsic omission error [29], table S7) with respect to the maximum sum of test sensitivity plus specificity (maxSSS) threshold, which is proposed for PO data [30,31]. It is calculated by Maxent and describes the proportion of test localities which fall in areas predicted as unsuitable after thresholding the continuous model output into a binary presence absence map.

References

[1] PAN & ILÖK. 2010 *Bewertung des Erhaltungszustandes der Arten nach Anhang II und IV der Fauna-Flora-Habitat-Richtlinie in Deutschland. Überarbeitete Bewertungsbögen der Bund-Länder-Arbeitkreise als Grundlage für ein bundesweites FFH-Monitoring.* Umpublished work comissioned by the German Federal Agency for Nature Conservation (BfN).

[2] Schulte, U., Veith, M. & Hochkirch, A. 2012 Rapid genetic assimilation of native wall lizard populations (*Podarcis muralis*) through extensive hybridization with introduced lineages. *Mol. Ecol.* **21**, 4313–4326.

[3] Busack, S., Lawson, R. & Arjo, W. 2005 Mitochondrial DNA, allozymes, morphology and historical biogeography in the *Podarcis vaucheri* (Lacertidae) species complex. *Amphib-reptil.* **26**, 239–256.

[4] Podnar, M., Haring, E., Pinsker, W. & Mayer, W. 2007 Unusual origin of a nuclear pseudogene in the Italian wall lizard: intergenomic and interspecific transfer of a large section of the mitochondrial genome in the genus *Podarcis* (Lacertidae). *J. Mol. Evol.* **64**, 308–320.

[5] Salvi, D., Harris, D. J., Kaliontzopoulou, A., Carretero, M. A. & Pinho, C. 2013 Persistence across Pleistocene ice ages in Mediterranean and extra-Mediterranean refugia: phylogeographic insights from the common wall lizard. *BMC Evol. Biol.* **13**, 147.

[6] Podnar, M. 2004 Mitochondrial phylogeography of the Dalmatian wall lizard, (Lacertidae). *Org. Divers. Evol.* **4**, 307–317.

[7] Nembrini, M. & Oppliger, A. 2003 Characterization of microsatellite loci in the wall lizard *Podarcis muralis* (Sauria: Lacertidae). *Mol. Ecol. Notes* **3**, 123–124.

[8] Heathcote, R. J. P., Dawson, D. A. & Uller, T. 2014 Characterisation of nine European wall lizard (*Podarcis muralis*) microsatellite loci of utility across sub-species. *Conserv. Genet. Resour.* **7**, 85–87.

[9] Boudjemadi, K., Martin, O., Simon, J.-C. & Estoup, A. 1999 Development and cross-species comparison of microsatellite markers in two lizard species, *Lacerta vivipara* and *Podarcis muralis*. *Mol. Ecol.* **8**, 518–520.

[10] Pinho, C., Sequeira, F., Godinho, R., Harris, D. J. & Ferrand, N. 2004 Isolation and characterization of nine microsatellite loci in *Podarcis bocagei* (Squamata: Lacertidae). *Mol. Ecol. Notes* **4**, 286–288.

[11] Schulte, U. 2008 *Die Mauereidechse. Erfolgreich im Schlepptau des Menschen.* Bielefeld: Laurenti-Verlag.

[12] Koen, E. L., Garroway, C. J., Wilson, P. J., Bowman, J. & Bersier, L.-F. 2010 The Effect of Map Boundary on Estimates of Landscape Resistance to Animal Movement. *PloS one* **5**, e11785.

[13] Schulte, U. 2008 *Die Mauereidechse. Erfolgreich im Schlepptau des Menschen.* Bielefeld: Laurenti-Verlag.

[14] Phillips, S. J., Dudík, M. & Schapire, R. E. 2004 A Maximum Entropy Approach to Species Distribution Modeling. In *Proceedings of the twenty-first international conference on Machine learning* (ed. C. Brodley), pp. 655–662. New York, NY: ACM.

[15] Phillips, S. J., Anderson, R. P. & Schapire, R. E. 2006 Maximum entropy modeling of species geographic distributions. *Ecol. Model.* **190**, 231–259.

[16] Phillips, S. J. & Dudík, M. 2008 Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**, 161–175.

[17] Merow, C., Smith, M. J. & Silander, J. A. 2013 A practical guide to MaxEnt for modeling species' distributions. What it does, and why inputs and settings matter. *Ecography* **36**, 1058–1069.

[18] Royle, J. A., Chandler, R. B., Yackulic, C. & Nichols, J. D. 2012 Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. *Methods Ecol Evol* **3**, 545–554.

[19] Yackulic, C. B., Chandler, R., Zipkin, E. F., Royle, J. A., Nichols, J. D., Campbell Grant, E. H., Veran, S. & O'Hara, R. B. 2013 Presence-only modelling using MAXENT. When can we trust the inferences? *Methods Ecol Evol* **4**, 236–243.

[20] Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E. & Yates, C. J. 2011 A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.* **17**, 43–57.

[21] Beninde, J., Feldmeier, S., Werner, M., Peroverde, D., Schulte, U., Hochkirch, A. & Veith, M. 2016 Cityscape genetics: structural vs. functional connectivity of an urban lizard population. *Mol. Ecol.* **25**, 4984–5000.

[22] Burnham, K. P. & Anderson, D. R. 2002 *Model Selection and Multimodel Inference: A Practical Information-theoretic Approach.* New York: Springer International Publishing.

[23] Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, Jaime R. García, Gruber, B., Lafourcade, B. & Leitão, P. J. *et al.* 2013 Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36**, 27–46.

[24] Brown, J. L. & Anderson, B. 2014 SDMtoolbox: a python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *Methods Ecol Evol* **5**, 694–700.

[25] Kramer-Schadt, S., Niedballa, J., Pilgrim, J. D., Schröder, B., Lindenborn, J., Reinfelder, V., Stillfried, M., Heckmann, I., Scharf, A. K. & Augeri, D. M. *et al.* 2013 The importance of correcting for sampling bias in MaxEnt species distribution models. *Diversity Distrib.* **19**, 1366–1379.

[26] Qiao, H., Peterson, A. T., Campbell, L. P., Soberón, J., Ji, L. & Escobar, L. E. 2015 NicheA. Creating Virtual Species and Ecological Niches in Multivariate Environmental Scenarios. *Ecography* **39**, 1–9.

[27] Elith, J. & Graham, C. H. 2009 Do they? How do they? Why do they differ? - on finding reasons for differing performances of species distribution models. *Ecography* **32**, 66–77.

[28] Lobo, J. M., Jiménez-Valverde, A. & Real, R. 2008 AUC. A misleading measure of the performance of predictive distribution models. *Global Ecol. Biogeogr.* **17**, 145–151.

[29] Anderson, R. P., Lew, D. & Peterson, A.T. 2003 Evaluating predictive models of species' distributions. Criteria for selecting optimal models. *Ecol. Model.* **162**, 211–232.

[30] Liu, C., White, M., Newell, G. & Pearson, R. 2013 Selecting thresholds for the prediction of species occurrence with presence-only data. *J. Biogeogr.* **40**, 778–789.

[31] Liu, C., Newell, G. & White, M. 2016 On the selection of thresholds for predicting species occurrence with presence-only data. *Ecol Evol* **6**, 337–348.
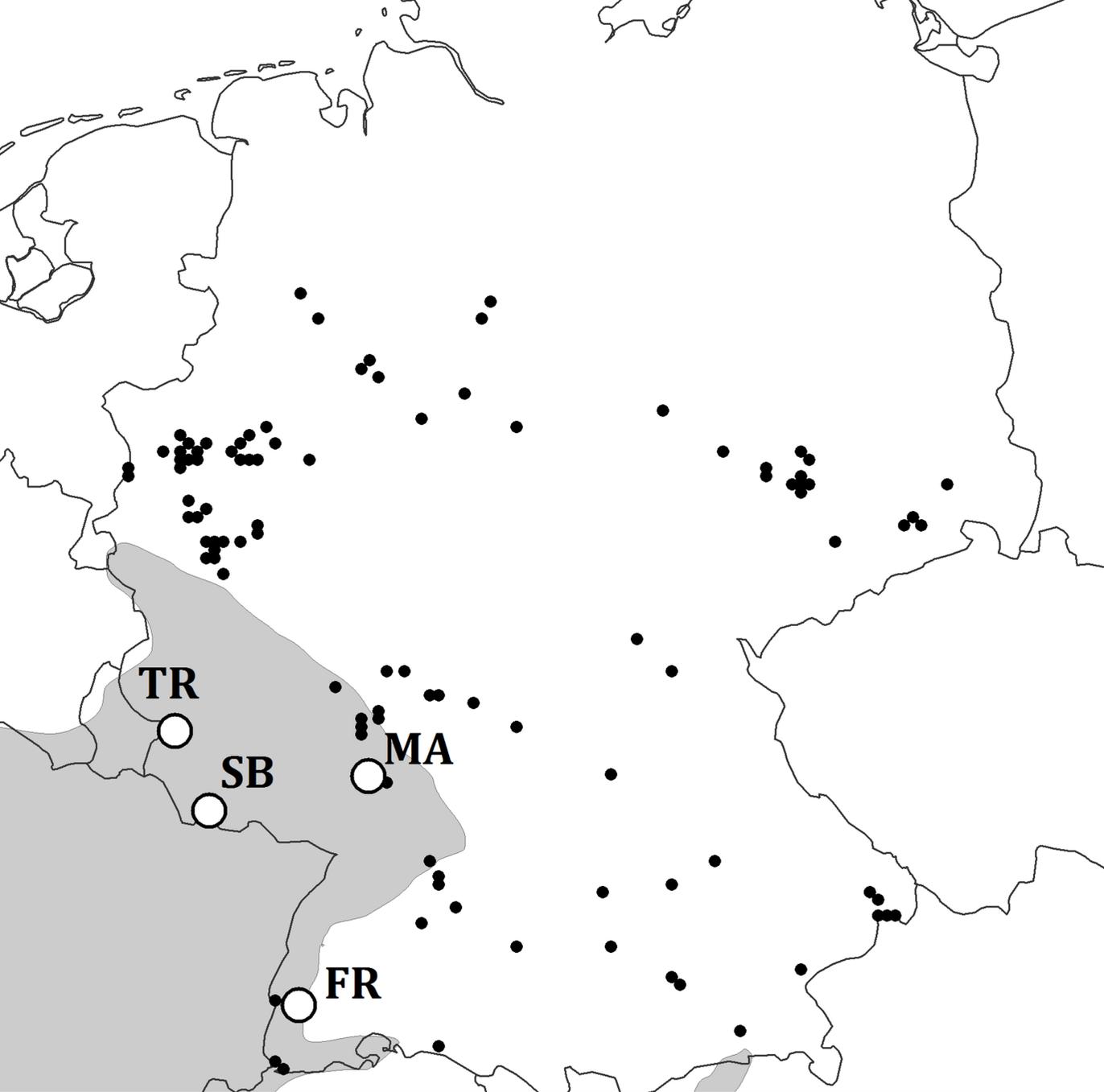
Figure S1: Introduced common wall lizard populations in Germany. The native range of the Eastern France evolutionary lineage is shaded in grey. White circles indicate cities sampled in this study (Trier, TR; Saarbrücken, SB; Mannheim, MA; Freiburg, FR). Black circles show additional locations known to harbour introduced common wall lizards.
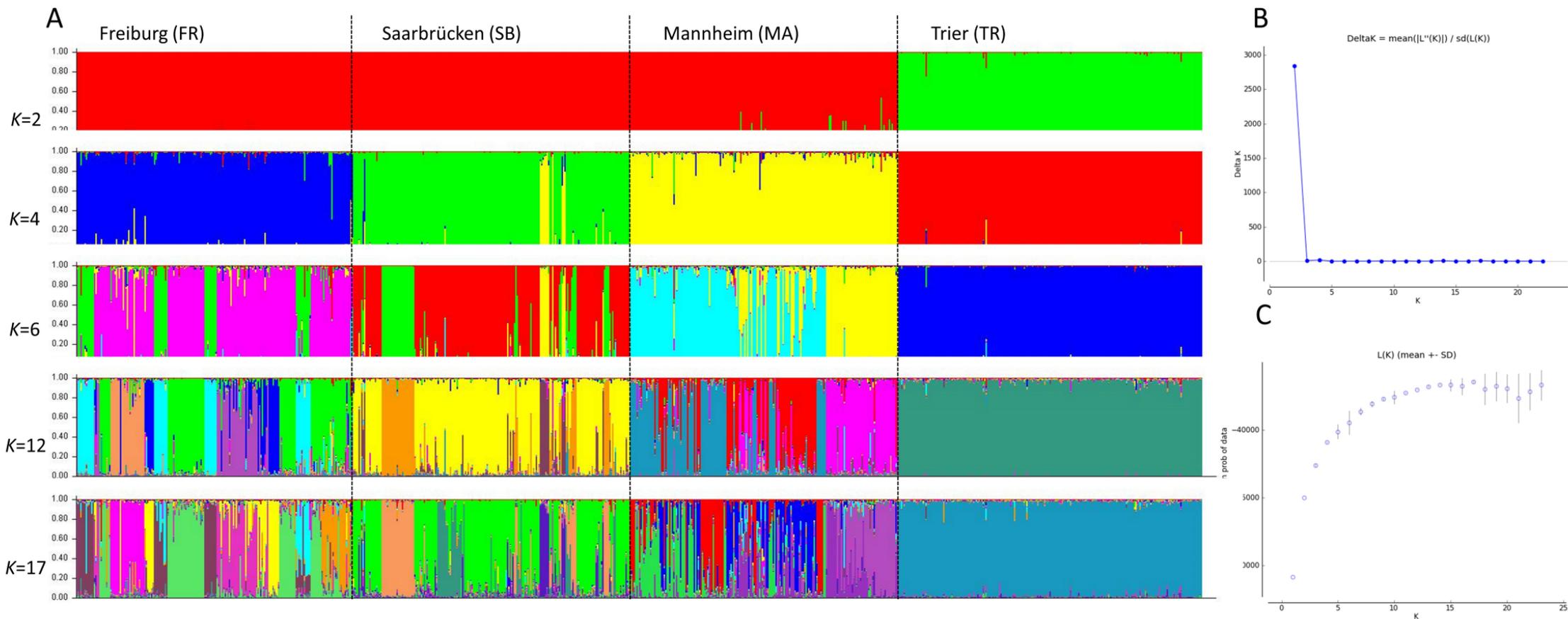
Figure S2: Bayesian clustering results for all 826 sampled individuals of the four cities (FR=203 individuals; SB=197; MA=203; TR=223). A: Barplots from STRUCTURE for *K*=2 (Δ*K*=max) until *K*=17 (L(*K*) max). Each of K clusters is shown in a different colour and each vertical bar represents an individuals probability of membership to *K* clusters. B & C: Results from STRUCTURE harvester, which depicts the likelihood of different values of *K*, calculated by different parameters of results of STRUCTURE runs. B: Plot of Δ*K* (as inferred following Evanno *et al.* 2005). C: Plot of L(*K*).
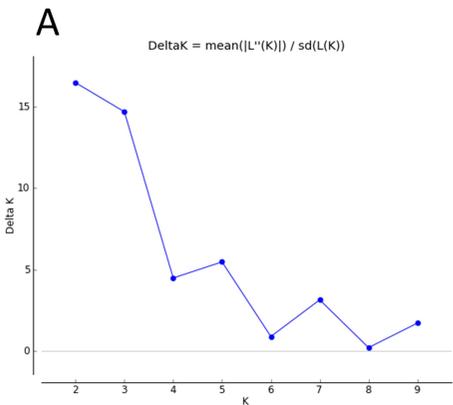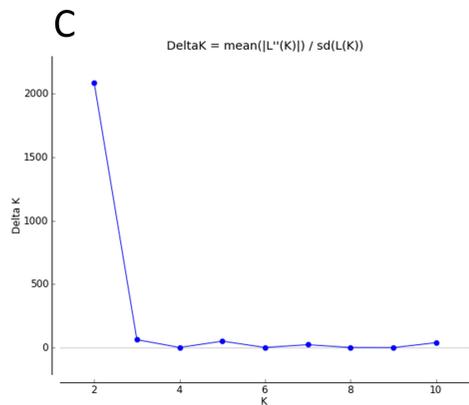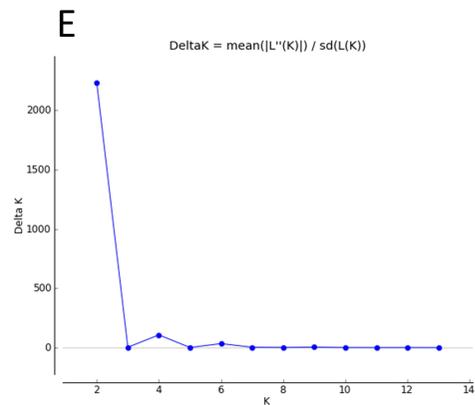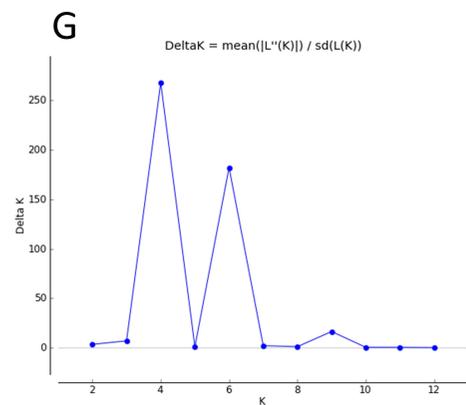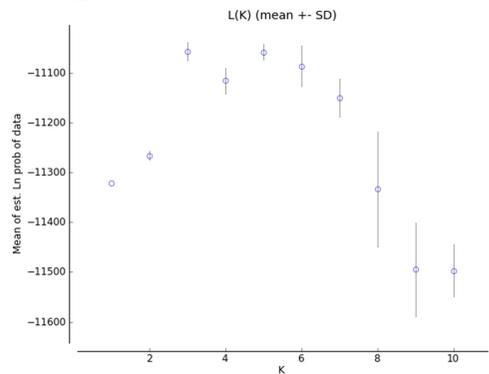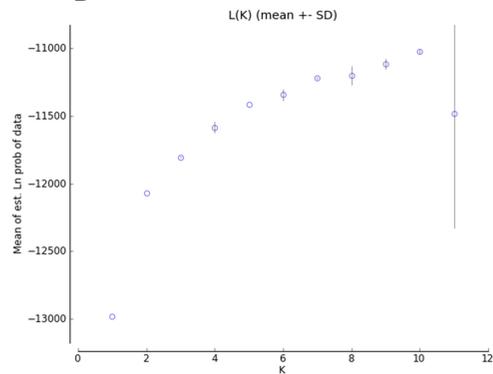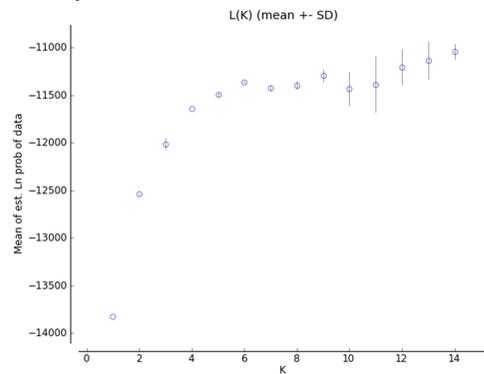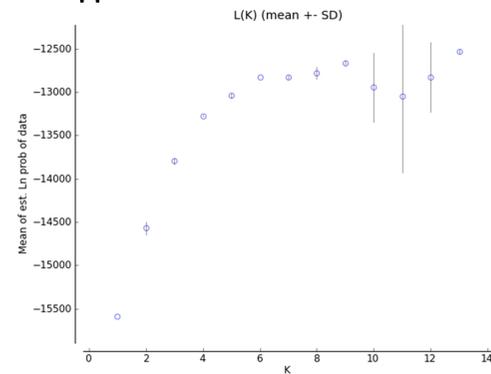
Figure S3: STRUCTURE harvester results, showing ΔK and L(K) for all cities of separate STRUCTURE analyses across 10 replicates per K. Results for Trier are shown in A+B; Saarbrücken in C+D; Mannheim in E+F; Freiburg in G+H.
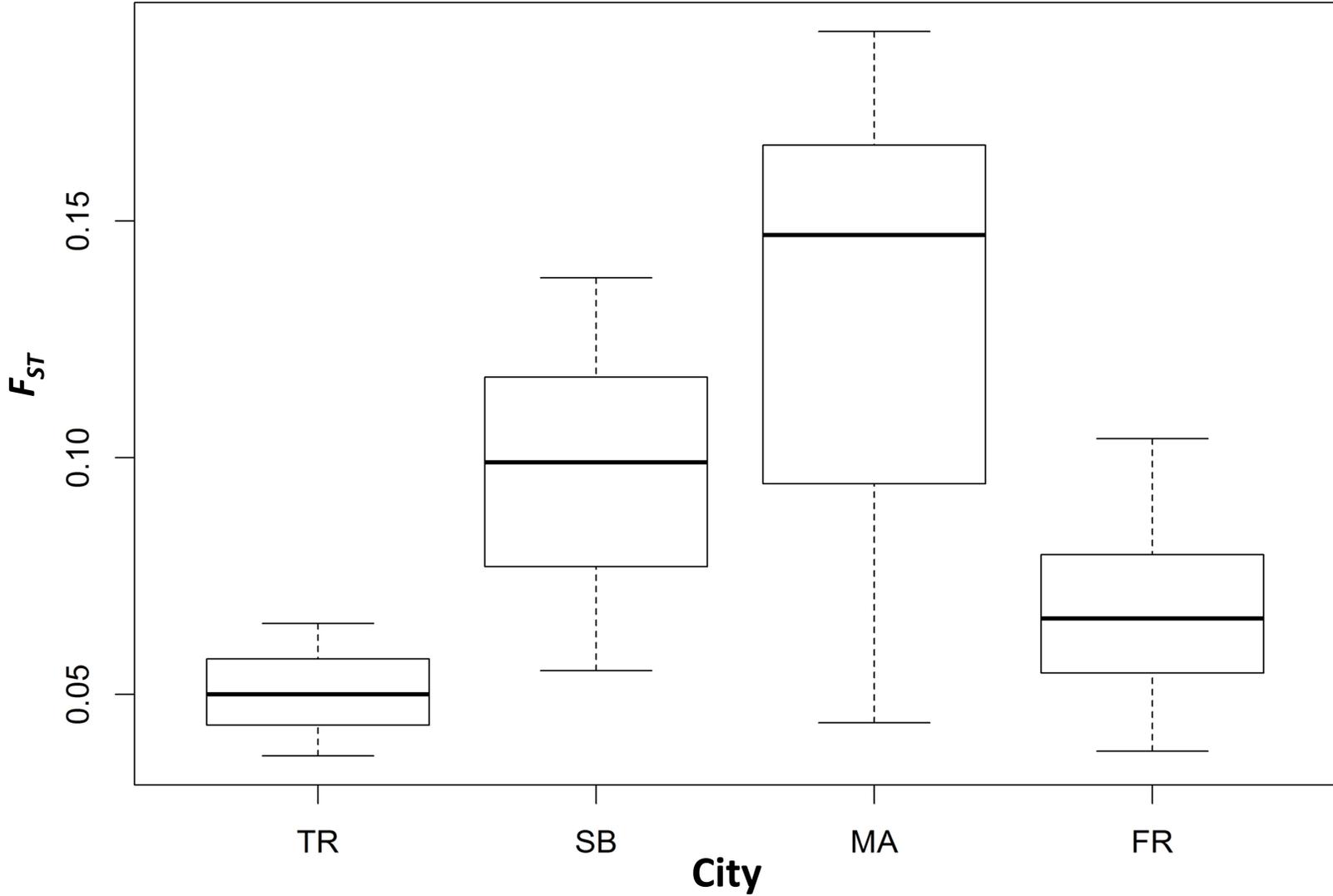
Figure S4: Boxplot of pairwise *F_{ST}*-values of all clusters within each city (calculated using GenAlEx). City labels refer to Trier (TR), Saarbrücken (SB), Mannheim (MA) and Freiburg (FR).
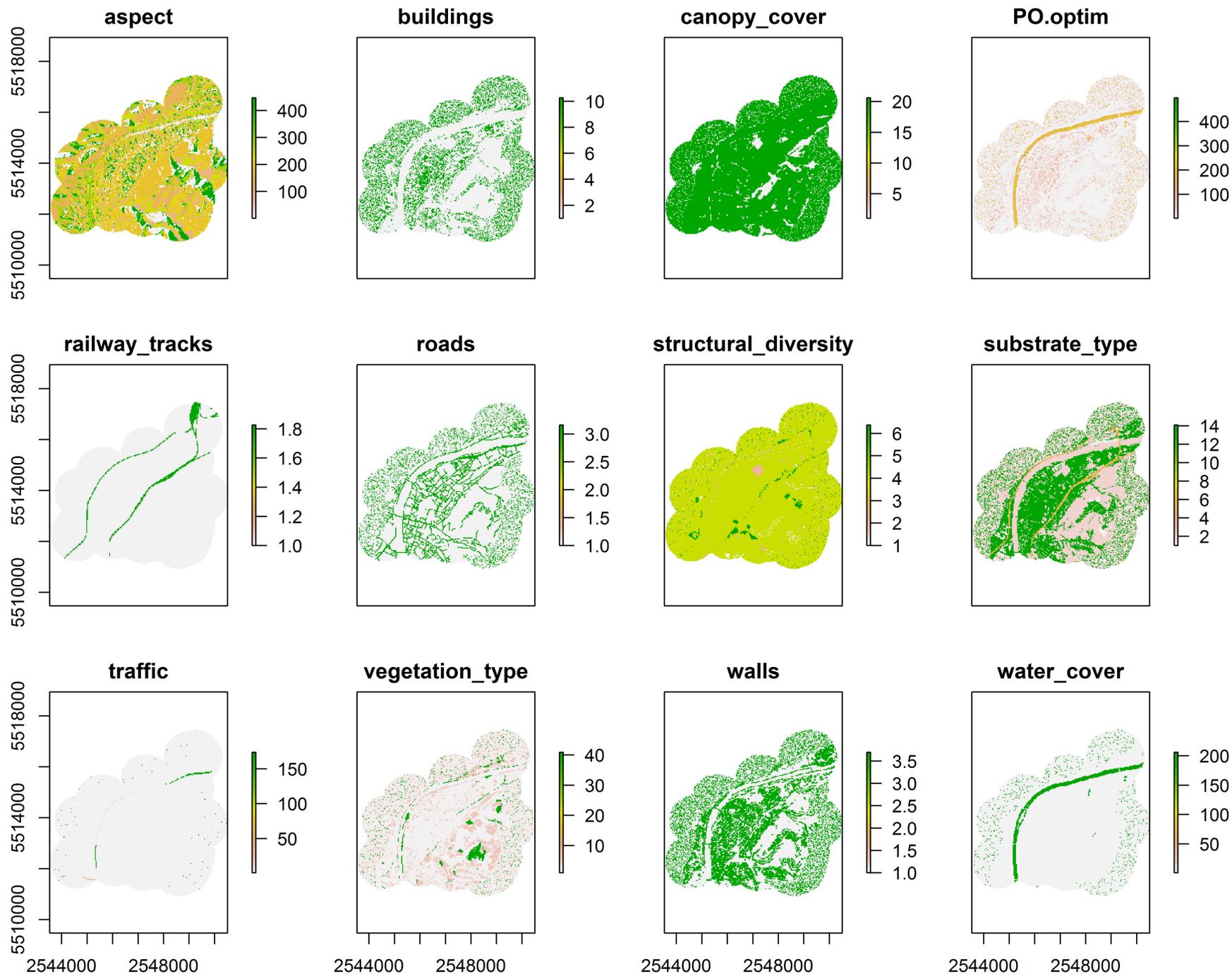
Figure S5: Optimized models of all environmental factors for the city of Trier. Landscape genetic optimization using ResistanceGA transforms original data values to optimally explain genetic distances of individuals. Various functions for transformation of data values are used to search a wide parameter space employing a genetic algorithm. Ranking of model performance is based on AICc ranks (see table S2). The environmental factors are described in detail in section 3 of the supplementary material. Colours indicate resistance, with higher values showing higher resistance to gene flow.
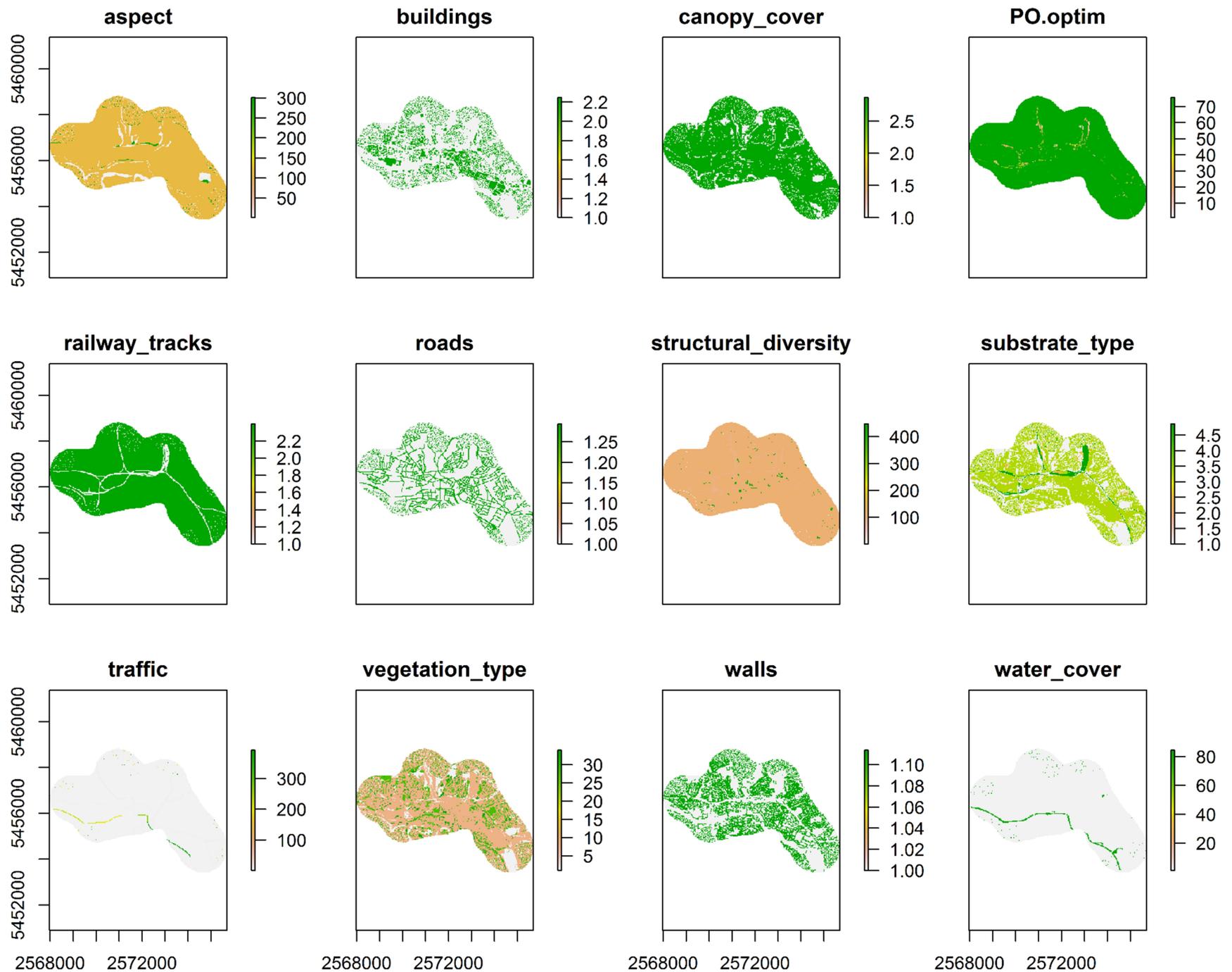
Figure S6: Optimized models of all environmental factors for the city of Saarbrücken. Landscape genetic optimization using ResistanceGA transforms original data values to optimally explain genetic distances of individuals. Various functions for transformation of data values are used to search a wide parameter space employing a genetic algorithm. Ranking of model performance is based on AICc ranks (see table S2). The environmental factors are described in detail in section 3 of the supplementary material. Colours indicate resistance, with higher values showing higher resistance to gene flow.
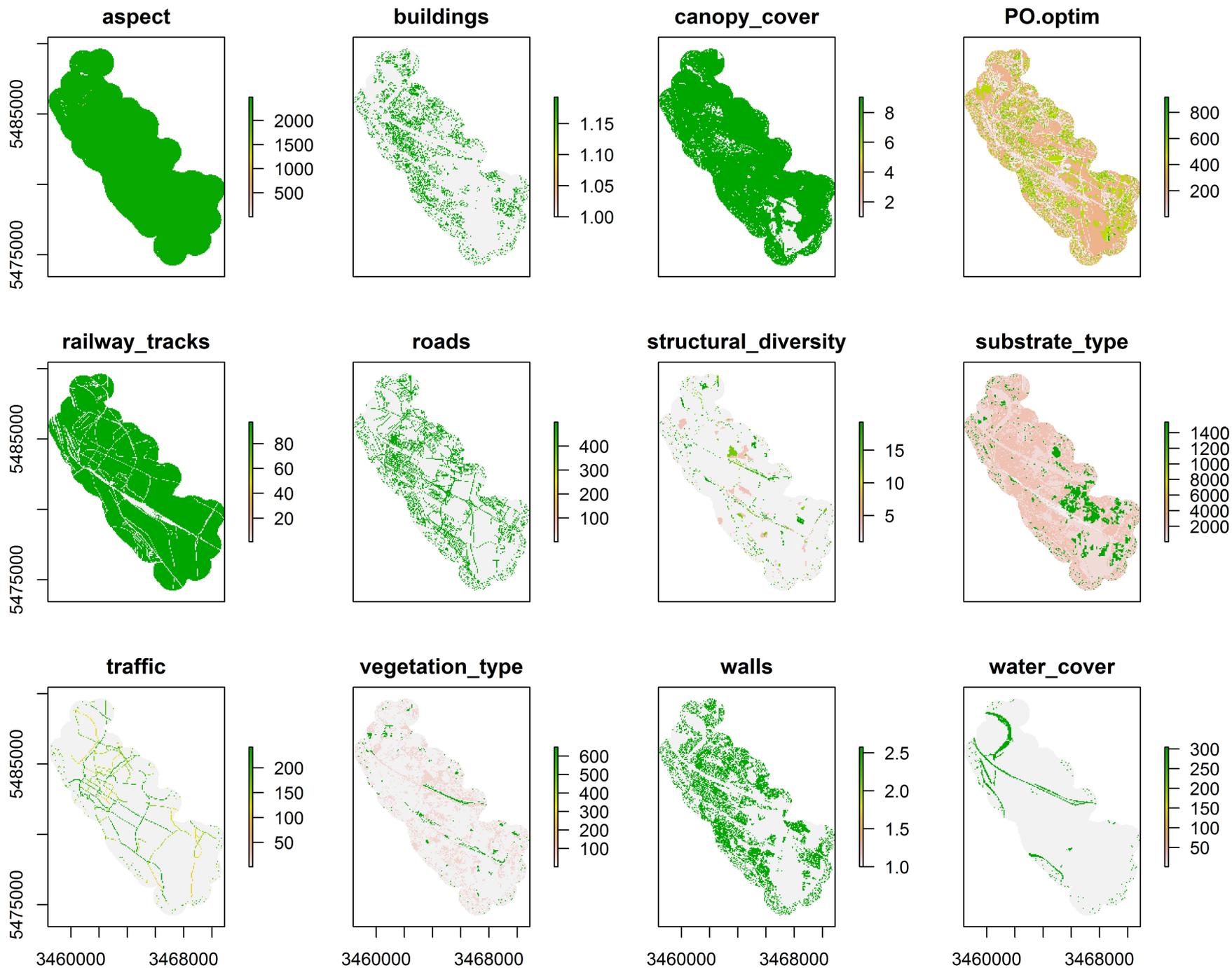
Figure S7: Optimized models of all environmental factors for the city of Mannheim. Landscape genetic optimization using ResistanceGA transforms original data values to optimally explain genetic distances of individuals. Various functions for transformation of data values are used to search a wide parameter space employing a genetic algorithm. Ranking of model performance is based on AICc ranks (see table S2). The environmental factors are described in detail in section 3 of the supplementary material. Colours indicate resistance, with higher values showing higher resistance to gene flow.
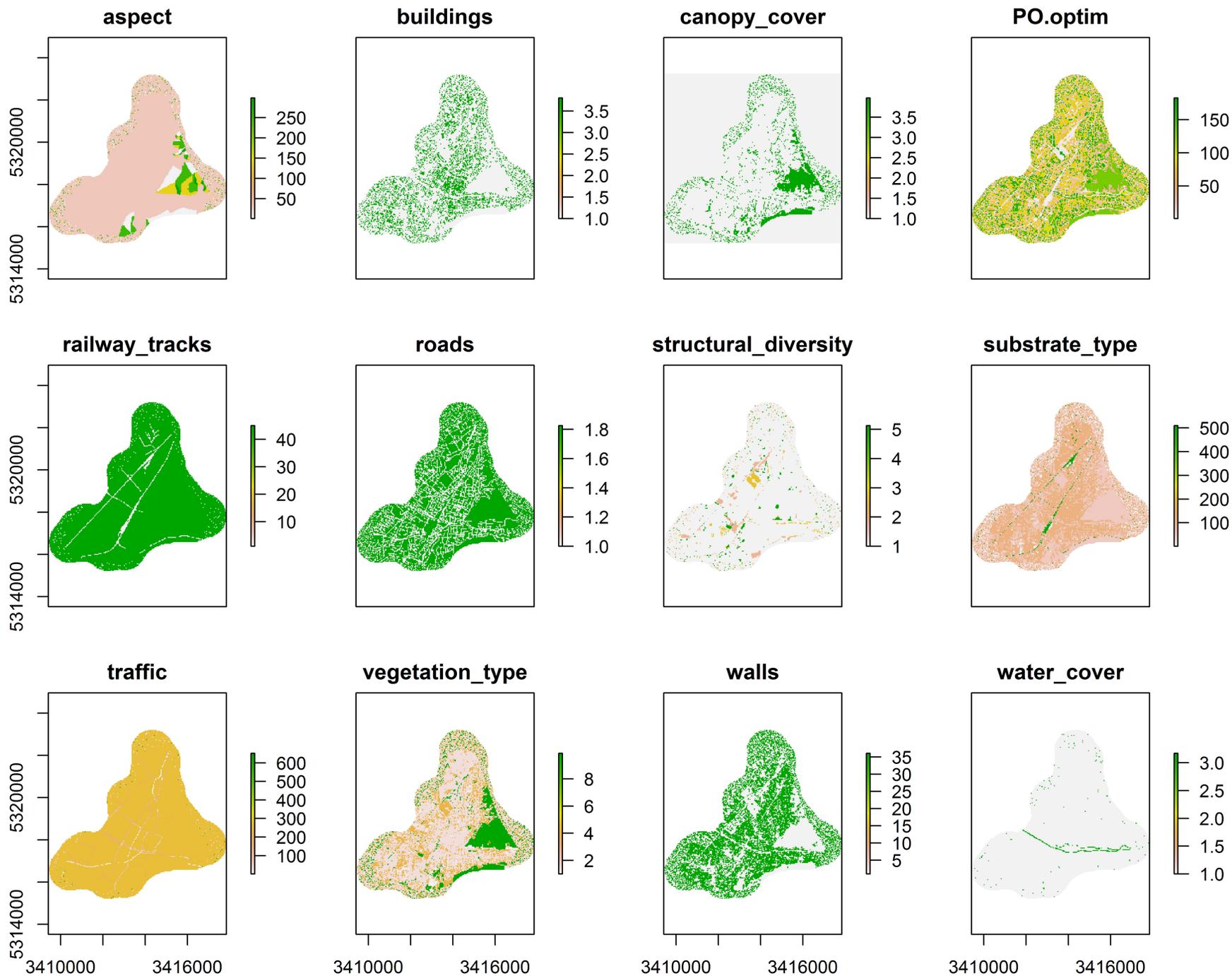
**Figure S8: Optimized models of all environmental factors for the city of Freiburg.** Landscape genetic optimization using ResistanceGA transforms original data values to optimally explain genetic distances of individuals. Various functions for transformation of data values are used to search a wide parameter space employing a genetic algorithm. Ranking of model performance is based on AICc ranks (see table S2). The environmental factors are described in detail in section 3 of the supplementary material. Colours indicate resistance, with higher values showing higher resistance to gene flow.
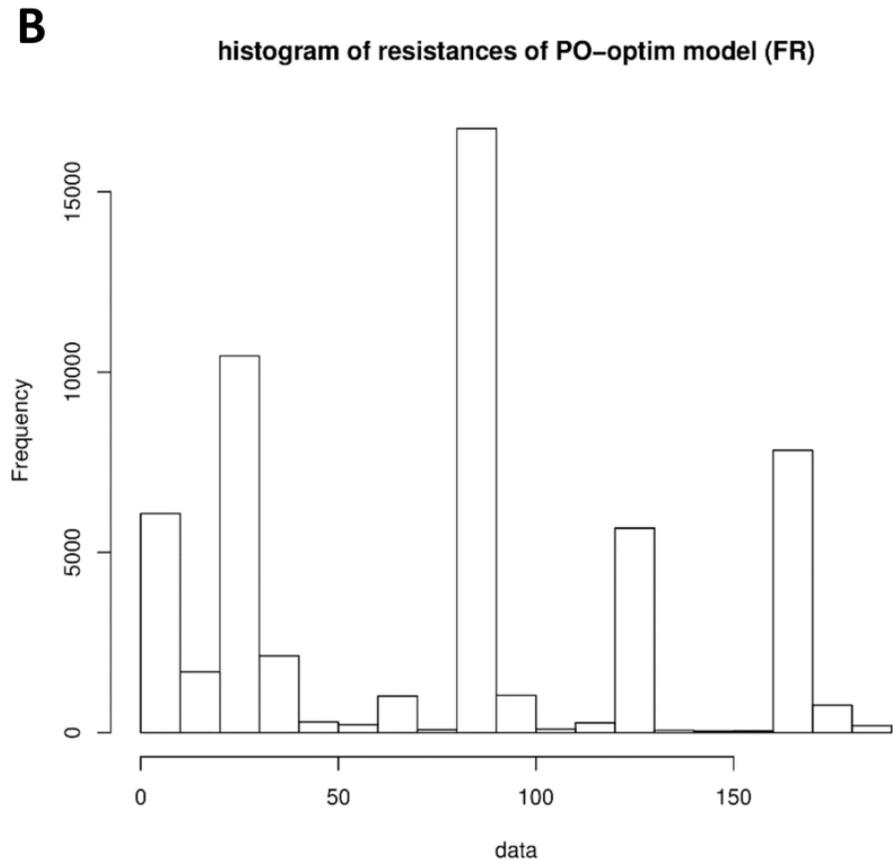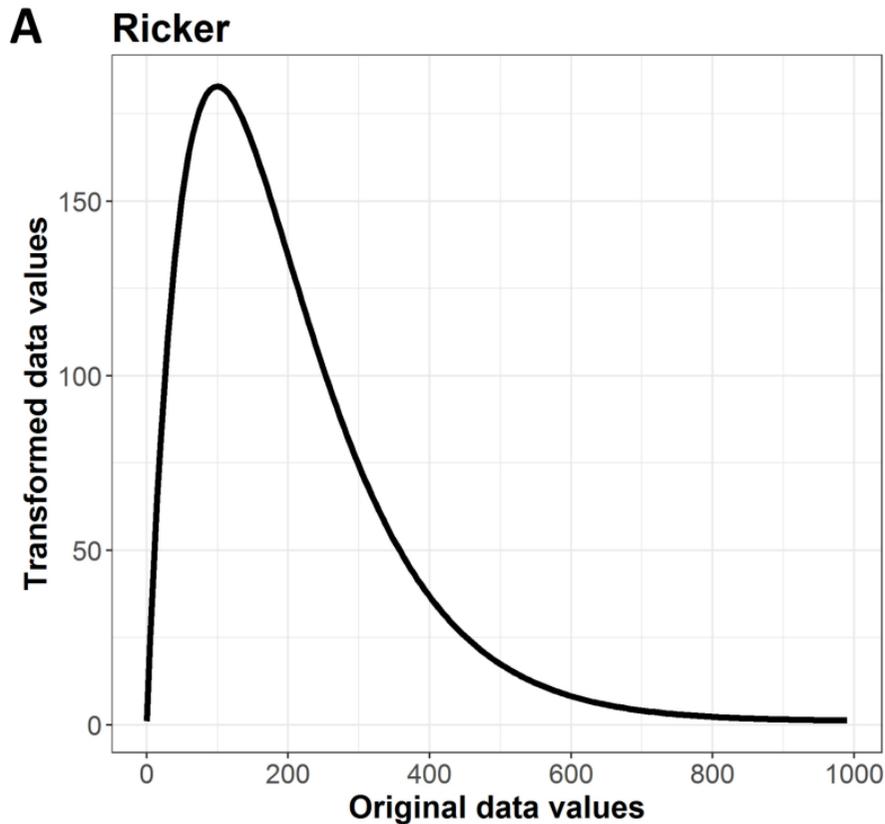
Figure S9: Details of the PO-optim model for Freiburg (FR), the model explaining gene flow best in Freiburg. A: Ricker function used to transform the species distribution model to the PO-optim model, following optimization with ResistanceGA. B: Histogram of resistance values of the PO-optim model.

Table S1: Composition of evolutionary lineages and quantification of genetic diversity for clusters of all cities. Lineage are codes with EF: Eastern France, SA: Southern Alps, VE: Venetian, TU: Tuscany, WF: Western France.

cities and clusters	N ind. (Q >0.8)	N admixed (Q <0.8)	haplotypes of ind. (Q >0.8)	new haplotype (0.2>Q >0.8)	pi-diversity (cytb)	allelic richness	He
Trier	131	92			0.2%		
TR_1	51	34	EF	-	0.1%	3.40	0.578
TR_2	40	36	EF	-	0.2%	4.50	0.688
TR_3	40	12	EF	-	0.1%	4.63	0.700
Saarbrücken	98	99			2.4%		
SB_1	10	5	EF, SA, WF	-	3.1%	3.27	0.670
SB_2	17	19	EF, SA	-	1.8%	4.60	0.780
SB_3	22	14	EF	-	0.1%	4.04	0.691
SB_4	15	4	EF, WF	SA	1.4%	3.37	0.665
SB_5	12	9	EF, SA, WF	-	2.5%	4.47	0.802
SB_6	17	10	EF, SA	-	0.6%	3.48	0.701
SB_7	6	10	EF, SA	-	2.5%	3.44	0.677
Mannheim	108	95			1.8%		
MA_1	9	24	SA	EF, VE	0.0%	3.46	0.690
MA_2	14	20	SA	VE, WE	0.0%	2.84	0.610
MA_3	30	39	EF, SA	-	0.4%	3.42	0.656
MA_4	13	9	EF, SA, WF	-	2.4%	3.97	0.700
MA_5	11	35	SA	EF	0.4%	3.25	0.668
MA_6	31	11	VE	EF, SA	0.0%	3.40	0.649
Freiburg	180	23			3.1%		
FR_1	17	7	SA, WF	EF	1.8%	3.55	0.692
FR_2	19	9	SA, TU, VE	EF	2.4%	4.45	0.738
FR_3	40	1	EF, SA, TU, VE	-	2.6%	5.13	0.786
FR_4	22	2	EF	TU	0.2%	3.80	0.734
FR_5	44	17	EF, TU	SA	2.0%	3.47	0.695
FR_6	27	10	EF, SA, TU, WF	-	2.5%	5.33	0.791

Table S2: Admixture-types of individuals in all cities at higher number of *K*'s (A) and Δ*K*=max (B). Admixture-types are defined by the genetic composition of parental clusters: type 0 of two same lineage clusters; type 1 of two different lineage clusters; type 2 of a hybrid swarm cluster and a single lineage cluster; type 3 of two hybrid swarm clusters.

A	Trier (TR)		Saarbrücken (SB)		Mannheim (MA)		Freiburg (FR)	
	N	prop.	N	prop.	N	prop.	N	prop.
admixture-type 0	75	100.0%	0	0.0%	7	10.3%	0	0.0%
admixture-type 1	0	0.0%	0	0.0%	15	22.1%	0	0.0%
admixture-type 2	0	0.0%	20	30.8%	44	64.7%	2	8.7%
admixture-type 3	0	0.0%	45	69.2%	2	2.9%	21	91.3%
total	75		65		68		23	

B	Trier (TR)		Saarbrücken (SB)		Mannheim (MA)		Freiburg (FR)	
	N	prop.	N	prop.	N	prop.	N	prop.
admixture-type 0	155	100.0%	0	0.0%	0	0.0%	0	0.0%
admixture-type 1	0	0.0%	0	0.0%	0	0.0%	0	0.0%
admixture-type 2	0	0.0%	0	0.0%	17	100.0%	3	17.6%
admixture-type 3	0	0.0%	55	100.0%	0	0.0%	14	82.4%
total	155		55		17		17	

Table S3: Landscape genetic results and scores of model fit of all cities. Results are sorted according to AICc rank within cities. R2m=marginal R²; R2c=conditional R².

City	environmental factor	k	AICc	ΔAICc	R2m	R2c
				scores of model fit		
Trier	PO-optim model	4	-47595.07		0.10	0.31
TR	Water cover	3	-47592.07	3.00	0.09	0.30
	Slope	4	-47542.06	53.02	0.13	0.35
	Substrate	5	-47516.33	78.74	0.09	0.31
	Traffic	4	-47516.01	79.06	0.05	0.27
	Canopy cover	3	-47483.31	111.76	0.05	0.27
	Walls	3	-47480.63	114.44	0.05	0.28
	Buildings	3	-47479.37	115.70	0.05	0.27
	Roads	3	-47477.07	118.01	0.05	0.28
	Railway tracks	3	-47475.45	119.62	0.04	0.26
	Structural diversity	5	-47475.36	119.71	0.04	0.26
	Distance	2	-47474.88	120.19	0.03	0.26
	Aspect	9	-47454.56	140.51	0.06	0.28
	Vegetation type	5	-47317.59	277.48	0.27	0.54
	Null	1	-46863.21	731.86	0.00	0.24
Saarbrücken	Vegetation type	5	1535.14		0.04	0.06
SB	Traffic	4	1542.43	7.29	0.03	0.06
	Water cover	3	1544.76	9.63	0.03	0.06
	Buildings	3	1554.77	19.63	0.03	0.06
	Distance	2	1555.59	20.45	0.03	0.06
	Railway tracks	3	1555.59	20.45	0.03	0.06
	Canopy cover	3	1556.43	21.30	0.03	0.06
	PO-optim model	4	1556.83	21.69	0.03	0.06
	Roads	3	1557.46	22.32	0.03	0.06
	Walls	3	1557.63	22.49	0.03	0.06
	Structural diversity	6	1558.98	23.84	0.03	0.06
	Aspect	6	1563.72	28.59	0.03	0.07
	Substrate	5	1582.81	47.67	0.03	0.07
	Null	1	1683.36	148.22	0.00	0.02
Mannheim	Water cover	3	-36899.77		0.66	0.98
MA	Substrate	5	-36653.92	245.85	0.61	0.88
	Railway tracks	3	-36450.10	449.67	0.61	0.87
	PO-optim model	4	-36275.89	623.88	0.48	0.69
	Vegetation type	5	-36201.37	698.40	0.66	1.00
	Traffic	4	-35479.32	1420.45	0.66	0.97
	Roads	3	-35426.25	1473.52	0.62	0.94
	Structural diversity	5	-35292.17	1607.60	0.56	0.83
	Canopy cover	3	-35286.24	1613.53	0.21	0.33
	Walls	3	-35259.30	1640.47	0.23	0.35
	Distance	2	-35253.92	1645.85	0.20	0.33
	Buildings	3	-35252.04	1647.73	0.20	0.33
	Aspect	6	-35249.45	1650.32	0.20	0.32
	Null	1	-32451.82	4447.95	0.00	0.14
Freiburg	PO-optim model	4	136040.94		0.41	0.57
FR	Railway tracks	3	136166.93	125.99	0.46	0.62
	Canopy cover	3	136323.47	282.53	0.37	0.51
	Vegetation type	5	136434.96	394.02	0.41	0.58
	Aspect	6	136470.40	429.46	0.41	0.58
	Traffic	4	136479.69	438.75	0.27	0.42

Walls	3	136494.34	453.40	0.43	0.69
Water cover	3	136517.31	476.37	0.27	0.40
Buildings	3	136534.27	493.33	0.25	0.41
Roads	3	136538.78	497.84	0.25	0.39
Distance	2	136538.80	497.86	0.24	0.39
Structural diversity	5	136657.24	616.30	0.29	0.46
Substrate	5	136753.03	712.09	0.38	0.62
Null	1	139753.32	3712.38	0.00	0.16

Table S4: Pearson correlations of environmental factors and the PO-optim model for the city of Freiburg (FR), which is the model best explaining gene flow.

	PO-optim model
Railway tracks	0.295
Substrate	-0.291
Structural diversity	-0.279
Buidlings	-0.250
Walls	-0.238
Canopy cover	0.165
Roads	0.146
Aspect	0.126
Vegetation type	0.050
Water cover	-0.032
Traffic	0.024

Table S5: Resistance values per subcategory of environmental factors following extraction from PO-optim for the city of Freiburg, FR.

environ.factor	#classes	class	min	max	median	N.cells
Aspect	5	1	1.368	178.289	129.604	2488
Aspect	5	4	2.483	178.289	129.604	1926
Aspect	5	7	1.237	182.924	82.431	34290
Aspect	5	8	2.043	178.289	129.604	300
Aspect	5	10	1.243	178.289	129.604	1410
Buildings	2	0	1.237	182.924	82.431	30667
Buildings	2	1	1.333	182.462	82.431	9747
Canopy cover	2	0	1.237	182.924	82.431	33717
Canopy cover	2	1	1.266	182.743	129.604	6697
Railway tracks	2	0	1.241	182.924	82.431	38517
Railway tracks	2	1	1.237	125.594	1.922	1897
Roads	2	0	1.239	182.924	82.431	27825
Roads	2	1	1.237	182.743	22.484	12589
Structural diversity	4	1	1.355	182.924	82.431	37246
Structural diversity	4	4	1.248	130.274	2.523	783
Structural diversity	4	7	1.237	7.244	2.201	927
Structural diversity	4	10	1.294	178.477	33.669	1458
Substrate	4	0	1.239	182.743	129.604	17110
Substrate	4	2	1.237	182.924	82.431	21789
Substrate	4	6	1.241	35.453	8.324	740
Substrate	4	10	1.299	10.719	1.922	775
Traffic	48	0	1.237	181.210	82.431	37719
Traffic	48	1	1.989	83.498	20.692	86
Traffic	48	2	1.432	174.282	19.039	172
Traffic	48	3	2.814	176.803	17.517	63
Traffic	48	4	1.413	70.399	16.362	42
Traffic	48	5	15.167	174.530	15.167	29
Traffic	48	6	14.276	91.808	14.276	56
Traffic	48	7	1.319	176.959	13.437	123
Traffic	48	8	12.745	12.745	12.745	10
Traffic	48	9	1.248	157.292	12.089	71
Traffic	48	10	1.726	155.411	11.555	34
Traffic	48	11	1.248	152.863	11.046	186
Traffic	48	12	1.307	151.575	10.639	105
Traffic	48	13	10.325	181.612	10.325	18
Traffic	48	14	1.273	181.839	10.021	37
Traffic	48	15	1.302	147.000	9.799	118
Traffic	48	17	1.432	182.387	9.511	47
Traffic	48	18	1.355	144.351	9.371	85
Traffic	48	19	1.465	144.351	9.371	120
Traffic	48	20	1.261	43.799	9.301	33
Traffic	48	21	1.429	182.387	9.371	59
Traffic	48	22	9.441	9.441	9.441	3
Traffic	48	23	9.511	145.015	9.511	41
Traffic	48	25	1.473	147.000	9.872	107
Traffic	48	26	1.676	148.316	10.096	22
Traffic	48	27	1.365	149.625	10.403	43
Traffic	48	29	1.844	153.504	11.213	49
Traffic	48	31	12.273	55.771	12.273	41
Traffic	48	32	12.939	53.098	12.939	41
Traffic	48	33	5.374	13.642	13.642	33

Traffic	48	36	16.611	16.611	16.611	19
Traffic	48	37	1.241	172.429	17.919	32
Traffic	48	38	3.752	19.476	19.476	19
Traffic	48	39	3.140	21.166	21.166	14
Traffic	48	41	1.365	180.463	25.363	37
Traffic	48	45	1.319	182.743	37.874	88
Traffic	48	47	1.629	134.976	2.071	31
Traffic	48	49	58.152	175.804	58.152	30
Traffic	48	50	1.765	179.749	64.918	44
Traffic	48	52	2.751	167.829	79.807	41
Traffic	48	55	3.613	179.901	106.793	45
Traffic	48	57	68.669	131.926	126.928	31
Traffic	48	62	9.946	178.477	170.453	55
Traffic	48	63	61.882	176.406	176.406	34
Traffic	48	64	14.493	180.188	180.188	29
Traffic	48	65	151.300	182.527	182.462	28
Traffic	48	67	2.483	181.361	108.047	41
Traffic	48	69	2.563	182.924	172.386	203
Vegetation type	4	0	1.237	182.924	79.807	23450
Vegetation type	4	3	1.310	182.527	97.640	1396
Vegetation type	4	7	1.289	182.743	169.422	9461
Vegetation type	4	10	1.239	129.604	129.604	6107
Walls	2	0	1.237	182.924	82.431	19413
Walls	2	1	1.239	182.743	82.431	21001
Water cover	2	-10	1.241	182.924	4.532	592
Water cover	2	0	1.237	182.743	82.431	39822

Table S6: Landscape genetic results of all cities, sorted by average AICc rank across cities.

Final rank	Model	average AICc rank
1	Water cover	3.5
2	PO-optim model	3.75
3	Traffic	5
4	Railway tracks	5.25
5	Vegetation type	6
6	Canopy cover	6.75
7	Substrate	8.25
8	Buildings	8.5
9	Walls	9
10	Roads	9.25
11	*Distance*	10
12	Structural diversity	11
13	Aspect	11.25
14	*Null*	15

Table S7: Evaluation metrics of the SDMs of all cities. Test AUC and test omission (proportion of misidentified test presences with respect to maxSSS threshold rule) are presented (mean and SD).

	SB			MA			FR			TR		
	mean	±	sd	mean	±	sd	mean	±	sd	mean	±	sd
Test AUC	0.792	±	0.058	0.825	±	0.067	0.852	±	0.032	0.852	±	0.041
Test omission	0.308	±	0.137	0.217	±	0.165	0.236	±	0.072	0.144	±	0.085