Supplementary materials

Physical and ecological isolation contribute to maintain genetic differentiation between fire salamander subspecies

Bernardo Antunes[1,2]
Guillermo Velo-Antón[3]
David Buckley[4,5]
Ricardo J. Pereira[6]
Iñigo Martínez-Solano[2*]

[1] Institute of Environmental Sciences, Jagiellonian University, Gronostajowa 7, 30-387 Kraków, Poland.
[2] Departamento de Biodiversidad y Biología Evolutiva. Museo Nacional de Ciencias Naturales MNCN-CSIC, c/José Gutiérrez Abascal 2, 28006 Madrid, Spain.
[3] CIBIO/InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos da Universidade do Porto. Instituto de Ciências Agrárias de Vairão. Rua Padre Armando Quintas 7. 4485-661 Vairão Portugal.
[4] Departamento de Biología (Genética). Facultad de Ciencias (Ed. Biología), Universidad Autónoma de Madrid (UAM). c/ Darwin 2, 28049-Madrid, Spain.
[5] Centro de Investigaciones en Biodiversidad y Cambio Global (CIBC-UAM). Edificio de Biología, Universidad Autónoma de Madrid. c/ Darwin 2, 28049-Madrid, Spain.
[6] Division of Evolutionary Biology, Faculty of Biology II, Ludwig-Maximilians-Universität München, Grosshaderner Strasse 2, 82152 Planegg-Martinsried, Germany.

*Corresponding author: Iñigo Martínez-Solano (inigomsolano@mncn.csic.es)
Fax: +34 91 564 5078

Dataset 11 loci, replicated analyses and results

After removing the two loci potentially containing null alleles (SST-A6-I and SalE2), we replicated some analyses for the Total and MTR ranges. Specifically, we re-ran STRUCTURE analyses and model selection, which included obtaining matrices of pairwise genetic differentiation, followed by univariate regression for physical isolation (IBD/R) and model selection in multiple regressions including both physical (IBD/R) and ecological (IBE) isolation matrices. STRUCTURE analyses for the Total range showed the same result as previous analyses, with two as the best clustering number (K=2, highest ΔK). For the MTR, results showed slight differences, with three being the best clustering number (K=3, highest ΔK) in contrast to the K=2 previously identified. However, this variation is only minor, as K=3 is among top ΔKs in previous analyses and posterior landscape genetic analyses showed exactly the same results, with selected models being almost identical (Figure S8, Tables S7 and S8). Results were thus confirmed to hold.

Supplementary Tables

Table S1. Details of the complete sample dataset, including sample ID, population ID, sampling locality, geographic coordinates (latitude and longitude), and samples that were genotyped with microsatellites and mtDNA lineages from Pereira et al. 2016.

(see associated File: "Antunes_et_al_2021_Table S1.xlx")

Table S2. Details of the 14 microsatellites used in this study. Information regarding multiplex arrangement, original published primers and fluorescently labelled oligonucleotides used as template for modified forward primers are displayed. The primer volume used to create a multiplex with a total volume of 100 µl (distilled H_2O plus primer and fluorescent label's volumes) is also represented (PVM). In this process, the forward and reverse primers were concentrated at 10 µM and 100 µM respectively.

Locus	Multiplex	Label*	Primer forward (5' – 3')	Primer reverse (5' – 3')	PVM (µl)
SST-A6-I [2]	Ssal1	NED	TTCAGTGCTCTTGCAGGTTG	AGTCTGCAAGGATAGAAAGATCG	2,8
SST-A6-II [2]	Ssal1	PET	ATTCTCTCTGACAAGGATTGTGG	GGTAGACAGACATCAAGGCAGAC	2
SalE14 [1]	Ssal1	VIC	GCTGCCCTCTCTGCCTACTGACCAT	GCCAAGACATGGAACACCCTCCCGC	1,6
Sal29 [1]	Ssal2	6-FAM	CTCTTTGACTGAACCAGAACCCC	GCCTGTCGGCTCTGTGTAACC	8
SST-B11 [2]	Ssal2	PET	TCAAACGGTGCCAAAGTTATTAG	TTAATTGGCAGTTTTCTTTCCAG	2
SalE12 [1]	Ssal2	VIC	CTCAGGAACAGTGTGCCCCAAATAC	CTCATAATTTAGTCTACCCTCCCAC	0,8
SST-C3 [2]	Ssal3	PET	CCGTTTGAGTCACTTCTTTCTTG	TTGCTTTACCAACCAGTTATTGTC	1,4
SalE7 [1]	Ssal3	NED	TTTCAGCACCAAGATACCTCTTTTG	CTCCCTCCATATCAAGGTCACAGAC	0,8
SalE5 [1]	Ssal3	6-FAM	CCACATGATGCCTACGTATGTTGTG	CTCCTGTTTACGCTTCACCTGCTCC	0,6
SalE2 [1]	Ssal3	VIC	CACGACAAAATACAGAGAGTGGATA	ATATTTGAAATTGCCCATTTGGTA	3
SalE06 [1]	Ssal4	VIC	GGACTCATGGTCACCCAGAGGTTCT	ATGGATTGTGTCGAAATAAGGTATC	1,2
Sal3 [1]	Ssal4	6-FAM	CTCAGACAAGAAATCCTGCTTCTTC	ATAAATCTGTCCTGTTCCTAATCAG	1,2
SalE8 [1]	Ssal4	NED	GCAAAGTCCATGCTTTCCCTTTCTC	GACATACCAAAGACTCCAGAATGGG	0,8
SST-G9 [2]	Ssal4	NED	CCTCGTCAGGGGTTGTAGG	CTTTCCAGGAAGAAACTGAGATG	0,8

Table S3. Occurrence records used for ecological niche modelling. Subspecies were defined based on mtDNA data, with 28 corresponding to *S. s. bejarae* and 10 to *S. s. almanzoris*. Note that these records were used in undifferentiated way in ENMs.

Subspecies	Sample ID	Latitude	Longitude
S. s. almanzoris	*PHS33*	40,15	-4,75
S. s. almanzoris	*IMS_747*	40,26	-5,12
S. s. almanzoris	*IMS_750*	40,27	-5,30
S. s. almanzoris	*IMS_718*	40,66	-4,13
S. s. almanzoris	*SS71*	40,79	-4,10
S. s. almanzoris	*IMS_712*	40,79	-3,83
S. s. almanzoris	*IMS_749*	40,86	-3,95
S. s. almanzoris	*IMS_465*	40,98	-3,81
S. s. almanzoris	*AHE web*	41,06	-3,67
S. s. almanzoris	*AHE web*	40,80	-3,96
S. s. bejarae	*AHE web*	39,31	-4,87
S. s. bejarae	*JAS4_9*	39,35	-4,40
S. s. bejarae	*AHE web*	39,44	-4,09
S. s. bejarae	*AHE web*	39,45	-5,21
S. s. bejarae	*AHE web*	39,48	-5,39
S. s. bejarae	*AHE web*	39,50	-5,11
S. s. bejarae	*AHE web*	39,52	-4,27
S. s. bejarae	*AHE web*	39,54	-4,79
S. s. bejarae	*AHE web*	39,54	-4,37
S. s. bejarae	*IMS_2173*	39,56	-4,59
S. s. bejarae	*JAS8_1*	39,57	-4,10
S. s. bejarae	*AHE web*	39,57	-4,10
S. s. bejarae	*JAS2_9*	39,59	-4,71
S. s. bejarae	*AHE web*	39,61	-4,96
S. s. bejarae	*AHE web*	39,61	-5,25
S. s. bejarae	*AHE web*	39,64	-5,42
S. s. bejarae	*1S*	39,72	-5,73
S. s. bejarae	*AHE web*	39,84	-5,99
S. s. bejarae	*AHE web*	39,86	-6,13
S. s. bejarae	*AHE web*	40,12	-5,99
S. s. bejarae	*AHE web*	40,19	-6,35
S. s. bejarae	*SS64*	40,22	-5,79
S. s. bejarae	*AHE web*	40,29	-6,15
S. s. bejarae	*IMS_1238*	40,32	-5,96
S. s. bejarae	*IMS_1075*	40,34	-4,36
S. s. bejarae	*SS69*	40,38	-4,44
S. s. bejarae	*SS121*	40,49	-6,06
S. s. bejarae	*SS120*	40,56	-4,29

Table S4. Average training area under the curve percent (AUC), and average percentage contribution of each variable, for the four ENMs (Five climatic variables, Landcover, Slope and All models).

Spatial data	Occurrence data	Training AUC	Bio3 (Isothermality)	Bio7 (Temperature Annual Range)	Bio8 (Mean Temperature of Wettest Quarter)	Bio16 (Precipitation of Wettest Quarter)	Bio17 (Precipitation of Driest Quarter)	Land cover (10 Categories)	Slope
Climate (5 variables)	*S. s. almanzoris* + *S. s. bejarae*	0.91 (±0.022)	35,8	19	21,2	10,2	13,7	-	-
Landcover (10 categories)	*S. s. almanzoris* + *S. s. bejarae*	0.786 (±0.031)	-	-	-	-	-	100,00	-
Slope	*S. s. almanzoris* + *S. s. bejarae*	0.884 (±0.028)	-	-	-	-	-	-	100,00
All (5 climatic variables + Slope + Landcover)	*S. s. almanzoris* + *S. s. bejarae*	0.966 (±0.007)	5,5	8,4	6,1	10,3	6,6	8,3	54,7

Table S5. Results from univariate regressions among matrices of genetic differentiation (pairwise FST) and matrices of physical isolation (IBD, IBR-Climate, IBR-Slope, IBR-Landcover and IBR-All). The coefficient of determination (r^2) and p-value (p) are displayed for MMRR models. The degrees of freedom (df) and conditional Akaike information criterion (AICc) are displayed for MLPE models.

	Univariate regressions	**Total**		**ICS**		**MTR**	
	for physical isolation models	r^2	p	r^2	p	r^2	p
MMRR	IBD	0,518	0,001	0,426	0,001	0,361	0,028
	IBR-Climate	0,489	0,001	0,261	0,005	**0,633**	0,001
	IBR-Landcover	0,506	0,001	0,446	0,001	0,379	0,022
	IBR-Slope	0,469	0,001	0,334	0,001	0,306	0,032
	IBR-All	**0,681**	0,001	**0,661**	0,001	0,306	0,020
		df	AICc	df	AICc	df	AICc
MLPE	IBD	4	225,922	4	63,829	4	59,313
	IBR-Climate	4	228,916	4	70,471	4	**40,950**
	IBR-Landcover	4	218,419	4	64,395	4	58,660
	IBR-Slope	4	234,311	4	65,696	4	62,793
	IBR-All	4	**210,246**	4	**57,244**	4	63,491

Table S6. Model importance (weight) of categorical variables in landcover-based ENMs, measured as probability of presence.

Landcover	Variable model importance
Artificial surfaces	0,17
Agricultural areas	0,11
Heterogeneous agricultural areas	0,17
Broad-leaved forest	0,54
Coniferous forest	0,61
Mixed forest	0,54
Scrub and/or herbaceous vegetation associations	0,49
Open spaces with little or no vegetation	0,38
Inland wetlands and Maritime wetlands	0,17
Inland waters	0,17

Table S7. Results from univariate regressions among matrices of genetic differentiation (pairwise FST) and matrices of physical isolation (IBD, IBR-Climate, IBR-Slope, IBR-Landcover and IBR-All) based in the 11 loci dataset. The coefficient of determination (r^2) and p-value (p) are displayed for MMRR models. The degrees of freedom (df) and conditional Akaike information criterion (AICc) are displayed for MLPE models.

	Univariate regressions for physical isolation models	**Total - 11 loci**		**MTR - 11 loci**	
		r^2	p	r^2	p
MMRR	IBD	0,485	0,001	0,399	0,003
	IBR-Climate	0,433	0,001	**0,705**	**0,001**
	IBR-Landcover	0,481	0,001	0,430	0,002
	IBR-Slope	0,430	0,001	0,344	0,013
	IBR-All	**0,666**	**0,001**	0,362	0,003
		df	AICc	df	AICc
MLPE	IBD	4	236,031	4	63,377
	IBR-Climate	4	246,600	4	**38,696**
	IBR-Landcover	4	227,235	4	61,674
	IBR-Slope	4	244,004	4	66,548
	IBR-All	4	**219,206**	4	66,153

Table S8. Best MMRR and MLPE models, for each range (see Methods) based on the 11-loci dataset. The coefficient of determination (r^2), regression coefficient (β) and p-value (p) are displayed for the MMRR model. The regression coefficient (β), degrees of freedom (df) and conditional Akaike information criterion (AICc) are displayed for MLPE models.

MMRR - Selected models for each region

Total - 11 loci			
Models	r^2	β	p
IBR-All	-	0,83	0,001
IBE-PC2-Total	-	0,17	0,035
IBE-PC3-Total	-	0,20	0,030
Global	0,713	-	0,001

MTR - 11 loci			
Models	r^2	β	p
IBR-Climate	0,705	1,65	0,001

MLPE - Selected models for each region

Total - 11 loci			
Models	β	df	AICc
IBR-All	0,70	-	-
IBE-PC1-Total	0,14	-	-
IBE-PC2-Total	0,27	-	-
Global	-	6	190,96

MTR - 11 loci			
Models	β	df	AICc
IBR-Climate	1,646	4	38,696

Supplementary Figures

Figure S1. Results of principal components analyses used to reduce data complexity in environmental variables. Numbers on the right of each panel represent the 18 populations used in landscape genetic analyses. Orange tones represent environmental dissimilarity (similar tones and highly contrasting tones represent low and high dissimilarity, respectively).

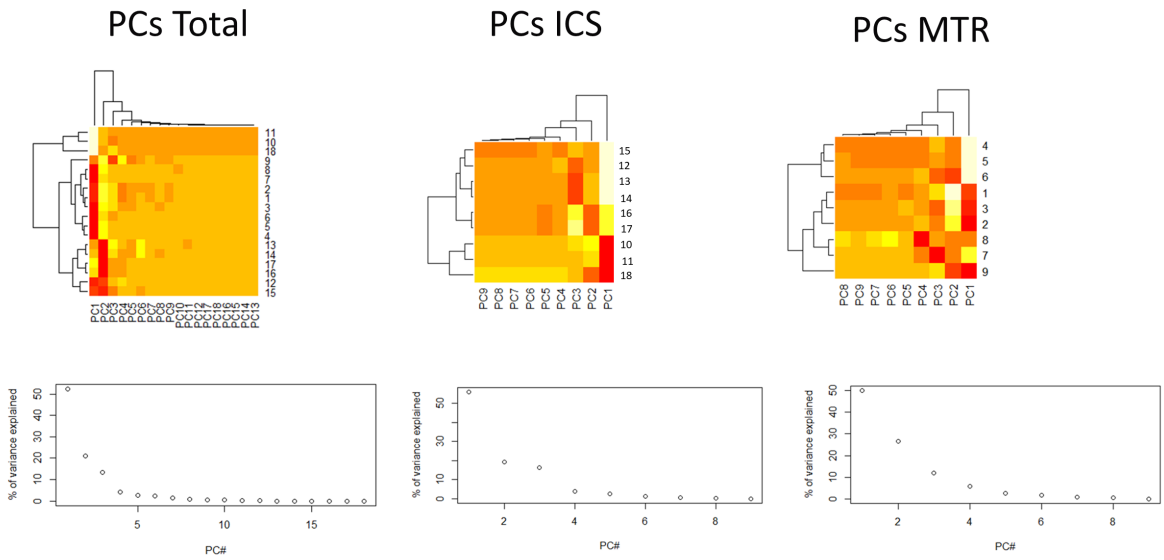

Figure S2. Variable weight from principal components analyses used to reduce data complexity in environmental variables for the three ranges. Bioclimatic variables follow original numeration (see WorldClim Version 1.4.). Ignore characters following variable ids (e.g., b1_t1_t reads as bioclimatic variable 1).

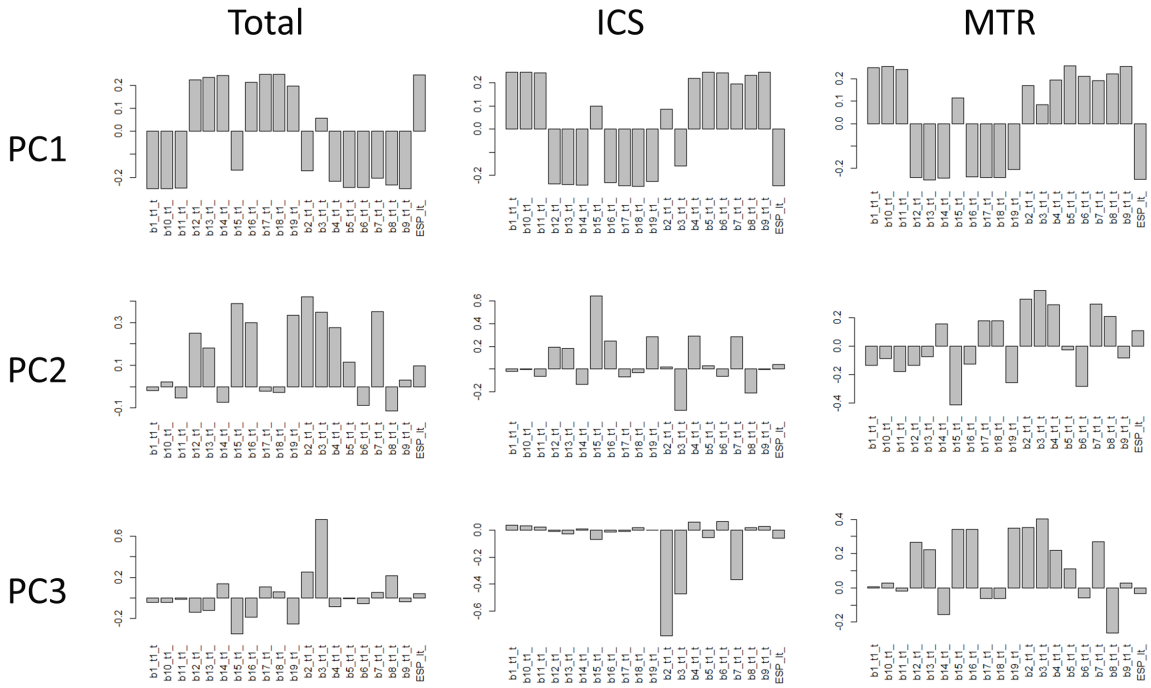

Figure S3. Distance clustering trees based on correlation (Pearson´s r) among matrices of landscape resistance and environmental dissimilarity. Green and red line highlight r =0.7 and r=0.9, respectively.

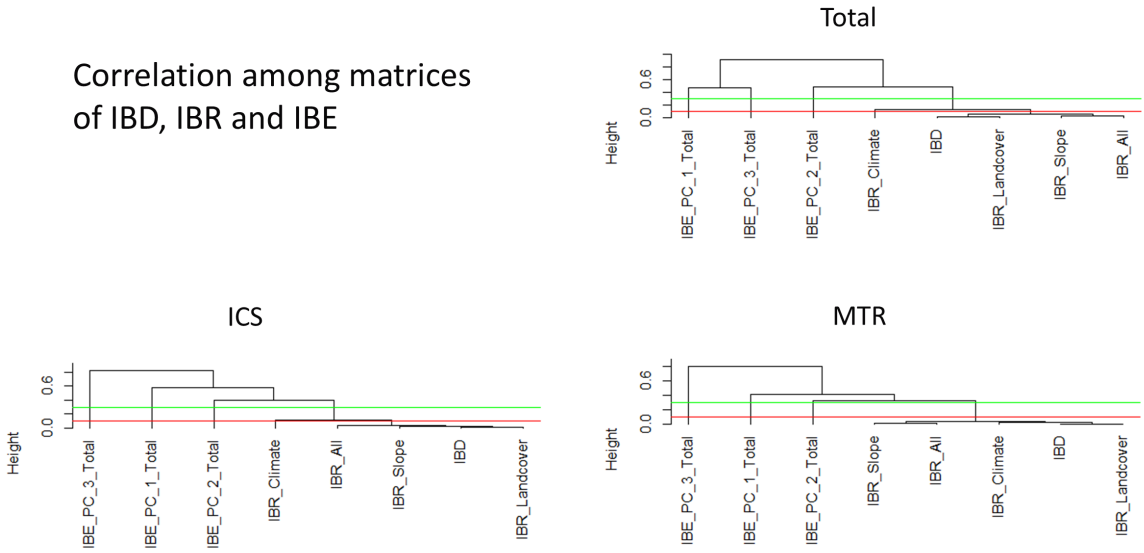

Figure S4. Map showing STRUCTURE results for Total range (K2). Pie charts, with size representing sample size, display population averages of ancestry proportions according to STRUCTURE results (K2, Total range). The associated barplot is shown below the map. Also, an additional barplot displays assignment of individuals to mtDNA lineages (when available; grey color indicates lack of mtDNA data).

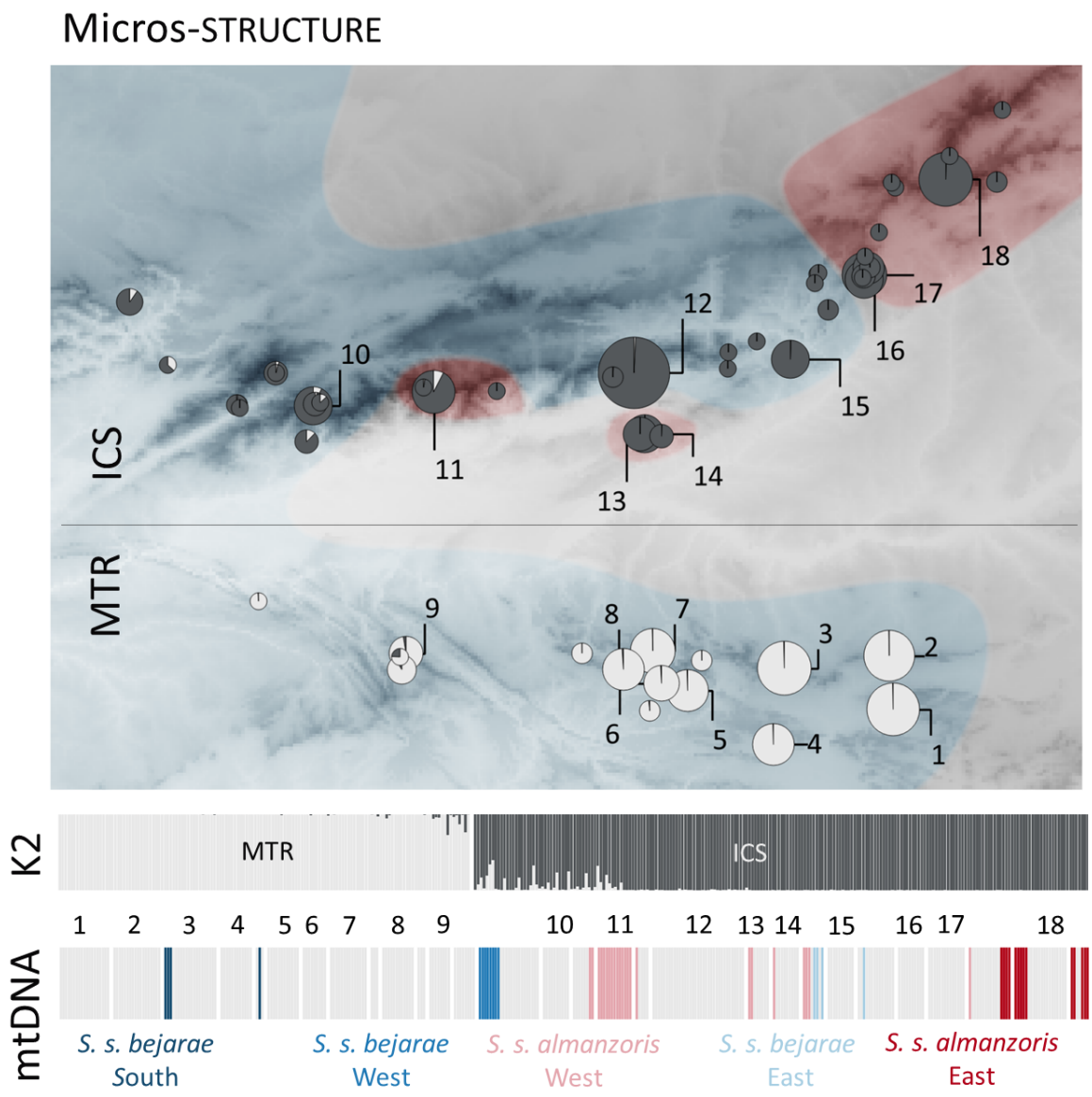

Figure S5. Trees representing environmental dissimilarity among populations used for landscape genetic analyses.

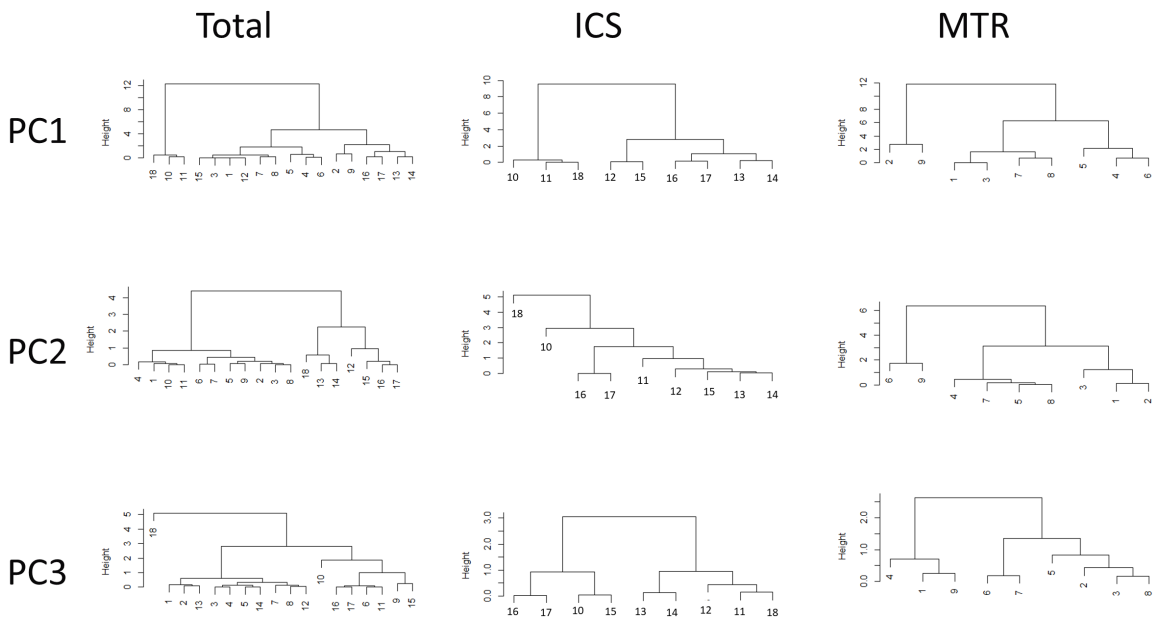

Figure S6. Information from STRUCTURE analyses.

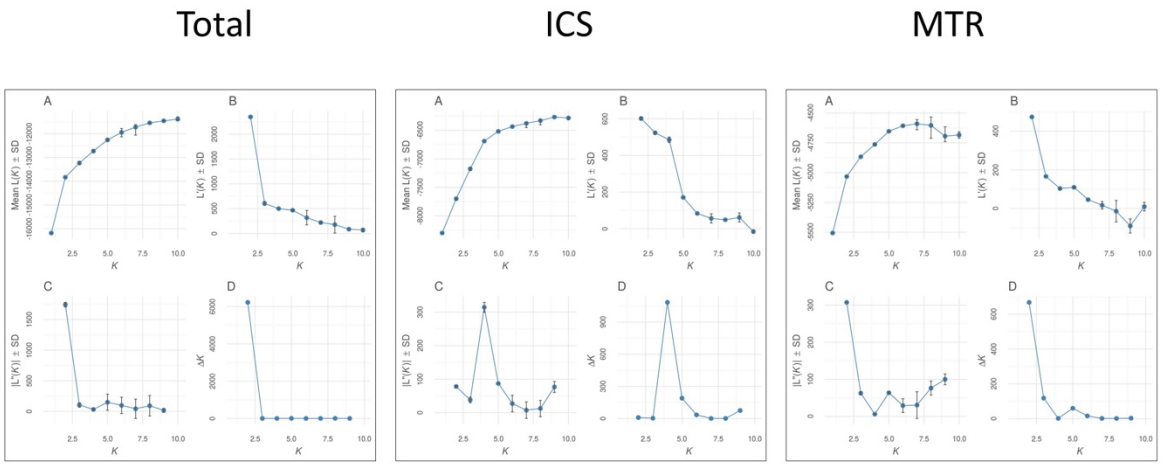

Figure S7. Current maps obtained from CIRCUITSCAPE highlighting connectivity among populations.

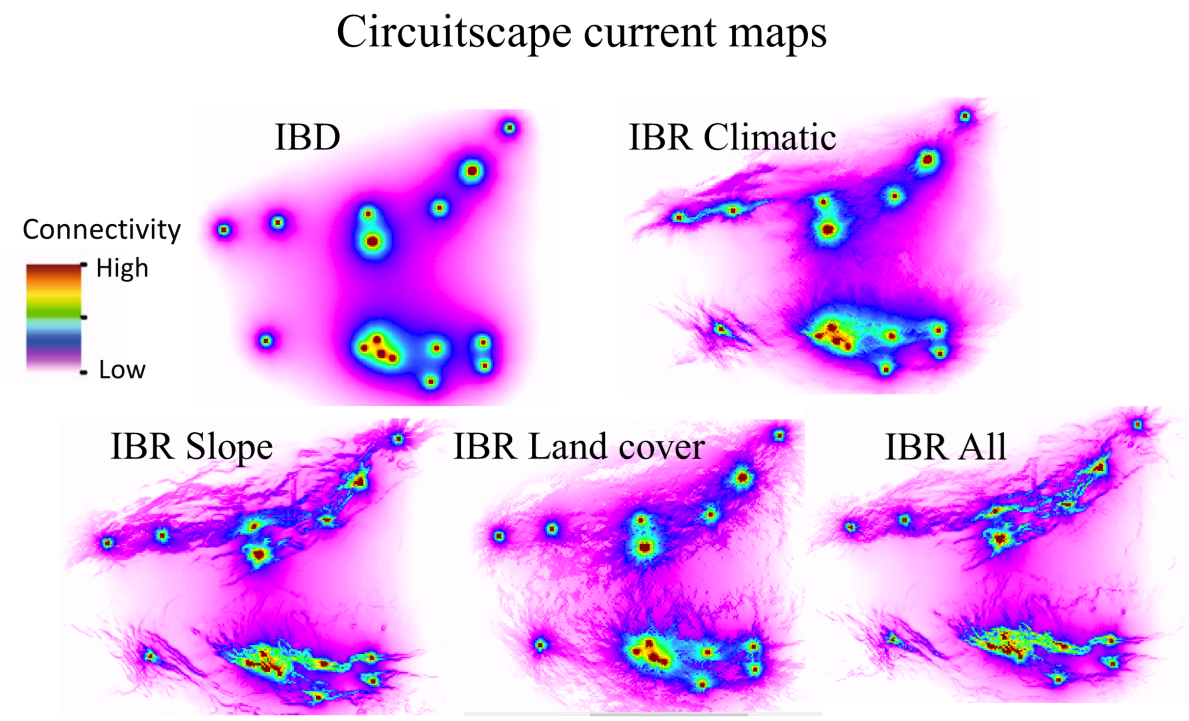

Figure S8. Information from STRUCTURE analyses and results for the 11 loci dataset (see details above).