

Hybridization patterns in two contact zones of grass snakes reveal a new Central European snake species

Carolin Kindler, Maxime Chèvre, Sylvain Ursenbacher, Wolfgang Böhme,
Axel Hille, Daniel Jablonski, Melita Vamberger & Uwe Fritz

Scientific Reports

Supplementary Information

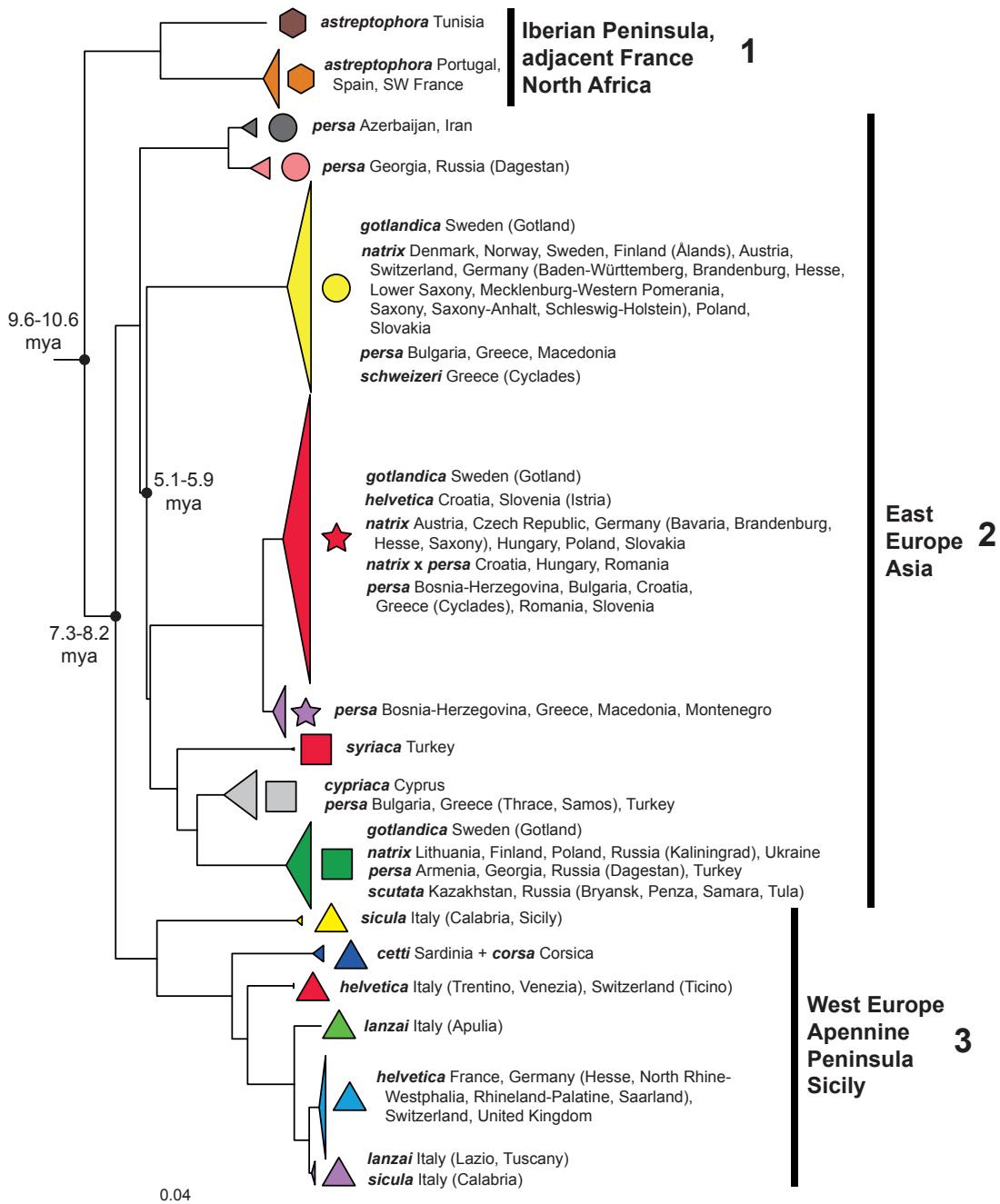


Figure S1. Simplified mitochondrial phylogeny for grass snakes (from Kindler *et al.* 2013, 2014 based on 1,984 bp of ND4 and cyt b). The three major clades 1-3 are highlighted; numbers at nodes indicate split ages according to Fritz *et al.* (2012). The new record for Gotland, Sweden, is added (clade with green square). Note the many mismatches between morphologically defined taxa and mitochondrial clades. The traditional identification of grass snakes from Istria (Kabisch 1999; Kreiner 2007) is in error and also conflicts with morphology (unpubl. data). The figure was created using ADOBE ILLUSTRATOR CS6 (<http://www.adobe.com/products/illustrator.html>).

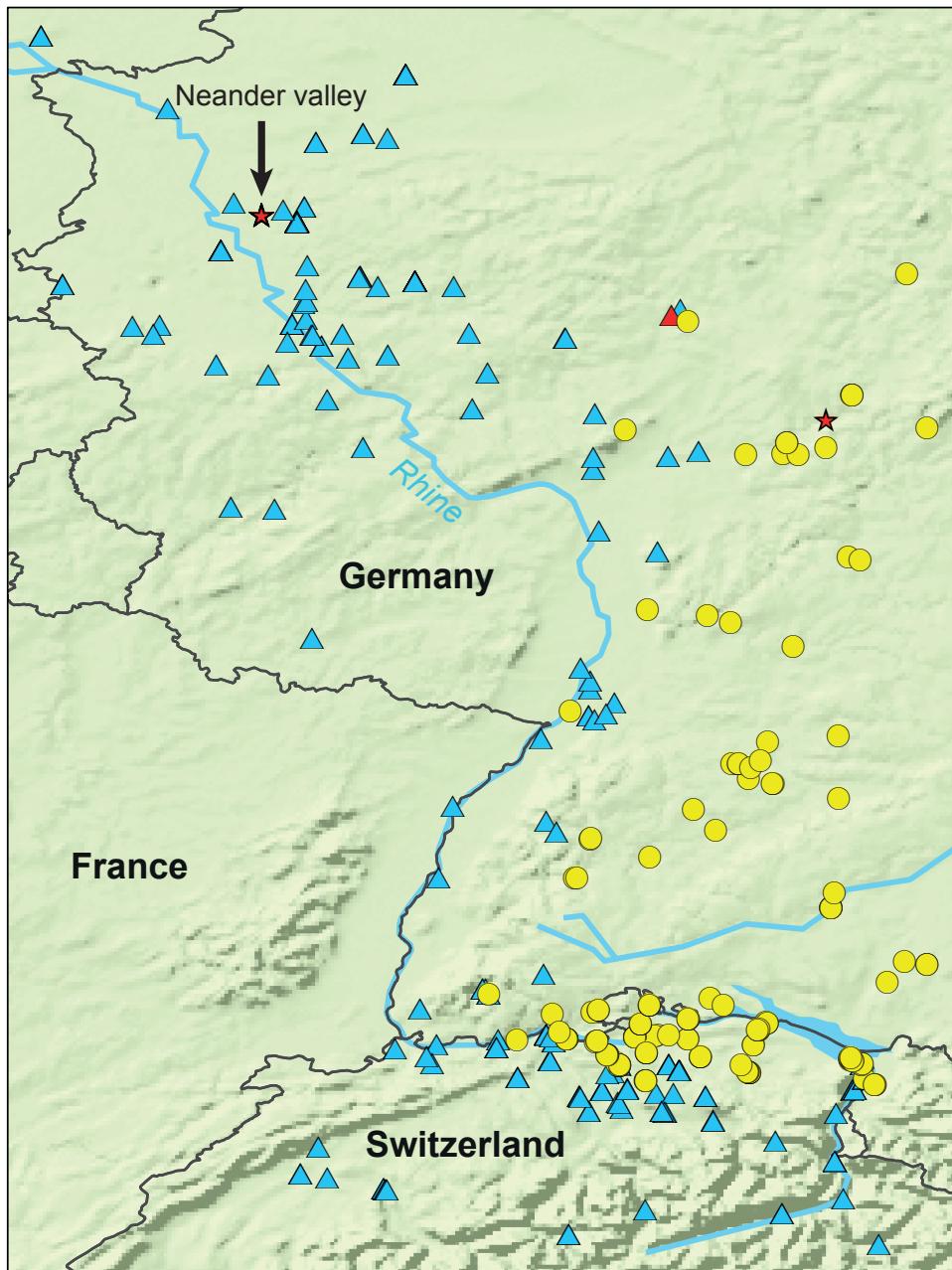
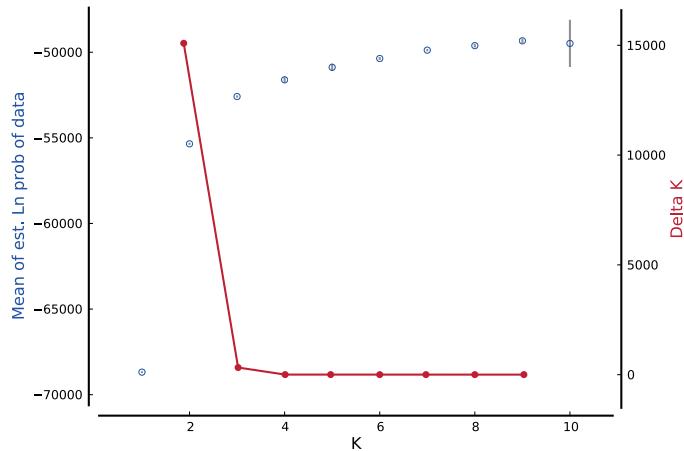
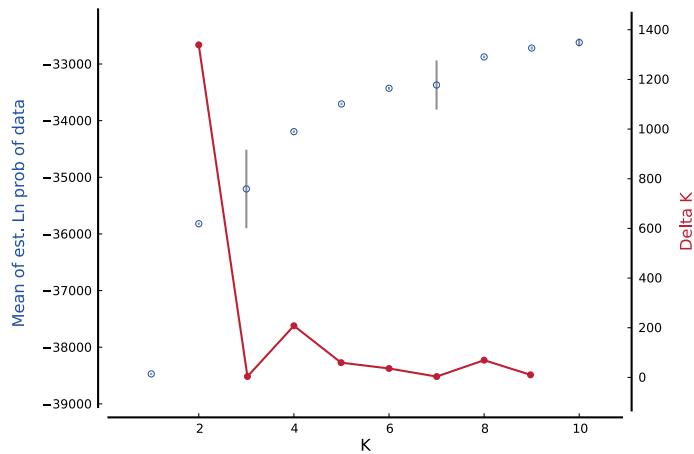


Figure S2. Distribution of mitochondrial haplotypes in the western contact zone.
Symbols correspond to Figures 1 and S1. Arrow highlights allochthonous population of the ‘red lineage’ in the Neander valley. Red triangle represents a non-native lineage from northern Italy. Map was created using ARCGIS 10.2 (<http://www.esri.com/arcgis>) and ADOBE ILLUSTRATOR CS6 (<http://www.adobe.com/products/illustrator.html>).

(a) *helvetica* and all eastern lineages



(b) All eastern lineages without impact of *helvetica*



(c) Yellow and red lineage without impact of adjacent lineages

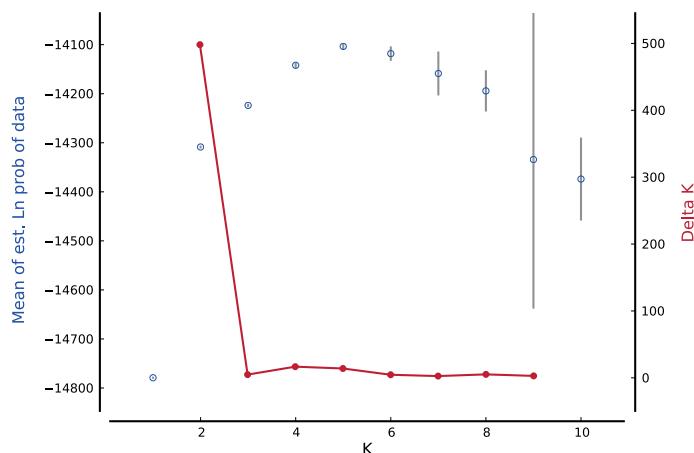


Figure S3. ΔK values and posterior probabilities for STRUCTURE runs for three data sets using STRUCTURE HARVESTER (Earl & vonHoldt, 2012). The modal value is always at $K=2$. $K=1$ could be excluded because of the higher posterior probabilities for $K=2$.

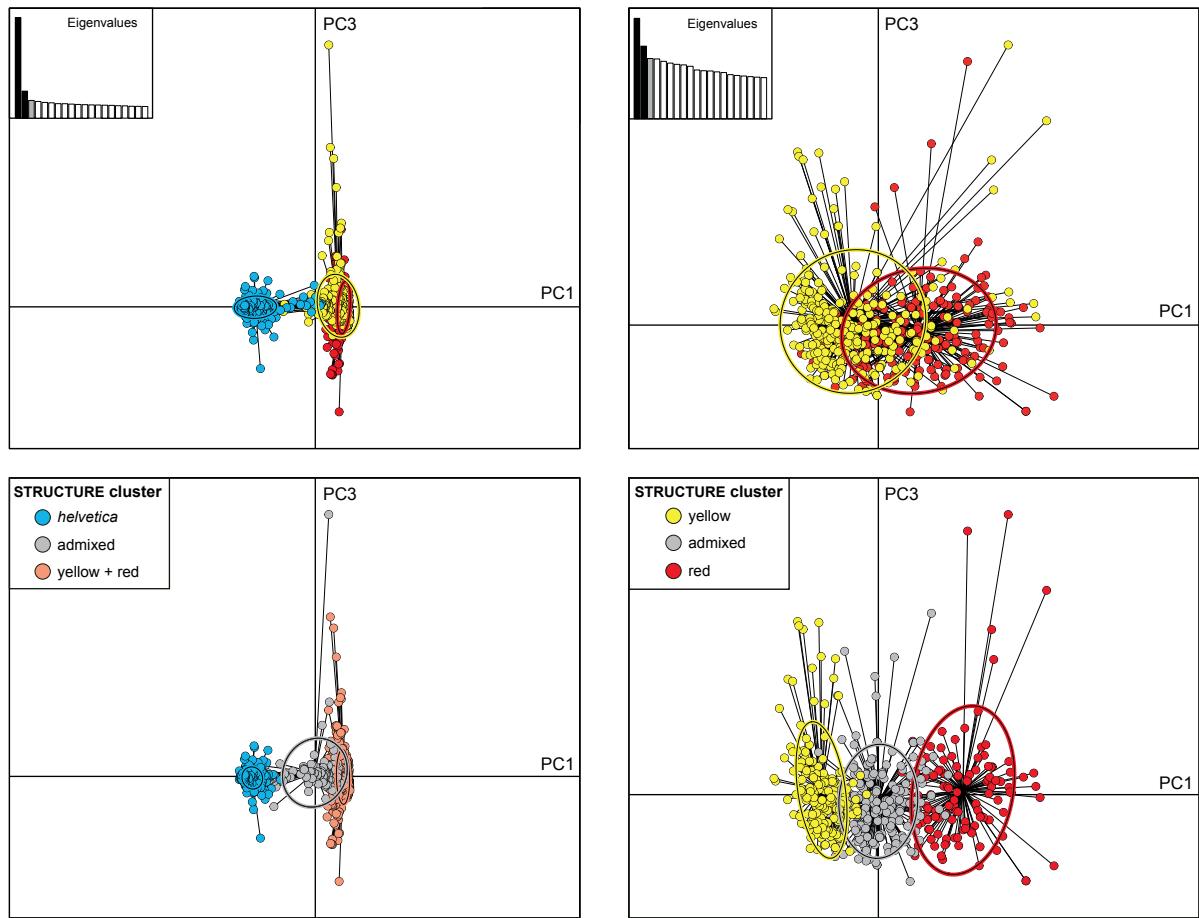


Figure S4. PCA axes 1–3 for microsatellite data of both contact zones. Samples are coloured according to mtDNA lineages (top) and STRUCTURE clusters (bottom). Admixed individuals were categorized according to HYBRIDLAD results. PCAs for the yellow and red lineages correspond to the samples from Figure 3c. Non-native samples were excluded. The oval outlines represent 95% confidential intervals. For *helvetica* and the eastern lineages (left), the x axis explains 16.6% and the y axis 2.9% of variation. For the eastern lineages (right), the x axis explains 3.8% and the y axis 2.4% of variation.

Museum acronyms of vouchers for Table S1:

BEV – Laboratoire de Biogéographie et Ecologie des Vertébrés, Centre d’Ecologie Fonctionnelle & Evolutive, Montpellier
 LMNM – Landesmuseum Natur und Mensch, Oldenburg
 MTD D – Museum of Zoology, Senckenberg Dresden (Herpetological Collection)
 MTD T – Museum of Zoology, Senckenberg Dresden (Tissue Collection)
 MWLK – Museum der Westlausitz, Kamenz
 NHMW – Naturhistorisches Museum Wien
 NME – Naturkundemuseum Erfurt
 SMF – Forschungsinstitut und Naturmuseum Senckenberg, Frankfurt a.M.
 SMNS – Staatliches Museum für Naturkunde Stuttgart
 SMNK – Staatliches Museum für Naturkunde Karlsruhe
 ZFMK – Zoologisches Forschungsmuseum Alexander Koenig, Bonn
 ZMB – Museum für Naturkunde Berlin
 ZMH – Zoologisches Museum Hamburg
 ZMUO – Zoologisk museum, Universitetet i Oslo
 ZSM – Zoologische Staatssammlung München
 ZSUM – Zoologische Sammlung der Philipps-Universität Marburg

Table S2. Akaike Information Criterion (AIC) scores for fitted clines under different models using the R package HZAR (Derryberry *et al.* 2014). Bold values in blue indicate the best-fit model from the 15-model comparison.

	Contact zone I		Contact zone II	
	Microsatellites	mtDNA	Microsatellites	mtDNA
Null Model	152.846	184.261	68.427	254.247
Model 1	38.966	17.566	11.670	74.262
Model 2	22.132	n/a	11.366	n/a
Model 3	17.449	21.839	15.249	78.504
Model 4	18.120	25.038	18.909	n/a
Model 5	17.692	n/a	18.865	82.518
Model 6	23.491	30.390	22.930	86.930
Model 7	29.795	21.169	15.795	78.360
Model 8	13.501	n/a	15.365	n/a
Model 9	19.169	26.013	18.782	82.610
Model 10	36.967	21.694	15.573	78.360
Model 11	26.258	n/a	15.353	n/a
Model 12	21.600	26.034	19.168	82.724
Model 13	15.214	21.694	15.111	78.360
Model 14	15.295	n/a	15.119	n/a
Model 15	20.469	25.999	19.306	82.733

Table S3. European Nucleotide Archive (ENA) accession numbers of ND4 and cyt b haplotypes.

ND4 haplotype	Accession number	cyt b haplotype	Accession number
<i>helvetica</i>			
h1	LT839092	h1	LT839229
h2	LT839093	h2	LT839230
h3	LT839094	h3	LT839231
h4	LT839095	h4	LT839232
h5	LT839096	h5	LT839233
h6	LT839097	h6	LT839234
h7	LT839098	h7	LT839235
h8	LT839099	h8	LT839236
h9	LT839100	h9	LT839237
h10	LT839101	h10	LT839238
h11	LT839102	h11	LT839239
h12	LT839103	h12	LT839240
		h13	LT839241
		h14	LT839242
		h15	LT839243
		h16	LT839244
		h17	LT839245
		h18	LT839246
		h19	LT839247
		h20	LT839248
		h21	LT839249
		h22	LT839250
		h23	LT839251
		h24	LT839252
		h25	LT839253
		h26	LT839254
		h27	LT839255
		h28	LT839256
		h29	LT839257
Yellow lineage (lineage 3 of Kindler <i>et al.</i> , 2013)			
y1	LT839104	y1	LT839258
y2	LT839105	y2	LT839259
y3	LT839106	y3	LT839260
y4	LT839107	y4	LT839261
y5	LT839108	y5	LT839262
y6	LT839109	y6	LT839263
y7	LT839110	y7	LT839264
y8	LT839111	y8	LT839265
y9	LT839112	y9	LT839266
y10	LT839113	y10	LT839267
y11	LT839114	y11	LT839268
y12	LT839115	y12	LT839269
y13	LT839116	y13	LT839270
y14	LT839117	y14	LT839271
y15	LT839118	y15	LT839272
y16	LT839119	y16	LT839273
y17	LT839120	y17	LT839274
y18	LT839121	y18	LT839275
y19	LT839122	y19	LT839276
y20	LT839123	y20	LT839277
y21	LT839124	y21	LT839278

• Table S3 continued

Yellow lineage (lineage 3 of Kindler <i>et al.</i> , 2013)			
y22	LT839125	y22	LT839279
y23	LT839126	y23	LT839280
y24	LT839127	y24	LT839281
y25	LT839128	y25	LT839282
y26	LT839129	y26	LT839283
y27	LT839130	y27	LT839284
y28	LT839131	y28	LT839285
y29	LT839132	y29	LT839286
y30	LT839133	y30	LT839287
y31	LT839134	y31	LT839288
y32	LT839135	y32	LT839289
y33	LT839136	y33	LT839290
y34	LT839137	y34	LT839291
y35	LT839138	y35	LT839292
y36	LT839139	y36	LT839293
y37	LT839140	y37	LT839294
y38	LT839141	y38	LT839295
y39	LT839142	y39	LT839296
y40	LT839143	y40	LT839297
y41	LT839144	y41	LT839298
y42	LT839145	y42	LT839299
y43	LT839146		
Red lineage (lineage 4 of Kindler <i>et al.</i> , 2013)			
r1	LT839147	r1	LT839300
r2	LT839148	r2	LT839301
r3	LT839149	r3	LT839302
r4	LT839150	r4	LT839303
r5	LT839151	r5	LT839304
r6	LT839152	r6	LT839305
r7	LT839153	r7	LT839306
r8	LT839154	r8	LT839307
r9	LT839155	r9	LT839308
r10	LT839156	r10	LT839309
r11	LT839157	r11	LT839310
r12	LT839158	r12	LT839311
r13	LT839159	r13	LT839312
r14	LT839160	r14	LT839313
r15	LT839161	r15	LT839314
r16	LT839162	r16	LT839315
r17	LT839163	r17	LT839316
r18	LT839164	r18	LT839317
r19	LT839165	r19	LT839318
r20	LT839166	r20	LT839319
r21	LT839167	r21	LT839320
r22	LT839168	r22	LT839321
r23	LT839169	r23	LT839322
r24	LT839170	r24	LT839323
r25	LT839171	r25	LT839324
r26	LT839172	r26	LT839325
r27	LT839173	r27	LT839326
r28	LT839174	r28	LT839327
r29	LT839175	r29	LT839328
r30	LT839176	r30	LT839329
r31	LT839177	r31	LT839330

• Table S3 continued

Red lineage (lineage 4 of Kindler *et al.*, 2013)

r32	LT839178	r32	LT839331
r33	LT839179	r33	LT839332
		r34	LT839333
		r35	LT839334
		r36	LT839335
		r37	LT839336
		r38	LT839337
		r39	LT839338
		r40	LT839339
		r41	LT839340
		r42	LT839341
		r43	LT839342
		r44	LT839343
		r45	LT839344
		r46	LT839345
		r47	LT839346
		r48	LT839347
		r49	LT839348
		r50	LT839349
		r51	LT839350
		r52	LT839351
		r53	LT839352
		r54	LT839353
		r55	LT839354
		r56	LT839355
		r57	LT839356
		r58	LT839357
		r59	LT839358
		r60	LT839359
		r61	LT839360
		r62	LT839361
		r63	LT839362
		r64	LT839363
		r65	LT839364
		r66	LT839365

Lilac lineage (lineage 5 of Kindler *et al.*, 2013)

l1	LT839180	l1	LT839366
l2	LT839181	l2	LT839367
l3	LT839182	l3	LT839368
l4	LT839183	l4	LT839369
l5	LT839184	l5	LT839370
l6	LT839185	l6	LT839371
l7	LT839186	l7	LT839372
l8	LT839187	l8	LT839373
l9	LT839188	l9	LT839374
l10	LT839189	l10	LT839375
l11	LT839190	l11	LT839376
l12	LT839191	l12	LT839377
l13	LT839192	l13	LT839378
l14	LT839193	l14	LT839379
l15	LT839194	l15	LT839380
l16	LT839195	l16	LT839381
l17	LT839196	l17	LT839382
l18	LT839197	l18	LT839383

• Table S3 continued

Lilac lineage (lineage 5 of Kindler <i>et al.</i> , 2013)			
l19	LT839198	l19	LT839384
l20	LT839199	l20	LT839385
		l21	LT839386
		l22	LT839387
Grey lineage (lineage 7 of Kindler <i>et al.</i> , 2013)			
gy1	LT839200	gy1	LT839388
gy2	LT839201	gy2	LT839389
gy3	LT839202	gy3	LT839390
gy4	LT839203	gy4	LT839391
gy5	LT839204	gy5	LT839392
gy6	LT839205	gy6	LT839393
gy7	LT839206	gy7	LT839394
gy8	LT839207	gy8	LT839395
gy9	LT839208	gy9	LT839396
gy10	LT839209	gy10	LT839397
gy11	LT839210	gy11	LT839398
		gy12	LT839399
		gy13	LT839400
Green lineage (lineage 8 of Kindler <i>et al.</i> , 2013)			
gn1	LT839211	gn1	LT839401
gn2	LT839212	gn2	LT839402
gn3	LT839213	gn3	LT839403
gn4	LT839214	gn4	LT839404
gn5	LT839215	gn5	LT839405
gn6	LT839216	gn6	LT839406
gn7	LT839217	gn7	LT839407
gn8	LT839218	gn8	LT839408
gn9	LT839219	gn9	LT839409
gn10	LT839220	gn10	LT839410
gn11	LT839221	gn11	LT839411
gn12	LT839222	gn12	LT839412
gn13	LT839223	gn13	LT839413
gn14	LT839224	gn14	LT839414
gn15	LT839225	gn15	LT839415
gn16	LT839226	gn16	LT839416
		gn17	LT839417
		gn18	LT839418
		gn19	LT839419
		gn20	LT839420
		gn21	LT839421
Lineage C of Kindler <i>et al.</i> (2013)			
c1	LT839227	c1	LT839422
		c2	LT839423
Lineage F of Kindler <i>et al.</i> (2013)			
f1	LT839228	f1	LT839424

Table S4. Used microsatellite loci. For primer sequences, annealing temperature and multiplex-sets, see Pokrant *et al.* (2016).

Locus	Repeat motif	Allele size range [bp]	Number of alleles	Original reference
Natnat09	(AC) ₂₂	80 - 144	25	Meister <i>et al.</i> (2009)
Natnat05	(GT) ₁₆	136 - 194	23	Meister <i>et al.</i> (2009)
μNt8new	(AC) ₁₅	75 - 123	20	Meister <i>et al.</i> (2009)
Nsμ3	(ATCT) ₁₄ ATC(CA) ₂₀	139 - 457	39	Prosser <i>et al.</i> (1999)
μNt3	(AC) ₁₆	111 - 163	24	Gautschi <i>et al.</i> (2000)
μNt7	(AC) ₁₇	164 - 212	25	Gautschi <i>et al.</i> (2000)
30	(CA) ₁₄	225 - 271	20	Burns & Houlden (1999)
Natnat11	(GA) ₁₃	102 - 228	25	Meister <i>et al.</i> (2009)
Natnat06	(GT) ₂₁	145 - 185	16	Meister <i>et al.</i> (2009)
Tbu A09	(AC) ₇	110 - 146	17	Sloss <i>et al.</i> (2012)
3TS	(GATA) ₁₉	186 - 270	20	Garner <i>et al.</i> (2002)
Eobμ1	(TG) ₂₁	120 - 142	12	Blouin-Demers & Gibbs (2003)
Eobμ13	(AC) ₂₀	118 - 162	19	Blouin-Demers & Gibbs (2003)

Table S5. Genetic diversity of STRUCTURE clusters based on 13 microsatellite loci. n number of individuals, n_A number of alleles, $n_{\bar{A}}$ average number of alleles per locus, n_P number of private alleles, AR allelic richness, H_O average observed heterozygosity, H_E average expected heterozygosity. F_{IS} inbreeding coefficient, F_{ST} fixation index. Individuals with mixed ancestries not considered. All F_{IS} and F_{ST} values were statistically significant.

Microsatellites									
Cluster	n	n_A	$n_{\bar{A}}$	n_P	AR	H_O	H_E	F_{IS}	F_{ST}
First STRUCTURE run									
<i>helvetica</i>	350	156	12.000	45	4.505	0.404	0.548	0.265	0.40
Eastern lineages*	953	230	17.692	119	5.761	0.501	0.612	0.182	
Second STRUCTURE run									
Yellow + red lineage	502	121	9.308	19	7.185	0.453	0.510	0.112	0.18
Adjacent lineages**	100	186	14.308	84	14.006	0.510	0.712	0.285	
Third STRUCTURE run									
Yellow	175	83	6.385	12	6.025	0.423	0.467	0.095	0.11
Red	122	102	7.846	31	7.722	0.501	0.555	0.098	

*Yellow, red, lilac, grey and green lineages

**Lilac, grey and green lineages

Table S6. Simulated data. Twenty samples of each parental group were chosen as pure parental genotypes. Using this data, 20 genotypes of each hybrid class (F_1 , F_2 and the two backcrosses) were modelled in HYBRIDLAB and analyzed with STRUCTURE. Individuals with Q values $\geq 92\%$ were reliably identified as pure *helvetica*, individuals with Q values $\geq 95\%$ as eastern grass snakes. Regarding the yellow and red lineages, the differentiation from backcrosses was more difficult. Although there was a misassignment rate of 5% for the pure yellow lineage, we decided to treat grass snakes with at least 80% cluster membership as pure yellow because otherwise the misidentification rate with backcrosses would be too high. Individuals with at least 83% were treated as pure red.

Parental group 1	Parental group 2	F_1 (referred to parental 1 cluster)	F_1 (referred to parental 2 cluster)	F_2 (referred to parental 1 cluster)	F_2 (referred to parental 2 cluster)	Backcross parental 1	Backcross parental 2	
<i>helvetica</i>		Eastern lineages						
Average Q score	0.965	0.971	0.493	0.507	0.467	0.533	0.772	0.765
SD	0.013	0.007	0.034	0.034	0.095	0.095	0.078	0.066
Minimum Q	0.921	0.954	0.444	0.444	0.328	0.319	0.634	0.661
Maximum Q	0.997	0.979	0.556	0.556	0.681	0.672	0.968	0.912
Misassignment	if $Q < 92\%:$ 0%	if $Q < 95\%:$ 0%	if $Q < 92\%:$ 0%	if $Q < 95\%:$ 0%	if $Q < 92\%:$ 0%	if $Q < 95\%:$ 0%	if $Q < 92\%:$ 5%	if $Q < 95\%:$ 0%
Yellow lineage		Red lineage						
Average Q score	0.880	0.890	0.480	0.520	0.580	0.420	0.740	0.720
SD	0.049	0.033	0.149	0.149	0.192	0.192	0.140	0.148
Minimum Q	0.770	0.825	0.191	0.310	0.206	0.150	0.442	0.353
Maximum Q	0.935	0.940	0.690	0.809	0.850	0.794	0.925	0.917
Misassignment	if $Q < 80\%:$ 5%	if $Q < 83\%:$ 0%	if $Q < 80\%:$ 0%	if $Q < 83\%:$ 0%	if $Q < 80\%:$ 15%	if $Q < 83\%:$ 0%	if $Q < 80\%:$ 40%	if $Q < 83\%:$ 30%

References

- Blouin-Demers, G. & Gibbs, H.L. Isolation and characterization of microsatellite loci in the black rat snake (*Elaphe obsoleta*). *Mol. Ecol. Notes* **3**, 98–99 (2003).
- Burns, E.L. & Houlden, B.A. Isolation and characterization of microsatellite markers in the broad-headed snake *Hoplocephalus bungaroides*. *Mol. Ecol.* **8**, 520–521 (1999).
- Derryberry, E.P., Derryberry, G.E., Maley, J.M. & Brumfield, R.T. HZAR: hybrid zone analysis using an R software package. *Mol. Ecol. Resour.* **14**, 652–663 (2014).
- Fritz, U., Corti, C. & Päckert, M. Mitochondrial DNA sequences suggest unexpected phylogenetic position of Corso-Sardinian grass snakes (*Natrix cetti*) and do not support their species status, with notes on phylogeography and subspecies delineation of grass snakes. *Org. Divers. Evol.* **12**, 71–80 (2012).
- Garner, T.W.J. *et al.* Geographic variation of multiple paternity in the common garter snake (*Thamnophis sirtalis*). *Copeia* **2002**, 15–23 (2002).
- Gautschi, B., Widmer, A. & Koella, J. Isolation and characterization of microsatellite loci in the dice snake (*Natrix tessellata*). *Mol. Ecol.* **9**, 2192–2193 (2000).
- Kabisch, K. *Natrix natrix* (Linnaeus, 1758) – Ringelnatter in *Handbuch der Reptilien und Amphibien Europas. Band 3/IIA: Schlangen II* (ed Böhme, W.), 513–580 (Aula-Verlag, 1999).
- Kindler, C. *et al.* Mitochondrial phylogeography, contact zones and taxonomy of grass snakes (*Natrix natrix*, *N. megalocephala*). *Zool. Scr.* **42**, 458–472 (2013).
- Kindler, C., Bringsøe, H. & Fritz, U. Phylogeography of grass snakes (*Natrix natrix*) all around the Baltic Sea: implications for the Holocene colonization of Fennoscandia. *Amphibia-Reptilia* **35**, 413–424 (2014).
- Kreiner, G. *Die Schlangen Europas* (Chimaira, 2007).
- Meister, B., Armbruster, F.J., Frauenfelder, N. & Bauer, B. Novel microsatellite loci in the grass snake (*Natrix natrix*) and cross-amplification in the dice snake (*Natrix tessellata*). *Mol. Ecol. Resour.* **9**, 604–606 (2009).
- Pokrant, F. *et al.* Integrative taxonomy provides evidence for the species status of the Ibero-Maghrebian grass snake *Natrix astreptophora*. *Biol. J. Linn. Soc.* **118**, 873–888 (2016).
- Prosser, M.R., Gibbs, H.L. & Weatherhead, P.J. Microgeographic population genetic structure in the northern water snake, *Nerodia sipedon sipedon* detected using microsatellite DNA loci. *Mol. Ecol.* **8**, 329–333 (1999).
- Sloss, B.L., Schuurman, G.W., Paloski, R.A., Boyle, O.D. & Kapfer, J.M. Novel microsatellite loci for studies of *Thamnophis* gartersnake genetic identity and hybridization. *Conserv. Genet. Resour.* **4**, 383–386 (2012).