Received: 15 October 2019 | Revised: 18 February 2020 | Accepted: 20 February 2020

DOI: 10.1111/zsc.12417

ORIGINAL ARTICLE

Zoologica Scripta | WILEY

Mitochondrial ghost lineages blur phylogeography and taxonomy of *Natrix helvetica* and *N. natrix* in Italy and Corsica

Nadine Schultze[1] | Cäcilia Spitzweg[1] | Claudia Corti[2] | Michel Delaugerre[3] | Matteo Riccardo Di Nicola[4] | Philippe Geniez[5] | Luca Lapini[6] | Cristiano Liuzzi[7] | Enrico Lunghi[2,8] | Nicola Novarini[9] | Orfeo Picariello[10] | Edoardo Razzetti[11] | Emilio Sperone[12] | Luca Stellati[13] | Leonardo Vignoli[13] | Marika Asztalos[1] | Carolin Kindler[1] | Melita Vamberger[1] | Uwe Fritz[1]

[1]Museum of Zoology (Museum für Tierkunde), Senckenberg Dresden, Dresden, Germany

[2]Sede "La Specola", Museo di Storia Naturale dell'Università di Firenze, Firenze, Italy

[3]Conservatoire du littoral, Bastia, Corsica, France

[4]Milano, Italy

[5]UMR 5175 CEFE, Biogéographie et Ecologie des Vertébrés, EPHE, PSL Université Recherche, Montpellier, France

[6]Sezione Zoologica, Museo Friulano di Storia Naturale, Udine, Italy

[7]Centro Studi de Romita, Bari, Italy

[8]Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing, China

[9]Museo di Storia Naturale di Venezia, Venezia, Italy

[10]Dipartimento di Biologia, Università di Napoli Federico II, Napoli, Italy

[11]Museo di Storia Naturale, Università di Pavia, Pavia, Italy

[12]Dipartimento di Biologia, Ecologia e Scienze della Terra, Università della Calabria, Rende, Italy

[13]Dipartimento di Scienze, Università degli Studi Roma Tre, Roma, Italy

Correspondence
Uwe Fritz, Museum of Zoology (Museum für Tierkunde), Senckenberg Dresden,

Abstract

Grass snakes are widely distributed across the Western Palearctic. Recent phylogeographic studies provided evidence that three distinct parapatric species exist. Two of these occur in Italy, *Natrix helvetica* and *N. natrix*, and a contact zone between both taxa has been suggested for north-eastern Italy. Moreover, previous investigations revealed for the Italian Peninsula a complex phylogeographic structure. Using mtDNA sequences and microsatellite loci, we examined the situation for mainland Italy, Sicily, Sardinia, and Corsica. Our study confirmed the occurrence of *N. natrix* in north-eastern Italy. Cline analyses revealed limited gene flow between *N. helvetica* and *N. natrix* across a narrow hybrid zone. Within *N. helvetica*, conflicting patterns of mitochondrial and nuclear genomic differentiation were revealed. Three nuclear genomic clusters were found; one of them corresponded to no fewer than five distinct and, in part, deeply divergent and ancient mitochondrial lineages from mainland Italy and Sicily. This cluster was paraphyletic with respect to the two remaining mitochondrial lineages, each of which matched with another nuclear genomic cluster (one from Corsica plus Sardinia and another one from western Europe north of the Alps). This unexpected pattern most likely results from mainly male-mediated gene flow and female philopatry combined with population-density-dependent processes such as 'high-density blocking'. With respect to taxonomy, we propose to synonymize *N. h. lanzai* Kramer, 1970 with *N. h. sicula* (Cuvier, 1829), acknowledging their lacking nuclear genomic differentiation. The studied hybrid zone of *N. h. helvetica* and *N. h. sicula* in Italy is wide, with a smooth cline for nuclear markers, supporting their subspecies status. We found no evidence for the distinctiveness of the two subspecies from Corsica (*N. h. corsa*) and Sardinia (*N. h. cetti*), suggesting their synonymy, but refrain from taxonomic conclusions because of small sample sizes and the endangered status of the Sardinian taxon.

Dresden, Germany.
Email: uwe.fritz@senckenberg.de

Funding information
Università della Calabria, Grant/Award Number: 2019, 0011888, 0001255, 3115, 42, 2A-2018-11-20-005 and 2B-2018-11-20-003; Chinese Academy of Sciences

KEYWORDS
cline analysis, female philopatry, male-mediated gene flow, microsatellites, mtDNA

1 | INTRODUCTION

Until recently, grass snakes were believed to represent one polytypic species (*Natrix natrix* sensu lato) with a large distribution range including most of Europe, northern Africa and extending eastwards to Central Asia (Kabisch, 1999). There has been much debate about the taxonomy of the morphologically highly diverse grass snakes, with some authors acknowledging up to 14 different subspecies (Kabisch, 1999; Kramer, 1970; Mertens & Wermuth, 1960). However, Thorpe (1979) recognized only four subspecies based on multivariate statistics of many morphological traits: *Natrix natrix natrix* in the distribution area east of the Rhine, *N. n. helvetica* west of the Rhine, *N. n. cetti* in Sardinia, and *N. n. corsa* in Corsica.

Several recent studies based on mitochondrial and nuclear DNA revealed conflicts between the traditionally recognized subspecies and genetic lineages and clusters (Fritz, Corti, & Päckert, 2012; Guicking, Lawson, Joger, & Wink, 2006; Kindler et al., 2013, 2017; Kindler, Bringsøe, & Fritz, 2014; Kindler, de Pous, et al., 2018; Kindler & Fritz, 2018; Kindler, Graciá, & Fritz, 2018; Pokrant et al., 2016), leading to the recognition of three distinct species within what was formerly known as *Natrix natrix* (Kindler et al., 2017; Pokrant et al., 2016), namely the common or eastern grass snake *N. natrix* (Linnaeus, 1758), the red-eyed grass snake *N. astreptophora* (Seoane, 1884), and the barred grass snake *N. helvetica* (Lacepède, 1789).

Natrix astreptophora is distributed across northern Africa and the Iberian Peninsula plus adjacent southern France. *Natrix helvetica* inhabits western Europe from the Pyrenees to the Rhine region. It also occurs on the Italian Peninsula, Sicily, Corsica, and Sardinia. *Natrix natrix* occurs eastwards of the Rhine region to Central Asia including Fennoscandia, the Balkans, and Asia Minor. *Natrix helvetica* and *N. natrix* meet in the Rhine region and hybridize there in a narrow contact zone, which has been studied in detail by Kindler et al. (2017).

Based on morphological and genetic differences, five subspecies of *N. helvetica* are currently recognized (Kindler & Fritz, 2018): The nominotypical subspecies *N. h. helvetica* (Lacepède, 1789) lives in western Europe (France, Britain, Benelux countries, western Germany, and Switzerland). Corsica and Sardinia are each inhabited by one subspecies, namely *N. h. cetti* Gené, 1839 in Sardinia and *N. h. corsa* (Hecht, 1930) in Corsica. The subspecies *N. h. sicula* (Cuvier, 1829) occurs in southern Italy (Calabria) and Sicily, while the remaining Italian Peninsula is inhabited by *N. h. lanzai* Kramer, 1970. This high subspecific diversity matches the complex biogeographic history of Italy. The Italian Peninsula plus the adjacent islands represent one of the major Mediterranean refuges and differentiation centres (Pedall, Fritz, Stuckas, Valdeón, & Wink, 2011; Schmitt, 2007; Vamberger et al., 2015) that has been shaped by ongoing complex geological transformations. Plate tectonics induced Alpine and Apennine orogeny; repeated sea-level fluctuations, associated with Pleistocene climatic cycles, resulted in submergence and recurrent shifts of the coastline, as well as island genesis (Marchetti, Soldati, & Vandelli, 2017). In particular, Corsica and Sardinia, which have been separated from the mainland by the Tyrrhenian Sea for most of their geological history, exhibit a high level of endemicity (Grill, Casula, Lecis, & Menken, 2007; Médail & Quézel, 1999; Salvi, Harris, Bombi, Carretero, & Bologna, 2010). This also applies to grass snakes in Corsica and Sardinia, which have long been known to be morphologically highly distinct (Hecht, 1930; Kabisch, 1999; Mertens & Wermuth, 1960). Based on their morphological distinctiveness, it was suggested that these grass snakes constitute the distinct species *Natrix cetti* with the two subspecies *N. cetti cetti* in Sardinia and *N. cetti corsa* in Corsica (Vanni & Cimmaruta, 2011). However, several studies based on mitochondrial and nuclear genetic evidence rejected this idea and assigned Corso-Sardinian grass snakes as subspecies to *N. helvetica* (Fritz et al., 2012; Kindler et al., 2013, 2017; Kindler, de Pous, et al., 2018; Kindler & Fritz, 2018).

With respect to mtDNA, the barred grass snake contains six distinct genealogical lineages, all of which occur in Italy, and four of them being endemic to Italy (Kindler et al., 2013, 2017; Kindler & Fritz, 2018). These lineages are not completely consistent with the currently recognized subspecies (Table 1).

Fossil-calibrated molecular clock calculations by Kindler, de Pous, et al. (2018) indicated that mitochondrial diversification within the barred grass snake commenced approximately 6.8 million years ago, with lineage A from Sicily and Calabria as the oldest branch. Lineage B, distributed in Corsica and Sardinia, with 3.9 million years is the second oldest lineage. Lineage C, occurring in the Po Plain in

TABLE 1 Subspecies and mitochondrial lineages of *N. helvetica* with their distribution ranges according to Kindler and Fritz (2018)

Subspecies	mtDNA lineage	Distribution
N. h. helvetica (Lacepède, 1789)	E	Western Europe from the Pyrenees to the Rhine region, Britain
N. h. cetti Gené, 1839	B	Sardinia
N. h. corsa (Hecht, 1930)	B	Corsica
N. h. lanzai Kramer, 1970	C, D, F	Mainland Italy except southern Calabria
N. h. sicula (Cuvier, 1829)	A	Southern Calabria, Sicily

northern Italy and in adjacent Switzerland, split from the remaining lineages 2.8 mya. Representatives of this lineage have recently also been recorded in Bavaria (southern Germany), suggesting Holocene transalpine dispersal (Glaw, Franzen, Oefele, Hansbauer, & Kindler, 2019). The youngest intraspecific divergence (approx. 0.3 million years) refers to the split between lineages E, corresponding to *N. h. helvetica*, and F of *N. h. lanzai*.

Many studies have emphasized that mitochondrial markers alone are insufficient for examining evolutionary processes of bisexual species because mtDNA is generally inherited in the maternal line, making the parallel application of biparentally inherited nuclear genomic markers mandatory (Ballard & Whitlock, 2004; Renoult, Geniez, Bacquet, Benoit, & Crochet, 2009; Thielsch, Knell, Mohammadyari, Petrusek, & Schwenk, 2017; Toews & Brelsford, 2012). Using 12 microsatellite loci combined with mtDNA, Kindler and Fritz (2018) obtained conflicting results with respect to the taxonomic differentiation of *N. helvetica*. These authors found a mismatch between mitochondrial lineages and nuclear genomic clusters on the Italian Peninsula. Two old and deeply divergent mitochondrial lineages (C, F) were not differentiated using STRUCTURE analyses and PCAs of microsatellite data. In contrast, grass snakes from Corsica and Sardinia shared the same mitochondrial lineage. Using microsatellite data for STRUCTURE analyses, the samples from the two islands were lumped together, but using PCAs, two distinct clusters were obtained. However, this study was compromised by small sample sizes and patchy coverage of the distribution area. One mitochondrial lineage (D) from mainland Italy was represented only by a single previously published GenBank sequence (Guicking et al., 2006), and microsatellite data were available for only two Sardinian and three Corsican snakes. In addition, grass snakes from north-eastern Italy, where *N. natrix* has been reported to occur (Lanza, 1983; Lapini, Dall'Asta, Bressi, Dolce, & Pellarini, 1999; Thorpe, 1979), were not studied by Kindler and Fritz (2018).

For the present study, we used an extended sampling from Italy and Corsica and reanalysed the genetic differentiation of grass snakes in this region. We followed the approach of Kindler et al. (2017) and employed 13 microsatellite loci and mtDNA sequences of up to 1,984 bp length. In particular, we aimed at answering the following questions: (a) Do genetic data confirm the presence of a contact zone between *N. helvetica* and *N. natrix* in north-eastern Italy and if so, to which extent do hybridization and gene flow occur between these two species? (b) Does the nuclear genomic differentiation of grass snakes in Italy and Corsica match that of their mitochondrial lineages? (c) Are grass snakes from Corsica genetically distinct from those in Sardinia? And (d) do our findings alter the taxonomy of Italian and Corsican grass snakes?

2 | MATERIALS AND METHODS

2.1 | Sampling

Samples of 190 grass snakes were analysed as the core data set of the present study. Of these, 100 samples were newly processed, and data for the remaining 90 grass snakes had been generated in our laboratory for previous studies (Fritz et al., 2012; Guicking et al., 2006; Kindler et al., 2013, 2017; Kindler, de Pous, et al., 2018; Kindler & Fritz, 2018; Pokrant et al., 2016). One hundred and forty-three samples originated from Italy and Corsica. For some analyses, 47 samples were added from adjacent areas (France, Switzerland, Slovenia, Croatia) to put the Italian samples into a broader framework. In addition, microsatellite data of 28 samples from France, generated by us for a yet unpublished study, were included for some STRUCTURE and cline analyses, and mtDNA sequences of another 38 samples, representing mitochondrial lineages of grass snakes from outside the study region, were added for phylogenetic tree calculations. A detailed list of all samples is provided in Tables S1 and S2. The 100 new samples (mainly tissue samples, some buccal swabs and shed skins) were processed according to laboratory procedures described in Kindler et al. (2013) and Pokrant et al. (2016).

For the phylogeographic analyses, two different marker systems were employed, which were previously used in other studies on grass snakes (Kindler et al., 2017; Kindler, de Pous, et al., 2018; Kindler & Fritz, 2018; Pokrant et al., 2016). Two mtDNA fragments were sequenced, one coding for part of the ND4 gene and adjacent tRNAs (tRNA-His, tRNA-Ser, tRNA-Leu; in total up to 867 bp), and the cytochrome *b* (cyt *b*) gene (up to 1,117 bp). For 12 samples, no mtDNA sequences could be obtained, and for some, only

one fragment could be sequenced. Sometimes, only incomplete sequences were obtained (ND4 + tRNAs: 313–716 bp, $n = 14$; cyt *b*: 372–1,045 bp, $n = 11$; Table S1).

As a second marker system, a set of 13 polymorphic microsatellite loci was applied for genotyping the samples. Microsatellite data could not be obtained for 30 samples. For detailed information on microsatellite loci, primers, and PCR conditions for mtDNA and microsatellites, see Tables S3 and S4, Kindler et al. (2013, 2017), and Pokrant et al. (2016).

2.2 | Mitochondrial sequence analyses and phylogenetic trees

Alignments of mtDNA sequences were compiled using BIOEDIT 7.0.5.2 (Hall, 1999). For phylogenetic tree calculations, a concatenated alignment of 1,984 bp length was used. Phylogenetic trees were built using Bayesian and Maximum Likelihood approaches. For doing so, three sequences for each terminal clade of *Natrix natrix*, *N. helvetica* and *N. astreptophora*, as identified by Kindler et al. (2013, 2017) and Kindler, de Pous, et al. (2018), were added to the alignment. Sequences of *N. maura*, *N. tessellata* and *Nerodia sipedon* were included as outgroups, using *N. sipedon* for tree rooting (Table S2). To determine the optimal partitioning scheme and evolutionary models, PARTITIONFINDER 2 (Lanfear, Frandsen, Wright, Senfeld, & Calcott, 2016) was run with linked branch lengths, the 'search = all' and 'models = all' options, and the Bayesian information criterion (BIC). This resulted in the selection of seven partitions, with each codon position of the two coding genes as one partition and the tRNAs merged in an extra partition. For the Maximum Likelihood (ML) approach, RAxML 7.2.8 (Stamatakis, 2006) was run with the default GTR + G model for all partitions. Five independent fast bootstrap algorithm searches were conducted to compute ML trees, starting from distinct randomized maximum parsimony trees. Subsequently, thorough bootstrap replicates were run until they converged with a cut-off of 1%. Convergence occurred after 2,900 replicates, and the resulting values were plotted against the best-scoring ML tree. In addition, Bayesian analyses were performed in MRBAYES 3.2.5 (Ronquist et al., 2012), using the best-fit models for each partition as determined by PARTITIONFINDER 2 (see Table S5). Two simultaneous independent runs were executed, each with four chains that ran for 10 million generations with every 500th generation sampled. Convergence was confirmed by average standard deviations of split frequencies approaching zero. A burn-in of 2.5 million generations was used to generate the final 50% majority-rule consensus tree. The mitochondrial clade for each new sample was identified using the tree results.

Alignments for each mtDNA block (only complete sequences) were then further examined using the TCS parsimony network algorithm (Clement, Posada, & Crandall, 2000) as implemented in POPART (http://popart.otago.ac.nz) to create haplotype networks. Previously identified haplotypes (Kindler et al., 2017, 2018) were added to the new Italian samples for network construction, to incorporate the full diversity of each mtDNA lineage.

2.3 | Microsatellite analyses

A set of 13 previously used microsatellite loci (Kindler et al., 2017; Kindler, de Pous, et al. 2018; Kindler & Fritz, 2018; Kindler, Graciá, et al., 2018; Pokrant et al., 2016) was subjected to unsupervised cluster analyses in STRUCTURE 2.3.2.1 (Falush, Stephens, & Pritchard, 2003; Pritchard, Stephens, & Donelly, 2000). STRUCTURE analyses were adjusted for the presence of null alleles because MICRO-CHECKER 2.2.3 (van Oosterhout, Hutchinson, Wills, & Shipley, 2004) detected null alleles for some loci. STRUCTURE was run using the admixture model and correlated allele frequencies. The Monte Carlo Markov chains ran for 1,000,000 generations with a burn-in of 250,000. Calculations were repeated ten times each for *K*s ranging from 1 to 10. As it is a known limitation of STRUCTURE to reveal only the uppermost level of genetic structuring (Evanno, Regnaut, & Goudet, 2005; Janes et al., 2017), a stepwise hierarchical approach was applied, with increasingly pruned data sets in consecutive STRUCTURE analyses (Evanno et al., 2005; Kalinowski, 2011; Puechmaille, 2016). For each round, the optimal number of clusters was estimated using the ΔK method (Evanno et al., 2005) as implemented in STRUCTURE HARVESTER (Earl & vonHoldt, 2012). STRUCTURE results were visualized in DISTRUCT 1.1 (Rosenberg, 2004). The hybrid status of individual samples of each round was inferred using simulations in HYBRIDLAB 1.0 (Nielsen, Bach, & Kotlicki, 2006).

Our first STRUCTURE analysis included all 160 samples. For HYBRIDLAB, 15 pure individuals of the local genetic lineage of each species (*Natrix helvetica*, *N. natrix*) that meet in north-eastern Italy were selected as parental genotypes (Table S1). Using these data, 15 genotypes of each hybrid class (F_1, F_2, backcrosses) were modelled. The HYBRIDLAB results were then subjected to STRUCTURE analyses for inferring the threshold for the best identification of pure individuals, F_1 and F_2 hybrids, and backcrosses, resulting in values of 93% and 91% with respect to *N. natrix* and *N. helvetica* (Table S6).

For a second STRUCTURE analysis, all samples of *N. natrix* and of all other grass snakes with genetic impact from *N. natrix* were excluded, so that only 112 samples of *N. helvetica* remained. To estimate the hybrid threshold for genetic lineages within *N. helvetica*, all 14 available Corso-Sardinian grass snake samples served as one parental group and 15 samples of the remaining snakes as the second. For each hybrid class, 15 genotypes were modelled and the simulated

data were processed in STRUCTURE, resulting in a threshold of 73% for snakes from Corsica and Sardinia and 90% for the others (Table S7).

For a third STRUCTURE round, the subset containing only samples of *N. helvetica* was further pruned. All those samples that corresponded to, or showed genetic impact of, the Corso-Sardinian cluster were removed. However, because STRUCTURE is known to be sensitive to uneven sample sizes (Puechmaille, 2016), additional pure samples of *N. h. helvetica* (lineage E) from south-eastern France were added to increase the number of individuals for this subspecies, yielding a total sample size of 120. To infer hybrid status for this round, 25 *N. h. helvetica* from France and Switzerland and 25 grass snakes from mainland Italy and Sicily were processed in HYBRIDLAB, and 25 genotypes for each hybrid class were modelled (Table S8). The resulting thresholds were 80% for *N. h. helvetica* and 86% for the grass snakes from mainland Italy and Sicily.

A final exploratory fourth STRUCTURE analysis included only those 49 samples from mainland Italy and Sicily that did not belong to the *N. h. helvetica* cluster and that showed no genetic impact from this cluster.

In addition to the STRUCTURE analyses, principal component analyses (PCAs) were performed using the R package ADEGENET (Jombart, 2008). STRUCTURE analyses are based on population genetic presumptions (Hardy–Weinberg equilibrium, linkage disequilibrium; Pritchard et al., 2000) and therefore prone to bias from uneven sample sizes. To validate STRUCTURE results, it is thus advisable to run additional PCAs that are less sensitive to sample size differences (Puechmaille, 2016). PCAs were run for the same four data sets as used for STRUCTURE.

2.4 | Cline analyses

Cline analyses are a powerful tool for understanding hybridization and genetic introgression between distinct taxa, in which the manner of transition of distinct character states, in our case, genetic identity, across a contact zone is analysed. In taxa with good dispersal capacity, wide and shallow clines indicate extensive gene flow and conspecificity. In contrast, narrow and steep clines are characteristic for incipient or complete reproductive isolation and suggest species status for the involved taxa (Barton, 1979; Derryberry, Derryberry, Maley, & Brumfield, 2014; Rieseberg, Whitton, & Gardner, 1999). Our cline analyses were calculated with the R package HZAR (Derryberry et al., 2014) for grass snakes from two contact zones in Italy using their mitochondrial identity and the STRUCTURE results for microsatellite data (*Q* values).

Two transects were selected. The first ran in west-east direction through north-eastern Italy (625 km), covering the contact zone between *Natrix helvetica* and *N. natrix* in north-eastern Italy. The second transect ran from south-eastern France through north-western Italy southwards to the region of Rome (1,300 km), covering the transition between the two microsatellite clusters of *N. helvetica*. For this second analysis, an extended sampling was used that included additional material from France to enlarge the number of pure *N. h. helvetica* samples. Using QGIS 3.4.5 (http://qgis.org), samples were pooled along the first transect into sections of 25 km length, including samples within 55 km distance on both sides of the transect. For the second transect, sections of 50 km length were chosen to pool samples along the transect, and samples within 100 km left and right of the transect were included. One sample of lineage G from Tuscany (MTD T 11977) was excluded from the cline analysis, as it is most likely non-native and introduced in this area. However, a sample of lineage C from Liguria (MSNVE 684) was included because grass snakes of this lineage are known to have crossed even the Alps (Glaw et al., 2019; Kindler & Fritz, 2018), and it is possible that lineage C dispersed naturally across the relatively low Apennine chain. For each section of the transects, the mean *Q* value of cluster membership (as determined by STRUCTURE) was calculated using microsatellite data. For mitochondrial data, the frequency of haplotypes of *N. helvetica* (north-eastern contact zone) or of lineage E (north-western contact zone) was calculated, opposed to haplotypes of *N. natrix* or other lineages of *N. helvetica*, respectively. These four data sets (two contact zones, mtDNA and microsatellite data for each) were then processed independently using the approach described in Kindler et al. (2017); the best cline model for each marker system was selected by the lowest AIC score (Table S9).

3 | RESULTS

3.1 | Mitochondrial differentiation in Italian and Corsican grass snakes

3.1.1 | Phylogenetic trees

Largely identical topologies were obtained under both tree building approaches (Figure 1). The only difference was that the Bayesian tree placed the well-supported clade E in an unresolved polytomy among sequences of lineage F, while the ML tree showed a distinct clade F.

In accordance with previous studies (Fritz et al., 2012; Guicking et al., 2006; Kindler et al., 2013; Kindler, de Pous, et al., 2018), the deeply divergent *Natrix astreptophora* constituted the sister taxon of a crown group formed by *N. natrix* and *N. helvetica*. Within *N. natrix*, all eight lineages identified by Kindler et al. (2013) were recovered. Many sequences from north-eastern Italy occurred in the same clade as haplotypes of lineage 4 of Kindler et al.

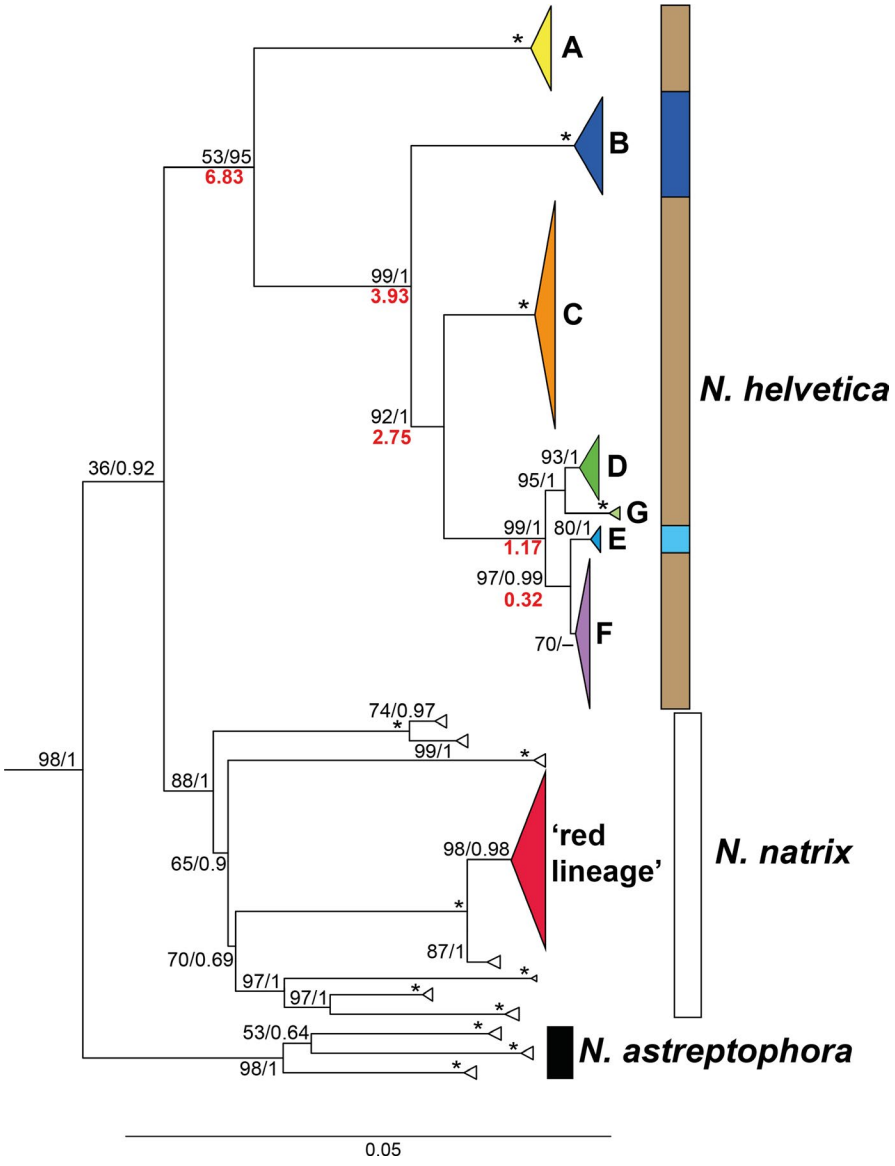

FIGURE 1 Maximum Likelihood tree based on mtDNA sequences of 1,984 bp length (cyt *b*, ND4 + tRNAs). Outgroups (*Nerodia sipedon*, *Natrix maura*, *N. tessellata*) removed for clarity. Terminal clades are collapsed to cartoons, coloured cartoons are clades present in Italy and Corsica. Black numbers at nodes are ML bootstrap values (3,000 bootstrap replicates) and Bayesian posterior probabilities. Asterisks (*) indicate maximum support under both approaches. Red numbers are divergence time estimates in million years from Kindler, de Pous, et al. (2018). The coloured bar for *N. helvetica* on the right shows congruence of mitochondrial clades and nuclear genomic clusters

(2013), the so-called 'red lineage' of *N. natrix* (Kindler & Fritz, 2018; Kindler, Graciá, et al., 2018), confirming that this species occurs in north-eastern Italy. A previously unknown clade (G) was discovered for *N. helvetica*, which was sister to clade D.

3.1.2 | Haplotype networks

All mtDNA clades were reflected in the parsimony networks as distinct haplotype clusters, both for ND4 + tRNAs and for the cyt *b* gene. The networks shown in Figure 2 exclude *Natrix natrix* because the sequences of Italian samples represented, besides previously identified haplotypes, only two new tip haplotypes for ND4 + tRNAs (r34, r35) that differed by one mutation step from the common haplotypes r4 and r16 (Table S10; see also the network in Kindler, Graciá, et al., 2018).

For ND4 + tRNAs of *N. helvetica* (Figure 2, top), a star-like cluster of 12 haplotypes of lineage E was located in the interior of the network. It was connected by a minimum of four mutation steps to the cluster D, and by a minimum of two steps to cluster F. Cluster G differed from clusters E and F each by a minimum of six steps. Cluster C differed from both clusters E and F by a minimum of 18 steps and from cluster G by 22 steps. Cluster A was connected to clusters B and E and differed by a minimum of 34 steps from cluster B and by a minimum of 41 steps from cluster E. Clusters B and E differed by a minimum of 19 steps. Cluster D comprised three haplotypes that differed by a maximum of two mutations, like the individual haplotypes of cluster E. Cluster F and the newly identified cluster G consisted each of two haplotypes that differed by one mutation. One of the haplotypes of cluster G was misinterpreted by Kindler and Fritz (2018) as belonging to clade F. Cluster C comprised three haplotypes that differed by a maximum of three mutations. Cluster A was formed by

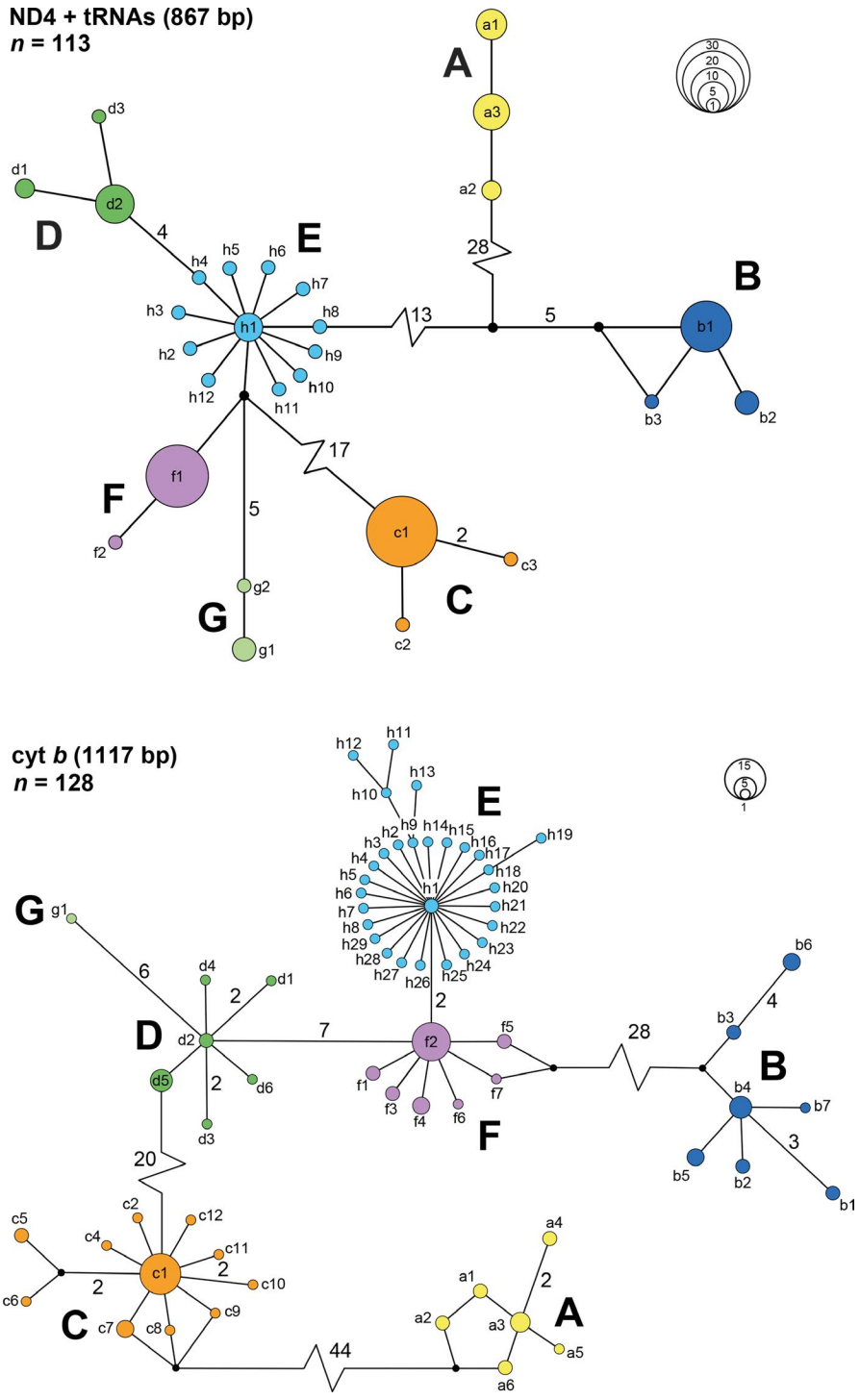

FIGURE 2 Parsimony networks of mtDNA sequences of *Natrix helvetica*. Symbol sizes correspond to number of samples per haplotype. Lines connecting haplotypes represent one mutation step, if not otherwise indicated by numbers along lines. Small black circles correspond to missing node haplotypes; colour coding according to mtDNA lineages (Figure 1)

three haplotypes differing from each other by a maximum of two steps. Cluster B contained a loop and consisted of three haplotypes that differed by a maximum of two mutations. One haplotype from each of the clusters B and F was newly identified in the present study (b3, f2), and two haplotypes from each of the clusters C, D, and G were new (c2, c3, d2, d3, g1, g2; European Nucleotide Archive accession numbers LR721654–LR721659, LR722601, LT900415).

A more complex network, with more haplotypes, was obtained for the cyt *b* gene (Figure 2, bottom). The network contained loops for three clusters (A, C, F), indicating alternative mutation pathways. Lineage A constituted the most distinct haplotype cluster that was connected over a minimum of 46 steps with cluster C. The latter cluster differed by a minimum of 20 mutations from cluster D, which in turn was connected by six steps with the only cyt *b* haplotype of lineage G. In addition, cluster D was connected by seven steps with cluster F, and cluster F by two steps with cluster E and by a minimum of 30 steps with cluster B. Cluster A comprised six haplotypes that differed by a maximum of four

mutations. Cluster C contained 11 haplotypes that differed by a maximum of five mutations. Cluster D comprised six haplotypes with a maximum of four mutations among them. Cluster F had five haplotypes differing by a maximum of two mutations. The star-like cluster E contained 29 haplotypes that differed from each other by a maximum of five steps, even though the majority of haplotypes differed only by one mutation from the central haplotype h1. Cluster B contained seven haplotypes that differed by a maximum of nine mutations. The following haplotypes were newly identified in the present study: a5, a6, b6, b7, c6-c12, d2-d6, f6, f7, and g1 (European Nucleotide Archive accession numbers LR722580–LR722598).

Additional information on uncorrected *p* distances between the haplotype clusters, and with respect to *N. natrix* from the study region, and information on variable and parsimony-informative sites can be found in the Supporting Information (Table S11).

3.1.3 | Geographic distribution of mitochondrial clades

In easternmost northern Italy (Friuli-Venezia Giulia, Veneto), haplotypes of the 'red lineage' of *Natrix natrix* (Kindler et al., 2013, 2017) were recorded (Figure 3). Previously, *N. natrix* was only known from east of the Piave River (Lapini et al., 1999). We found several records for the red lineage west of the Piave River, near Treviso and Venice; the westernmost records are far beyond the Piave River, from close to the Po estuary and Ferrara (Emilia-Romagna). Except for one record of clade C of *N. helvetica* in the region of Udine, amidst records of the red lineage, it is obvious that the mitochondrial haplotypes of clade C are largely parapatrically distributed, and the contact zone of the red lineage and clade C largely matches in the north with the Piave River, even though our new records reveal that it extends further south-westwards and beyond the Po River.

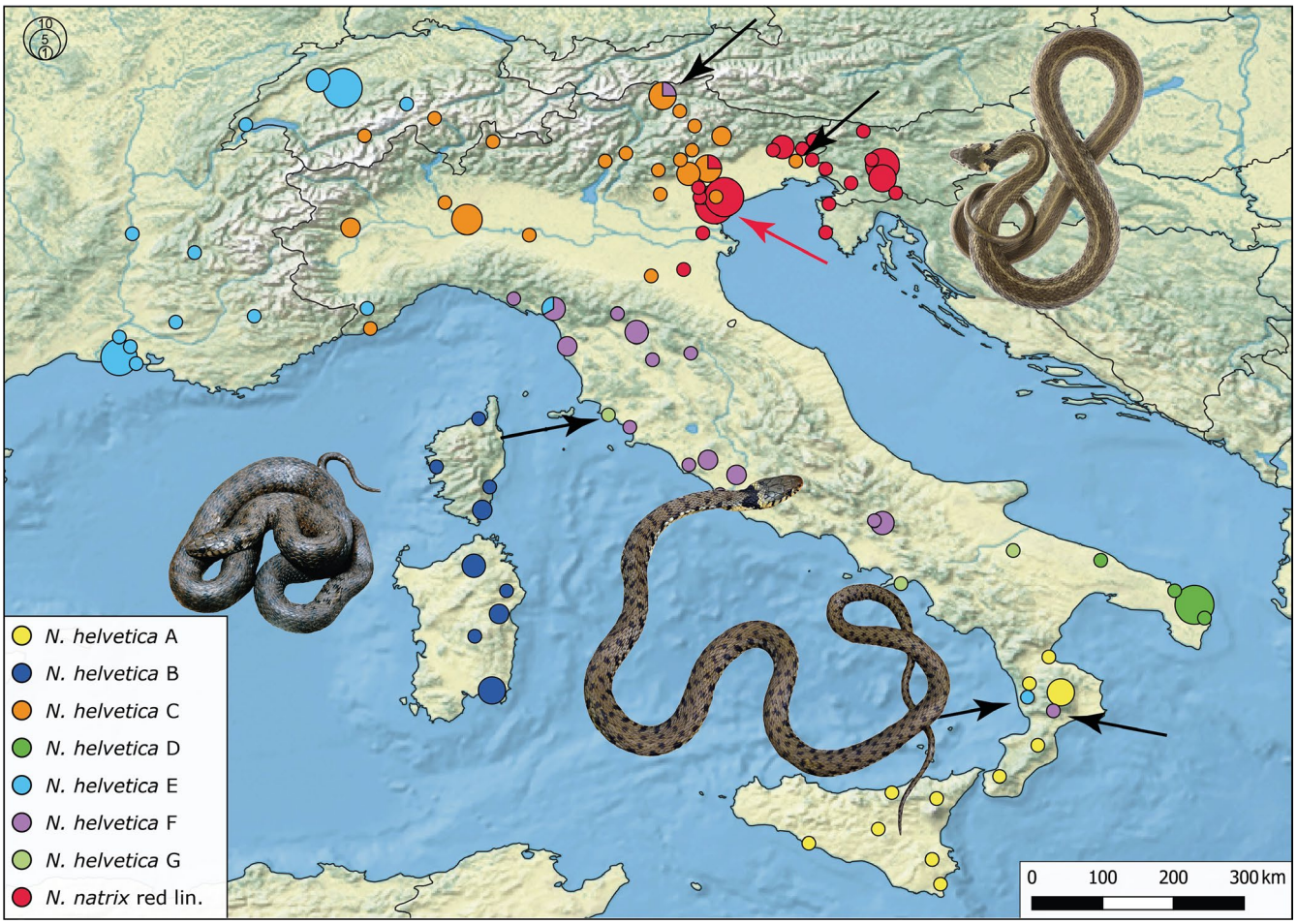

FIGURE 3 Geographic distribution of mtDNA clades for 172 grass snakes. Samples from neighbouring sites were pooled; circle sizes indicate number of samples per site. Black arrows highlight records representing most likely translocated grass snakes. The red arrow indicates the approximate location of the Piave River. Colour coding for lineages as in Figure 1. Snake photographs show from left to right: *Natrix helvetica corsa* (Galéria, Corsica; photo: P. Geniez), *N. h. sicula* (Altamura, Bari, Apulia; photo: M. Di Nicola), and *N. natrix* (striped morphotype, Isola della Cona, Gorizia, Friuli-Venezia Giulia; photo: M. Di Nicola)

Except for the very east, clade C of *N. helvetica* is widely distributed in northern Italy and occurs in the Padan Plain and along the southern slope of the Alps. Grass snakes of this clade crossed the Alps and occur in Switzerland (Kindler & Fritz, 2018) and Bavaria (Glaw et al., 2019). Another record of clade C (Taggia, Liguria) lies along the Mediterranean coast, beyond the Po Plain and the Apennine Mountains, in close proximity to a record of clade E of *N. helvetica*. In the face of the dispersal ability of grass snakes across Alpine passes (Glaw et al., 2019; Kindler & Fritz, 2018), this record could represent a natural occurrence. Clade E, widely distributed north-west of the Alps in western Europe, was also recorded from another sample from Tuscany (Carrara), in close proximity of grass snakes representing clade F. Clade F is widely distributed and was recorded, besides Liguria and Tuscany, from Campania and Lazio. An isolated record of clade F from Trentino-Alto Adige/Südtirol lies among records for clade C. Even from the same site (Mules, Campo di Trens = Mauls, Freienfeld), another sample corresponded to clade C, suggestive of a translocated individual. Likewise, another completely isolated record for clade F from Calabria seems questionable (see below). Two records of the newly discovered clade G are from Campania and Basilicata, and amidst records of clade F, another record of clade G was found in Tuscany. Clade D, previously only known from a single individual from southern Apulia (Guicking et al., 2006), was confirmed by several new records from this region (Salentine Peninsula). All samples from Sicily and most from Calabria corresponded to clade A. For Calabria, also one grass snake each harbouring clade F and E was found. All samples from Corsica and Sardinia represented clade B.

3.2 | Nuclear DNA structuring and admixture

3.2.1 | STRUCTURE analyses

Four different STRUCTURE analyses were executed. Since STRUCTURE is known to often find only the uppermost hierarchical level of genetic clustering (Evanno et al., 2005; Janes et al., 2017; Kalinowski, 2011), the software was run for four data sets to examine for hidden differentiation.

For the first analysis representing the uppermost hierarchical level (Figure 4a), all 160 samples were included. The ΔK method identified $K = 2$ as the optimal number of clusters (Figure S1). One cluster corresponded to *Natrix helvetica*, and the other, to *N. natrix*. A restricted amount of hybridization between the two species was detected for the north-eastern Italian Veneto region. In addition, two genotypically pure *N. natrix* from there had introgressed mitochondrial haplotypes of *N. helvetica* (Table S1: MSNVE 24716, MSNVE 24717).

For the second analysis (Figure 4b), all samples of pure *N. natrix* or with genetic impact from *N. natrix* were removed to examine structuring within *N. helvetica*. For this data set ($n = 112$), the ΔK method revealed three as the optimal number of clusters (Figure S2). One cluster corresponded to all samples from Corsica and Sardinia (mitochondrial clade B); the second, to samples from north-western Italy and adjacent France and Switzerland (mitochondrial clade E); and the third, to samples from the Italian Peninsula and Sicily (remaining mitochondrial clades of *N. helvetica*). Many grass snakes from north-western Italy and north and west of the Alps had hybrid signatures, reflecting admixture between the north-western cluster (mitochondrial clade E) and the cluster from mainland Italy and Sicily. Grass snakes from Corsica and Sardinia showed no genetic impact from any other clusters, while for a few snakes from southern France and southern Italy admixture with Corsican and Sardinian snakes was inferred. In view of the geographic location of the respective collection sites, this is difficult to explain and could represent a methodological artefact.

The third analysis examined structuring within the Italian Peninsula and Sicily and excluded all samples of the Corso-Sardinian cluster and the few mainland samples with putative impact from Corso-Sardinian grass snakes. Since STRUCTURE is known to be sensitive against uneven sample sizes (Puechmaille, 2016), additional samples representing *N. h. helvetica* from south-eastern France (mitochondrial clade E) were added to achieve comparable sample sizes for the individual mitochondrial lineages, resulting in a total number of 120 processed samples. The ΔK method revealed two as the optimal number of clusters (Figure S3). Despite the removal of the Corso-Sardinian cluster and its putative hybrids, no further structuring was detected and the results for the Italian Peninsula and Sicily resembled the second STRUCTURE analysis (Figure 4c). However, it is noteworthy that genetic impact from the peninsular cluster was traceable deep into the Rhône region of southern France.

For the fourth STRUCTURE analysis (Figure S4), only *N. helvetica* samples from mainland Italy and Sicily were included that did not belong to the north-western cluster corresponding to *N. h. helvetica* and that showed no genetic impact from this cluster. For this data set ($n = 49$), the ΔK method favoured a subdivision in two clusters again (Figure S5). However, all samples then had mean proportions for cluster membership (Q values) close to 0.5. The results were obviously geographically and biologically meaningless (Figure S4), indicating that the samples from mainland Italy and Sicily represent only one collective cluster. This was also corroborated by the log probability values for K (Figure S5), with the highest value for $K = 1$.

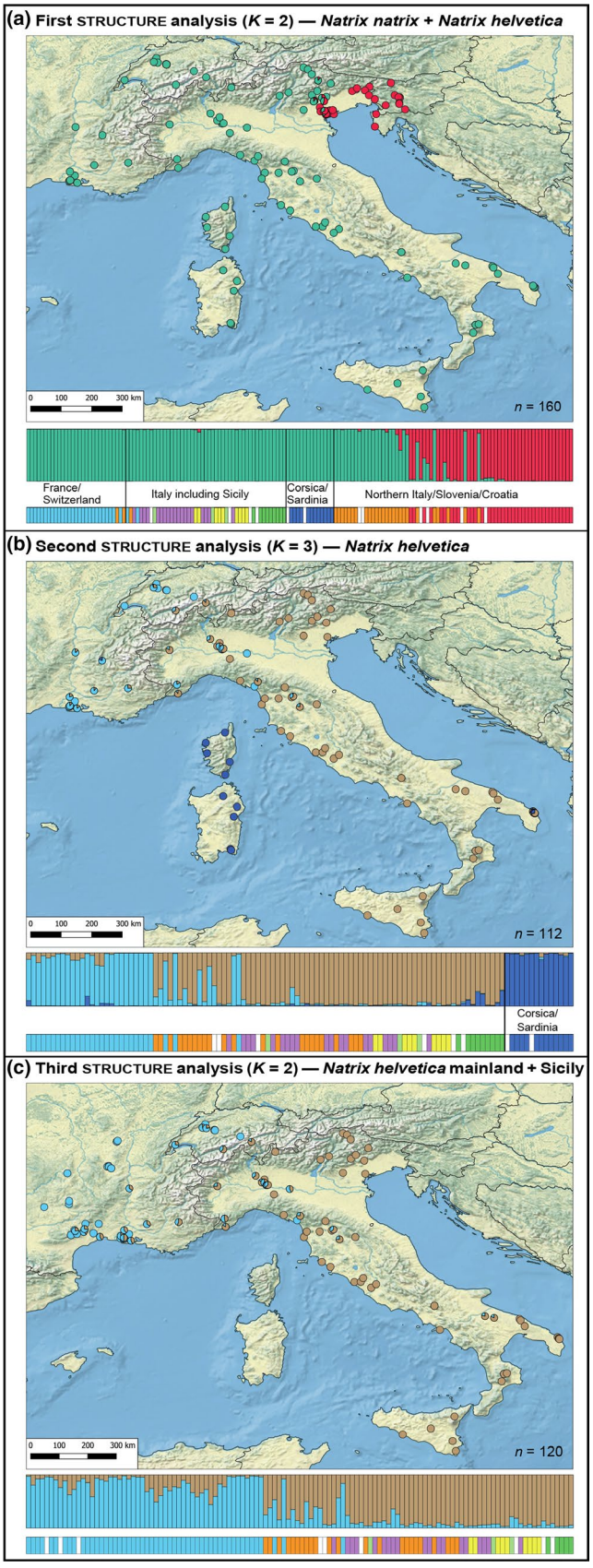

FIGURE 4 Genotypic structuring of grass snakes in Italy and Corsica and the adjacent mainland. Symbol colours indicate STRUCTURE cluster membership; admixed ancestries are depicted as pie chart sectors according to *Q* values. In the bar plots below the maps, each sample is represented by two bars, indicating its inferred cluster membership (upper bar) and its mitochondrial lineage (lower bar). Mitochondrial lineages colour-coded as in Figures 1–3; white bars indicate missing mtDNA data. Samples are arranged from west to east within the individual regions as defined in the bar plots

Additional information on variation of the studied microsatellite loci and population genetic parameters of the STRUCTURE clusters can be found in the Supporting Information (Tables S12–S16).

3.2.2 | Principal Component Analyses (PCAs)

The PCAs supported our STRUCTURE results, with respect to both the cluster numbers and admixed individuals. Snakes identified as admixed in STRUCTURE analyses were also generally intermediate between PCA clusters. In the PCA using the whole data set, *Natrix natrix* and continental *N. helvetica* were distinct and corresponded to two clusters. Grass snakes from Corsica and Sardinia constituted a third cluster that was distinct using the first two principal components (Figure S6a). The cluster assignment of the few hybrid individuals (*N. helvetica* × *N. natrix*) suggests that gene flow is bidirectional, with approximately the same number of hybrids clustering with each parental species. The only two individuals with cytonuclear discordance were two genotypically pure *N. natrix* with introgressed mitochondrial haplotypes of *N. helvetica*. In the PCA for *N. helvetica* only (Figure S6b), a new tripartite division emerged. Corso-Sardinian grass snakes again formed a distinct cluster. However, using the first two principal components, the other *N. helvetica* were divided into two clusters. One corresponded to the nominotypical subspecies *N. h. helvetica* (mitochondrial lineage E), and the other cluster to the remaining Italian grass snakes. Another PCA without Corso-Sardinian grass snakes but with additional data for the nominotypical subspecies (Figure S6c) confirmed, in perfect agreement with the STRUCTURE analyses (Figure 4), not only the differentiation but also extensive hybridization between *N. h. helvetica* and barred grass snakes from mainland Italy plus Sicily. The last PCA using only *N. helvetica* from mainland Italy plus Sicily without genetic impact from *N. h. helvetica* corroborated the absence of phylogeographic structuring across these regions (Figure S6d).

3.3 | Cline analyses

The first transect (625 km length) runs through north-eastern Italy from west to east across the Italian contact zone of *Natrix helvetica* and *N. natrix*. For both marker systems, the cline from *N. helvetica* to *N. natrix* is steep and with very similar cline centres and widths (Figure S7, left). For microsatellite data, the cline centre is 335.5 km (95% confidence

interval 282.8–352.8 km) from the reference locality used as the starting point. The cline width is 87.6 km (95% confidence interval 36.2–293.3 km). For mtDNA, the cline centre is located at nearly the same site, 345.4 km (95% confidence interval 314.7–361.5 km) from the reference locality, with a cline width of 66.7 km (95% confidence interval 30.1–211.7 km).

The second transect (1,300 km length) runs from south-eastern France through north-western Italy to the region of Rome, across the contact zone of *N. h. helvetica* and the peninsular genetic cluster of *N. helvetica*. For microsatellites, the cline is smooth and the cline centre lies 534.9 km (95% confidence interval 353.1–720.8 km) from the starting locality (Figure. S7, right). The cline has a considerable width of 1,065.4 km (95% confidence interval 691.2–1,359.8 km), corresponding to most of the studied transect. For mtDNA, the cline is much steeper and the centre is located 588.9 km (95% confidence interval 467.1–698.3 km) from the reference locality. The cline width for mtDNA is 331.5 km (95% confidence interval 172.2–599.7 km). The cline centres for both markers approximately match with the Alpine main chain.

4 | DISCUSSION

Our investigation is the first detailed examination of the genetic identity of grass snakes from Italy and Corsica. Previous studies (Kindler et al., 2013; Kindler & Fritz, 2018) were compromised by patchy sampling and small sample sizes. Nevertheless, the unexpectedly high level of genetic divergences made it clear that further investigations are needed for understanding the phylogeography and taxonomy of grass snakes in this region. Kindler and Fritz (2018) suggested that the following subspecies of *Natrix helvetica* should be recognized from the study region and beyond: *Natrix helvetica helvetica* (western Europe, corresponding to mitochondrial lineage E), *N. h. lanzai* (Po drainage and Italian Peninsula except southern Calabria, tentatively identified with populations harbouring the mitochondrial lineages C, D, and F), *N. h. sicula* (southern Calabria and Sicily, mitochondrial lineage A), and provisionally *N. h. cetti* (Sardinia, mitochondrial lineage B) and *N. h. corsa* (Corsica, mitochondrial lineage B).

Only very few samples could be analysed of the latter two subspecies, with weak support for their validity. Since the Sardinian subspecies is imperilled and morphologically distinctive (Thorpe, 1979), Kindler and Fritz (2018) refrained from premature taxonomic conclusions that may jeopardize conservation efforts. Lineage D from Apulia was only known by a single previously published GenBank sequence (Guicking et al., 2006). Grass snakes harbouring lineages C, D, and F were, despite deep mitochondrial divergence, tentatively assigned to the same subspecies, either because nuclear genomic evidence was unavailable (lineage D) or because analyses of microsatellite loci did not reveal differences (lineages C and F; Kindler & Fritz, 2018).

In the present study, we examined the genetics of grass snakes from north-eastern Italy for the first time. Based on morphological evidence, the common grass snake (*Natrix natrix*) was recorded east of the Piave River (Lapini et al., 1999). We confirmed this finding genetically and clarified that the involved genetic lineage of *N. natrix* is the so-called 'red lineage' of Kindler et al. (2017). This lineage also occurs in neighbouring Slovenia (Kindler et al., 2013, 2017). Our genetic data, however, also revealed that the Piave River does not form the distribution border for the red lineage. Its westernmost records lie approximately 20 km west of the Piave River in the Veneto region. To the south-west, additional records from the Po Delta and the Emilia-Romagna suggest a much wider distribution range than previously thought. The red lineage had its glacial refugia in the southern Balkans (Kindler, Graciá, et al., 2018) and is surely a Holocene immigrant to Italy that met in the Veneto region the barred grass snake (*N. helvetica*), which survived the Pleistocene glacial cycles in Italy (Kindler et al., 2013; Kindler & Fritz, 2018). Unlike the well-documented contact zone of *N. helvetica* and *N. natrix* in the Rhine region (Kindler et al., 2017; Schultze, Laufer, Kindler, & Fritz, 2019), this Italian contact zone had not been examined genetically before (Kindler & Fritz, 2018).

According to our cline analyses, the Italian contact zone is narrow (approx. 90 km based on microsatellite data and approx. 70 km based on mtDNA), with an abrupt genetic transition from one species to the other (Figure S7). Based on morphological evidence alone (multivariate analyses of many characters), Thorpe (1979) had suggested a much wider contact zone in this region, a conclusion that conflicts with our data. In its width, the Italian contact zone of the two species resembles that in the Rhine region. Like there, we detected localized admixture, with paternal genotypes occurring along with hybrids and introgressed animals (Figure 4, centre) in the contact zone, which therefore qualifies for a bimodal hybrid zone (Jiggins & Mallet, 2000). However, gene flow could be bidirectional (Figure 4; Figure S6), whereas in the Rhine region, gene flow is largely unidirectional from *N. helvetica* into *N. natrix* (Kindler et al., 2017).

In contrast to these narrow hybrid zones, the second contact zone studied here shows a pattern of wide-reaching gene flow within the two mainland clusters of *N. helvetica* (Figure S7), matching with a unimodal hybrid zone (Jiggins & Mallet, 2000). The high degree of nuclear genomic admixture corresponds to a much earlier stage of the speciation process, implying genetic amalgamation of formerly allopatric units after secondary contact. Yet, for mtDNA a steeper cline was found, in agreement with the hypothesis of female philopatry combined with population-density-dependent processes such as 'high-density blocking' (Waters, Fraser, & Hewitt, 2013; see below).

In contrast to *N. natrix*, *N. helvetica* is an old endemic in our study region and shows a remarkable diversity with several deeply divergent mitochondrial lineages. This differentiation has been attributed to the existence of several glacial refugia (Kindler et al., 2013; Kindler & Fritz, 2018). In addition to the previously identified lineages, we found a new Italian lineage (clade G), which is phylogenetically sister to clade D, previously known only from a single GenBank sequence derived from a grass snake from southern Apulia (Guicking et al., 2006). We confirmed the occurrence of lineage D for this region. The discovery of lineage G, recorded from two putatively native sites in Campania and Basilicata and one putatively translocated snake (see below), brings the number of mitochondrial lineages of *N. helvetica* to seven, which is an unrivalled diversity among all other amphibians and reptiles in the region (see the review in Pedall et al., 2011 and Bisconti, Porretta, Arduino, Nascetti, & Canestrelli, 2018; Canestrelli, Sacco, & Nascetti, 2012; Canestrelli, Salvi, Maura, Bologna, & Nascetti, 2012; Maura, Salvi, Bologna, Nascetti, & Canestrelli, 2014; Vamberger et al., 2015). All seven lineages occur in mainland Italy, Corsica, Sardinia, and Sicily, even though lineage E is mainly distributed in western Europe beyond the Alps (Kindler et al., 2017; Kindler, Graciá, et al., 2018).

In general, our new records for the mitochondrial lineages of *N. helvetica* (Figure 3) refine the picture drawn by Kindler and Fritz (2018). However, there are a few records with mismatched mitochondrial lineages (Figure 3: black arrows) that most likely represent either translocated snakes, as known for several other cases (Arnold, 2019; Böhme & Grell, 2013; Dubey et al., 2017; Kindler et al., 2017; Vamberger et al., 2015), or incorrectly labelled collection material (Table S1). If these problematical records are disregarded, the mitochondrial lineages of *N. helvetica* are distributed largely parapatrically. In the north-west of the peninsula, two lineages (E and F) meet, that is in the region where we found evidence for broad-scale gene flow. Lineage A is restricted to Sicily and Calabria, and lineage B, to Corsica and Sardinia. Lineage C occurs in the southern Alps and in the western and central Po Plain. This lineage has crossed the Alps in Switzerland (Kindler & Fritz, 2018; this study) and presumably Austria; the recently published northernmost records are known from Bavaria (Glaw et al., 2019). Lineage D occurs in southernmost Apulia. The newly identified lineage G is found north of lineage D, with records on both the Tyrrhenian and Adriatic sides of the Italian Peninsula, suggestive of female dispersal across the relatively low Apennine chain in this region (Figure 3). In the light of our new records, the contact zone of lineages A and F that had been postulated for Calabria (Kindler et al., 2013) is questionable because the record of lineage F refers to one of the doubtful samples mentioned above. Lineage F has a wide distribution in the western Italian Peninsula and meets lineage E in northern Tuscany.

According to a fossil-calibrated molecular clock (Kindler, de Pous, et al., 2018), the divergence of the mitochondrial lineages of *N. helvetica* commenced approximately 7 million years ago, with the split between lineage E and F being the youngest (approx. 320,000 years ago). The oldest divergence refers to lineage A from Sicily and Calabria. Lineage B (Corsica, Sardinia), the second oldest lineage, branched off from the remaining mitochondrial lineages approximately 4 million years ago. According to these estimates (Figure 1), several mitochondrial lineages predate the Pleistocene considerably, and others originated during the severe glacial cycles of the Pleistocene that started 1.2–0.8 million years ago (Ehlers & Gibbard, 2011).

However, our results for Italian grass snakes based on nuclear genomic data are clearly conflicting (Figures 1, 4; Figure S6). Despite the pronounced mitochondrial structure of *N. helvetica*, we found much less differentiation with respect to microsatellite loci. Our STRUCTURE analyses and PCAs assigned the *N. helvetica* samples to only three clusters, in sharp contrast to the mitochondrial data that correspond to seven deeply divergent, and in part very old, lineages. One microsatellite cluster included samples from Corsica and Sardinia, another cluster included those of the nominotypical subspecies *N. h. helvetica* (mtDNA lineage E) from north-western Italy and beyond, and the third cluster comprised all remaining barred grass snakes from Italy (Figure 4b). Thus, snakes harbouring the oldest mitochondrial lineage (A), with an inferred age of approximately 7 million years, are lumped together with the majority of the other mitochondrial lineages, among them also lineage F, that diverged from lineage E only during the Pleistocene (320,000 years ago). Despite this recent divergence, lineage E constitutes a distinct nuclear genomic cluster. In other words, one of the youngest mitochondrial lineages (E) matches a distinct nuclear genomic cluster, whereas another nuclear genomic cluster comprises snakes representing no fewer than five distinct mitochondrial lineages, among them the oldest lineage.

Grass snakes are highly mobile (Meister, Ursenbacher, & Baur, 2012), with comparatively large home ranges of 0.18–9.41 ha (Reading & Jofré, 2009: England) to 25–40 ha (Madsen, 1984: Sweden; Wisler, Hofer, & Arlettaz, 2008: Switzerland). We therefore hypothesize that effective geographic barriers restrict dispersal and gene flow, facilitating the distinctness of the current three nuclear genomic clusters. Corsica and Sardinia are separated from the mainland by the Tyrrhenian Sea, largely preventing exchange with continental populations of *N. helvetica*. For the two remaining clusters, the Alps act as a barrier, even though grass snakes have been reported in altitudes of up to 1,900 m a.s.l. (Switzerland; Kindler & Fritz, 2018) or even up to 2,300 m a.s.l. (French Alps; Piedmont, Italy; East Tyrol, Austria; Kabisch, 1999). However, it seems likely that during glacial periods, the barrier function of the Alps was much

stronger, implying that *N. h. helvetica* (lineage E), whose glacial refuge was most likely in southern France (Kindler, Graciá, et al., 2018), invaded the Italian Peninsula only in the Holocene. Simultaneously, lineage C would have expanded its range across the Alps to the north reaching southern Bavaria (Glaw et al., 2019).

This complicated pattern of mismatched mitochondrial and nuclear genomic differentiation results from the interplay of repeated dispersal and vicariance events during the geological history of our study region. The model of 'refugia within (Pleistocene) refugia' (Gómez & Lunt, 2007) only incompletely explains the observed mitochondrial structure. What is today the Italian Peninsula was an island chain during the Pliocene (5.3–2.6 mya; Marchetti et al., 2017), suggesting that the deeply divergent mitochondrial lineages evolved in allopatry in this island setting. After the formation of the peninsula, the younger lineages originated from glacial range disruptions. However, repeated range expansions during climatically favourable periods seem to have erased divergence on the nuclear genomic level, resulting in the present nuclear genomic uniformity in continental Italy and Sicily, so that only the mitochondrial lineages as 'ghosts of past' bear witness to the former divergence.

But how can the survival of so many mitochondrial lineages be explained? Our scenario implies philopatry in females and largely male-mediated gene flow. Despite the considerable dispersal capability of grass snakes as mirrored by large home ranges (Madsen, 1984; Meister et al., 2012; Reading & Jofré, 2009; Wisler et al., 2008), female grass snakes appear to be relatively stationary. It has been shown that female grass snakes use the same home range in successive years, with some site fidelity for oviposition (Madsen, 1984). Also, population-density-dependent processes, such as high-density blocking (Waters, Fraser, & Hewitt, 2013), are likely to have played a role in preventing establishment of another mitochondrial lineage in a preoccupied habitat. In any case, to the best of our knowledge, no comparably extreme example of conflicting nuclear and mitochondrial evidence became known to science (see the reviews in Salvi, Pinho, & Harris, 2017; Toews & Brelsford, 2012). The situation resembles that in sea turtles (Cheloniidae; Bowen & Karl, 2007; Clusa et al., 2018) and some North American freshwater turtles (*Graptemys*, Emydidae; Praschag, Ihlow, Flecks, Vamberger, & Fritz, 2017; Thomson, Spinks, & Shaffer, 2018), taxa also displaying pronounced female philopatry, but in both turtle groups the divergences of the maternal lineages are much shallower than in *N. helvetica*, with Pleistocene divergence estimates (Bowen & Karl, 2007; Thomson et al., 2018). However, old mitochondrial 'ghost lineages' are known from certain European wall lizard species (Renoult et al., 2009; Salvi et al., 2017), but these cases involve much less mitochondrial lineages.

5 | TAXONOMY

Based on our microsatellite evidence, the recognition of the subspecies *Natrix helvetica lanzai* and *N. h. sicula*, as suggested by Kindler and Fritz (2018), is not supported. These authors tentatively identified *N. h. lanzai* with barred grass snakes harbouring haplotypes of lineages C, D, and F and *N. h. sicula* with snakes possessing haplotypes of lineage A. Based on the samples available in their study, the two subspecies matched with two distinct clusters in PCAs for microsatellite loci, whereas they were lumped together in STRUCTURE runs. Using our expanded sampling, however, all Sicilian and peninsular Italian *N. helvetica* are not distinct with respect to their microsatellites, neither in STRUCTURE analyses nor in PCAs. Thus, if a subspecies is understood as an evolutionarily significant unit that differs both in nuclear and in mitochondrial DNA from other such units, but that is still capable of large-scale gene flow (Kindler & Fritz, 2018), the distinction of two subspecies for Sicilian and peninsular Italian grass snakes is not justified because all analysed samples represent a nuclear genomic continuum. The oldest available name for this subspecies is *Natrix helvetica sicula* (Cuvier, 1829), rendering *N. h. lanzai* Kramer, 1970 a junior synonym.

Further research is needed to better understand the complex morphological variation of barred grass snakes from Sicily and mainland Italy. Up to four or five subspecies have been suggested to occur there by some authors (Kramer, 1970; Lanza, 1983), while others recognized only two (Mertens & Wermuth, 1960) or assigned all populations from these regions to what is now the nominotypical subspecies of *N. helvetica* (Thorpe, 1979). The geographic distributions of some of the putative subspecies match the distinct mitochondrial lineages of *N. helvetica* to some extent, suggesting that some morphological heritage of the formerly allopatric populations might have survived until today.

While the morphological variation of Sicilian and mainland Italian populations obviously is complex and contradictory, barred grass snakes from Sardinia and Corsica have been unanimously regarded as two distinct subspecies based on their peculiar morphology (Kabisch, 1999; Kreiner, 2007; Lanza, 1983; Mertens, 1957; Mertens & Wermuth, 1960; Thorpe, 1979). With respect to mtDNA, the two taxa from Sardinia and Corsica together represent the deeply divergent mitochondrial lineage B (Kindler et al., 2013; Kindler & Fritz, 2018). However, using limited sampling, Kindler et al. (2013) found mtDNA sequences of Sardinian ($n = 3$) and Corsican grass snakes ($n = 4$) to be not reciprocally monophyletic. Using 11 samples of *N. h. cetti* and five of *N. h. corsa*, we found the same pattern, with one shared haplotype each for ND4 + tRNAs (b1) and for cyt *b* (b4) between specimens from both islands. Using microsatellite data, the few samples of each subspecies studied by Kindler and Fritz

(2018) were lumped together in STRUCTURE analyses (*N. h. cetti*: $n = 2$; *N. h. corsa*: $n = 3$), but PCA retrieved each subspecies as a distinct cluster. Our reanalyses using more, but still few, samples (*N. h. cetti*: $n = 9$; *N. h. corsa*: $n = 5$) resulted in only one cluster for the snakes from the two islands, both for STRUCTURE analyses and for PCAs. This suggests that Sardinian and Corsican grass snakes do not represent genetically distinct taxa and that both should be lumped together under the oldest available name *Natrix helvetica cetti* Gené, 1839. Because Sardinian grass snakes are very rare and listed as 'Critically Endangered' in the IUCN Red List (European Reptile and Amphibian Specialist Group, 1996), this conclusion could have negative implications for conservation. Therefore, we recommend studying more samples from each nominal subspecies before drawing a taxonomic decision.

6 | CONCLUSIONS

Our study confirmed the contact zone of *N. helvetica* and *N. natrix* in north-eastern Italy. There, barred grass snakes harbouring the mitochondrial lineage C meet the 'red lineage' of the common grass snake. Hybridization occurs in a narrow zone of less than 90 km width, with concordant steep clines for nuclear and mitochondrial markers. Gene flow may be bidirectional, in contrast to the hybrid zone of the two species in the Rhine region. *Natrix natrix* has a wider distribution in north-eastern Italy than previously known and occurs also west of the Piave River and south of the Po River.

Within barred grass snakes, we found conflicting patterns of mitochondrial and nuclear genomic differentiation. There are seven distinct mitochondrial lineages, some of them deeply divergent and considerably predating Pleistocene glaciations. In contrast, only three nuclear genomic clusters were detected. One of these clusters matches no fewer than five mitochondrial lineages, among them the oldest lineage (approx. 7 million years old), corresponding to grass snakes from Sicily and Calabria, and four other, much younger, continental Italian lineages. Another nuclear genomic cluster represents grass snakes from Corsica and Sardinia harbouring the second oldest mitochondrial lineage. The third cluster corresponds to grass snakes from western Europe north of the Alps with a young mitochondrial lineage of Pleistocene age that is sister to one of the young Italian lineages. This situation makes the cluster from Sicily and mainland Italy paraphyletic with respect to mitochondrial lineages.

Barred grass snakes from Corsica (*N. h. corsa*) and Sardinia (*N. h. cetti*) were found to be genetically undifferentiated, with respect to both nuclear and mitochondrial markers. The two islands were repeatedly connected during the Pliocene and during Pleistocene low sea-level stands (Marchetti et al., 2017), and the last connection between Corsica and Sardinia was only interrupted by the rising sea level in the early Holocene. As a consequence, grass snakes of the two islands most likely repeatedly formed a contiguous population system, suggesting that their morphological differences are the result of recent divergence. However, larger sample sizes from both islands should be studied genetically before taxonomic conclusions are drawn, not least because *N. h. cetti* is Critically Endangered (European Reptile and Amphibian Specialist Group, 1996).

The mismatched mitochondrial and nuclear genomic differentiation of *N. helvetica* represents an extreme case apparently resulting from female philopatry and male-mediated gene flow. Acknowledging the lack of nuclear genomic differentiation, we propose that *N. h. lanzai* Kramer, 1970 is relegated into the synonymy of *N. h. sicula* (Cuvier, 1829).

ACKNOWLEDGEMENTS

Francesco Luigi Leonetti and Gianni Giglio (Università della Calabria, Rende) assisted in Calabria in the field. Specimens collected and donated to the MSNVE by Roberto Basso, Alessandro Bellese, Paola Gazzola, Jacopo Richard, Enrico Romanazzi, Massimo Semenzato, and Elena Zamprogno contributed substantially to this work; further samples for this study were donated by Henrik Bringsøe. Fieldwork and sampling in Italy was permitted by the Ministero dell'ambiente e della tutela del territorio e del mare (Prot. 2019.0011888, Prot. 0001255/PNM) and the Regione Sardegna (Prot. N. 3115 Rep. N. 42); fieldwork and sampling in Corsica was permitted by the Direction Régionale de l'Environnement de l'Aménagement et du Logement, Corse-du-Sud (Arrête nos 2A-2018-11-20-005 and 2B-2018-11-20-003). Enrico Lunghi is supported by the Chinese Academy of Sciences President's International Fellowship Initiative for postdoctoral researchers. Ralf Britz (Senckenberg Dresden) critically read a draft of this study. Heiko Stuckas (Senckenberg Dresden) helped with some population genetic calculations. All genetic work was conducted in the Senckenberg Dresden Molecular Laboratory (SGN-SNSD-Mol-Lab).

ORCID

Cäcilia Spitzweg https://orcid.org/0000-0001-9297-7633
Claudia Corti https://orcid.org/0000-0001-6015-0092
Michel Delaugerre https://orcid.org/0000-0001-5447-260X
Matteo Riccardo Di Nicola https://orcid.org/0000-0001-7199-0804
Enrico Lunghi https://orcid.org/0000-0002-4228-2750
Emilio Sperone https://orcid.org/0000-0002-3180-7611
Leonardo Vignoli https://orcid.org/0000-0002-1599-092X
Melita Vamberger https://orcid.org/0000-0002-1404-2469
Uwe Fritz https://orcid.org/0000-0002-6740-7214

REFERENCES

Arnold, A. (2019). Streifen- und Barrenringelnattern (*Natrix n. persa* und *N. helvetica*) in der Elster-Luppe-Aue zwischen Leipzig und Schkeuditz. *Elaphe*, *5*, 74–79.

Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, *13*, 729–744. https://doi.org/10.1046/j.1365-294X.2003.02063.x

Barton, N. H. (1979). Gene flow past a cline. *Heredity*, *43*, 333–339. https://doi.org/10.1038/hdy.1979.86

Bisconti, R., Porretta, D., Arduino, P., Nascetti, G., & Canestrelli, D. (2018). Hybridization and extensive mitochondrial introgression among fire salamanders in peninsular Italy. *Scientific Reports*, *8*, 13187. https://doi.org/10.1038/s41598-018-31535-x

Böhme, W., & Grell, O. (2013). Ringelnattern, *Natrix natrix* (Linnaeus, 1758), auf Sylt (Schleswig-Holstein): Eingeschleppt aus verschiedenen Ländern? *Faunistisch-ökologische Mitteilungen*, *9*, 311–317.

Bowen, B. W., & Karl, S. A. (2007). Population genetics and phylogeography of sea turtles. *Molecular Ecology*, *16*, 4886–4907. https://doi.org/10.1111/j.1365-294X.2007.03542.x

Canestrelli, D., Sacco, F., & Nascetti, G. (2012). On glacial refugia, genetic diversity, and microevolutionary processes: Deep phylogeographical structure in the endemic newt *Lissotriton italicus*. *Biological Journal of the Linnean Society*, *105*, 42–55. https://doi.org/10.1111/j.1095-8312.2011.01767.x

Canestrelli, D., Salvi, D., Maura, M., Bologna, M. A., & Nascetti, G. (2012). One species, three Pleistocene evolutionary histories: Phylogeography of the Italian crested newt, *Triturus carnifex*. *PLoS ONE*, *7*, e41754. https://doi.org/10.1371/journal.pone.0041754

Clement, M., Posada, D., & Crandall, K. A. (2000). TCS: A computer program to estimate gene genealogies. *Molecular Ecology*, *9*, 1657–1659. https://doi.org/10.1046/j.1365-294x.2000.01020.x

Clusa, M., Carreras, C., Cardona, L., Demetropoulos, A., Margaritoulis, D., Rees, A. F., … Pascual, M. (2018). Philopatry in loggerhead turtles *Caretta caretta*: Beyond the gender paradigm. *Marine Ecology Progress Series*, *588*, 201–213. https://doi.org/10.3354/meps12448

Derryberry, E. P., Derryberry, G. E., Maley, J. M., & Brumfield, R. T. (2014). HZAR: Hybrid zone analysis using an R software package. *Molecular Ecology Resources*, *14*, 652–663. https://doi.org/10.1111/1755-0998.12209

Dubey, S., Ursenbacher, S., Schuerch, J., Golay, J., Aubert, P., & Dufresnes, C. (2017). A glitch in the *Natrix*: Cryptic presence of alien grass snakes in Switzerland. *Herpetology Notes*, *10*, 205–208.

Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*, 359–361. https://doi.org/10.1007/s12686-011-9548-7

Ehlers, J., & Gibbard, P. (2011). Quaternary Glaciation. In V. P. Singh, P. Singh, & U. K. Haritashya (Eds.), *Encyclopedia of snow, ice and glaciers* (pp. 873–882). Dordrecht, Netherlands: Springer. https://doi.org/10.1007/978-90-481-2642-2_423

European Reptile and Amphibian Specialist Group. (1996). *Natrix natrix* ssp. *cetti. The IUCN Red List of Threatened Species, 1996*, e.T14364A4436077. https://doi.org/10.2305/IUCN.UK.1996.RLTS.T14364A4436077.en

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, *14*, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, *164*, 1567–1587.

Fritz, U., Corti, C., & Päckert, M. (2012). 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. *Organisms Diversity & Evolution*, *12*, 71–80. https://doi.org/10.1007/s13127-011-0069-8

Glaw, F., Franzen, M., Oefele, M., Hansbauer, G., & Kindler, C. (2019). Genetischer Erstnachweis, Verbreitung und südalpine Herkunft der Barrenringelnatter (*Natrix helvetica* ssp.) in Bayern. *Zeitschrift für Feldherpetologie*, *26*, 1–20.

Gómez, A., & Lunt, D. H. (2007). Refugia within refugia: Patterns of phylogeographic concordance in the Iberian Peninsula. In S. Weiss, & N. Ferrand (Eds.), *Phylogeography of Southern European refugia* (pp. 155–188). Dordrecht, The Netherlands: Springer. https://doi.org/10.1007/1-4020-4904-8_5

Grill, A., Casula, P., Lecis, R., & Menken, S. (2007). Endemism in Sardinia. In S. Weiss, & N. Ferrand (Eds.), *Phylogeography of Southern European refugia* (pp. 273–296). Dordrecht, The Netherlands: Springer. https://doi.org/10.1007/1-4020-4904-8_10

Guicking, D., Lawson, R., Joger, U., & Wink, M. (2006). Evolution and phylogeny of the genus *Natrix* (Serpentes: Colubridae). *Biological Journal of the Linnean Society*, *87*, 127–143. https://doi.org/10.1111/j.1095-8312.2006.00561.x

Hall, T. A. (1999). BIOEDIT: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, *41*, 95–98.

Hecht, G. (1930). Systematik, Ausbreitungsgeschichte und Ökologie der europäischen Arten der Gattung *Tropidonotus* (Kuhl) H. Boie. *Mitteilungen aus dem Zoologischen Museum Berlin*, *16*, 244–393.

Janes, J. K., Miller, J. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., Cullingham, C. I., & Andrew, R. L. (2017). The $K = 2$ conundrum. *Molecular Ecology*, *26*, 3594–3602. https://doi.org/10.1111/mec.14187

Jiggins, C. D., & Mallet, J. (2000). Bimodal hybrid zones and speciation. *Trends in Ecology and Evolution*, *15*, 250–255. https://doi.org/10.1016/S0169-5347(00)01873-5

Jombart, T. (2008). ADEGENET: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*, 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Kabisch, K. (1999). *Natrix natrix* (Linnaeus, 1758) – Ringelnatter. In W. Böhme (Ed.), *Handbuch der Reptilien und Amphibien Europas. Band 3/II A – Schlangen II* (pp. 513–580). Wiebelsheim, Germany: Aula-Verlag.

Kalinowski, S. T. (2011). The computer program STRUCTURE does not reliably identify the main genetic clusters within species: Simulations and implications for human population structure. *Heredity*, *106*, 625–632. https://doi.org/10.1038/hdy.2010.95

Kindler, C., Böhme, W., Corti, C., Gvoždík, V., Jablonski, D., Jandzik, D., … Fritz, U. (2013). Mitochondrial phylogeography, contact zones and taxonomy of grass snakes (*Natrix natrix*, *N. megalocephala*). *Zoologica Scripta*, *42*, 458–472. https://doi.org/10.1111/zsc.12018

Kindler, C., Bringsøe, H., & Fritz, U. (2014). Phylogeography of grass snakes *(Natrix natrix)* all around the Baltic Sea: implications for the Holocene colonization of Fennoscandia. *Amphibia-Reptilia*, *35*, 413–424. https://doi.org/10.1163/15685381-00002962.

Kindler, C., Chèvre, M., Ursenbacher, S., Böhme, W., Hille, A., Jablonski, D., … Fritz, U. (2017). Hybridization patterns in two contact zones of grass snakes reveal a new Central European snake species. *Scientific Reports*, *7*, 7378. https://doi.org/10.1038/s41598-017-07847-9

Kindler, C., de Pous, P., Carranza, S., Beddek, M., Geniez, P., & Fritz, U. (2018). Phylogeography of the Ibero-Maghrebian red-eyed grass snake (*Natrix astreptophora*). *Organisms Diversity & Evolution*, *18*, 143–150. https://doi.org/10.1007/s13127-017-0354-2

Kindler, C., & Fritz, U. (2018). Phylogeography and taxonomy of the barred grass snake (*Natrix helvetica*), with a discussion of the subspecies category in zoology. *Vertebrate Zoology*, *68*, 269–281.

Kindler, C., Graciá, E., & Fritz, U. (2018). Extra-Mediterranean glacial refuges in barred and common grass snakes (*Natrix helvetica*, *N. natrix*). *Scientific Reports*, *8*, 1821. https://doi.org/10.1038/s41598-018-20218-2

Kramer, E. (1970). Revalidierte und neue Rassen der europäischen Schlangenfauna. *Lavori della Società Italiana di Biogeografia,1*, 667–676. https://doi.org/10.21426/B61110535

Kreiner, G. (2007). *Schlangen Europas*. Frankfurt am Main, Germany: Edition Chimaira.

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2016). PARTITIONFINDER 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, *34*, 772–773. https://doi.org/10.1093/molbev/msw260

Lanza, B. (1983). *Guide per il riconoscimento delle specie animali delle acque interne italiane. Anfibi, Rettili (Amphibia, Reptilia)*. Verona, Italy: Stamperia Valdonega, Consiglio Nazionale delle Ricerche.

Lapini, L., Dall'Asta, A., Bressi, N., Dolce, S., & Pellarini, P. (1999). *Atlante corologico degli Anfibi e dei Rettili del Friuli-Venezia Giulia*. Udine, Italy: Edizioni del Museo Friulano di Storia Naturale.

Madsen, T. (1984). Movements, home range size and habitat use of radio-tracked grass snakes (*Natrix natrix*) in southern Sweden. *Copeia*, *1984*, 707–713. https://doi.org/10.2307/1445153

Marchetti, M., Soldati, M., & Vandelli, V. (2017). The great diversity of Italian landscapes and landforms: Their origin and human imprint. In I. M. Soldati, & M. Marchetti (Eds.), *World geomorphological landscapes. Landscapes and landforms of Italy* (pp. 7–20). Basel, Switzerland: Springer. https://doi.org/10.1007/978-3-319-26194-2_2

Maura, M., Salvi, D., Bologna, M. A., Nascetti, G., & Canestrelli, D. (2014). Northern richness and cryptic refugia: Phylogeography of the Italian smooth newt *Lissotriton vulgaris meridionalis*. *Biological Journal of the Linnean Society*, *113*, 590–603. https://doi.org/10.1111/bij.12360

Médail, F., & Quézel, P. (1999). Biodiversity hotspots in the Mediterranean Basin: Setting global conservation priorities. *Conservation Biology*, *13*, 1510–1513. https://doi.org/10.1046/j.1523-1739.1999.98467.x

Meister, B., Ursenbacher, S., & Baur, B. (2012). Grass snake population differentiation over different geographic scales. *Herpetologica*, *68*, 134–145. https://doi.org/10.1655/herpetologica-D-11-00036.1

Mertens, R. (1957). Die Amphibien und Reptilien Korsikas. *Senckenbergiana biologica*, *38*, 175–192.

Mertens, R., & Wermuth, H. (1960). *Die Amphibien und Reptilien Europas (Dritte Liste, nach dem Stand vom 1. Januar 1960)*. Frankfurt am Main, Germany: Verlag Waldemar Kramer.

Nielsen, E. E., Bach, L. A., & Kotlicki, P. (2006). HYBRIDLAB (version 1.0): A program for generating simulated hybrids from population samples. *Molecular Ecology Notes*, *6*, 971–973. https://doi.org/10.1111/j.1471-8286.2006.01433.x

Pedall, I., Fritz, U., Stuckas, H., Valdeón, A., & Wink, M. (2011). Gene flow across secondary contact zones of the *Emys orbicularis* complex in the Western Mediterranean and evidence for extinction and re-introduction of pond turtles on Corsica and Sardinia (Testudines: Emydidae). *Journal of Zoological Systematics and Evolutionary Research*, *49*, 44–57. https://doi.org/10.1111/j.1439-0469.2010.00572.x

Pokrant, F., Kindler, C., Ivanov, M., Cheylan, M., Geniez, P., Böhme, W., & Fritz, U. (2016). Integrative taxonomy provides evidence for the species status of the Ibero-Maghrebian grass snake *Natrix astreptophora*. *Biological Journal of the Linnean Society*, *118*, 873–888. https://doi.org/10.1111/bij.12782

Praschag, P., Ihlow, F., Flecks, M., Vamberger, M., & Fritz, U. (2017). Diversity of North American map and sawback turtles (Testudines: Emydidae: *Graptemys*). *Zoologica Scripta*, *46*, 675–682. https://doi.org/10.1111/zsc.12249

Pritchard, J. K., Stephens, M., & Donelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*, 945–959.

Puechmaille, S. J. (2016). The program STRUCTURE does not reliably recover the correct population structure when sampling is uneven: Subsampling and new estimators alleviate the problem. *Molecular Ecology Resources*, *16*, 608–627. https://doi.org/10.1111/1755-0998.12512

Reading, C. J., & Jofré, G. M. (2009). Habitat selection and range size of grass snakes *Natrix natrix* in an agricultural landscape in southern England. *Amphibia-Reptilia*, *30*, 379–388. https://doi.org/10.1163/156853809788795164

Renoult, J. P., Geniez, P., Bacquet, P., Benoit, L., & Crochet, P.-A. (2009). Morphology and nuclear markers reveal extensive mitochondrial introgressions in the Iberian wall lizard species complex. *Molecular Ecology*, *18*, 4298–4315. https://doi.org/10.1111/j.1365-294X.2009.04351.x

Rieseberg, L. H., Whitton, J., & Gardner, K. (1999). Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics*, *152*, 713–727.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., … Huelsenbeck, J. P. (2012). MRBAYES 3.2: Efficient Bayesian phylogenetic inference and model selection across a large model space. *Systematic Biology*, *61*, 539–542. https://doi.org/10.1093/sysbio/sys029

Rosenberg, N. A. (2004). DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Notes*, *4*, 137–138. https://doi.org/10.1046/j.1471-8286.2003.00566.x

Salvi, D., Harris, D. J., Bombi, P., Carretero, M. A., & Bologna, M. A. (2010). Mitochondrial phylogeography of the Bedriaga's rock lizard, *Archaeolacerta bedriagae* (Reptilia: Lacertidae) endemic to Corsica and Sardinia. *Molecular Phylogenetics and Evolution*, *56*, 690–697. https://doi.org/10.1016/j.ympev.2010.03.017

Salvi, D., Pinho, C., & Harris, D. J. (2017). Digging up the roots of an insular hotspot of genetic diversity: Decoupled mito-nuclear histories in the evolution of the Corsican-Sardinian endemic lizard *Podarcis tiliguerta*. *BMC Evolutionary Biology*, *17*, 63. https://doi.org/10.1186/s12862-017-0899-x

Schmitt, T. (2007). Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in Zoology*, *4*, 11. https://doi.org/10.1186/1742-9994-4-11

Schultze, N., Laufer, H., Kindler, C., & Fritz, U. (2019). Distribution and hybridisation of barred and common grass snakes (*Natrix*

helvetica, *N. natrix*) in Baden-Württemberg, South-western Germany. *Herpetozoa*, *32*, 229–236. https://doi.org/10.3897/herpetozoa.32.e38897

Stamatakis, A. (2006). RAxML-VI-HPC: Maximum Likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, *22*, 2688–2690. https://doi.org/10.1093/bioinformatics/btl446

Thielsch, A., Knell, A., Mohammadyari, A., Petrusek, A., & Schwenk, K. (2017). Divergent clades or cryptic species? Mito-nuclear discordance in a *Daphnia* species complex. *BMC Evolutionary Biology*, *17*, 227. https://doi.org/10.1186/s12862-017-1070-4

Thomson, R. C., Spinks, P. Q., & Shaffer, H. B. (2018). Molecular phylogeny and divergence of the map turtles (Emydidae: *Graptemys*). *Molecular Phylogenetics and Evolution*, *121*, 61–70. https://doi.org/10.1016/j.ympev.2017.11.012

Thorpe, R. S. (1979). Multivariate analysis of the population systematics of the ringed snake, *Natrix natrix* (L.). *Proceedings of the Royal Society of Edinburgh*, *78B*, 1–62. https://doi.org/10.1017/S0269727000012 94X

Toews, D. P. L., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, *21*, 3907–3930. https://doi.org/10.1111/j.1365-294X.2012.05664.x

Vamberger, M., Stuckas, H., Sacco, F., D'Angelo, S., Arculeo, M., Cheylan, M., … Fritz, U. (2015). Differences in gene flow in a twofold secondary contact zone of pond turtles in southern Italy (Testudines: Emydidae: *Emys orbicularis galloitalica*, *E. o. hellenica*, *E. trinacris*). *Zoologica Scripta*, *44*, 233–249. https://doi.org/10.1111/zsc.12102

van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., & Shipley, P. (2004). MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, *4*, 535–538. https://doi.org/10.1111/j.1471-8286.2004.00684.x

Vanni, S., & Cimmaruta, R. (2011). *Natrix cetti* Gené, 1839. In C. Corti, M. Capula, L. Luiselli, E. Razzetti, & R. Sindaco (Eds.), *Fauna d'Italia, Reptilia* (pp. 538–545). Bologna, Italy: Calderini.

Waters, J. M., Fraser, C. I., & Hewitt, G. M. (2013). Founder takes all: Density-dependent processes structure biodiversity. *Trends in Ecology & Evolution*, *28*, 78–85. https://doi.org/10.1016/j.tree.2012.08.024

Wisler, C., Hofer, U., & Arlettaz, R. (2008). Snakes and monocultures: Habitat selection and movements of female grass snakes (*Natrix natrix* L.) in an agricultural landscape. *Journal of Herpetology*, *42*, 337–346. https://doi.org/10.1670/07-027.1

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Schultze N, Spitzweg C, Corti C, et al. Mitochondrial ghost lineages blur phylogeography and taxonomy of *Natrix helvetica* and *N. natrix* in Italy and Corsica. *Zool Scr*. 2020;00:1–17. https://doi.org/10.1111/zsc.12417