Molecular Phylogenetics and Evolution 184 (2023) 107787

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Genomics reveals broad hybridization in deeply divergent Palearctic grass and water snakes (*Natrix* spp.)

Yannis Schöneberg [a,b], Sven Winter [a,c], Oscar Arribas [d], Matteo Riccardo Di Nicola [e], Maya Master [f], John Benjamin Owens [f], Michail Rovatsos [g], Wolfgang Wüster [f], Axel Janke [a,b,h], Uwe Fritz [i,*]

[a] Senckenberg Biodiversity and Climate Research Centre (BiK-F), Senckenberganlage 25, 60325 Frankfurt am Main, Germany
[b] Institute for Ecology, Evolution and Diversity, Goethe University, Max-von-Laue-Straße 9, 60325 Frankfurt am Main, Germany
[c] Research Institute of Wildlife Ecology, University of Veterinary Medicine Vienna, Savoyenstraße 1, 1160 Vienna, Austria
[d] IES Castilla, Junta de Castilla, Castilla y León, 42003 Soria, Spain
[e] IRCCS San Raffaele Hospital, Unit of Dermatology, Via Olgettina 60, 20132 Milan, Italy
[f] Molecular Ecology and Evolution at Bangor (MEEB), School of Natural Sciences, Bangor University, Environment Centre Wales, Bangor LL57 2UW, Wales, UK
[g] Department of Ecology, Faculty of Science, Charles University, Viničná 7, 12844 Praha 2, Czech Republic
[h] LOEWE-Centre for Translational Biodiversity Genomics (TBG), Senckenberg Nature Research Society, Senckenberganlage 25, 60325 Frankfurt am Main, Germany
[i] Senckenberg Dresden, Museum of Zoology, A. B. Meyer Building, 01109 Dresden, Germany

ARTICLE INFO

Keywords:
Genomics
Hybridization
Molecular clock
Natricidae
Reptilia
Speciation
Squamata

ABSTRACT

Understanding speciation is one of the cornerstones of biological diversity research. Currently, speciation is often understood as a continuous process of divergence that continues until genetic or other incompatibilities minimize or prevent interbreeding. The Palearctic snake genus *Natrix* is an ideal group to study speciation, as it comprises taxa representing distinct stages of the speciation process, ranging from widely interbreeding parapatric taxa through parapatric species with very limited gene flow in narrow hybrid zones to widely sympatric species. To understand the evolution of reproductive isolation through time, we have sequenced the genomes of all five species within this genus and two additional subspecies. We used both long-read and short-read methods to sequence and de-novo-assemble two high-quality genomes (*Natrix h. helvetica*, *Natrix n. natrix*) to their 1.7 Gb length with a contig N50 of 4.6 Mbp and 1.5 Mbp, respectively, and used these as references to assemble the remaining short-read-based genomes. Our phylogenomic analyses yielded a well-supported dated phylogeny and evidence for a surprisingly complex history of interspecific gene flow, including between widely sympatric species. Furthermore, evidence for gene flow was also found for currently allopatric species pairs. Genetic exchange among these well-defined, distinct, and several million-year-old reptile species emphasizes that speciation and maintenance of species distinctness can occur despite continued genetic exchange.

1. Introduction

The publication of Darwin's (1859) "Origin of Species" propelled the species to the status of the key unit of biodiversity, and the process of species formation and diversification to a research theme of fundamental importance for evolutionary and systematic biology. Traditionally, a central role has been ascribed to the development of reproductive isolation during the evolutionary divergence process (e.g., the Biological Species Concept; see Mayr, 1942; Coyne and Orr, 2004; Zachos, 2016). The effects of reproductive isolation, such as interrupted gene flow and genetic diversification, are recorded in the genome of an organism and can be recovered and analyzed. However, it is only today that technological advances in the field of genomics have given us the analytical tools required to access this information to further our understanding of speciation and evolution—a central theme of evolutionary research.

Recent evidence challenges the traditional view of the central importance of reproductive isolation, as increasing numbers of cases of extensive gene flow between distinct species are being documented. Population genetic and genomic approaches in vertebrates, including amphibians, squamates, turtles, and mammals (Vamberger et al., 2015;

* Corresponding author.
E-mail address: uwe.fritz@senckenberg.de (U. Fritz).

https://doi.org/10.1016/j.ympev.2023.107787
Received 31 October 2022; Received in revised form 3 April 2023; Accepted 13 April 2023
Available online 18 April 2023

Kindler et al., 2017; Kumar et al., 2017; Pabijan et al., 2017; Pöschel et al., 2018; Winter et al. 2018; Schield et al. 2019; Burbrink et al. 2020; Asztalos et al., 2021a; Dufresnes et al., 2021; Pavón-Vázquez et al. 2021) have found ample evidence for gene flow across species limits. This suggests that the traditional understanding of a species is insufficient for comprehending its true nature and how it is formed during evolution. Consequently, the understanding of speciation, the central process in the diversification, composition, and evolution of biodiversity, and the species category itself, remain vague. Moreover, while whole genome sequencing promises to provide the deepest insights into the speciation process, studies using this approach have remained largely confined to mammals until now (e.g., Kumar et al., 2017; Winter et al. 2018; but see Yang et al., 2021 for a study on wall lizards).

Currently, the understanding of speciation is that it reflects a gradual process, a "speciation continuum." However, once distinct species are capable of sympatric occurrence, the end of that speciation continuum has been reached, even though the integration of foreign genomic components acquired via hybridization may continue. The existence of speciation with gene flow (Nosil, 2008; Arnold, 2015) has long been known, but its systematic study has only now been enabled by the advent of genomics. However, current speciation research has remained heavily concentrated on mammals, for which most genomes have been sequenced for various basic and applied research questions (Coimbra et al., 2021; Figueiró et al., 2017; Nater et al., 2017; Árnason et al., 2018; Supple and Shapiro, 2018; Zhou et al., 2018). However, mammals represent only a small fraction of life's diversity.

In the present study, we focus on an ideal reptile model for speciation research: grass snakes and water snakes of the genus *Natrix*. The five *Natrix* species are widely distributed and abundant across the Western Palearctic (Speybroeck et al., 2020), and two of them (*Natrix natrix*, *N. tessellata*) extend east into Central Asia (Sindaco et al., 2013). These snakes are ideal for studying speciation because they contain the entire range of the speciation continuum. Besides parapatrically distributed and broadly hybridizing subspecies, there are parapatrically distributed species still capable of hybridization across narrow geographic contact zones, and there are fully diverged species that occur in wide sympatry without obvious evidence for hybridization (Kindler et al., 2013, 2017, 2018b; Sindaco et al., 2013; Pokrant et al., 2016; Kindler and Fritz,

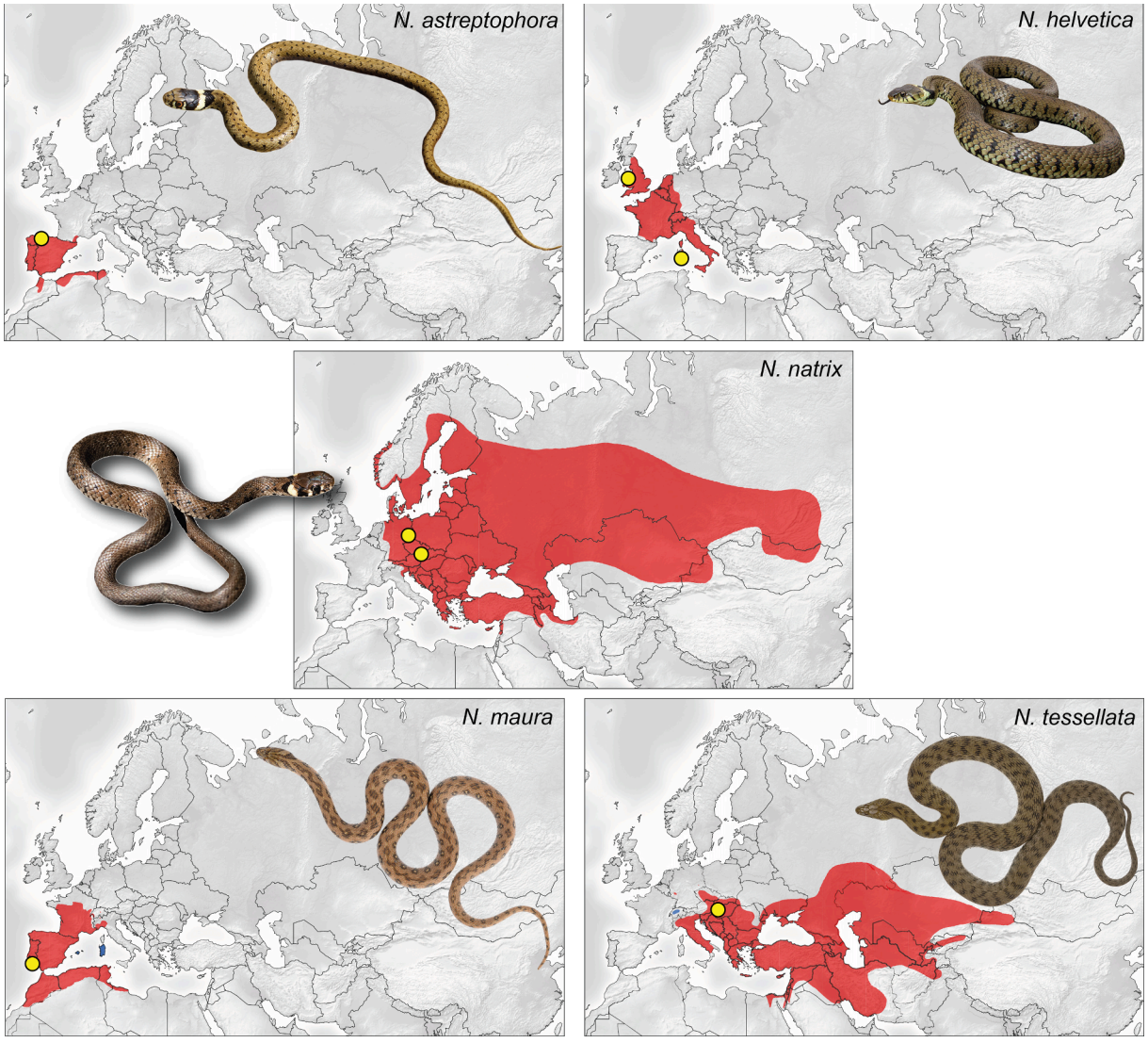

Fig. 1. Distribution ranges and sampling sites (yellow circles) for *Natrix* species. Top and center: grass snakes, bottom: water snakes. Native distribution ranges in red; distribution of naturalized populations of *N. maura* (Balearic Islands, Sardinia) and *N. tessellata* (Switzerland) in blue. Distribution ranges according to Gruschwitz et al. (1999), Schätti (1999), Kindler et al. (2017, 2018b), Schultze et al. (2019, 2020), and Asztalos et al. (2020, 2021a, b). Photos: *N. astreptophora* and *N. natrix* – H. Bringsøe; *N. helvetica* – W. Wüster, *N. maura* and *N. tessellata* – M. R. Di Nicola. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2018).

The two species of water snakes (*Natrix maura, N. tessellata*) are semiaquatic, mostly piscivorous, and occur in freshwater habitats like creeks, rivers, ponds, and lakes (Gruschwitz et al. 1999; Schätti 1999; Geniez, 2015), but with largely mutually exclusive distribution ranges (Fig. 1). The viperine snake (*N. maura*) lives in northwestern Africa, the Iberian Peninsula, and parts of western Europe (Schätti, 1999; Geniez, 2015), whereas the dice snake (*N. tessellata*) ranges from Central Europe and Italy to Central Asia (Gruschwitz et al., 1999; Geniez, 2015). Their distributions overlap only in northwestern Italy. Despite their similar morphology and natural history, no hybrids are known, neither from nature nor captivity (Mebert et al., 2011).

The three grass snake species (*N. astreptophora*, *N. helvetica*, *N. natrix*) also occupy largely parapatric distributions (Fig. 1), but each is widely sympatric with *N. maura* or *N. tessellata* or both. Hybridization between grass snakes and either of the two water snakes is generally believed to be extremely rare (Mebert et al., 2011; de la Vega et al., 2021). Grass snakes are less strictly bound to aquatic habitats and have a wider food spectrum than the two water snakes (Kabisch, 1999). The red-eyed grass snake (*N. astreptophora*) occurs in northwestern Africa and the Iberian Peninsula. North of the Pyrenees, in southwestern France, lies a narrow hybrid zone with *N. helvetica*, where both the parental species and hybrids are found (Pokrant et al. 2016; Asztalos et al. 2020). The barred grass snake (*N. helvetica*) is distributed in Western Europe north of the Pyrenees, including Britain and most of Italy. *Natrix helvetica* hybridizes in three geographically distinct narrow contact zones with the third grass snake species, the common grass snake (*N. natrix*), namely (1) in the Rhine region, (2) in northeastern Italy, and (3) in southern Bavaria and adjacent western Austria (Kindler et al., 2017; Schultze et al., 2019, 2020; Asztalos et al., 2021b). *Natrix natrix* has the largest distribution range of all *Natrix* snakes and occurs from Central Europe and Fennoscandia through the Balkan Peninsula and Asia Minor to Lake Baikal (Kindler et al., 2017), i.e., across a region spanning approximately 6,300 km in an east–west direction.

Phylogenetic analyses of mtDNA sequences suggested that the three grass snake species (*N. astreptophora*, *N. helvetica*, *N. natrix*) together constitute the sister taxon of *N. tessellata* and that *N. maura* represents their successive sister taxon (Guicking et al., 2006; Kindler et al., 2018b). Based on constant sequence divergence rates, Guicking et al. (2006) estimated that *N. maura* branched off 27–18 million years ago (Mya), whereas grass snakes and *N. tessellata* only diverged 22–13 Mya. Using a fossil-calibrated molecular clock, Kindler et al. (2018b) inferred that *N. astreptophora* diverged 10.6 Mya, whereas *N. helvetica* and *N. natrix* separated 8.6 Mya.

Recently, a phylogeographic study (Asztalos et al., 2021a) revealed frequent hybridization between two widely sympatric species (*N. natrix* and *N. tessellata*), suggesting that hybridization between sympatric *Natrix* species is more common than previously thought. This implies intricate evolutionary processes, involving gene flow between ancient lineages. However, all studies available until now only relied on evidence from mitochondrial DNA (mtDNA) sequences and microsatellite loci, preventing an in-depth analysis of the detected hybridization patterns.

To obtain deeper insights into the divergence and hybridization processes of grass and water snakes, we generated seven genome assemblies across the five *Natrix* species and used our data to infer the radiation and divergence times of the individual taxa and the evolutionary impact of hybridization.

2. Materials and methods

2.1. Data generation

Genomes of all five *Natrix* species were sequenced. Two of the species were represented by two subspecies each (*N. astreptophora*, *N. h. helvetica*, *N. h. cetti*, *N. maura*, *N. n. natrix*, *N. n. vulgaris*, *N. tessellata*). All grass snake sampling localities were several hundred kilometers distant from any documented contact or hybrid zones involving the respective species (Fig. 1; Suppl. 1 Table S1). The sample for *N. n. vulgaris* was taken from the name-bearing museum type specimen of this taxon (Fritz and Schmidtler, 2020). DNA was extracted from frozen and ethanol-preserved tissue and blood using a standard phenol/chloroform method, except for *N. n. natrix,* for which DTAB was used.

For two taxa, *Natrix h. helvetica* and *N. n. natrix*, continuous long-read data were generated on a PacBio Sequel II at the Radboud University Medical Center, Nijmegen (Netherlands), following library preparation using the PacBio SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, Menlo Park, CA, USA). Additionally, 150 bp paired-end libraries with an average insert size of 350 bp were prepared using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs Inc., Ipswich, MA, USA) and sequenced on an Illumina NovaSeq6000 at Novogene (Cambridge, UK) to an approximate 30-fold coverage. For *N. n. natrix*, tissue samples from brain, heart, liver, and gonads of specimen MTD 20734 were stored in RNAlater prior to RNA extraction using a TRIzol-RNA extraction protocol. The RNA isolates were then sent to Novogene for library preparation and sequencing.

For the remaining species, 150 bp paired-end libraries with an average insert size of 350 bp were prepared using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs Inc., Ipswich, MA, USA) and sequenced on an Illumina NovaSeq6000 at Novogene (Cambridge, UK) to an approximate 30-fold coverage. All generated data were uploaded to NCBI and are accessible through BioProject PRJNA803021.

2.2. Assembly

De-novo genome assemblies of *N. h. helvetica* and *N. n. natrix* were generated from PacBio long-reads using the assemblers wtdbg2 v2.5 and Raven v1.4.0 (Ruan and Li, 2020; Vaser and Šikić, 2021) with default parameters. For *N. n. natrix*, Flye v2.8.1 (Kolmogorov et al., 2019) was additionally used. The best assembly for each species was chosen for downstream analyses (i.e., the Raven assembly for *N. h. helvetica*; Flye for *N. n. natrix*) based on gene set completeness analyses with BUSCO v5.0.0 using the Sauropsida_odb10 dataset and contig statistics calculated with Quast v5.0.2 (Okonechnikov et al., 2016; Kriventseva et al., 2019; Seppey et al., 2019). The best assembly of *N. h. helvetica* was polished twice by mapping the long-read data against the assembly with Minimap2 v2.17 before running Racon v1.4.3 (Vaser et al., 2017; Li, 2018). Consecutively, two rounds of short-read-polishing were performed to correct base-level errors in the assembly. For each round, the short-read data were mapped against the polished long-read assembly using bwa mem v0.7.17 and sorted by samtools v1.9 prior to polishing with Pilon v1.23 (Li and Durbin, 2009; Walker et al., 2014; Danecek et al., 2021). *Natrix n. natrix* was polished only using Pilon, as Flye already included a long-read polishing step. The quality of the polished assemblies was checked again by Quast v5.0.2 and BUSCO v5.0.0 (Gurevich et al., 2013; Kriventseva et al., 2019; Seppey et al., 2019). Furthermore, mapping and coverage statistics were calculated with Qualimap v2.2.2a after mapping the short-read and long-read data against the assemblies using bwa mem v0.7.17 and Minimap2 v2.17, respectively (Li and Durbin, 2009; Okonechnikov et al., 2016). Additionally, Blobtools v1.1.1 was run, together with a BLASTN v2.9.0 search against the NCBI nucleotide database, to detect possible contamination in the assembly (Laetsch and Blaxter, 2017; Altschul et al., 1990). The two reference assemblies are accessible from NCBI through BioProject PRJNA803021.

2.3. Repeat masking

To prevent repeats in the genome from interfering with downstream analyses, repeats were masked in both de-novo assemblies. RepeatModeler v2.0.1, which itself utilizes Recon v1.08 and RepeatScout

v1.0.6, was used to generate de-novo repeat libraries for each assembly (Bao and Eddy, 2002; Price et al., 2005; Flynn et al., 2019). These repeat libraries were then used in RepeatMasker v4.1.1 to "hard-mask" interspersed repeats and "soft-mask" low-complexity repeats in two subsequent runs to account for low-complexity repeats that might be parts of genes (Smit et al., 2013–2015).

2.4. Phylogenetic analyses

2.4.1. Genome fragment analysis and dated phylogeny

A genome fragment (GF) analysis, as described in Coimbra et al. (2021), was performed (scripts available at https://github.com/rtfcoimbra/Coimbra-et-al-2021_CurrBiol; scripts and procedure described under "Nuclear phylogenomic inference" in this repository). Reference-based assemblies were generated for all specimens (except *N. h. helvetica*), including the outgroup *Thamnophis elegans*, by mapping the respective short-read data against the de-novo assembly of *N. h. helvetica* using bwa mem v0.7.17, and subsequent conversion of the mapping file into a sorted and indexed BAM-Format using the samtools v1.12 commands fixmate, sort, and index (Langmead and Salzberg, 2012; Danecek et al., 2021). The short reads for the outgroup were downloaded from the Vertebrate Genome Project (genomeark/species/Thamnophis_elegans/rThaEle1/genomic_data/10x/). Mapping rates and qualities were evaluated with Qualimap v2.2.2 (Okonechnikov et al., 2016). Repetitive regions in the assemblies were removed based on the *N. h. helvetica* assembly with "hard-masked" interspersed repeats using bedtools v2.30.0 and samtools v1.13 (Quinlan and Hall, 2010; Danecek et al., 2021). To generate the consensus sequences, we first calculated the per base sequencing depth using Sambamba v0.8.1 (Tarasov et al., 2015) and then called the consensus sequence using ANGSD v0.935, reducing mapping bias by using "–iupacRatio 0.33" (Korneliussen et al., 2014). As all reference-based assemblies had the same reference, no additional alignment step was necessary. Sets of genome fragments with a length of 20 kb to 470 kb were created with a 50-kb-step size. An approximately unbiased (AU) tree topology test was run in IQ-tree v2.1.4 to determine the minimum fragment length with which alternative topologies can be significantly rejected (Shimodaira, 2002; Minh et al., 2020). ASTRAL v5.7.7 (Zhang et al., 2018) was used to calculate a multispecies coalescent tree from 860 individual ML trees based on 320 kb long genome fragments. The individual ML trees were inferred using IQ-tree v2.1.4 (Minh et al., 2020). Internal branches were evaluated for support of alternative groupings by a quartet analysis (Sayyari et al., 2018).

Divergence time estimation on the topology obtained using a GF analysis was performed with MCMCtree as implemented in PAML v4.9 (Yang, 2007). This analysis was based on a reduced dataset of 430 genome fragments due to computational limitations of MCMCtree. Each genome fragment was randomly sampled from the 860 genome fragments, each 320 kb long. The large amount of data prevented the application of a Bayesian method. Two nodes were calibrated using hard lower and upper boundaries by setting both the left and right tail probabilities to 1e-300. The oldest *Natrix* fossil from Sardinia (Pliocene, 3.6 million years; Delfino et al., 2011) was included to constrain the divergence between the European mainland lineage represented by *N. h. helvetica* and the Sardinian subspecies *N. h. cetti* with a minimum of 3.6 million years and a maximum of 5.3 million years. The maximum corresponds to the end of the Miocene, coinciding with the termination of the Mediterranean salinity crisis and the interruption of the land bridge between Corsosardinia and mainland Italy (Krijgsman et al., 1999). The oldest record of a representative of the grass snake lineage, the extinct species *N. longivertebrata* (MN4 Mammal Neogene Zone 16.9–16.0 million years; Vasilyan et al., 2022), was used to calibrate the divergence between *N. tessellata* and the grass snake clade to a minimum of 16.0 Mya (Suppl. 1 Table S2). As the maximum for this node, the base of the Burdigalian (20.4 Mya), the respective Early Miocene age to which MN4 belongs, was chosen.

2.4.2. Network analysis and gene flow

We used SplitsTree v4.17.1 (Huson and Bryant, 2006) to construct a phylogenetic network from ML trees of 860 genome fragments to examine these for phylogenetic conflict. Then, we tested the support for different numbers of reticulations with SNAQ as implemented in PhyloNetworks v0.15.3 (Solís-Lemus et al., 2017) and PhyloNet v3.8.2 (Wen et al., 2018) using the individual GF trees. We tested the support of 0–5 reticulations using SNAQ, and of 0–10 reticulations using PhyloNet. The PhyloNetworks results were depicted using PhyloPlots v1.0.0 (https://github.com/cecileane/PhyloPlots.jl). Both analyses account for incomplete lineage sorting (ILS). Furthermore, to examine whether potential phylogenetic conflict is caused by gene flow, we ran a HyDe analysis. HyDe v0.4.3 was run on the concatenated reference-based assemblies and the reference genome of *N. h. helvetica* (Blischak et al., 2018).

2.5. Examining reference bias

To account for reference bias (e.g., ingroup or outgroup bias), we mapped all short read data against the assembly of *T. elegans* as described above for the reference-based assemblies. Then, we reran the GF analysis, dating the phylogeny, and all gene flow analyses using the new reference-based assemblies. Methods, results, and discussion are explained in Supplementary Materials File 2.

3. Results

3.1. Assembly

The assemblies of *N. n. natrix* and *N. h. helvetica* yielded genomes with a length of 1.74 Gbp and 1.78 Gbp, contig N50 lengths of 1.5 and 4.6 Mbp, and BUSCO gene set completeness of 91.0% and 92.6%, respectively (Suppl. 1 Tables S3-S5). The Blobtools analysis revealed negligible contamination (Suppl. 1 Figs. S1 and S2). Repeat masking identified > 50% of the genome sequence as repeats in both de-novo assemblies, with LINES, LTR elements, and DNA transposons being the most frequent classes, while unclassified repeats account for approximately 20% (Suppl. 1 Table S6). The reference-based assemblies are characterized by a high mapping quality against *N. h. helvetica* (>90%; Table 1), and hence sufficient for phylogenetic analyses and the estimation of divergence times.

3.2. Phylogeny and divergence times

The AU tree topology test for the best genome fragment length showed that 320 kbp is the shortest fragment length at which a topology can be significantly ($p < 0.05$) supported over a set of alternative topologies (Suppl. 1 Figs. S3 and S4). With this information, 860 orthologous genome fragments of 320 kb length were created for each taxon. The resulting multispecies coalescent tree (Fig. 2) recovered *N. maura* as the sister taxon to a clade containing the remaining *Natrix* species. Within this clade, *N. tessellata* was sister to the grass snake species, and within the grass snakes, *N. astreptophora* was the sister taxon to a clade consisting of the two subspecies of *N. helvetica*. This *N. astreptophora* + *N. helvetica* clade constituted the sister group to another clade containing the two subspecies of *N. natrix*.

A Quartet Frequency analysis (Suppl. 1 Fig. S5) revealed that almost all branches of the topology in Fig. 2 showed no conflict. The only uncertainty involved branch 1, i.e., the relationships among *N. h. helvetica*, *N. h. cetti*, and *N. astreptophora*.

The overall well-resolved phylogenetic topology was used to infer divergence times based on two fossil calibration points (Fig. 2). It was estimated that *Natrix* diverged from the outgroup *Thamnophis* in the late Eocene (35.8 Mya, 95% confidence interval [CI]: 28.8–52.1 Mya). Diversification of *Natrix* commenced when *N. maura* branched off in the early Miocene (21.5 Mya, CI: 17.3–31.3 Mya), closely followed by

Table 1
Qualimap report for all reference-based assemblies used in the genome fragment analysis. All species show good mapping qualities, even the outgroup mapped with over 90% of the reads. All values rounded to one digit after the decimal point.

Taxon	Coverage mean	Coverage std	GC [%]	Mapping quality mean	Insert size median	Mapped reads [%]
N. astreptophora	32.8	83.9	42.7	37.9	353	99.4
N. h. helvetica	32.1	82.0	44.5	46.1	1217	99.6
N. h. cetti	47.9	180.4	42.2	41.7	366	99.6
N. maura	22.9	277.5	43.3	34.7	327	98.7
N. n. natrix	28.6	140.8	42.6	37.2	314	99.2
N. n. vulgaris	21.3	226.5	43.2	37.9	237	94.8
N. tessellata	23.8	385.0	41.8	36.4	316	98.9
T. elegans	76.7	833.5	40.8	31.2	337	90.7

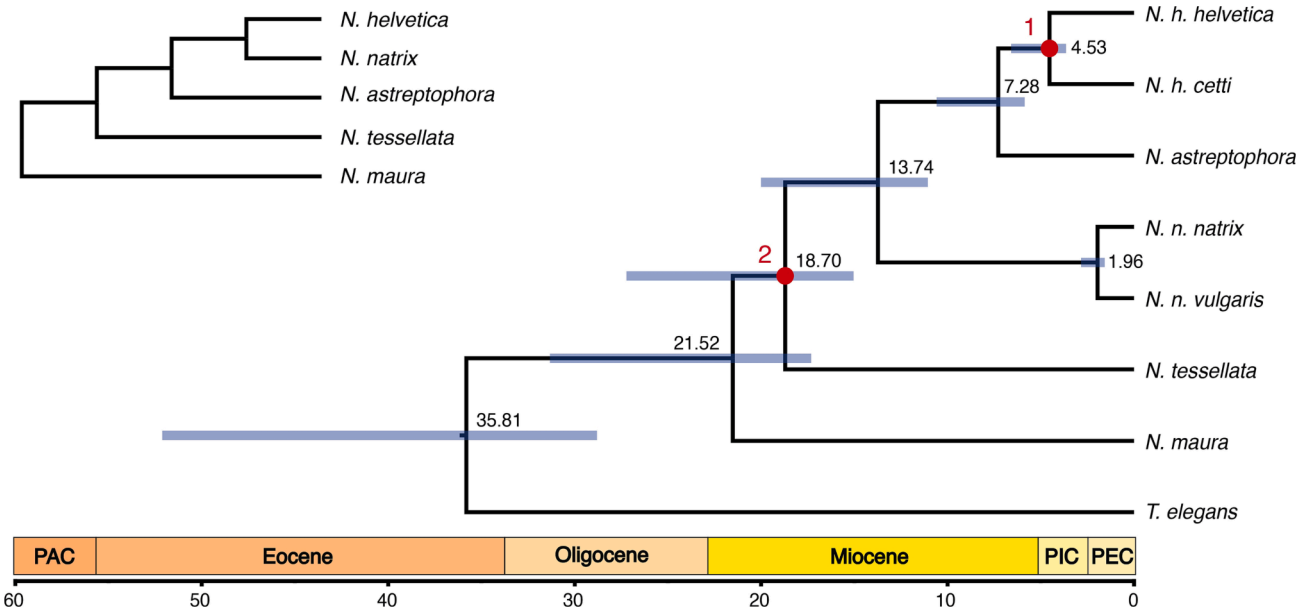

Fig. 2. Dated ML phylogeny using the sequence data from 430 genome fragments of 320 kb length in a MCMCTree analysis. The calibration point (1) for the split between *N. h. helvetica* and its subspecies was set to a minimum of 3.6 million years and a maximum of 5.3 million years. The second calibration point (2) for the divergence of the grass snake clade and *N. tessellata* was set to a minimum of 16.0 Mya and a maximum of 20.4 Mya. The fossil calibration points are marked with red dots. The error bars show the 95% confidence interval. The time bar is in million years. Values were rounded to the second decimal position. Abbreviations: PAC: Paleocene; PIC: Pliocene; PEC: Pleistocene. The small tree in the upper left corner shows the mitochondrial topology, according to Guicking et al. (2006) and Kindler et al. (2018b). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

N. tessellata at approx. 18.7 Mya (CI: 15.1–27.2 Mya). With the beginning of the Serravallian (middle Miocene; 13.7 Mya, CI: 11.1–20.0 Mya), *N. natrix* diverged as the most basal grass snake species. Its subspecies *N. n. vulgaris* separated from the nominotypical taxon in the early Pleistocene (approx. 2.0 Mya, CI: 1.6–2.9 Mya). In the upper Miocene (7.3 Mya, CI: 5.9–10.6 Mya), the Ibero-Maghrebian *N. astreptophora* split from the western European *N. helvetica.* The Sardinian subspecies of the latter, *N. h. cetti*, was estimated to have diverged in the Pliocene (4.5 Mya, CI: 3.6–6.6 Mya), with a confidence interval embracing the end of the Messinian salinity crises when the land connection between Corsosardinia and mainland Italy was interrupted.

3.3. Network analysis and gene flow

To test for phylogenetic conflict among the ML trees inferred from the 860 genome fragments, the 860 individual ML topologies were combined into a phylogenetic consensus network analysis (Suppl. 1 Fig. S6). A reticulation shows phylogenetic conflict over the relative placement of *N. astreptophora, N. h. helvetica*, and *N. h. cetti*, which is supported by 22% of the genome fragments. Another reticulation, supported by 2% of the genome fragments, also indicates conflict over the relative placement of *N. maura* and *N. tessellata* (Suppl. 1 Fig. S6). In both cases, missing or random phylogenetic information can be excluded as a reason for the reticulation because the p-AU analyses show that 320 kbp long genome fragments yield robust, well-supported topologies. Each branch of the ML trees based on the 860 genome fragments was significantly supported, indicating that the conflict is either caused by incomplete lineage sorting or introgression/gene flow.

The analysis using SNAQ supported three reticulations (Fig. 3; Suppl. 1 Table S7), suggesting introgression (1) between *N. h. helvetica* and *N. astreptophora,* (2) between the last common ancestor of *N. helvetica* + *N. astreptophora* and *N. natrix,* and (3) between the two water snakes *N. maura* and *N. tessellata*. PhyloNet did not support any reticulations, but in the light of the other gene flow analyses, we consider this an outlier (Suppl. 1 Table S8).

Despite the limited conflict in the network analyses and from quartet scoring, gene flow analysis using HyDe showed extensive gene flow, and there were only few taxon pairs between which no gene flow was identified: between *N. n. natrix* and *N. tessellata*, *N. h. cetti*, or *N. astreptophora*, between *N. astreptophora* and *N. n. vulgaris* or *N. h. cetti*, or between *N. tessellata* and *N. maura* (Fig. 4; Suppl. 1 Table S9).

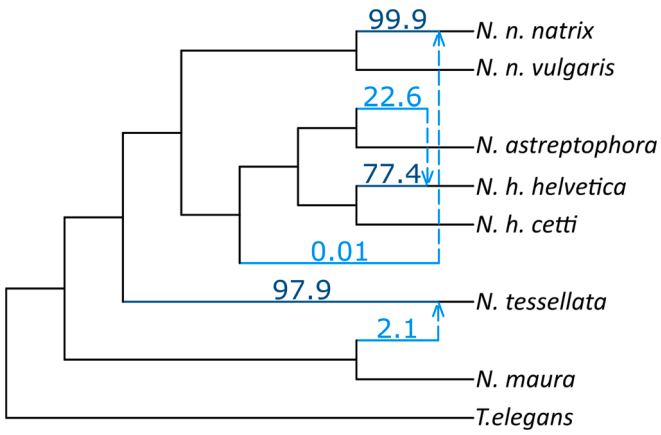

Fig. 3. The best-supported SNAQ result showed three reticulations. Analyses with a higher number of reticulations did not improve the log-likelihood. Reticulations and favored placements are shown in percent. *Natrix h. helvetica*, for example, was placed in 22.6% of cases as sister to *N. astreptophora* and in 77.4 % as sister to *N. h. cetti*.

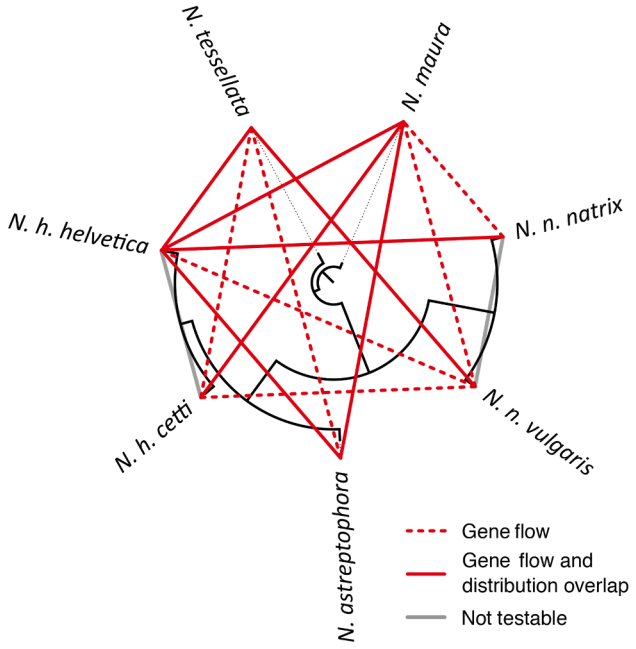

Fig. 4. Results of the HyDe analyses indicating complex and widespread hybridization among *Natrix* species. Red lines indicate significant evidence for hybridization between the connected taxa. Grey lines indicate species combinations where testing for hybridization was not possible due to the topology. Dotted lines link allopatric taxa; solid lines, sympatric taxa or parapatric taxa with abutting ranges with or without known hybrid zones. The phylogeny is shown as underlying black tree. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

4.1. Genomes

We sequenced and assembled the genomes of all five *Natrix* species (*N. astreptophora*, *N. helvetica*, *N. maura*, *N. natrix*, *N. tessellata*), with two species represented by two subspecies each (*N. n. natrix*, *N. n. vulgaris*; *N. h. helvetica*, *N. h. cetti*). We used this data to infer speciation and gene flow. Comparable previous studies using genomes focused mainly on mammals (e.g., Winter et al., 2018; Árnason et al., 2018; Coimbra et al., 2021), while studies on reptiles remain rare (Vilaça et al., 2021; Yang et al., 2021). Thus, our investigation provides valuable additional genome data from another major vertebrate lineage, squamate reptiles.

An important finding was the unexpected extent of interspecific gene flow, suggesting the evolution and maintenance of distinct species-level taxa despite gene flow and the continued capability for hybridization. This is in line with two other recent studies that reported hybridization in ancient reptile lineages (wall lizards and sea turtles, Vilaça et al., 2021; Yang et al., 2021; see below under 4.2 Phylogeny and hybridization).

The sizes of the *Natrix* genomes (approx. 1.7 Gbp) resemble previously published colubrid and natricid genomes (1.3–1.8 Gbp; Suppl. 1 Table S10). The relatively high contig N50 and BUSCO values indicate well-assembled de-novo genomes that are suitable for phylogenomic, speciation, and gene flow studies. In addition, our investigation includes a genome sequenced from the name-bearing museum type specimen of a grass snake taxon (*Natrix natrix vulgaris*, neotype, Natural History Museum Vienna, NMW 36405:2), which will serve as a valuable genomic standard for this subspecies.

4.2. Phylogeny and hybridization

With respect to the phylogenetic relationships of *Natrix* species, our results conflict with the previously published phylogenetic reconstructions using mtDNA sequences. According to mtDNA, *N. astreptophora* represents the sister lineage of a clade comprising *N. helvetica* and *N. natrix* (Guicking et al., 2006; Kindler et al., 2013). However, the analyses of genomic data reveal that *N. natrix* is the sister group of *N. astreptophora* + *N. helvetica* (Fig. 2). These conflicting relationships correspond to swapped branches in the phylogeny, suggesting mitochondrial capture due to hybridization.

The inference of ancient and ongoing hybridization is also supported by gene flow analyses. The network analysis (Suppl. 1 Fig. S6) revealed conflicting branching patterns for *N. astreptophora*, *N. h. helvetica*, and *N. h. cetti*, and *N. maura*, *N. tessellata,* and the outgroup *T. elegans.* The SNAQ analysis, which accounts for incomplete lineage sorting, confirmed these findings, and identified additional evidence for introgression between the three grass snake species (Fig. 3). The reticulation between *N. maura* and *N. tessellata* is particularly interesting because it is the only species combination for which no hybrids have been documented to date, not even from captive breeding (Mebert et al., 2011).

For most other species combinations, hybrids had been reported based on microsatellite and mtDNA sequence data: *Natrix astreptophora* × *N. h. helvetica* (Pokrant et al., 2016; Asztalos et al., 2020); *N. helvetica* × *N. natrix* (Kindler et al., 2017; Schultze et al., 2019, 2020; Asztalos et al., 2021b); *N. natrix* × *N. tessellata* (Asztalos et al., 2021a). Additionally, there is morphological evidence for hybridization between *N. astreptophora* and *N. maura* (de la Vega et al., 2021). Even the Sardinian subspecies *N. h. cetti*, thought to be completely isolated since the flooding of the Mediterranean Basin after the Messinian salinity crisis (5.3 mya, Kindler et al. 2018b), shows genomic evidence for hybridization with *N. natrix* (Figs. 3 and 4), suggestive of a very old hybridization event.

To examine the signals of gene flow further, we conducted a HyDe analysis, which has been shown to deliver robust results even under complex hybridization scenarios involving multiple hybridization events (Kong and Kubatko, 2021).

The first notable finding is evidence of hybridization between *N. maura* and all other taxa except for *N. tessellata* (Fig. 4). This suggests hybridization between *N. maura* and the ancestor of the extant grass snakes (*N. astreptophora*, *N. helvetica*, *N natrix*). However, there is also morphological evidence for extant hybridization of *N. maura* and *N. astreptophora* (de la Vega et al., 2021), underlining the possibility of multiple and continued hybridization events over an extended time period.

According to our data, *N. tessellata* seems to hybridize or have

hybridized in the past with all studied grass snake taxa except *N. n. natrix*. This finding is interesting because the *N. n. natrix* specimen we used for genome sequencing was collected from outside the distribution of *N. tessellata*. In contrast, our specimen of *N. n. vulgaris* came from a site where *N. tessellata* is also expected to occur. Although this might be a false negative for *N. n. natrix*, it could also indicate recent hybridization between sympatric *N. n. vulgaris* and *N. tessellata*, but not with *N. n. natrix*, which occurs completely outside the distribution range of *N. tessellata* (Asztalos et al., 2021a). This is in line with recent findings using microsatellite and mtDNA data, showing that *N. tessellata* continues to hybridize with sympatric *N. natrix* to the present day (Asztalos et al., 2021a).

However, *N. tessellata* also shows evidence for hybridization with fully allopatric taxa (*N. astreptophora*, *N. h. helvetica*, *N. h. cetti*), suggesting that the last common ancestor of *N. astreptophora* and *N. helvetica* hybridized with *N. tessellata*. Recent hybridization between *N. tessellata* and another subspecies of *N. helvetica* (*N. h. sicula*) cannot be excluded for mainland Italy. However, if the signal for hybridization in the allopatric taxa *N. astreptophora* or *N. h. helvetica* originated from this, introgression extending over hundreds of kilometers would have to be invoked. Moreover, in the face of the putative isolation of the island subspecies *N. h. cetti* for >5 million years (Kindler et al. 2018b), an ancient hybridization of *N. tessellata* with the last common ancestor of *N. astreptophora* and *N. helvetica* seems more likely.

For the three parapatric grass snake species, narrow hybridization zones have been characterized using microsatellite and mtDNA data (*Natrix astreptophora* × *N. h. helvetica*, Pokrant et al., 2016; Asztalos et al., 2020; *N. helvetica* × *N. natrix*, Kindler et al., 2017; Schultze et al., 2019, 2020; Asztalos et al., 2021b). Each hybrid zone is typically less than 50 km wide and characterized by steep genetic clines, with parental genotypes occurring syntopically with hybrids. Our genome-based results are in line with these findings, with the continental subspecies *N. h. helvetica* showing evidence for gene flow with *N. natrix* and *N. astreptophora*. However, the island taxon *N. h. cetti*, thought to be isolated since the end Miocene, also shows evidence for hybridization with *N. natrix*, suggestive of a very old hybrid signature. This implies that the genomes of the individual *Natrix* species are impacted by a series of subsequent hybridization events that contribute to a continued enrichment with foreign DNA.

Thus, the present evidence suggests a complex evolutionary history of the grass and water snakes, in which hybridization substantially contributed to the current genomic identity. Considering recent studies using microsatellites and mtDNA data (Pokrant et al., 2016; Kindler et al., 2017; Schultze et al., 2019, 2020; Asztalos et al., 2020, 2021b), there seems to be an intricate history of ancient and recent hybridization events. Our samples used for genome sequencing were collected in sites far away from the present-day hybrid zones of the respective parapatric grass snake species. Therefore, our data should be largely unaffected by very recent hybridization events of grass snake species (Fig. 1). To fully comprehend these evolutionary processes, gene flow on the population level of all species needs to be studied. The genomes presented here will be a valuable resource for such a population-genomic analysis.

Our genomic analysis of the genus *Natrix* thus demonstrates that this genus may be an ideal model clade for studying speciation under gene flow. The individual *Natrix* species have a complex history of introgression over extended timescales until the present day and a wide range of divergence ages. The inferred divergence ages of 21.5–7.3 Mya of the extant *Natrix* species (Fig. 2) considerably predate the onset of the Pleistocene glaciations (Knudsen et al., 2020), indicating that the Pleistocene climatic fluctuations were not the primary drivers of the diversification. The age of the *Natrix* lineages involved in hybridization is considerable and exceeds those of Mediterranean wall lizards (*Podarcis* spp.), for which recently similar intricate patterns of genomic introgression across species boundaries have been inferred based on genome sequence data (Yang et al., 2021). For marine turtles even older lineages have been suggested to hybridize (Vilaça et al., 2021). However, the divergence time estimates of this study conflict with much younger dates inferred by another recent study using a fossil-calibrated molecular clock (Thomson et al., 2021). In any case, the divergence ages of the *Natrix* lineages involved in hybridization are much older than those of mammals known to hybridize (10 Mya in whales and 5 Mya in bears; Kumar et al., 2017; Árnason et al., 2018).

Our inferred divergence dates for *Natrix* species, based on genomic data, closely resemble those based on mtDNA divergence (Guicking et al., 2006; Kindler et al., 2018a, 2018b), despite the phylogenetic conflict within the grass snake clade. Both molecular clocks support that divergence of *Natrix* commenced in the early Miocene. The paleontological record (Rage and Augé, 1993; Ivanov, 2001) and extant diversity (Deepak et al., 2022) suggest that natricids originated in Asia and that the *Natrix* lineage dispersed to the western Palearctic, where the genus *Natrix* diversified.

In line with our divergence time estimate, Guicking et al. (2006) suggested that the most basal species, *N. maura*, evolved as a consequence of a range expansion into Africa approximately 20 Mya, when the Afro-Arabian and Eurasian plates collided. Shortly later, *N. tessellata* seems to have diverged. Divergence of *N. tessellata* and diversification of the grass snake clade (*N. astreptophora*, *N. helvetica*, *N. natrix*) seem to correlate with the Miocene Climate Optimum during the Middle Miocene Transition, a period of many climatic instabilities (Flower and Kennett, 1994), which could have led to repeated range restrictions of the possibly widespread ancestor of *N. tessellata* and grass snakes.

5. Conclusions

The genomic identity of the individual *Natrix* species suggests a history of subsequent hybridization events that contributed to a continued enrichment with foreign DNA, despite species divergence. The transition between subspecies still capable of wide-ranging gene flow and distinct species that maintain their identities despite continued hybridization is not well understood. Further population-level studies focusing on a characterization of this transition using whole genome data may contribute to a better understanding of divergence processes under gene flow, making *Natrix* species an interesting model for evolutionary biology.

CRediT authorship contribution statement

Yannis Schöneberg: Data curation, Formal analysis, Investigation, Visualization, Writing – original draft. **Sven Winter:** Data curation, Investigation, Supervision, Writing – review & editing. **Oscar Arribas:** Resources, Writing – review & editing. **Matteo Riccardo Di Nicola:** Resources. **Maya Master:** Resources. **John Benjamin Owens:** Resources. **Michail Rovatsos:** Resources. **Wolfgang Wüster:** Resources, Writing – review & editing. **Axel Janke:** Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. **Uwe Fritz:** Project administration, Resources, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All generated data are accessible at NCBI through BioProject PRJNA803021.

Acknowledgements

The present study was funded by the Centre for Translational

Biodiversity Genomics (LOEWE-TBG) through the program "LOEWE–Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz" of the Hessian Ministry of Higher Education, Research, and Arts. We are grateful to Georgios Georgalis, Krister Smith, and Martin Ivanov for insightful discussions of fossil calibration points, and Nico Fuhrmann for his opinions on our results. Henrik Bringsøe donated two photos shown in Fig. 1. We especially thank the anonymous reviewers for their constructive feedback to our study.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2023.107787.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Árnason, Ú., Lammers, F., Kumar, V., Nilsson, M.A., Janke, A., 2018. Whole-genome sequencing of the blue whale and other rorquals finds signatures for introgressive gene flow. Sci. Adv. 4, eaap9873. https://doi.org/10.1126/sciadv.aap9873.

Arnold, M.L., 2015. Divergence with genetic exchange. Oxford University Press, Oxford, UK.

Asztalos, M., Schultze, N., Ihlow, F., Geniez, P., Berroneau, M., Delmas, C., Guiller, G., Legentilhomme, J., Kindler, C., Fritz, U., 2020. How often do they do it? An in-depth analysis of the hybrid zone of two grass snake species (*Natrix astreptophora* and *Natrix helvetica*). Biol. J. Linn. Soc. 131, 756–773. https://doi.org/10.1093/biolinnean/blaa152.

Asztalos, M., Ayaz, D., Bayrakcı, Y., Afsar, M., Tok, C.V., Kindler, C., Jablonski, D., Fritz, U., 2021a. It takes two to tango – Phylogeography, taxonomy and hybridization in grass snakes and dice snakes (Serpentes: Natricidae: *Natrix natrix*, *N. tessellata*). Vert. Zool. 71, 813–834. https://doi.org/10.3897/vz.71.e76453.

Asztalos, M., Glaw, F., Franzen, M., Kindler, C., Fritz, U., 2021b. Transalpine dispersal: Italian barred grass snakes in southernmost Bavaria—This far but no further! J. Zool. Syst. Evol. Res. 59, 1136–1148. https://doi.org/10.1111/jzs.12471.

Bao, Z., Eddy, S.R., 2002. Automated de novo identification of repeat sequence families in sequenced genomes. Genome Res. 12, 1269–1276.

Blischak, P.D., Chifman, J., Wolfe, A.D., Kubatko, L.S., 2018. HyDe: A Python package for genome-scale hybridization detection. Syst. Biol. 67, 821–829. https://doi.org/10.1093/sysbio/syy023.

Burbrink, F.T., Gehara, M., McKelvy, A.D., Myers, E.A., 2020. Resolving spatial complexities of hybridization in the context of the gray zone of speciation in North American ratsnakes (*Panterophis obsoletus* complex). Evolution 75, 260–277. https://doi.org/10.1111/evo.14141.

Coimbra, R.T.F., Winter, S., Kumar, V., Koepfli, K.-P., Gooley, R.M., Dobrynin, P., Fennessy, J., Janke, A., 2021. Whole-genome analysis of giraffe supports four distinct species. Curr. Biol. 31, 2929–2938.e5. https://doi.org/10.1016/j.cub.2021.04.033.

Coyne, J.A., Orr, H.A., 2004. Speciation. Sinauer Associates, Sunderland, MA.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H., 2021. Twelve years of SAMtools and BCFtools. GigaScience 10, giab008. https://doi.org/10.1093/gigascience/giab008.

Darwin, C., 1859. The origin of species by means of natural selection. John Murray, London.

de la Vega, J.P.G., Barnestein, J.A., del Mármol, G.M., Mebert, K., 2021. Hybridization between *Natrix astreptophora* and *Natrix maura*: Potential cases from Andalusia. Spain. Bol. Asoc. Herpetol. Esp. 32, 45–53.

Deepak, V., Cooper, N., Poyarkov, N.A., Kraus, F., Burin, G., Das, A., Narayanan, S., Streicher, J.W., Smith, S.-J., Gower, D.J., 2022. Multilocus phylogeny, natural history traits and classification of natricine snakes (Serpentes: Natricinae). Zool. J. Linn. Soc. 195, 279–298. https://doi.org/10.1093/zoolinnean/zlab099.

Delfino, M., Bailon, S., Pitruzzella, G., 2011. The Late Pliocene amphibians and reptiles from "Capo Mannu D1 Local Fauna" (Mandriola, Sardinia, Italy). Geodiversitas 33, 357–382. https://doi.org/10.5252/g2011n2a10.

Dufresnes, C., Brelsford, A., Jeffries, D.L., Mazepa, G., Suchan, T., Canestrelli, D., Nicieza, A., Fumagalli, L., Dubey, S., Martínez-Solano, I., Litvinchuk, S.N., Vences, M., Perrin, N., Crochet, P.-A., 2021. Mass of genes rather than master genes underlie the genomic architecture of amphibian speciation. Proc. Natl. Acad. Sci. U.S.A. 118, e2103963118. https://doi.org/10.1073/pnas.2103963118.

Figueiró, H.V., Li, G., Trindade, F.J., Assis, J., Pais, F., Fernandes, G., Santos, S.H.D., Hughes, G.M., Komissarov, A., Antunes, A., Trinca, C.S., Rodrigues, M.R., Linderoth, T., Bi, K., Silveira, L., Azevedo, F.C.C., Kantek, D., Ramalho, E., Brassaloti, R.A., Villela, P.M.S., Nunes, A.L.V., Teixeira, R.H.F., Morato, R.G., Loska, D., Saragüeta, P., Gabaldón, T., Teeling, E.C., O'Brien, S.J., Nielsen, R., Coutinho, L.L., Oliveira, G., Murphy, W.J., Eizirik, E., 2017. Genome-wide signatures of complex introgression and adaptive evolution in the big cats. Sci. Adv. 3, e1700299.

Flower, B.P., Kennett, J.P., 1994. The middle Miocene climatic transition: East Antarctic ice sheet development, deep ocean circulation and global carbon cycling. Palaeogeogr. Palaeoclimatol. Palaeoecol. 108, 537–555. https://doi.org/10.1016/0031-0182(94)90251-8.

Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., Smit, A.F., 2019. RepeatModeler2: Automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. USA 117, 9451–9457. https://doi.org/10.1073/pnas.1921046117.

Fritz, U., Schmidtler, J.F., 2020. The Fifth Labour of Heracles: Cleaning the Linnean stable of names for grass snakes (*Natrix astreptophora*, *N. helvetica*, *N. natrix* sensu stricto). Vert. Zool. 70, 621–665. https://doi.org/10.26049/VZ70-4-2020-07.

Geniez, P., 2015. Serpents d'Europe, d'Afrique du Nord et du Moyen-Orient. Delachaux et Niestlé, Paris.

Gruschwitz, M., Lenz, S., Mebert, K., Laňka, V., 1999. *Natrix tessellata* (Laurenti, 1768) – Würfelnatter. In: Böhme, W. (Ed.), Handbuch der Reptilien und Amphibien Europas. Band 3/IIA: Schlangen II. Aula-Verlag, Wiebelsheim, pp. 581–644.

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

Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: Quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075. https://doi.org/10.1093/bioinformatics/btt086.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267. https://doi.org/10.1093/molbev/msj030.

Ivanov, M., 2001. Changes in the composition of the European snake fauna during the Early Miocene and at the Early/Middle Miocene transition. Paläontol. Z. 74, 563–573. https://doi.org/10.1007/BF02988162.

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

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

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

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

Kindler, C., de Pous, P., Carranza, S., Beddek, M., Geniez, P., Fritz, U., 2018b. Phylogeography of the Ibero-Maghrebian red-eyed grass snake (*Natrix astreptophora*). Org. Divers. Evol. 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. Vert. Zool. 68, 253–267.

Knudsen, M.F., Nørgaard, J., Grischott, R., Kober, F., Egholm, D.L., Hansen, T.M., Jansen, J.D., 2020. New cosmogenic nuclide burial-dating model indicates onset of major glaciations in the Alps during Middle Pleistocene Transition. Earth Planet. Sci. Lett. 549, 116491 https://doi.org/10.1016/j.epsl.2020.116491.

Kolmogorov, M., Yuan, J., Lin, Y., Pevzner, P.A., 2019. Assembly of long, error-prone reads using repeat graphs. Nat. Biotechnol. 37, 540–546. https://doi.org/10.1038/s41587-019-0072-8.

Kong, S., Kubatko, L.S., 2021. Comparative performance of popular methods for hybrid detection using genomic data. Syst. Biol. 70, 891–907. https://doi.org/10.1093/sysbio/syaa092.

Korneliussen, T.S., Albrechtsen, A., Nielsen, R., 2014. ANGSD: analysis of next generation sequencing data. BMC Bioinform. 15, 356. https://doi.org/10.1186/s12859-014-0356-4.

Krijgsman, W., Hilgen, F.J., Raffi, I., Sierro, F.J., Wilson, D.S., 1999. Chronology, causes and progression of the Messinian salinity crisis. Nature 400, 652–655. https://doi.org/10.1038/23231.

Kriventseva, E.V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F.A., Zdobnov, E.M., 2019. OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic Acids Res. 47, D807–D811. https://doi.org/10.1093/nar/gky1053.

Kumar, V., Lammers, F., Bidon, T., Pfenninger, M., Kolter, L., Nilsson, M.A., Janke, A., 2017. The evolutionary history of bears is characterized by gene flow across species. Sci. Rep. 7, 46487. https://doi.org/10.1038/srep46487.

Laetsch, D.R., Blaxter, M.L., 2017. BlobTools: Interrogation of genome assemblies. F1000Research 6, 1287. https://doi.org/10.12688/f1000research.12232.1.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923.

Li, H., 2018. Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Mayr, E., 1942. Systematics and the Origin of Species from the Viewpoint of a Zoologist. Columbia University Press, New York, N.Y.

Mebert, K., Trapp, B., Dall'Asta, A., Velenský, P., Böhme, W., 2011. Hybrids between *Natrix tessellata* and *N. natrix/maura*. Mertensiella 18, 154–156.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: New models and efficient methods for

phylogenetic inference in the genomic era. Mol. Biol. Evol. 37, 1530–1534. https://doi.org/10.1093/molbev/msaa015.
Nater, A., Mattle-Greminger, M.P., Nurcahyo, A., Nowak, M.G., de Manuel, M., Desai, T., Groves, C., Pybus, M., Sonay, T.B., Roos, C., Lameira, A.R., Wich, S.A., Askew, J., Davila-Ross, M., Fredriksson, G., de Valles, G., Casals, F., Prado-Martinez, J., Goossens, B., Verschoor, E.J., Warren, K.S., Singleton, I., Marques, D.A., Pamungkas, J., Perwitasari-Farajallah, D., Rianti, P., Tuuga, A., Gut, I.G., Gut, M., Orozco-terWengel, P., van Schaik, C.P., Bertranpetit, J., Anisimova, M., Scally, A., Marques-Bonet, T., Meijaard, E., Krützen, M., 2017. Morphometric, behavioral, and genomic evidence for a new orangutan species. Curr. Biol. 27, 3487–3498.e10. https://doi.org/10.1016/j.cub.2017.09.047.
Nosil, P., 2008. Speciation with gene flow could be common. Mol. Ecol. 17, 2103–2106. https://doi.org/10.1111/j.1365-294X.2008.03715.x.
Okonechnikov, K., Conesa, A., García-Alcalde, F., 2016. Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 32, 292–294. https://doi.org/10.1093/bioinformatics/btv566.
Pabijan, M., Zieliński, P., Dudek, K., Stuglik, M., Babik, W., 2017. Isolation and gene flow in a speciation continuum in newts. Mol. Phylogenet. Evol. 116, 1–12. https://doi.org/10.1016/j.ympev.2017.08.003.
Pavón-Vázquez, C.L., Brennan, I.G., Keogh, J.S., 2021. A comprehensive approach to detect hybridization sheds light on the evolution of the earth's largest lizards. Syst. Biol. 70, 877–890. https://doi.org/10.1093/sysbio/syaa102.
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*. Biol. J. Linn. Soc. 118, 873–888. https://doi.org/10.1111/bij.12782.
Pöschel, J., Heltai, B., Graciá, E., Franch Quintana, M., Velo-Antón, G., Arribas, O., Valdeón, A., Wink, M., Fritz, U., Vamberger, M., 2018. Complex hybridization patterns in European pond turtles (*Emys orbicularis*) in the Pyrenean Region. Sci. Rep. 8, 15925. https://doi.org/10.1038/s41598-018-34178-0.
Price, A.L., Jones, N.C., Pevzner, P.A., 2005. De novo identification of repeat families in large genomes. Bioinformatics 21 (Suppl. 1), i351–i358. https://doi.org/10.1093/bioinformatics/bti1018.
Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033.
Rage, J.-C., Augé, M., 1993. Squamates from the Cainozoic of the western part of Europe. A review. Rev. Paleobiol. 7, 199–216.
Ruan, J., Li, H., 2020. Fast and accurate long-read assembly with wtdbg2. Nat. Methods 17, 155–158. https://doi.org/10.1038/s41592-019-0669-3.
Sayyari, E., Whitfield, J.B., Mirarab, S., 2018. DiscoVista: Interpretable visualizations of gene tree discordance. Mol. Phylogenet. Evol. 122, 110–115. https://doi.org/10.1016/j.ympev.2018.01.019.
Schätti, B., 1999. *Natrix maura* (Linnaeus, 1758) – Vipernnatter. In: Böhme, W. (Ed.), Handbuch der Reptilien und Amphibien Europas. Band 3/IIA: Schlangen II. Aula-Verlag, Wiebelsheim, pp. 483–503.
Schield, D.R., Perry, B.W., Adams, R.H., Card, D.C., Jezkova, T., Pasquesi, G.I.M., Nikolakis, Z.L., Row, K., Meik, J.M., Smith, C.F., Mackessy, S.P., Castoe, T.A., 2019. Allopatric divergence and secondary contact with gene flow: A recurring theme in rattlesnake speciation. Biol. J. Linn. Soc. 128, 149–169. https://doi.org/10.1093/biolinnean/blz077.
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.
Schultze, N., Spitzweg, C., Corti, C., Delaugerre, M., Di Nicola, M.R., Geniez, P., Lapini, L., Liuzzi, C., Lunghi, E., Novarini, N., Picariello, O., Razzetti, E., Sperone, E., Stellati, L., Vignoli, L., Asztalos, M., Kindler, C., Vamberger, M., Fritz, U., 2020. Mitochondrial ghost lineages blur phylogeography and taxonomy of *Natrix helvetica* and *N. natrix* in Italy and Corsica. Zool. Scr. 49, 395–411. https://doi.org/10.1111/zsc.12417.
Seppey, M., Manni, M., Zdobnov, E.M., 2019. BUSCO: Assessing genome assembly and annotation completeness. Methods Mol. Biol. 1962, 227–245. https://doi.org/10.1007/978-1-4939-9173-0_14.
Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51, 492–508. https://doi.org/10.1080/10635150290069913.
Sindaco, R., Venchi, A., Grieco, C., 2013. The Reptiles of the Western Palearctic, Volume 2: Annotated Checklist and Distributional Atlas of the Snakes of Europe, North Africa, Middle East and Central Asia, with an Update to Volume 1. Edizioni Belvedere, Latina.
Smit, A., Hubley, R., Green, P., 2013-2015. RepeatMasker Home Page. http://www.repeatmasker.org/. Accessed 10 December 2020.
Solís-Lemus, C., Bastide, P., Ané, C., 2017. PhyloNetworks: a package for phylogenetic networks. Mol. Biol. Evol. 34, 3292–3298. https://doi.org/10.1093/molbev/msx235.
Speybroeck, J., Beukema, W., Dufresnes, C., Fritz, U., Jablonski, D., Lymberakis, P., Martínez-Solano, I., Razzetti, E., Vamberger, M., Vences, M., Vörös, J., Crochet, P.-A., 2020. Species list of the European herpetofauna – 2020 update by the Taxonomic Committee of the Societas Europaea Herpetologica. Amphibia-Reptilia 41, 139–189. https://doi.org/10.1163/15685381-bja10010.
Supple, M.A., Shapiro, B., 2018. Conservation of biodiversity in the genomics era. Genome Biol. 19, 131. https://doi.org/10.1186/s13059-018-1520-3.
Thomson, R.C., Spinks, P.Q., Shaffer, H.B., 2021. A global phylogeny of turtles reveals a burst of climate-associated diversification on continental margins. Proc. Natl. Acad. Sci. USA 118, e2012215118. https://doi.org/10.1073/pnas.2012215118.
Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., Prins, P., 2015. Sambamba: Fast processing of NGS alignment formats. Bioinformatics 31, 2032–2034. https://doi.org/10.1093/bioinformatics/btv098.
Vamberger, M., Stuckas, H., Sacco, F., D'Angelo, S., Arculeo, M., Cheylan, M., Corti, C., Lo Valvo, M., Marrone, F., Wink, 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*). Zool. Scr. 44, 233–249. https://doi.org/10.1111/zsc.12102.
Vaser, R., Šikić, M., 2021. Time- and memory-efficient genome assembly with Raven. Nat. Comput. Sci. 1, 332–336. https://doi.org/10.1038/s43588-021-00073-4.
Vaser, R., Sović, I., Nagarajan, N., Šikić, M., 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 27, 737–746. https://doi.org/10.1101/gr.214270.116.
Vasilyan, D., Čerňanský, A., Szyndlar, Z., Mörs, T., 2022. Amphibian and reptilian fauna from the early Miocene of Echzell Germany. . Foss. Rec. 25, 99–145. https://doi.org/10.3897/fr.25.83781.
Vilaça, S.T., Piccinno, R., Rota-Stabelli, O., Gabrielli, M., Benazzo, A., Matschiner, M., Soares, L.S., Bolten, A.B., Bjorndal, K.A., Bertorelle, G., 2021. Divergence and hybridization in sea turtles: inferences from genome data show evidence of ancient gene flow between species. Mol. Ecol. 30, 6178–6192. https://doi.org/10.1111/mec.16113.
Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., Earl, A.M., 2014. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PloS One 9, e112963.
Wen, D., Yu, Y., Zhu, J., Nakhleh, L., 2018. Inferring phylogenetic networks using PhyloNet. Syst. Biol. 67, 735–740. https://doi.org/10.1093/sysbio/syy015.
Winter, S., Fennessy, J., Janke, A., 2018. Limited introgression supports division of giraffe into four species. Ecol. Evol. 8, 10156–10165. https://doi.org/10.1002/ece3.4490.
Yang, Z., 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591. https://doi.org/10.1093/molbev/msm088.
Yang, W., Feiner, N., Pinho, C., While, G.M., Kaliontzopoulou, D., Harris, D.J., Salvi, D., Uller, T., 2021. Extensive introgression and mosaic genomes of Mediterranean endemic lizards. Nat. Comm. 12, 2762. https://doi.org/10.1038/s41467-021-22949-9.
Zachos, F.E., 2016. Species Concepts in Biology. Springer International, Switzerland.
Zhang, C., Rabiee, M., Sayyari, E., Mirarab, S., 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinform. 19 (Suppl. 6), 153. https://doi.org/10.1186/s12859-018-2129-y.
Zhou, X., Guang, X., Di Sun, Xu, S., Li, M., Seim, I., Jie, W., Yang, L., Zhu, Q., Xu, J., Gao, Q., Kaya, A., Dou, Q., Chen, B., Ren, W., Li, S., Zhou, K., Gladyshev, V.N., Nielsen, R., Fang, X., Yang, G., 2018. Population genomics of finless porpoises reveal an incipient cetacean species adapted to freshwater. Nat. Commun. 9, 1276. https://doi.org/10.1038/s41467-018-03722-x.