

## Article

# Genomic and mitochondrial evidence of ancient isolations and extreme introgression in the four-lined snake

Evanthia THANOU <sup>a,b,\*</sup>, Panagiotis KORNILIOS<sup>a,b,c</sup>, Petros LYMBERAKIS<sup>d</sup>, and Adam D. LEACHÉ<sup>a</sup>

<sup>a</sup>Department of Biology, University of Washington, Seattle, WA, USA, <sup>b</sup>The Molecular Ecology Backshop, Loutraki, Greece, <sup>c</sup>Institute of Evolutionary Biology (CSIC - Universitat Pompeu Fabra), Barcelona, Spain and <sup>d</sup>Natural History Museum of Crete, University of Crete, Irakleio, Greece

\*Address correspondence to Evanthia Thanou. E-mail: thanouevanthia@gmail.com.

Handling editor: Wen Wang

Received on 7 January 2019; accepted on 11 April 2019

## Abstract

Comparing mitochondrial and genomic phylogenies is an essential tool for investigating speciation processes, because each genome carries different inheritance properties and evolutionary characteristics. Furthermore, mitonuclear discordance may arise from ecological adaptation, historic isolation, population size changes, and sex-biased dispersal. Closely related taxa are expected to experience gene flow; however, this may not be true for insular populations or populations isolated in refugia. The four-lined snake *Elaphe quatuorlineata* has a fragmented distribution, separating populations of the Italian and Balkan Peninsulas, whereas several insular Aegean populations of significantly smaller body size (Cyclades island group and Skyros Island, Greece) are currently considered distinct subspecies. We constructed the species-tree phylogeny of this species utilizing genome-wide single nucleotide polymorphisms and a gene-tree based on complete cytochrome b sequences, aiming to detect convergence and discrepancies between biparentally and maternally inherited genomes. Population structuring, phylogenetic patterns and migration events among geographically defined lineages supported our hypothesis of isolation in multiple sub-refugia. Where biogeographical barriers did not restrict subsequent dispersal, extensive genetic exchange occurred between mainland Balkan populations. This process has led to the mitochondrial sweep of an ancestral mitolineage that survived only in peripheral (East Greece) and insular populations (North Cyclades and Skyros). The Central Cyclades represent an ancient lineage for both molecular markers that emerged almost 3.3 Mya. Considering their distinct morphology, insular *E. quatuorlineata* populations should be the future focus of an extensive sampling, especially since the mitonuclear discordance observed in this species could be related to ecological adaptations, such as the island-dwarfism phenomenon.

**Key words:** Aegean Islands, *Elaphe quatuorlineata* species-tree, genome-wide SNPs, mitochondrial introgression, phylogeography, refugia within refugia

The molecular traits of the mitochondrial genome (mtDNA), mainly the absence of recombination, the uniparental inheritance and a relatively constant mutation rate, have led to its extensive use in the reconstruction of time-calibrated, maternal phylogenies for the past

three decades (Ballard and Whitlock 2004). In recent years, massively parallel sequencing allowed for the collection of an overwhelming amount of genome-wide data that was expected to complement the mitochondrial reconstructions. However, it seems that the

discordance between the maternal and biparental phylogenies is as common as their congruence (Rubinoff and Holland 2005; Toews and Brelsford 2012) and it is not always straightforward to understand the underlying evolutionary histories, when different datasets support incongruous patterns (Sloan et al. 2017). Due to the different inheritance properties and evolutionary characteristics of each genome, comparing mtDNA and genomic phylogenies is an essential tool for representing the complexity of speciation processes (Rubinoff and Holland 2005), especially since mitonuclear incompatibilities have been identified as an important reproductive barrier (Sloan et al. 2017).

Speciation is typically thought to occur in the absence of gene flow, nevertheless, hybridization and genomic introgression between closely related species is far from rare (Mallet 2005). On the other hand, conspecific populations are expected to experience extensive gene flow but if they are respectively isolated, they may undergo evolutionary processes that promote speciation by favoring small population size and the fixation of rare alleles (founder and bottleneck events), or local ecological adaptations. This particularly applies for insular populations or populations that survived in distinct refugia, as well as populations living in differentiated environments (Bierne et al. 2013). Biparentally inherited genetic markers are present as two different alleles in heterozygotic individuals, thus evolutionary processes (e.g., selection or population-size changes) and historical events (e.g., isolation of populations or migration events) affect the distribution of allele frequencies and analyses of such markers can trace back this effect and reveal fine patterns in population clustering and genetic admixture.

Gene flow is expected to be restored after isolated populations reach secondary contact, but genetic exchange between them is usually not symmetrical, showing a directional movement from one population into the other (Toews and Brelsford 2012; Sloan et al. 2017). Moreover, as the equilibrium of migration–selection–drift among populations changes through space and time, patterns of genetic heterogeneity may not necessarily be reflected across the whole genome (Bierne et al. 2013; Ravinet et al. 2017). For example, introgression of mtDNA from one population into another may occur without nuclear gene flow (Toews and Brelsford 2012). This may be a consequence of selective pressure for a locally adapted mitotype (Ballard and Whitlock 2004; Chan and Levin 2005) or/and the effect of non-adaptive mechanisms, such as disparities in range size or abundance between hybridizing populations (Funk and Omland 2003), rapid lineage sorting of the mtDNA due to its four-fold smaller effective population size (Chan and Levin 2005; Currat et al. 2008) and sex-biased dispersal (Excoffier et al. 2008; Toews and Brelsford 2012).

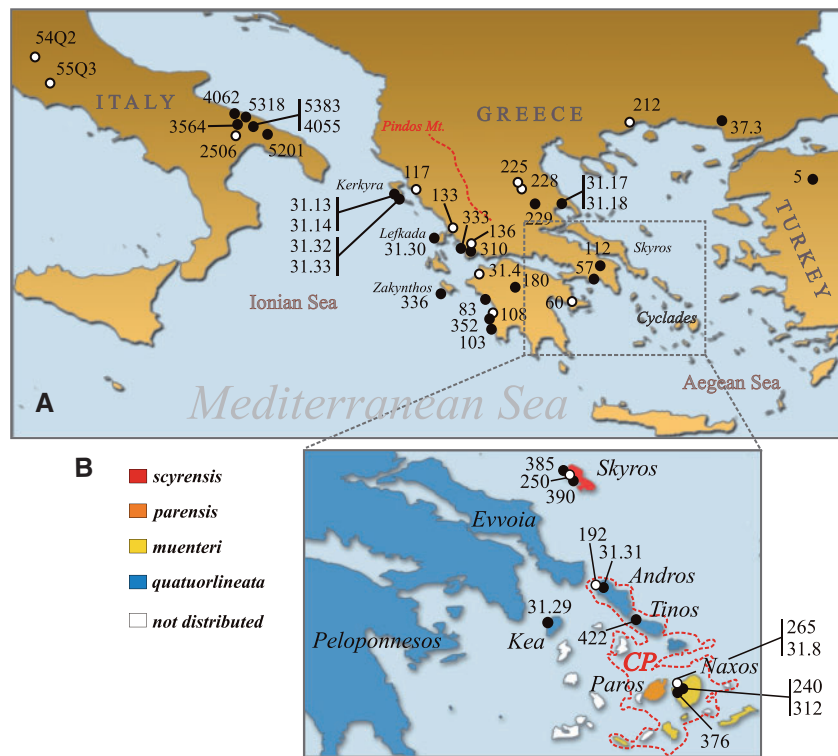
Phylogeographic studies based on mitochondrial and nuclear markers suggest that secondary contact between parapatric forms can lead to genetic introgression across the species boundaries and in several cases mtDNA introgression was found between European populations of the grass snake species-complex (Kindler et al. 2013), Italian fire salamanders (Bisconti et al. 2018), and asp vipers (Barbanera et al. 2009). Most of the divergence between incipient European species has occurred during a succession of contraction/expansion cycles, dramatically affecting the range, size, and isolation of their ancestral populations (Hewitt 2011). Climate oscillations related to the glacial/interglacial cycles of the Quaternary repeatedly forced several European species to shift their distributions, driving the divergence of closely related lineages during isolation in refugia, such as the southern parts of the European Peninsulas, and their postglacial secondary contact (Hewitt 2011; Nieto-Feliner 2011).

In this sense, European species provide ideal natural systems to study complex speciation processes, such as the genetic exchange between incompletely isolated taxa that reach secondary contact. Particularly interesting are those species that 1) are composed of recently evolved and closely related lineages, 2) include populations both widely distributed and regionally confined, covering a range of ecological conditions and differing in the degree of geographic isolation and 3) present marked differentiation in phenotypic traits possibly related to fitness (Friis et al. 2016). Amphibians and reptiles are best candidate study-organisms because ectotherms are expected to be more severely affected by past climatic oscillations, but also because closely related taxa found syntopically can exhibit niche-divergence or differentiation in life-history traits that reduces gene flow between them (Wielstra and Arntzen 2012).

The four-lined snake, *Elaphe quatuorlineata* (Lacépède, 1789), is a protected European species (Crnobrnja-Isailović et al. 2009) that shows a fragmented distribution with a gap separating populations in the southern parts of the Italian and Balkan Peninsulas. It also occurs in several Aegean Islands and some of these insular populations show unique morphological features, regarding the color patterns of the young snakes, body size, and number of scales (Cattaneo 1998; 1999). Based on these differences, several subspecies are described (Figure 1b); *E. q. parensis* (Cattaneo 1999) (endemic of Paros Island, Cyclades island group, central Aegean), *E. q. muenteri* (Bedriaga 1882) (several other Cycladic islands), *E. q. scyrensis* (Cattaneo 1999) (endemic of Skyros Island, northern Aegean), and *E. q. quatuorlineata* (Bonnaterre 1790) which is found on the mainland and adjacent islands. Animals described as *muenteri* and *scyrensis*, and to some extent *parensis*, are smaller (120–140 cm, up to 160 cm for *parensis*; Cattaneo 1998, 1999), whereas *quatuorlineata* is large and often exceeds 200 cm in length.

The first insight on the four-lined snake's phylogeny was based on mitochondrial markers and corroborated the existence of cryptic speciation (Kornilios et al. 2014). The three major maternal lineages corresponded to the Cyclades islands, Skyros Island and Italy/Balkans. Several subclades were also found; one for each Peninsula and within the southern Balkans. Finally, molecular evidence suggested past population restriction and subsequent rapid expansion for this species and especially its southern Balkan populations. So the mitochondrial phylogeny, although unresolved in some cases, showed a geographical pattern linked to known Italian and Balkan refugia. Additionally it supported a deep mitochondrial differentiation among the subspecies *muenteri*, *scyrensis* and *quatuorlineata*, but since the first two were not sister-lineages, the smaller body size of these snakes might reflect the island-dwarfism phenomenon rather than common ancestry.

In this study, we revisit the four-lined snake diversification by utilizing genome-wide markers, generated with the double-digest Restriction-site Associated DNA approach (ddRAD), and complete cytochrome b sequences (*cytb*). Our aim was to construct the first genomic phylogeny of the species and then compare patterns and detect convergence and discrepancies between the biparentally and uniparentally inherited genomes. Analyses of the genome-wide markers were used to detect population size changes and historical events of isolation and migration, and reveal population clustering and genetic admixture among clusters. We identified genetic clusters within *E. quatuorlineata* and subsequently tested their distinctiveness with phylogenetic methods and species-tree analyses. We expected convergence in genomic and mitochondrial divergence-patterns for populations that have undergone long periods of genetic isolation or show restricted gene flow with their current



**Figure 1.** (A) Map of the Aegean region with sampling localities and codes of *E. quatuorlineata* specimens (see also Supplementary Table S1). Black points represent specimens for which both mtDNA sequences of the complete *cytb* gene and genome-wide SNPs were obtained, whereas for specimens with white points only *cytb* was available. (B) Map of the Cyclades island group with the respective sampling localities and codes, and the distribution of the currently described subspecies (*E. quatuorlineata quatuorlineata*, *E. q. muensteri*, *E. q. parensis*, and *E. q. scyrensis*). The approximate geographical position of the Pindos Mt. Range and the outline of the Cyclades Plateau (CP) during the late Pleistocene (modified by Kapsimalis et al. 2009) are shown.

conspecifics. In this case, diversification patterns would coincide with ancestral refugia (e.g., the Italian and Balkan Peninsulas) and/or currently isolated insular populations (e.g., Cyclades and Skyros Islands). On the contrary, populations that were historically restricted and subsequently expanded, as is the case of the mainland populations in the south Balkans, would exhibit incongruent genetic patterns for maternal and biparental genomes and possibly show asymmetric genetic introgression due to secondary contact.

## Material and Methods

### Sampling

We originally obtained 70 specimens of *E. quatuorlineata* and its sister-species *E. sauromates* (Pallas 1811) from museum collections (Supplementary Table S1). Due to the low DNA quality of some specimens, amplification, and/or library preparation was not successful in all cases. However, the final mitochondrial (51 specimens) and genomic (30 specimens) datasets (Figure 1) satisfy our sampling strategy, which was designed to represent the major mitochondrial lineages found in Kornilios et al. (2014) and the most important biogeographical regions: the Italian Peninsula, the Ionian islands, West and East Continental Greece (in respect to the Pindos Mt. Range; Figure 1), Peloponnesos, Skyros Island, and the Cyclades island group.

### Molecular data

#### DNA extraction and mtDNA sequencing

We extracted high molecular weight DNA using the Qiagen DNEasy extraction kit. We used Polymerase Chain Reaction (PCR)

to amplify the complete *cytb* gene with primers L14919 (de Queiroz et al. 2002) and H16064 (Burbrink et al. 2000), and a protocol that involved an initial cycle of denaturation at 95°C for 5 min, and 35 subsequent cycles at 94°C for 1 min, 50°C for 1 min and 72°C for 1 min. Sequences were obtained using both primers of the amplification procedure (GENEWIZ), aligned in ClustalX 2.0.12 (Larkin et al. 2007) and deposited in GenBank.

#### ddRAD library preparation and sequencing

We collected ddRAD data following previously published protocols, adaptors, and indices (Peterson et al. 2012). Specifically, we double-digested 500 ng of genomic DNA for each sample with 20 units each of 2 restriction enzymes (SbfI: restriction site 5'-CCTGCAGG-3'; MspI: restriction site 5'-CCGG-3', New England Biolabs) for 6 h at 37°C. Fragments were purified with Sera Mag Magnetic Beads (Millipore Sigma) before ligation of barcoded Illumina adaptors. Samples with unique adaptors were pooled, and each pool of 8 samples was size-selected for fragments in the range of 415–515 bp, after accounting for adaptor length, using a Pippin Prep system (Sage Sciences). Illumina multiplexing read indices were ligated to individual samples using a Phusion polymerase kit (high-fidelity Taq polymerase, New England Biolabs) and the final pools were again purified before their fragment size distribution and concentration were determined on an Agilent 2200 TapeStation (Agilent Technologies). A final quantitative PCR (qPCR) was performed to determine sequenceable library concentrations, before sequencing on a single Illumina HiSeq 4000 lane.

under a 50 bp single-end read protocol (Genomics Sequencing Laboratory, UC Berkeley).

#### ddRAD bioinformatics and single nucleotide polymorphism data collection

We processed raw Illumina reads using the program ipyrad 0.7.8 (Eaton 2014). We demultiplexed samples using their unique barcode and adaptor sequences, and reduced each read to 39 bp after removing the 6 bp restriction site overhang and the 5 bp barcode. Sites with Phred quality scores under 99% (Phred score = 20) were changed into “N” characters, and reads with  $\geq 10\%$  N’s were discarded. Samples that passed the quality filters but had  $< 50,000$  reads were excluded from further analyses.

Within the ipyrad pipeline, the filtered reads for each sample were clustered using VSEARCH 2.4.3 (Rognes et al. 2016) and aligned with MUSCLE 3.8.31 (Edgar 2004). We assembled the ddRADseq data using a relatively stringent clustering threshold of 92% in order to reduce the risk of combining paralogs. As an additional filtering step, consensus sequences that had low coverage ( $< 10$  reads), excessive undetermined or heterozygous sites ( $> 4$ ) or too many haplotypes ( $> 2$  for diploids) were discarded. The consensus sequences were clustered across samples using the within-sample clustering threshold (92%). Again, alignment was done with MUSCLE, applying a paralog filter that removes loci with excessive shared heterozygosity among samples (paralog filter = 200). The number of single nucleotide polymorphisms (SNPs) per locus was set to a maximum of 10 to minimize the possibility of returning paralogs. We generated final datasets by testing 4 levels of missing data, expressed as the minimum number of individuals with data for a given locus: 100% (i.e., all loci are present for all samples), 90% (each locus present in at least 90% of the samples), 75%, and 50%, respectively. Higher levels of missing data will increase the number of loci and SNPs in the final data matrix but can also negatively influence downstream analyses. For different analyses, we generated different types of final data matrices including the entire ddRAD locus (variable and invariant sites), all variable sites (SNPs), or 1 random SNP from each putatively unlinked locus (uSNPs), using ipyrad’s branching method to generate the appropriate dataset for the specific set of samples.

#### Phylogenetic analyses, population clustering, molecular dating

##### Mitochondrial gene tree and estimation of divergence times

Partitioned Maximum Likelihood (ML) analysis was carried out in IQ-TREE 1.4.3 (Nguyen et al. 2015), with the “partitionfinder” and “Auto” options (Chernomor et al. 2016) that determined the best partitioning scheme and the best-fit substitution model for each partition ([the Hasegawa-Kishino-Yano model with invariable positions (HKY+I) for the 1<sup>st</sup> and HKY for the 2<sup>nd</sup> and 3<sup>rd</sup> codon position]. Nodal support of the tree was tested via the standard non-parametric bootstrap approach with 1,000 alignments (Felsenstein 1985), the ultrafast bootstrap (UFBoot) approximation with 10,000 alignments (UFBoot; Minh et al. 2013), 10,000 replicates of the Shimodaira-Hasegawa-like approximate likelihood ratio test (SH-aLRT; Guindon et al. 2010), and the approximate Bayes test (abayes; Anisimova et al. 2011).

Since our aim was to reconstruct the phylogeography of *E. quatuorlineata* in the Aegean, we did not use dated paleogeographic events of this region, to avoid circular reasoning. Instead, we used “external” calibration age-constraints, combining our *cytb*

sequences with 44 published sequences from GenBank to produce the phylogeny of colubrid and natricid snakes (Supplementary Table S2). The same approach has been successfully used before (Kuriyama et al. 2011; Wood et al. 2011; Kyriazi et al. 2013) and its validity for molecular dating was thoroughly tested in the case of *E. quatuorlineata/sauromates* by Kornilios et al. (2014). Here, we used the age of three fossil records (the earliest *Lampropeltis*, *Pantherophis*, and *Thamnophis* fossils, respectively), two of which are situated within the Colubrinae, and one within the Natricinae. We incorporated the fossil ages by choosing prior age distributions so that the youngest age of the distribution corresponded to the youngest possible age at which that lineage existed and a standard deviation so that 95% of the log-normal distribution was younger than the oldest age of appearance.

Analyses ran in BEAST 1.8.4 (Drummond et al. 2012) under an uncorrelated log-normal relaxed molecular clock with a Yule prior on rates of cladogenesis. In order to conform to the Yule model, we ran the analysis on a reduced dataset including one sample for each main mtDNA clade (see results). We employed the HKY model of substitution with independent parameters for each codon position. We performed 4 runs with a chain length of  $5 \times 10^7$  iterations and a burn-in of 25%, which were checked for convergence (Effective Sample Size (ESS)  $> 200$ ) in TRACER 1.6 (Rambaut and Drummond 2009), combined in LogCombiner, and then produced the chronogram with TreeAnnotator and FigTree.

#### Genomic DNA analyses

**Population genetic structure:** We used the 100% uSNPs dataset (0% of missing data, unlinked SNPs) to study population genetic structure and investigate the number and geographic distribution of genetically distinct groups. We applied this strict cut-off on missing data in order to avoid overestimation of genetic clusters, since population structure analyses that are based on the estimation of allele frequencies can be particularly sensitive to the presence of rare alleles (Falush et al. 2007). First, we ran a discriminant analysis of principal components (DAPC) with the adegenet package in R (Jombart 2008; Jombart et al. 2010). A model of admixture is not required in this analysis, so it is expected to perform better when organisms exhibit continuous spatial population structure (Royer et al. 2010). We set the maximum number of clusters to 12, used the *optim.a.score* function to optimize the number of principal components (PCs) and the Bayesian inference criterion (BIC) to determine the number of independent genetic clusters (K) with the highest probability.

Second, we ran STRUCTURE 2.3.4 (Pritchard et al. 2000) in order to also detect possible admixture among population clusters. We originally tested the probability of different numbers (1–12) of K assuming correlated versus independent allele frequencies and admixture versus no admixture models. We ran each test with 3 independent replicates ( $2.5 \times 10^4$  iterations each, 50% burn-in) and evaluated the stability of the likelihood and admixture proportion (Q score) values within and among replicates as an indication of convergence. The optimal K (Kopt) was then estimated for each test over all runs by examining log-likelihood per K [ $\ln P(X/K)$ ] (Pritchard et al. 2000) and Evanno’s  $\Delta K$  (Evanno et al. 2005) with the CLUMPAK online web server (Kopelman et al. 2015), its “greedy” option and 2,000 random input orders. Kopt, the stability of alignments for each K and the respective probabilities of the clustering, was compared among tests. Although Kopt did not vary, the admixture model with correlated allele frequencies (Fst with mean value of  $0.01 \pm 0.05$ ) showed a higher probability. Conclusively,



final analyses were performed under these assumptions, with 5 runs of  $5 \times 10^5$  iterations each (50% burn-in). The degree of admixture between each subpopulation ( $a = 0.01 \pm 0.05$ ) and the distribution of allele frequencies ( $\lambda = 0.46$  for  $K=1$ ) were inferred from our data. We applied a hierarchical approach, analyzing first all samples of *E. quatuorlineata* to detect major clusters and then performing independent analyses in order to investigate further population structuring within each major cluster.

**SNP phylogenies and coalescence species-trees.** We constructed preliminary phylogenomic trees to evaluate datasets with different levels of missing data. The 90% or 75% datasets returned topologically identical trees, although the latter gave trees with slightly stronger clade-support. Allowing for more missing data (50% dataset) resulted in unresolved topologies and/or lower nodal support. We also tested the effect of admixed individuals on tree topologies, and run tests after removing individuals at different levels of admixture. Accordingly, downstream analyses ran on the 75% dataset, rooting trees with *E. sauromates* (the sister-species of *E. quatuorlineata*; Lenk et al. 2001) and removing admixed individuals if their membership probability to a given cluster was  $Q < 0.95$  (or  $Q = 0.85$  for the case of Italy where all individuals were admixed, see details in “Results” section).

A phylogenomic ML tree was constructed using the concatenated ddRAD loci with IQ-TREE. We used the “Auto” option for best-fit substitution model and tested nodal support via 1,000 bootstrap alignments and 1,000 SH-aLRT tests. We also applied the coalescent approach SVDquartets 1.0 (Chifman and Kubatko 2014) implemented in PAUP\*4.0b10 (Swofford 2003) which infers trees for subsets of 4 samples and estimates the species tree using a quartet assembly method. We evaluated all possible quartets, with and without prior assignment to populations (as resulted from the clustering analyses), and estimated statistical support with non-parametric bootstrapping of 1,000 replicates.

Finally, we estimated a species-tree under the Bayesian multispecies coalescent framework of SNAPP 1.3 (Bryant et al. 2012) implemented in BEAST2 2.8 (Bouckaert et al. 2014). We used the 75% uSNPs dataset (without *E. sauromates*) transformed into a biallelic format. Individuals were assigned to populations (“species”) based on the clustering results and the phylogenomic reconstructions. SNAPP can be computationally demanding, depending on the number of individuals included in the analysis (Bryant et al. 2012), so we used a reduced dataset (maximum of 3 individuals per “species”) and removed most admixed individuals (see “Results” section). Mutation rates ( $u, v$ ) were both fixed at 1.0 and the coalescent rate was set to 10. For the Yule model, we specified mean values for the two priors representing the speciation rate lambda ( $\lambda = \alpha \times \beta = 1,500$ ; set as  $\alpha = 2, \beta = 750$ ) and theta ( $\theta = \alpha/\beta = 2 \times 10^{-4}$  set as  $\alpha = 40, \beta = 200,000$ ) as broad gamma distributions. We estimated the tree height (maximum observed divergence between any pair of taxa divided by two) using our own ddRAD dataset containing all variable and constant characters and then utilized the script *pyule* (<https://github.com/joaks1/pyule>) to determine the mean value of  $\lambda$ . For the mean value of  $\theta$ , we estimated the percentage of polymorphic sites within each of the defined populations, so our value implies 0.02% variation between two randomly sampled individuals in a population.

We performed 2 independent runs with a chain length of  $3 \times 10^6$  generations, sampling every 1,000 generations. We checked convergence (ESS > 200) and determined burn-in (25%) with TRACER and evaluated the posterior distribution of trees with DensiTree. We combined the runs using LogCombiner and created a maximum

clade credibility (MCC) tree with TreeAnnotator, visualized with FigTree.

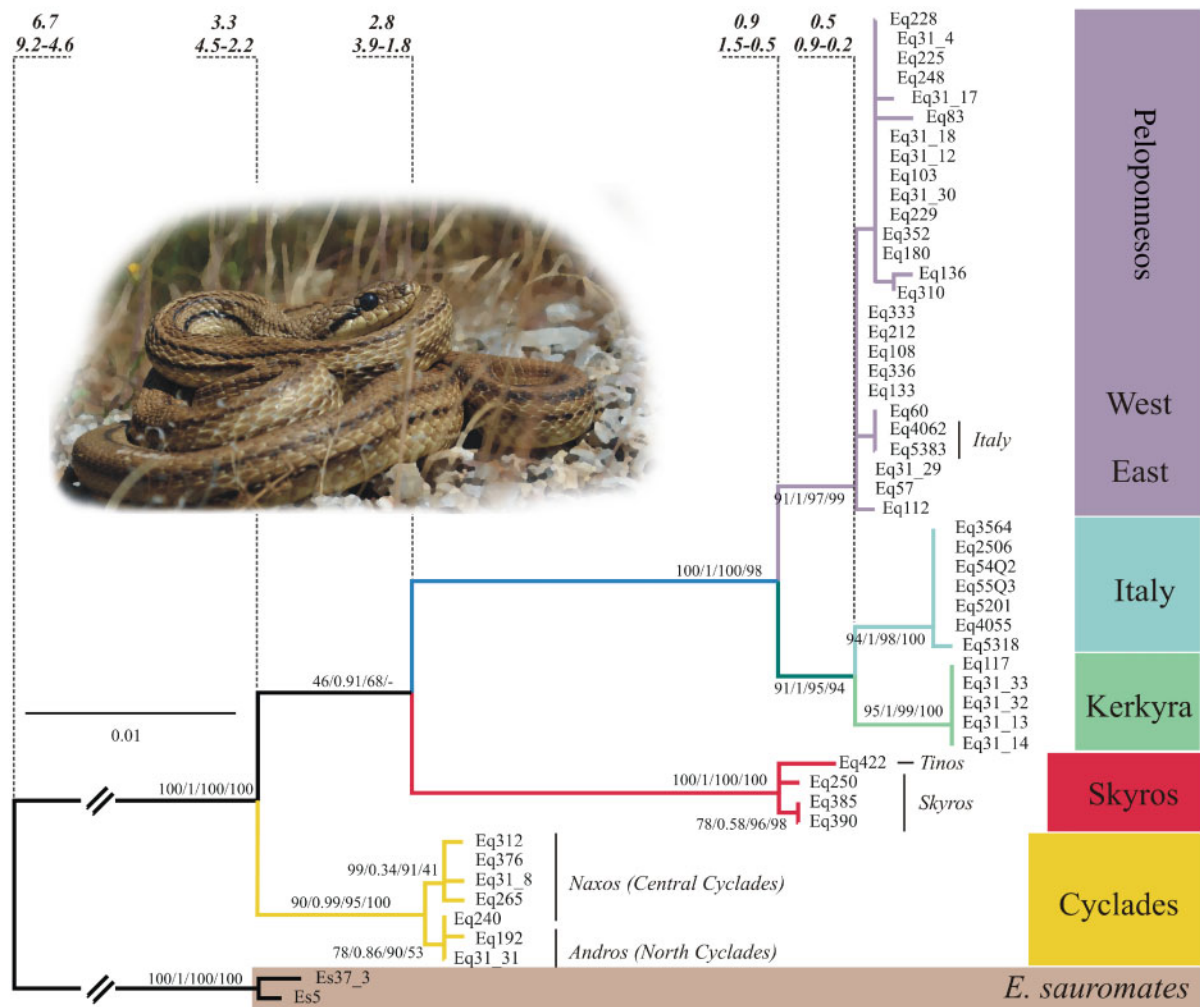
**Estimation of migration events.** To investigate the admixture history within *E. quatuorlineata*, we applied a population-tree inference model as implemented in TREEMIX 1.3 (Pickrell and Pritchard 2012). In this analysis, a bifurcating ML population-tree is inferred with leaves (tips) representing extant populations, internal nodes representing ancestral ones and connections between nodes (migration points) interpreted as migration events that led to admixture in the leaf population. TREEMIX assumes that allele-frequency differences between populations are solely caused by genetic drift (Pickrell and Pritchard 2012) and was originally designed for whole-genome datasets in population studies and possibly shorter time-scales and different evolutionary models in respect to species-tree inference methods (Wu 2015). Typically, this analysis is performed on SNPs aligned against a reference genome, split into shorter “blocks” and rearranged in repeats, in order to correct for linkage disequilibrium (Pickrell and Pritchard 2012). In *de novo* built ddRAD datasets this cannot be fully accomplished because there is no certain way of recognizing loci that might be linked. To balance this uncertainty, we followed the common practice (e.g., Spinks et al. 2014; Alter et al. 2017) and included only what are considered to be unlinked SNPs (one randomly chosen SNP per ddRAD locus). We assigned individuals to “populations,” following the results of the population clustering and the phylogenomic and species-tree analyses. We used the 75% biallelic uSNPs, including all the four-lined snake samples and performed several preliminary tests in order to investigate the effect of different population clustering and the presence of admixed individuals, progressively adding migration events ( $m$ ) up to a total of 8. Although rooting trees is suggested for the TREEMIX inference (Pickrell and Pritchard 2012), we also considered unrooted trees, in order to avoid forcing a topological constrain that might artificially split populations of common ancestry.

## Results

### Mitochondrial gene tree and estimation of divergence times

Our final *cytb* dataset (51 sequences of 1,117 bp) included several new localities for *E. quatuorlineata* from the Ionian Islands (Kerkyra, Zakynthos, and Lefkada), Aegean Islands (Naxos, Tinos, Andros, and Kea), Skyros Island, and Peloponnesos, as well as *E. sauromates* sequences from northeast Greece and Asia Minor (Figure 1, Supplementary Table S1, GenBank Accession Numbers MH444309–MH444359).

Since different measures of nodal support have different interpretations, we considered nodes strongly supported when SH-aLRT/abays  $\geq 80\%$ , UFboot  $\geq 95\%$ , and BS  $> 80\%$  (Minh et al. 2017). In this context and in line with preliminary analyses (Kornilios et al. 2014), the mitochondrial phylogeny revealed three strongly supported clades within *E. quatuorlineata*, although relationships among them were not fully resolved (Figure 2). The first clade included most populations from the Cyclades islands (Naxos from Central Cyclades and Andros from North Cyclades). The next clade included all individuals from Skyros (Northern Aegean) and one individual from Tinos (North Cyclades) and the last clade all remaining populations. The last clade further split into three sub-clades, grouping 1) individuals from Kerkyra and its most adjacent continental region, 2) all the Italian samples, but two (Eq4062 and Eq5383) and 3) these two Italian samples and all the remaining



**Figure 2.** ML phylogenetic tree based on the complete *cytb* dataset. Coding in terminal nodes refer to samples in Figure 1 and Supplementary Table S1. Numbers in branches give the respective nodal support estimated as standard bootstrap values/ultrafast bootstrap values/SH-like approximate likelihood ratio test/approximate Bayes posterior probabilities. The time of divergence (mean and 95% HPD intervals values in millions of years, Mya) are also given for nodes that were represented in the molecular-clock analysis (Supplementary Table S1, Figure S1). Inset: Typical morphotype corresponding to *E. quatuorlineata quatuorlineata* (Photo by P.K.).

populations from Peloponnesos, East and West continental Greece (assigning east and west populations with respect to the Pindos Mt. Range; Figure 1). The phylogenetic pattern revealed in the case of Tinos was unexpected. Despite its geographical position within the North Cyclades (Figure 1), the mitochondrial haplotype found in Tinos was very distinct from the other Cycladian haplotypes and differed by 3–4 mutations from the haplotypes of Skyros, thus resulting in its peculiar placement within the Skyros clade instead of the Cyclades clade (Figure 2). Individuals from Kerkyra and Italy were well-differentiated from all the remaining mainland populations. Excluding Kerkyra, extremely low levels of mtDNA variation were found within the Balkans.

The time-calibrated tree included one sequence from each clade found in the mitochondrial tree (Supplementary Figure S1). The split between *E. quatuorlineata* and *E. sauromates* occurred in the Late Miocene [6.7 Ma, 95% Highest Posterior Density (HPD) intervals: 4.5–9.2 Ma]. The *E. quatuorlineata* radiation, which led to the three major clades, occurred in the mid-Pliocene [time of most recent common ancestor (TMRCA) 3.2 Ma, 4.4–2.2 Ma]. During the

Pleistocene (TMRCA 0.9 Ma, 0.5–1.5 Ma) the last clade further split into three sub-clades.

### Analyses of genome-wide SNPs

Illumina sequencing of ddRAD libraries originally included 33 individuals but three samples (Eq31.29, Eq5383, Eq60) were removed due to their low number of reads (< 50,000). For the remaining samples (28 *E. quatuorlineata* and 2 *E. sauromates*), the number of reads per individual ranged between 0.8 and 7.4 million (mean = 2.6). The size of the *E. quatuorlineata* dataset with no missing data was 2,321 total filtered loci (and 2,052 when *E. sauromates* was included). Allowing 25% of missing data increased the number of loci to 11,803 (11,931 including *E. sauromates*). The maximum number of SNPs found per locus was 8. A detailed presentation of the number of loci, SNPs, or uSNPs produced through different data filtering is given in Tables 1 and 2 and details on the type and size of each of the datasets used in our final analyses are presented below.

**Genetic clusters.** DAPC analysis included all *E. quatuorlineata* samples and returned five or four clusters as equally supported

**Table 1.** Summary of the ddRAD data matrices as resulted from the iPyRAD pipeline

iPyRAD parameters	Minimum % of individuals for a given locus (total number of individuals included during filtering)	100 (28)	90 (30)	75 (30)	50 (30)
Filtering Statistics	Number of prefiltered loci	45,915	47,670	47,670	47,670
	Number of filtered loci	2,321	7,543	11,931	17,640
	Total number of base pairs	93,032	302,558	478,587	707,588
	Number of SNPs	725	3,572	5,649	8,339
	Number of uSNPs	531	2,518	3,929	5,602

This includes information on the number of prefiltered and filtered loci, the total number of base pairs, and the number of SNPs and unlinked uSNPs, given for 4 different datasets. Each dataset was built with different percentages of missing data, expressed as % of individuals for a given locus: 100% (0% missing data, i.e., all loci present for all samples), 90%, 75%, and 50% (respectively 10%, 25%, and 50% missing data, i.e., all loci present for at least 90%, 75%, and 50% of the samples). The number of samples included during the filtering for each dataset is given in parentheses (*E. quatuorlineata* samples with or without *E. saurosmates*).

**Table 2.** The ddRAD datasets (size and number of included individuals shown in parentheses) used in the respective final analyses

Analyses	Type of dataset	Size of dataset (individuals included)
DAPC	uSNPs	531 (28)
STRUCTURE	uSNPs	531 (28)
		646 (24)
ML after concatenation	total base pairs (bp)	477,705 (21)
SVDquartets	uSNPs	2,430 (21)
SNAPP	biallelic uSNPs	2,612 (16)
TREEMIX	uSNPs	2,628 (26)

DAPC analysis was performed to investigate population structure and STRUCTURE analysis to further identify individuals with genetic admixture within *E. quatuorlineata* (including or excluding the samples from the Cyclades islands). Phylogenomic trees were constructed with the concatenated ML and SVDquartets methods after removing admixed individuals and with *E. saurosmates* as the outgroup. Species-tree inferred with SNAPP used a reduced dataset of non-admixed *E. quatuorlineata* individuals and migration events were inferred with TREEMIX, after removing only 2 admixed *E. quatuorlineata* individuals. Further information on which individuals were removed or retained in every case is presented in the text.

optimal number of clusters (Figure 3A). In the first case (Figure 3B and 3C), the five clusters were 1) Central Cyclades, 2) North Cyclades, 3) populations from West Continental Greece and adjacent islands, Italy, and Peloponnesos, 4) populations from East Continental Greece, and 5) Skyros. In the second case, the four-clusters were the same as above with the exception of either Skyros grouped with East Continental Greece populations (Figure 3D) or North Cyclades grouped together with the Central Cyclades (Figure 3E).

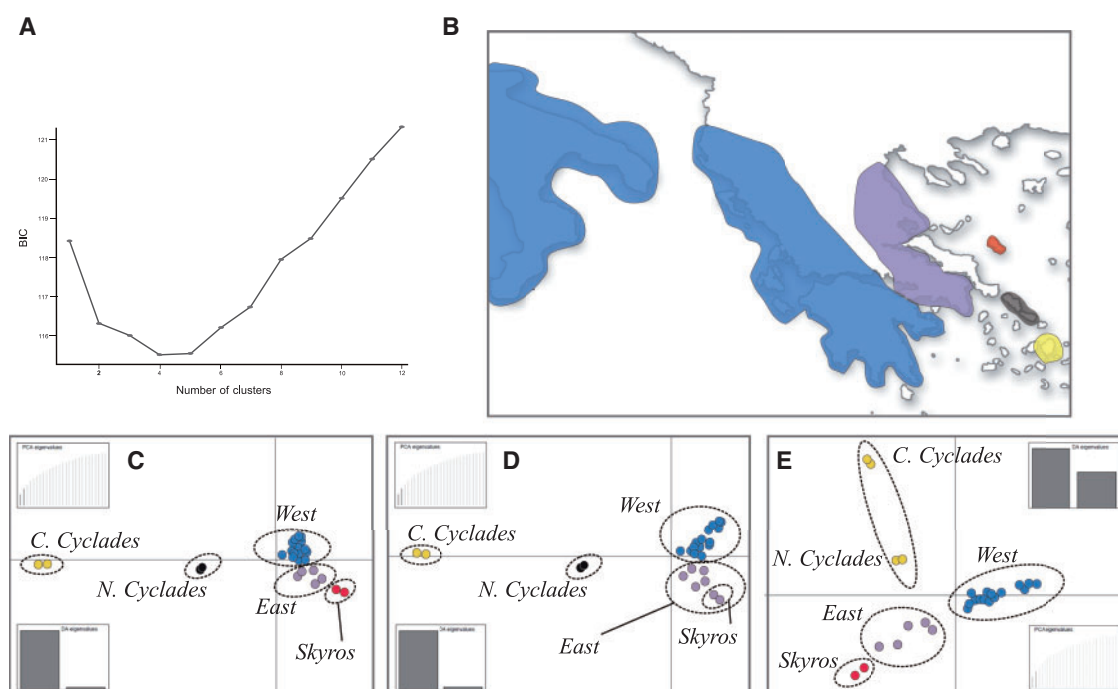
Analysis with STRUCTURE (Figure 4A) ran first on the same dataset, including all *E. quatuorlineata* samples. Five independent runs converged to  $K_{opt} = 2$ , that is, one cluster for the Central Cyclades and one for all other populations, with the North Cyclades individuals showing admixture with the Central Cyclades cluster ( $Q$  values of 0.26 and 0.19, respectively). Then we repeated the analysis, after excluding all samples from the Cyclades islands. The remaining populations clustered in  $K_{opt} = 5$ , corresponding to five geographical regions 1) Italy, 2) Kerkyra, 3) Peloponnesos and West Continental Greece 4) East Continental Greece and 5) Skyros. With the exception of those from Kerkyra and Skyros, all other individuals were more or less admixed, for example, all the Italian individuals showed admixture with the pure Kerkyra cluster (ranging from

0.14 to 0.35), whereas the remaining Ionian Islands (Lefkada and Zakynthos) and West Continental Greece were found admixed with elements of all western clusters.

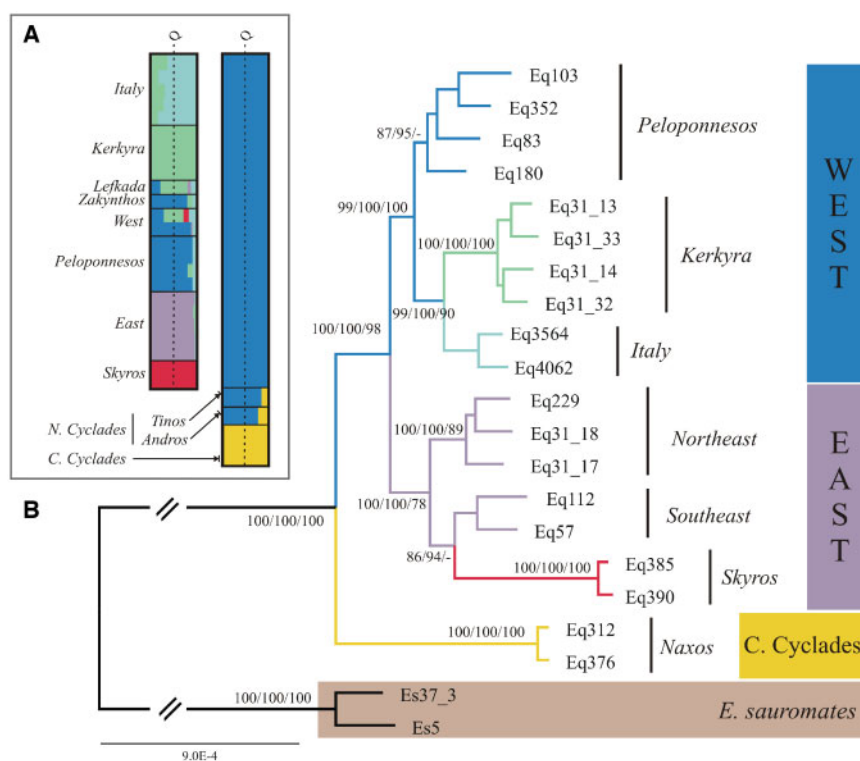
**Phylogenomic and species-trees.** All our tree analyses returned the same strongly supported phylogeny that split *E. quatuorlineata* into three major clades. The first one to split was the Cyclades islands, followed by a more recent split into an East clade with all populations from East Continental Greece and Skyros, and a West clade with all the remaining populations (Figures 4B and 5A). These major groupings and the relationships among them were retained irrespectively of the presence of admixed individuals (Supplementary Figures S2 and S3). However, in our final analyses, admixed individuals that showed a membership probability to a given cluster of  $Q < 0.95$  were removed, with the exception of two individuals from Italy. Since all Italian specimens were found to be admixed but still represent a distinct genetic cluster (Figure 4A), we included the two Italian individuals with the highest membership probability ( $Q = 0.85$ ) in all final analyses in order to obtain information on the phylogenetic position of this Italian cluster.

The final ML analysis of the concatenated ddRAD loci (21 individuals) detected further sub-groups within the East and West clade with strong to moderate support (Figure 4B). The East clade included two subgroups (one for northeast populations and one for the southeast populations + Skyros). The West included three for Peloponnesos, Italy, and Kerkyra and supported the closer relationship of the last two. The final SVDquartets analysis (including the same 21 individuals without prior assignment to populations) was mostly in agreement with the concatenated ML tree. Their difference was that the SVDquartets tree did not support a closer relationship of Skyros with the southeastern populations from Continental Greece, but rather an unresolved relationship among three eastern sub-clades (Northeast, Southeast, and Skyros). Also, the monophyly of Peloponnesos was not supported (Figure 4B).

Two species-tree coalescence-based analyses were conducted with a priori assignment to populations, that is, Central Cyclades, Skyros, Southeast and Northeast, Peloponnesos, Kerkyra and Italy; SNAPP with a reduced dataset of 16 individuals and SVDquartets with the aforementioned 21 individuals assigned to populations. In our final SNAPP analysis we included the two Italian individuals with  $Q = 0.85$ , given that the SNAPP model assumes a lack of gene flow but can accommodate incomplete lineage sorting (ILS) (Bryant et al. 2012). For the Italian samples 1) it is reasonable to assume that there is no current gene flow between Italian and Greek populations, given their allopatric distribution, 2) the membership probability to their cluster was lower than  $Q < 0.95$  but still relatively

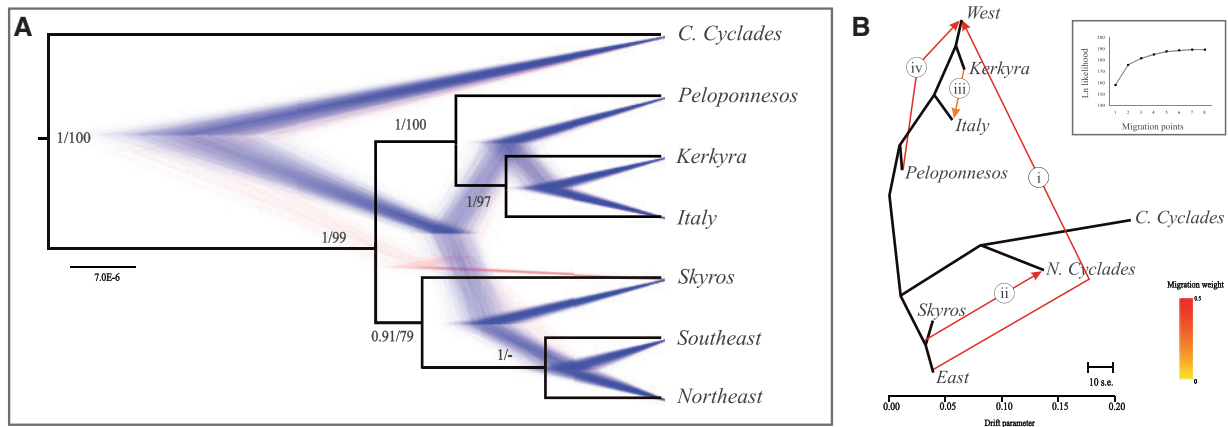


**Figure 3.** Results of the DAPC analysis for the population clustering of *E. quatuorlineata* based on the uSNPs dataset. **(A)** Plot of the ln likelihood values estimated under the BIC to the number of clusters (up to  $K=12$ ) found. The plot shows an almost equal support for  $K=4$  and  $K=5$ . **(B)** Map showing the approximate geographical distribution of 5 clusters and **(C)** scatterplot of the two first components inferred for  $K=5$ . **(D)** and **(E)** Respective scatterplots inferred for  $K=4$ . Insets in each scatterplot represent the number of principal component analysis (PCA) and discriminant analysis (DA) eigenvalues retained in each analysis.



**Figure 4.** **(A)** Hierarchical population clustering with STRUCTURE using 2 uSNPs datasets; including all *E. quatuorlineata* and excluding individuals from Cyclades islands (Naxos, Andros and Tinos). Q represents the membership probability of each individual to a given cluster. **(B)** ML phylogenomic tree based on the concatenated ddRAD loci dataset, including non-admixed *E. quatuorlineata* individuals (see text for details) and *E. sauromates* as outgroup. Numbers in branches give the respective nodal support estimated as standard bootstrap values/SH-like approximate likelihood ratio/bootstrap values estimated for the same nodes inferred by the SVDquartets analysis.





**Figure 5. (A)** Coalescent species-tree for *E. quatuorlineata* produced by SNAPP analysis on the uSNPs dataset, transformed into biallelic format. The MCC tree is over-imposed on the posterior distribution of the 95% HPD set of tree-topologies graph. Numbers in branches give the respective support estimated as posterior probability values for the MCC tree/bootstrap values estimated for the same nodes inferred by the SVDquartets analysis. **(B)** TREEMIX admixture graph with 4 migration events (i–iv). The strength (migration weight as the fraction of ancestry derived from the migration source) and directionality of the inferred introgression is indicated by the colored arrows. The plot of the ln likelihood values to the number of migration edges ( $m$ ) up to  $m=8$  is also given.

high and 3) preliminary testing showed that including them did not alter the relationships among the remaining individuals. According to our final analyses, in both SNAPP and SVDquartets trees (Figure 5A) the Central Cyclades split first and all other populations form its sister-clade, which further split into an East and a West group. Within the West group, Peloponnesos is sister to the Kerkyra + Italy clade. In the East group, SNAPP supports the monophyly of the Southeast + Northeast clade but SVDquartets cannot resolve their relationship with respect to Skyros.

**Migration estimated with TREEMIX.** The number and directionality of estimated migration points were the same irrespectively of the included admixed individuals. Unrooted trees and trees rooted with Central Cyclades showed only minor differences in likelihood values. North Cyclades and the western Greek mainland plus two adjacent islands (Lefkada and Zakynthos) were treated as independent “populations” to test whether genomic immigration could explain their admixed pattern. Final analysis included a total of 26 individuals; we retained all the admixed individuals of North Cyclades, Italy and western Greek mainland (forming the populations of N. Cyclades, Italy and West, respectively) and removed the two individuals from Lefkada and Zakynthos islands. This dataset resulted in the best-fit unrooted tree with the highest likelihood (Figure 5B), which reached a plateau after including four migration points 1) from East to West, 2) from a pre-ancestral population of East and Skyros to North Cyclades, 3) from Kerkyra to Italy and 4) from Peloponnesos to West.

## Discussion

### The ancient lineage of the Cyclades Plateau

Mitochondrial phylogenies of *E. quatuorlineata* (Kornilios et al. 2014; this study, Figure 2) depicted three major lineages corresponding to the Central Aegean Islands of the Cyclades group that diverged first, the North Aegean island of Skyros (also including the North Cycladian Island of Tinos in this study) and all the remaining Italian and Balkan populations. Population clustering and phylogenetic analyses of genome-wide SNPs portrayed a more complex geographical pattern. Genomic results also identified the Central Cyclades as a genetically homogenous, very distinct lineage that was the first to diverge from the common ancestor of *E. quatuorlineata*

(Figures 4B and 5A). North and Central Cyclades populations share the Cycladian mtDNA haplotypes; however, North Cyclades were found admixed between the Cycladian and mainland genomic clusters (Figures 4A and 5B).

Speciation within *E. quatuorlineata* began in the mid-Pliocene (3.3 Mya) with the split of the central Aegean Islands’ lineage, according to our time-calibrated mitochondrial phylogeny (Figure 2, Supplementary Figure S1). During the Pliocene, the central Aegean Islands formed a continuous landmass, the Cyclades Plateau (CP; Figure 1B), separated from the mainland in the west but connected with it in the north (Dermitzakis and Papanikolaou 1981). It remained so until some 3.5 Mya, when the land bridge between today’s Evvoia and Andros Islands submerged (Anastasakis et al. 2006) and never re-established. This vicariant event isolated the ancestral population of the Cyclades and led to the differentiation of the Cycladian lineage. Similar and synchronous diversification patterns are also observed in other reptiles (e.g., Ursenbacher et al. 2008; Kyriazi et al. 2013). Pleistocenic paleogeographic reconstructions (Kapsimalis et al. 2009) show the eventual breaking of the CP into two smaller island groups: one north (including Andros and Tinos) and one south (including Paros and Naxos) (Figure 1b), until the final separation of contemporary islands.

The paleogeography of the region and the genomic and mitochondrial data (Kornilios et al. 2014; present study) suggest that the Cycladian lineage is probably distributed in all the central Aegean Islands where the four-lined snake occurs (Figure 1B). This lineage, which corresponds to the morphologically distinct *muensteri* morphotype and possibly also *parensis*, has remained isolated for more than 3 My. The Cycladian lineage represents the sister-lineage to all the remaining four-lined snakes, thus suggesting that *E. quatuorlineata* should be best treated as a species-complex with at least two sister-taxa.

### Ancient mitogenome retained in insular populations

Increasing our sampling in this study, we revealed an unexpected pattern for Tinos island. Despite its geographical position within the North Cyclades (Figure 1), the mitochondrial haplotype found in Tinos was very distinct from the other Cycladian haplotypes and placed within the Skyros clade instead of the Cyclades clade (Figure 2). The placement of Tinos and Skyros appears to be in

conflict between the two phylogenies. According to the mitochondrial phylogeny, the next phylogenetic event after the split of Cyclades is the separation of Skyros. Our time-calibrated mtDNA tree dated this event at ~2.8 Mya (Figure 2, Supplementary Figure S1). Skyros Island has been isolated from the mainland during the past 4.5–5.0 My (Dermitzakis 1990) which would suggest an early overseas dispersal. However, genomic data did not support an early diversification event. Clustering analyses grouped Skyros together with all east mainland populations or as a distinct group close to the east mainland (Figure 3B–E). Phylogenomic and species-tree analyses (Figures 4B and 5A) nested Skyros within the clade of mainland populations from East Greece. Therefore, it is very unlikely that the Skyros population has remained genetically isolated for the past 3.0 My, as suggested by the mitochondrial dating. The overall genetic pattern better fits a recent transmarine dispersal of individuals with an ancestral mitogenome that colonized the island. Overseas dispersal is not uncommon in reptiles, and snakes have been known to overcome sea barriers, especially during periods when sea level is low due to climatic factors (Nagy et al. 2003; Kyriazi et al. 2013).

Skyros was probably colonized from east mainland but we cannot accurately place the source of this colonization based on our phylogenomic reconstructions (Figures 4 and 5) and we cannot rule out colonization from an unsampled area. The best candidate is the nearby island of Evvoia, the closest land west of Skyros, where *E. quatuorlineata* is also distributed (Figure 1B). The geographical position of Evvoia being the most proximate land to the North Cyclades further explains the unexpected mitochondrial resemblance between Skyros and Tinos. If this hypothesis is correct, then Evvoia could also be the source of the re-colonization of the North Cyclades. These islands are expected to harbor two mitochondrial types, the “Central Cyclades” type and the “Evvoia/Skyros” type and admixed genomic DNA with elements from the mainland and the Central Cyclades. Our results from STRUCTURE (Figure 4A) and TREEMIX (Figure 5B) strongly support this hypothesis, showing admixture and a migration event from an ancestral eastern population to the North Cyclades. The alternative scenario of a direct colonization of North Cyclades from Skyros would be biogeographically unprecedented and extreme, whereas it is not backed up by our phylogenomic reconstructions.

Under this light, the peculiar mtDNA placement of Tinos may not be so peculiar after all but rather gives additional evidence of introgression in both the maternal and biparental genome. We propose that it is the absence of differentiated mtDNA in east mainland populations in our present data that conceal the mitochondrial divergence of *E. quatuorlineata* into a West and an East lineage. In this context, the mitochondrial lineage of Skyros and Tinos actually represents an ancestral East mainland lineage that split some 3.0 Mya (Figure 2). This ancient lineage could be entirely extinct; otherwise it may still be restricted in eastern mainland populations and could be found in future sampling.

### Expansion from refugia and genomic introgression

Within the third clade of the mitochondrial gene-tree (Figure 2), individuals from Kerkyra and Italy formed two sister-lineages and were well-differentiated from all the remaining populations distributed throughout the Balkan Peninsula (see also Kornilios et al. 2014). In addition, according to our genomic data and contrary to the mtDNA, the mainland populations were clearly clustered in an east and a west group, both including sub-clusters that also followed a geographical pattern and showed more or less extended admixture between them (Figures 4A and 5B). However, following the

aforementioned reasoning of an eastern mtDNA lineage represented here only by Skyros–Tinos, we can conclude that the maternal phylogeny is in line with the genomic analyses and both support the differentiation of the continental populations into an east and a west group. This east-west differentiation probably reflects the role of the Pindos Mt. (Figure 1A), a mountain range which descends from the western Balkans into Greece and has acted as a barrier between west and east populations of many animal taxa distributed in the Balkans, including reptiles (e.g., Ursenbacher et al. 2008; Psonis et al. 2017).

The west group includes the four-lined snakes of the Italian and the west Balkan Peninsulas clustered in respective genetic subgroups which coincide with the three mitochondrial lineages of Italy, Kerkyra and the Greek mainland (Figures 2–5). This split is dated during the Pleistocene (1.5–0.5 Mya; Figure 2) and may reflect the restriction of ancestral populations in refugia linked to the glacial/interglacial Pleistocene cycles. The Italian and Balkan Peninsulas have played a well-known role as Pleistocene refugia associated with the diversification of European species (Hewitt 2011). Our genomic analyses identified pure or almost pure clusters (Figure 4A) that correspond well to these two major refugia and further suggest the existence of multiple refugia-within-refugia in the Balkans (Nieto-Feliner 2011), such as Kerkyra and Peloponnesos. The latter is a known refugial area and an important biodiversity hotspot, with several endemic taxa, as well as intraspecific phylogenetic lineages (Thanou et al. 2014 and references therein). Kerkyra and/or the northwestern adjacent mainland emerges as a putative northern sub-refugium, where a differentiated lineage survived and subsequently expanded northwards and southwards, given that expansion toward the east was probably restricted by the Pindos Mt. Despite the distributional gap between them, populations in Italy and Greece share a few common mitochondrial haplotypes (Figure 2), genomic admixture (Figure 4A), and a migration point from Kerkyra to Italy (Figure 5B). There is historical evidence of human-mediated dispersal from Greece to Italy in recent times that may have induced this genetic admixture (Kornilios et al. 2014).

Finally, an extensive pattern of admixture was found among *E. quatuorlineata* of southwestern Greece. This area shares nuclear genomes from two “pure” clusters, a southern one (Peloponnesos) and a northern one (Kerkyra) (Figure 4A) and according to TREEMIX (Figure 5B), two migration points from Peloponnesos and East are directed there. This implies the conjunction of genetically pure populations that met in this area as they expanded toward each other. The eastern source of genomic introgression is a currently undefined sub-refugium in the east (e.g., Evvoia) and seems to have expanded to the west of continental Greece, colonized Skyros in the northeast and re-colonized North Cyclades in the southeast. Overall, the phylogeographic patterns inferred by merging mtDNA and genomic results suggest that the divergence within *E. quatuorlineata* was driven by cycles of isolation and secondary contact among several differentiated ancient populations. The subsequent genetic exchange ranged from extensive to very restricted, depending on the paleogeographic connections and the landscape features of the region.

### Mitochondrial sweep across the mainland

In contrast to the West/East divergence revealed by the analysis of biparentally inherited genome-wide SNPs (Figures 3–5), all Peloponnesian, East and West Greek populations share the same maternally inherited mitogenome (Figure 2). This “west”

mitochondrial type seems to be extensively distributed throughout continental Greece and the entire Balkan Peninsula (see also Kornilios et al. 2014). Some degree of mitonuclear discordance may be attributed to genetic drift and ILS and the fact that individual gene-trees may deviate from the species-tree. Even so, ILS cannot sufficiently explain the pattern revealed here, since other mitochondrial lineages (e.g., Italy and Kerkyra) have been well-differentiated in the same time and in concordance for mitochondrial and nuclear data. Furthermore, mitonuclear discordance that arises from ILS is not expected to leave any predictable biogeographic pattern, so ILS can usually be ruled out when strong geographic inconsistencies are found (Funk and Omland 2003; Toews and Brelsford 2012), as is the case of *E. quatuorlineata*. Given the restricted geographical distribution of the ancient “east” mitochondrial type (currently sampled in Skyros and Tinos), the discrepancies between our SNP and mtDNA phylogenies can be explained by an extreme case of mitochondrial introgression (such as mitochondrial sweep or capture) throughout the mainland populations.

Genetic exchange may not be symmetrical between incompletely isolated populations that reach contact (Sloan et al. 2017). For neutral alleles, introgression usually shows a directional movement from the stable (or local) population into the expanding (invader) one, due to the small number of immigrants. Excoffier et al. (2008) proposed three mechanisms: 1) progressive dilution of the invader’s genome by the local population’s genome, 2) contrasting population dynamics (growth vs. decline of the population size for the invader and local population, respectively) and 3) increasing frequency of local alleles into the invader’s growing population due to genetic drift and the continuous supply of new copies through hybridization. In any case, individuals in the area of contact have admixed genomes but as expansion progresses, the invading population will eventually lose any alleles transferred from the local population.

However, the mitochondrial replacement of an entire population with the mtDNA of another population may also occur, due to selective pressure (Ballard and Whitlock 2004; Chan and Levin 2005) or non-adaptive evolutionary mechanisms that preferentially affect cytoplasmic genomes (Sloan et al. 2017). As for non-adaptive mechanisms, lineage sorting is more rapid for mtDNA, so mitochondrial replacement is more common than the replacement of sections of the nuclear genome (Chan and Levin 2005; Currat et al. 2008). Secondary contact between genetically distinct, hybridizing populations, which show great differences in range size or abundance, can also promote mtDNA introgression (Funk and Omland 2003). Finally, sex-biased dispersal may also cause greater introgression for uniparentally inherited genomic parts that are associated with the least dispersing sex, for example, the mtDNA in species for which females disperse less. Local females hybridize with dispersing males and although the male parental genome is transmitted to future generations, the offsprings will always contain the local mitogenome (Excoffier et al. 2008; Toews and Brelsford 2012).

In the case of *E. quatuorlineata*, the prevalence of the “west” mitogenome over the “east” suggests that an east population acted as the invader population and a west as the local, assuming adaptive neutrality (Excoffier et al. 2008; Currat et al. 2008). This is in agreement with the migration event reported by the TREEMIX analysis of our genomic markers (Figure 5B), which showed introgression from the East into the West populations. Expanding from its respective sub-refugium, the ancestral eastern population invaded the west parts of continental Greece, already inhabited by four-lined snakes of the “west” mitochondrial type. As the invader’s front

moved toward the west, “east” was gradually replaced by the “west” mitochondrial type, due to asymmetric mitochondrial introgression. Following the typical pattern of mitochondrial sweep, the “west” type is currently fixed throughout the mainland and the “east” type is probably rare or even extinct at the east parts of continental Greece, retained only in peripheral insular populations (Skyros).

If females are the least dispersing sex for the four-lined snake, the maternally inherited mtDNA is expected to experience far greater introgression than the biparentally inherited genomic SNP (Excoffier et al. 2008). The dispersal behavior of *E. quatuorlineata* is not known and in fact few snake species have been studied to reveal any general movement or dispersal patterns. However, capture–recapture, mortality rate, and population genetics studies show that male-biased dispersal could be common for snakes, particularly during the mating season when they actively search for partners (Pernetta et al. 2011 and references therein). Particularly for rat-snakes, there is evidence that females show a higher site fidelity to their hibernacula and mate-searching males travel longer distances (Bonnet et al. 1999; Blouin-Demers et al. 2005).

A growing number of studies report extensive mtDNA introgression between closely related taxa of mammals (e.g., Beysard et al. 2011), birds (e.g., Friis et al. 2016; Wang et al. 2018), amphibians (e.g., Wielstra and Arntzen 2012; Bisconti et al. 2018), lizards (e.g., McGuire et al. 2007), and snakes (Barbanera et al. 2009; Kindler et al. 2013). In the majority of studies, this can be attributed mostly to the geographical isolation and secondary contact of the taxa involved (Toews and Brelsford 2012). Discordant patterns between mtDNA and nuDNA can also arise if selection for mtDNA variants varies geographically (Bierne et al. 2013). Some empirical studies found evidence of mitochondrial capture between closely related species of amphibians and reptiles that could be associated to the thermal adaptation of these poikilotherms (Weisrock et al. 2005; Sequeira et al. 2011; Bryson et al. 2014). Although the observed pattern and directionality of mitochondrial introgression in *E. quatuorlineata* fits well with our assumption of genetic neutrality, the effect of natural selection cannot be dismissed.

Overall, our hypothesis of distinct genetic lineages within the four-lined snake that were, respectively, isolated in multiple sub-refugia is supported. For the mainland populations throughout the Balkans, where biogeographical barriers did not restrict dispersal, gene flow is restored, leading to extensive genetic exchange between them and most possibly mitochondrial sweep. Peripheral insular populations (Cyclades, Skyros, and possibly Evvoia) still act as refugia for the survival of ancient genetic polymorphism. Considering their distinct morphology and in order to provide the basis for taxonomic and conservation decisions, these populations should be the future focus of an extensive sampling. Moreover, their morphological differences need to be further addressed, especially under the light of adaptation related to island-dwarfism and the possibility of an adaptive selection driving mitonuclear discordance.

## Authors’ Contributions

E.T., P.K., and A.D.L. designed the study. E.T., P.K., and P.L. collected samples and retrieved museum loans. E.T. and P.K. undertook all the laboratory and molecular data analyses. E.T. drafted the manuscript, and all authors contributed to manuscript revisions and approved the final version for publication.

## Acknowledgments

The authors would wish to thank Roberto Sindaco and Cristiano Liuzzi (Museo Civico di Storia Naturale di Carmagnola, Torino, Italy), Manolis Papadimitrakīs (Natural History Museum of Crete, Irakleio, Greece), Jiří Moravec (National Museum, Prague, Czech Republic), Sinos Giokas (Zoological Museum of the University of Patras, Patras, Greece), and Yusuf Kumluca and Çetin Ilgaz (Zoological collection of the Department of Biology, Dokuz Eylül University, İzmir, Turkey) who kindly helped them with museum loans. They are also grateful to Kevin Epperly for his help in the production of genomic libraries.

## Funding

P.K. was supported by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant [grant number: 656006, project acronym: CoPhyMed]. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation Grant.

## Supplementary Material

Supplementary material can be found at <https://academic.oup.com/cz>.

## References

- Alter SE, Munshi-South J, Stiasny ML, 2017. Genomewide SNP data reveal cryptic phylogeographic structure and microallopatric divergence in a rapid-adapted clade of cichlids from the Congo River. *Mol Ecol* 26:1401–1419.
- Anastasakis G, Piper DJW, Dermitzakis MD, Karakitsios V, 2006. Upper Cenozoic stratigraphy and paleogeographic evolution of Myrtoon and adjacent basins, Aegean Sea, Greece. *Mar Petrol Geol* 23:353–369.
- Anisimova M, Gil M, Dufayard JF, Dessimoz C, Gascuel O, 2011. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst Biol* 60:685–699.
- Ballard JO, Whitlock MC, 2004. The incomplete natural history of mitochondria. *Mol Ecol* 13:729–744.
- Barbanera F, Zuffi MAL, Guerrini M, Gentili A, Tofanelli S et al., 2009. Molecular phylogeography of the asp viper *Vipera aspis* (Linnaeus, 1758) in Italy: evidence for introgressive hybridization and mitochondrial DNA capture. *Mol Phylogenet Evol* 52:103–114.
- Beysard M, Perrin N, Jaarola M, Heckel G, Vogel P, 2011. Asymmetric and differential gene introgression at a contact zone between two highly divergent lineages of field voles *Microtus agrestis*. *J Evol Biol* 25:400–408.
- Bierne N, Gagnaire P-A, David P, 2013. The geography of introgression in a patchy environment and the thorn in the side of ecological speciation. *Curr Zool* 59:72–86.
- Bisconti R, Porretta D, Arduino P, Nascetti G, Canestrelli D, 2018. Hybridization and extensive mitochondrial introgression among fire salamanders in peninsular Italy. *Sci Rep* 8:13187.
- Blouin-Demers G, Gibbs HL, Weatherhead PJ, 2005. Genetic evidence for sexual selection in black ratsnakes, *Elaphe obsoleta*. *Anim Behav* 69:225–234.
- Bonnet X, Naudeau G, Shine R, 1999. The dangers of leaving home: dispersal and mortality in snakes. *Biol Conserv* 89:39–50.
- Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH et al., 2014. Beast 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 10:e1003537.
- Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, Roy-Choudhury A, 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Mol Biol Evol* 29:1917–1932.
- Bryson R, Smith BT, Nieto-Montes de Oca A, García-Vázquez UO, Riddle BR, 2014. The role of mitochondrial introgression in illuminating the evolutionary history of Neartic treefrogs. *Zool J Linn Soc* 172:103–116.
- Burbrink FT, Lawson R, Slowinski JB, 2000. Mitochondrial DNA phylogeography of the polytypic North American rat snake *Elaphe obsoleta*: a critique of the subspecies concept. *Evolution* 54:2107–2118.
- Cattaneo A, 1998. Gli anfibi e rettili delle isole greche di Skyros, Skopelos e Alonissos (Sporadi settentrionali.). *Atti Soc It Sci Nat Museo Civ Stor Nat Milano* 139:127–149.
- Cattaneo A, 1999. Variabilità e sottospecie di *Elaphe quatuorlineata* (Lacépède) nelle piccole isole Egee (Serpentes: colubridae.). *Atti Soc It Sci Nat Museo Civ Stor Nat Milano* 140:119–139.
- Chan KMA, Levin SA, 2005. Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inheritance DNA. *Evolution* 59:720–729.
- Chernomor O, von Haeseler A, Minh BQ, 2016. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst Biol* 65:997–1008.
- Chifman J, Kubatko L, 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30:3317–3324.
- Crnobrnja-Isailović J, Ajtic R, Vogrin M, Corti C, Pérez-Mellado V et al., 2009. *Elaphe quatuorlineata*. In: *The IUCN Red List of Threatened Species*. version 2012.2 [last accessed on 2018 May 18]. Available from: <http://www.iucnredlist.org>.
- Curat M, Ruedi M, Petit RJ, Excoffier L, 2008. The hidden side of invasions: massive introgression by local genes. *Evolution* 62:1908–1920.
- de Queiroz A, Lawson R, Lemos-Espinal JA, 2002. Phylogenetic relationships of North American garter snakes (*Thamnophis*) based on four mitochondrial genes: how much DNA sequence is enough? *Mol Phylogenet Evol* 22:315–329.
- Dermitzakis MD, Papanikolaou DJ, 1981. Paleogeography and geodynamics of the Aegean region during the Neogene. *Annal Geolog Pays Hellen* 4:245–289.
- Dermitzakis MD, 1990. Paleogeography, geodynamic processes and event stratigraphy during the late Cenozoic of the Aegean area. *Accad Naz Lincei* 85:263–288.
- Drummond AJ, Suchard MA, Xie D, Rambaut A, 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29:1969–1973.
- Eaton DAR, 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics* 30:1844–1849.
- Edgar RC, 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797.
- Evanno G, Regnaut S, Goudet J, 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620.
- Excoffier L, Foll M, Petit RJ, 2008. Genetic consequences of range expansions. *Annu Rev Ecol Evol Syst* 40:481–501.
- Falush D, Stephens M, Pritchard JK, 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes* 7:574–578.
- Felsenstein J, 1985. Confidence limits on phylogenies: an approach using the Bootstrap. *Evolution* 39:783–791.
- Friis G, Aleixandre P, Rodriguez-Estrélla R, Navarro-Sigüenza AG, Milá B, 2016. Rapid postglacial diversification and long-term stasis within the songbird genus Junco: phylogeographic and phylogenomic evidence. *Mol Ecol* 25:6175–6195.
- Funk DJ, Omland KE, 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from Animal Mitochondrial DNA. *Annu Rev Ecol Evol Syst* 34:397–423.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W et al., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59:307–321.
- Hewitt GM, 2011. Quaternary phylogeography: the roots of hybrid zones. *Genetica* 139:617–638.
- Jombart T, 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403–1405.
- Jombart T, Devillard S, Balloux F, 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet* 11:94–108.



- Kapsimalis V, Pavlopoulos K, Panagiotopoulos I, Drakopoulou P, Vandarakis D, Sakelariou D, Anagnostou C, 2009. Geoarchaeological challenges at the Cyclades continental shelf (Aegean Sea). *Ann Geomorphol* 53:169–190.
- Kindler C, Bohme W, Corti C, Gvozdk V, Jablonski D et al., 2013. Mitochondrial phylogeography, contact zones and taxonomy of grass snakes (*Natrix natrix*, *N. megaloccephala*). *Zool Scripta* 42:458–472.
- Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I, 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Res* 15:1179–1191.
- Kornilios P, Thanou E, Lymberakis P, Sindaco R, Liuzzi C et al., 2014. Mitochondrial phylogeography, intraspecific diversity and phenotypic convergence in the four-lined snake (Reptilia, Squamata). *Zool Scripta* 43: 149–160.
- Kuriyama T, Brandley MC, Katayama A, Mori A, Honda M et al., 2011. A time-calibrated phylogenetic approach to assessing the phylogeography colonization history and phenotypic evolution of snakes in the Japanese Izu Islands. *J Biogeogr* 38:259–271.
- Kyriazi P, Kornilios P, Nagy ZT, Poulakakis N, Kumlutaş Y et al., 2013. Comparative phylogeography reveals distinct colonization patterns of Cretan snakes. *J Biogeogr* 40:1143–1155.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA et al., 2007. ClustalW and ClustalX version 2.0. *Bioinformatics* 23:2947–2948.
- Lenk P, Joger U, Wink M, 2001. Phylogenetic relationships among European ratsnakes of the genus *Elaphe* Fitzinger based on mitochondrial DNA sequence comparisons. *Amphibia-Reptilia* 22:329–339.
- Mallet J, 2005. Hybridization as an invasion of the genome. *Trends Ecol Evol* 20:229–237.
- McGuire JA, Linkem CW, Koo MS, Hutchison DW, Lappin AK et al., 2007. Mitochondrial introgression and incomplete lineage sorting through space and time: phylogenetics of *Crotaphytid* lizards. *Evolution* 61: 2879–2897.
- Minh BQ, Nguyen MAT, von Haeseler A, 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol* 30:1188–1195.
- Minh BQ, Trifinopoulos J, Schrempf D, Schmidt HA, 2017. IQ-TREE version 1.6.0: Tutorials and Manual. Phylogenomic software by maximum likelihood. Available from: <http://www.iqtree.org>, accessed 25 June 2018.
- Nagy ZT, Joger U, Wink M, Glaw F, Vences M, 2003. Multiple colonization of Madagascar and Socotra by colubrid snakes: evidence from nuclear and mitochondrial gene phylogenies. *Proc R Soc B Biol Sci* 270: 2613–2621.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ, 2015. IQTREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol Biol Evol* 32:268–274.
- Nieto-Feliner G, 2011. Southern European glacial refugia: a tale of tales. *Taxon* 60:365–372.
- Pernetta AP, Allen JA, Beebe TJ, Reading CJ, 2011. Fine-scale population genetic structure and sex-biased dispersal in the smooth snake *Coronella austriaca* in southern England. *Heredity* 107:231–238.
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE, 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* 7:e37135.
- Pickrell JK, Pritchard JK, 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet* 8:e1002967.
- Pritchard JK, Stephens M, Donnelly P, 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
- Psonis N, Antoniou A, Oleg K, Jablonski D, Boyan P et al., 2017. Hidden diversity in the *Podarcis tauricus* (Sauria, Lacertidae) species subgroup in the light of multilocus phylogeny and species delimitation. *Mol Phylogenet Evol* 106:6–17.
- Rambaut A, Drummond AJ, 2009. Tracer. ver. 1. 5. [Computer software]. Edinburgh, UK: Institute of Evolutionary Biology, University of Edinburgh. Available from: <http://tree.bio.ed.ac.uk/software/tracer>, accessed 10 August 2018.
- Ravinet M, Faria R, Butlin RK, Galindo J, Bierre N et al., 2017. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *J Evol Biol* 30:1450–1477.
- Rognes T, Flouri T, Nichols B, Quince C, Mahé F, 2016. A versatile open source tool for metagenomics. *PeerJ* 4:e2584.
- Royal CD, Novembre J, Fullerton SM, Goldstein DB, Long JC et al., 2010. Inferring genetic ancestry: opportunities, challenges, and implications. *Am J Hum Genet* 86:661–673.
- Rubinoff D, Holland BS, 2005. Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Syst Biol* 54:952–961.
- Sequeira F, Sodr  D, Ferrand N, Bernardi JA, Sampaio I et al., 2011. Hybridization and massive mtDNA unidirectional introgression between the closely related Neotropical toads *Rhinella marina* and *R. schneideri* inferred from mtDNA and nuclear markers. *BMC Evol Biol* 11:264–269.
- Sloan DB, Havird JC, Sharbrough J, 2017. The on-again, off-again relationship between mitochondrial genomes and species boundaries. *Mol Ecol* 26: 2212–2236.
- Spinks PQ, Thomson RC, Shaffer HB, 2014. The advantages of going large: genome-wide SNPs clarify the complex population history and systematics of the threatened western pond turtle. *Mol Ecol* 23:2228–2241.
- Swofford DL, 2003. *PAUP\*: Phylogenetic Analysis Using Parsimony, Version 4.0 b10*. Sunderland, MA: Sinauer Associates Inc.
- Thanou E, Giokas S, Kornilios P, 2014. Phylogeography and genetic structure of the slow worms *Anguis cephalonica* and *Anguis graeca* (Squamata: anguidae) from the southern Balkan Peninsula. *Amphibia-Reptilia* 35:263–269.
- Toews DPL, Brelsford A, 2012. The biogeography of mitochondrial and nuclear discordance in animals. *Mol Ecol* 21:3907–3930.
- Ursenbacher S, Schweiger S, Tomović L, Crnobrnja-Isailović J, Fumagalli L et al., 2008. Molecular phylogeography of the nose-horned viper *Vipera ammodytes*: evidence for high genetic diversity and multiple refugia in the Balkan Peninsula. *Mol Phylogenet Evol* 46:1116–1128.
- Wang W, Wang Y, Lei FY, Wang H, Chen J, 2018. Incomplete lineage sorting and introgression in the diversification of Chinese spot-billed ducks and mallards. *Curr Zool* <https://doi.org/10.1093/cz/zoy074>.
- Weisrock DWK, Hozak H, Larson A, 2005. Phylogeographic analysis of mitochondrial gene flow and introgression in the salamander *Plethodon shermani*. *Mol Ecol* 14:1457–1472.
- Wielstra B, Arntzen JW, 2012. Postglacial species displacement in *Triturus* newts deduced from asymmetrically introgressed mitochondrial DNA and ecological niche models. *BMC Evol Biol* 12:161–172.
- Wood DA, Vandergast AG, Lemos-Espinal JA, Fisher RN, Holycross AT, 2011. Refugial isolation and divergence in the Narrowheaded Gartersnake species complex *Thamnophis rufipunctatus* as revealed by multilocus DNA sequence data. *Mol Ecol* 20:3856–3878.
- Wu Y, 2015. A coalescent-based method for population tree inference with haplotypes. *Bioinformatics* 31:691–698.