Title:
Genome-wide markers untangle the green-lizard radiation in the Aegean Sea and support a rare biogeographical pattern

Short running title: *Lacerta* phylogeography in the Aegean

Article type: Research paper

Authors' names:
Panagiotis Kornilios[1,2,3†]; Evanthia Thanou[1,3]; Petros Lymberakis[4]; Çetin Ilgaz[5,6]; Yusuf Kumlutaş[5,6]; Adam Leaché[1,7]

Authors' research addresses:
1 Department of Biology, University of Washington, Box 351800, Seattle, WA 98195-1800, USA
2 Institute of Evolutionary Biology (CSIC - Universitat Pompeu Fabra), Passeig Marítim de la Barceloneta 37-49, E-08003 Barcelona, Spain
3 The Molecular Ecology Backshop, G. Lekka 16, GR–26300, Loutraki, Greece
4 Natural History Museum of Crete, University of Crete, Knossou Ave., GR-71409 Irakleio, Crete, Greece
5 Department of Biology, Faculty of Science, Dokuz Eylül University, 35160 Buca-İzmir, Turkey
6 Research and Application Center for Fauna Flora, Dokuz Eylul University, 35160 Buca-İzmir, Turkey.
7 Burke Museum of Natural History and Culture, University of Washington, Box 353010, Seattle, WA 98195-3010, USA

† **Corresponding author**: Department of Biology, University of Washington, Box 351800, Seattle, WA 98195-1800, USA, email: korniliospanagiotis@gmail.com, korniliospan@yahoo.gr

Word count: Abstract – 276
Main Text + references = 7741
Estimated number of pages required by figures and tables: 4

Abstract

Aim: The Aegean Sea constitutes a major biogeographic barrier between the European and Asian continents and several models of diversification in the Aegean have been documented. Here we test three of those models for the Aegean green-lizards (*Lacerta trilineata–pamphylica* group): Vicariance vs. Overland Dispersal vs. Island Stepping-stone Dispersal. We investigate these hypotheses and complement our knowledge on the impact of the Aegean Barrier on east Mediterranean taxa.

Location: Aegean Sea, east Mediterranean

Taxon: *Lacerta* lizards

Methods: We analysed ddRAD loci (double-digest restriction-site-associated DNA) to estimate species-trees under coalescent models and maximum likelihood trees using concatenation. We performed hierarchical population structure analyses and inferred ancestral distribution-areas. We also sequenced the complete cytochrome *b* gene and produced a time-calibrated mtDNA gene-tree tree to conduct a critical comparison with previous studies.

Results: Aegean green-lizards diverged into four main groups in parallel during the Late Pliocene with distributions to the East and West of the Aegean. The Eastern group includes *Lacerta pamphylica* and East Aegean *L. trilineata*, while the Western group contains the Central Cyclades populations and the remaining populations of the Balkan Peninsula. The Aegean green-lizards' ancestor occurred in Anatolia, while the West lineage ancestor occurred in the Central Cyclades islands, revealing a dispersal between the two regions. The radiations of all major green-lizard groups, including *trilineata*+*pamphylica*, occurred in parallel in the Late Pliocene.

Main Conclusions: In contrast to previously suggested biogeographical hypotheses for the group, based on mtDNA, the Island Stepping-stone Dispersal scenario is strongly supported. Green lizards offer a rare paradigm of diversification in the Aegean, where populations largely expanded their geographical distribution and crossed the Aegean Barrier by using the central Aegean islands as stepping stones.

KEYWORDS: Aegean Sea barrier, Anatolia, ddRAD, East Mediterranean, Genome wide SNPs, Lacertidae, Mid-Aegean Trench, Overseas dispersal, phylogeography, SNAPP coalescence

1. INTRODUCTION

1.1. The Aegean Sea: a biogeographic barrier and a bridge

The Aegean Sea constitutes a major contemporary barrier to biotic exchange between the continents of Asia and Europe. However, this has not always been the case. During the early and middle Miocene, this region was in fact a united landmass called the Agäis or Aegaeis (Creutzburg, 1963). This land included what we recognise today as the southern Balkan Peninsula in the west, Anatolia in the east and the Aegean itself, which at that time formed an unbroken plain (Figure 1). This configuration permitted cross-continental dispersal and uninterrupted geographic distributions of flora and fauna throughout the area.

The fragmentation of this landmass and the formation of one of the most important biogeographical barriers in the region derived from the northward movement and subduction of the African plate beneath the Eurasian plate (Le Pinchon & Angelier, 1981). One important consequence of this collision was the intrusion of the sea, which began approximately 12 Ma east of today's Crete (Creutzburg, 1963), moved northward and was complete some 8-9 Ma (Dermitzakis, 1990) (Figure 1). This sea barrier, often termed the Mid-Aegean Trench (MAT; Poulakakis et al., 2003), represented the first form of the Aegean Sea. Trenches are long, narrow depressions on the seafloor that form at the meeting points of tectonic plates. For example, the Hellenic Trench was formed in the south of the Aegean region from the collision of the African and Eurasian plates (Le Pinchon & Angelier, 1981). Since the Mid-Aegean Trench is not technically a trench, we refer to it as the Aegean Barrier (AB).

Palaeogeographic reconstructions some 8 Ma show that the eastern and western parts of the Aegean were separated by the AB, while the present central Aegean islands formed a united landmass connected to the west mainland, i.e. the present Balkan Peninsula. Crete was also isolated but in the form of a chain of smaller islands (Dermitzakis & Papanikolanou, 1981; Figure 1). During the late Pliocene, the palaeogeographic reconstructions converge to great extent but with some differentiations. Dermitzakis and Papanikolanou (1981) present a map of the Aegean at approximately 3.5 Ma, which shows no connections between the east and the west (the AB remained undisrupted), and the central Aegean islands were still forming a landmass connected to the mainland. Creutzburg (1963) presents a map of the late Pliocene, with no precise dates in Ma, according to which there was a narrow landbridge that connected present Anatolia and the central Aegean landmass (the AB was disrupted), while the latter was separated from the mainland by a narrow sea strait.

In both studies, Crete remained isolated, while the Peloponnese was disconnected from the rest of the Balkan Peninsula (Figure 1). Finally, Anastasakis, Piper, Dermitzakis and Karakitsios (2006) show that during the early Pliocene, the central Aegean landmass was connected to the mainland, but it was disconnected during the late Pliocene, again with no precise dates in Ma. The consensus of all these depictions is that until the late Pliocene the east and west Aegean most probably remained isolated by the AB and if there were actual landbridges between them, they must have been very limited in space and time. Additionally, the central Aegean islands formed a large landmass connected to the present Balkan Peninsula until the transition from the late to early Pliocene, when they were disconnected. During the Pleistocene, the geological configurations of the Aegean were similar to the present ones. However, the oscillating climatic conditions during the glacial and interglacial periods led to sea-level fluctuations, which connected and disconnected islands with their neighbouring mainlands. Accordingly, the central Aegean islands at times formed larger islands or were even united in one landmass, which never re-connected with the Balkan mainland because of the great sea depth in this area (Anastasakis et al., 2006).

In a biogeographical context, animal taxa have either historically not overcome the AB and are confined to one of the two sides, or can be found on both sides. In the latter cases, five models of diversification can be observed:

1. Vicariance: an old distribution becomes fragmented because of the AB formation.
2. Overland Dispersal: Dispersal from one side to the other following a route in the north around the AB.
3. Single Crossings: Taxa have the bulk of their distribution on one side but appear on a single or few locations on the other.
4. Human-mediated introductions to one or multiple locations.
5. Island Stepping-stone Dispersal from one side of the AB to the other and subsequent expansion. This rare pattern has been inferred for very few invertebrates in recent studies (Allegrucci, Trucchi & Sbordoni, 2011; Kornilios, Thanou, Kapli, Parmakelis & Chatzaki, 2016), but so far it has not been observed in terrestrial vertebrates.

1.2. Study system and biogeographical hypotheses

Reptiles and especially lizards have long been used as models for the study of speciation processes and phylogeographical patterns, with the family Lacertidae being

the most commonly studied group (Camargo, Sinervo & Sites, 2010). The west Eurasian genus *Lacerta* includes eight recognized species forming three species-groups. The east Mediterranean *L. trilineata* group comprises *L. media* Lantz & Cyrén, 1920, which is morphologically and genetically distinct (Ahmadzadeh et al., 2013), *L. pamphylica* Schmidtler, 1975 and *L. trilineata* Bedriaga, 1886.

The *trilineata+pamphylica* clade is an ideal model to test biogeographical hypotheses regarding the Aegean region. It presents significant morphological and genetic variation, has a large distribution on both sides of the AB and is found on all major island-groups (Figure 1). Mitochondrial phylogenies converge to an eastern (Anatolian) origin of the group, but with contradicting conclusions regarding the timing and mode of divergence (Ahmadzadeh et al., 2013; Sagonas et al., 2014). There is no prior knowledge on whether Aegean green-lizards are good or bad dispersers, but limited studies on other *Lacerta* species point to low dispersal capacities and a male-biased dispersal mode (Böhme, Schneeweiss, Fritz, Schlegel & Berendonk, 2006).

Here we test three alterative hypotheses regarding the historical biogeography of this group to draw broader conclusions on the patterns observed in the Aegean region, focusing on the role of the AB: Vicariance (V) vs. Overland Dispersal (OD) vs. Island Stepping-stone Dispersal (ISD) (see Appendix S1 in Supporting Information for a schematic presentation of the three models). Since the group presents an old radiation and is largely distributed on both sides of the AB (Ahmadzadeh et al., 2013; Sagonas et al., 2014), the human mediated and single-crossing scenaria are excluded. We test biogeographic models, by investigating the population structure, phylogeny, and biogeography of the species group using independent loci from across the genome, with a double-digest restriction-site-associated DNA sequencing approach (ddRADseq; Peterson, Weber, Kay, Fisher & Hoekstra, 2012). We compare our results with an expanded mtDNA gene tree for straightforward comparisons with published studies.

2. MATERIALS AND METHODS

2.1. Sampling

Sampling was designed to represent genetic diversity from *L. trilineata* and *L. pamphylica*, all involved biogeographical regions and all known mitochondrial lineages (Figure 1b). Specimen data (working codes, sampling localities and GenBank Accession Numbers) are given in Table 1, while sampling localities are also shown in Figure 1.

2.2. Genomic data: ddRAD library preparation, sequencing and bioinformatics

We collected ddRADseq data following the protocol, barcode-adaptors and indices of Peterson et al. (2012). Briefly, we double-digested genomic DNA with enzymes SbfI and MspI. Fragments were ligated with barcoded Illumina adapters, samples were pooled and, after each pool of eight samples was size-selected for fragments in the range of 415 - 515 bp, they were ligated with Illumina multiplexing indices. Sequencing was done on a single Illumina HiSeq 4000 lane (50 bp single-end read).

We processed raw Illumina reads using the program iPyRAD v.0.7.8 (Eaton, 2014). We demultiplexed samples using their unique barcode and index, and reduced each read to 39 bp after removal of the 6 bp restriction site overhang and the 5 bp barcode. Within the iPyRAD pipeline, the filtered reads for each sample were clustered using VSEARCH v.2.4.3 (Rognes, Flouri, Nichols, Quince & Mahé, 2016) and aligned with MUSCLE v.3.8.31 (Edgar, 2004).

We generated final datasets using three thresholds for the minimum number of individuals with data for a given locus: D0 (0% missing data, i.e. all loci present for all samples), D10 (10% missing data, all loci present for at least 90% of the samples) and D25 (25% missing data). Two types of final data-matrices were generated for different downstream analyses that included either the entire ddRAD locus (variable and invariant sites combined: “ddRAD” datasets) or by choosing one random SNP from each putatively unlinked locus (“uSNP” datasets). Each time a new dataset was generated the iPyRAD pipeline ran again for the specific set. Details regarding the parameters used for de-multiplexing and a summary of all resulting datasets from the iPyRAD pipeline are provided in Appendix S2, while a summary of this information is given in Table 2.

2.3. Genomic data: genetic clusters and admixture

Genetic structure within *trilineata+pamphylica* was inferred with two approaches using the uSNP-D0 datafile: Discriminant Analysis of Principal Components (DAPC; Jombart, Devillard & Balloux, 2010) using the R package ADEGENET (Jombart, 2008) and the Bayesian clustering analysis implemented in STRUCTURE v.2.3.4 (Pritchard, Stephens, & Donnelly, 2000). The optimal number of clusters (from 1 to 12) was estimated with the find.cluster function in ADEGENET, and the Bayesian Information Criterion (BIC) was used to select the optimal number of groups. To avoid overfitting, we used the a-score function to determine the appropriate number of principal components.

We used STRUCTURE to infer the number of genetic clusters K (1 to 12) and potential admixture. Analyses were performed with five runs of 500,000 iterations each (250,000 burn-in), with correlated allele frequencies and under the admixture model (Falush, Stephens, & Pritchard, 2003). We processed runs with the 'greedy' option and 2,000 random input orders in the CLUMPAK online web server (Kopelm, Mayzel, Jakobsson, Rosenberg & Mayrose, 2015) and evaluated the optimal K following Evanno, Regnaut & Goudet (2005). We used a hierarchical approach to determine whether additional structure was present in inferred groups by repeating STRUCTURE analyses using subsets of the data (see details in Results). We repeated this procedure until no additional population structure was supported ($K = 1$).

2.4. Genomic data: coalescence species trees and concatenated ddRAD loci

We estimated a species tree under the Bayesian multispecies coalescent framework of SNAPP v1.3 (SNP and AFLP Package for Phylogenetic analysis; Bryant, Bouckaert, Felsenstein, Rosenberg & RoyChoudhury, 2012) implemented in BEAST2 v2.4 (Bayesian Evolutionary Analysis Sampling Trees; Bouckaert et al., 2014). To avoid model violations (SNAPP assumes no gene flow), we excluded admixed individuals (membership probability <99% according to STRUCTURE). The final dataset included unlinked biallelic SNPs (biallelic uSNP-D10), no outgroup and the genetic-clustering results used for population assignments. Since SNAPP is computationally intensive, each population included 2-5 individuals, to a total of 22. Mutation rates (u, v) were both fixed at 1.

SNAPP uses a Yule prior with parameter lambda (λ) representing the speciation rate. For the λ prior we used a broad gamma distribution with a mean value of 1,000, set as $\alpha \times \beta$ ($\alpha=2$, $\beta=500$). Since λ determines the prior expected height of the tree, we used our non-reduced ddRAD dataset, containing all variable and constant characters, to

estimate the tree height (maximum observed divergence between any pair of taxa divided by two). We then utilised Jamie Oaks' script pyule (https://github.com/joaks1/pyule) to determine the mean value of λ.

The theta (θ) prior was also set as a gamma distribution with a mean value of 0.001, defined as α/β (α=25, β=25,000). Using this ratio, and since α/β^2 represents the variance, we ensured that the Standard Deviation (SD) was 0.0002. A prior mean θ =0.001 implies 0.1% variation between two randomly sampled alleles in a population. The values used (mean and SD) were estimated from the non-reduced dataset (percentage of polymorphic sites within each of the defined populations).

We performed two independent runs with a chain length of $6*10^6$ generations, sampling every 1,000 generations. We checked convergence (ESS>200) and determined burnin (10%) with TRACER v1.6 (Rambaut & Drummond, 2007). A maximum clade credibility tree (MCC) was summarized with TreeAnnotator.

Phylogenomic relationships among individuals and populations were also inferred using the coalescent method SVDquartets v.1.0 (Chifman & Kubatko, 2014) implemented in PAUP* v.4.0b10 (Phylogenetic Analysis Using Parsimony; Swofford, 2003). We evaluated all possible quartets, with and without prior assignment to populations and used non-parametric bootstrapping with 1,000 replicates for the statistical support. Final analyses ran without admixed individuals, *L. viridis* as outgroup and the uSNP-D25 dataset.

A maximum likelihood (ML) tree was also constructed using the concatenated ddRAD loci with IQ-TREE1.4.3 (Nguyen, Schmidt, von Haeseler & Minh, 2015). We used the "Auto" option for best-fit substitution model and tested nodal support via SH-aLRT tests (Shimodaira–Hasegawa approximate Likelihood Ratio Test; 1,000 replicates) (Guindon et al., 2010) and 1,000 ultrafast bootstrap alignments (Minh, Nguyen & von Haeseler, 2013). We tested several datasets and final analyses ran without admixed individuals, with an outgroup and the ddRAD-D25 dataset.

2.5. Genomic data: biogeographical analysis

To infer ancestral distribution areas, we employed the Bayesian binary Markov chain Monte Carlo (BBM) implemented in RASP v.3.2 (Reconstruct Ancestral State in Phylogenies; Yu, Harris, Blair & He, 2015). This analysis does not require an ultrametric tree nor considers the branch lengths, while it accepts trees with polytomies. In this context, we used as input the tree from the SVDquartets analysis of individuals, after pruning the outgroup. We assigned the terminal nodes to seven geographical areas that were defined based on the distribution of lineages, the geography and

palaeogeography of the region (Figure 1b). We run two independent analyses with ten MCMC chains for 10^7 generations under the JC+G model, states sampled every 100 generations, 25% burnin and ancestral ranges allowed to include up to four areas.

2.6. Mitochondrial DNA: gene tree analysis and estimation of divergence times

The complete mitochondrial cytochrome *b* gene (*cytb*) was PCR-amplified with primers L14910 and H16064 (Burbrink, Lawson & Slowinski, 2000), following standard procedures (e.g. Kyriazi et al., 2013). New sequences were combined with published ones to construct a dataset that included 41 ingroup sequences with a length of 1,137 bp.

The mtDNA phylogeny was assessed with ML carried out in IQ-TREE. The analysis ran with the "partitionfinder" and "Auto" options that determine the best partitioning scheme and the best-fit substitution model for each partition (Chernomor, von Haeseler & Minh, 2016). Nodal support was tested via 10,000 SH-aLRT tests, 10,000 ultrafast bootstrap alignments and 100 standard bootstrap alignments. We included *L. media* and rooted the tree with *L. agilis*.

To infer divergence times and in order to investigate the discrepancies between published studies, we combined our *cytb* sequences with sequences from GenBank to generate the datasets of Ahmadzadeh et al. (2013) and Sagonas et al. (2014). Specifically, we included the same six *Gallotia* and four *Timon* species, but added *Timon lepidus* since the split between this and its sister-species *T. nevadiensis* is the appropriate node to calibrate. Regarding the *Lacerta* representatives, we downloaded all available complete or near-complete *cytb* sequences from all eight species and removed redundant haplotypes. We generated a 1,137 bp dataset of 346 haplotypes and built a ML tree in IQ-TREE. This was, in turn, used to delimit mtDNA clusters with mPTP (Multi-rate Poisson tree processes; Kapli et al., 2017) and generate a final dataset of 45 sequences by including one representative from each mtDNA cluster, in order to conform to the Yule model of cladogenesis. We also performed a critical re-analysis of the Sagonas et al. (2014) dataset (Appendix S3).

To time-calibrate the phylogeny, we used the same four calibration points and prior probability distributions as in both published studies (details in Appendix S3). Analyses were performed in BEAST v1.8.4 (Drummond, Suchard, Xie & Rambaut, 2012), under the uncorrelated lognormal relaxed-clock approach with a Yule tree prior, four runs of $5*10^7$ generations and sampling every 1,000 steps. Analyses were carried out in CIPRES science gateway (Miller, Pfeiffer & Schwartz, 2010).

3. RESULTS

The ddRAD assemblies included a total of 34 samples, including two *L. viridis* individuals used as outgroups (Table 1). The mtDNA ML phylogeny was built on 47 complete *cytb* sequences (41 ingroup), while 31 additional sequences were used in the final molecular clock analysis (Table 1). Details on the ddRAD datasets that were used in the final analyses are shown in Table 2 and Appendix S2.

3.1. Genetic clusters and admixture

DAPC returned *K*=8 as the optimal number of clusters (Figure 2b), with a mean assignment probability of 0.97. Cluster analysis with STRUCTURE (Figure 2c) supported *K*=2, but the hierarchical procedure detected additional population genetic structure resulting in a total of nine groups. The two clusters grouped East and West Aegean populations, respectively. The East group was further divided into three groups: *L. pamphylica*, Lesvos island and all remaining East *L. trilineata* populations. Admixture between the last two was found for one of the Turkish specimens located across the island of Lesvos (Q=0.67). Analysis of the West group also returned three groups: the Central Cyclades islands, all specimens from Peloponnesos and Crete, and all remaining mainland and island populations. Further genetic clustering was supported within each of the last two groups. In the first case, populations from Peloponnesos and Crete were assigned into two respective clusters, and in the second, three more clusters were identified, the West Cyclades islands, East mainland Greece and adjacent islands, and West mainland Greece and islands. Two individuals from West mainland Greece showed admixture between the last two clusters (Q=0.85 and Q=0.67).

3.2. Phylogenomic trees and ancestral distributions

The MCC tree from SNAPP (Figure 2a) could not resolve the phylogenetic position of the Cretan populations relative to the Peloponnesian ones and the position of the West Cyclades populations relative to the mainland Greece ones. The SVDquartets species-tree favoured the relationship between Crete and east Peloponnesos (Figure 2a). The SVDquartets tree with individuals as terminal branches was very similar to the concatenated ML tree (Figure 3). In all ddRAD-based trees (Figures 2 and 3), two major monophyletic groups are identified, East and West of the Aegean. The East shows a clear split between *L. pamphylica* and *L. trilineata*. In the West, Central Cyclades populations branch first with all others monophyletic. These, in turn, form a southern

clade (Peloponnesos and Crete) and a northern one (mainland Greece, adjacent islands, West Cyclades).

According to the ancestral-area reconstructions (Figure 4), using the topology from SVDquartets, the ancestor of *trilineata+pamphylica* occurred in Anatolia (probability 0.80), while the ancestor of the West lineage occurred in the Central Cyclades islands (probability 0.76), revealing a dispersal event between the two regions.

3.3. Mitochondrial gene-tree and divergence times

The ML analysis ran with three partitions (per-codon position of *cytb*) and the resulting mtDNA gene-tree supports four major clades including *L. pamphylica*, Central Cyclades, East Aegean and West Aegean (Figure 3). The clade containing *L. pamphylica* and Central Cyclades conflicts with the topologies supported by the ddRADseq data, which place *L. pamphylica* sister to eastern populations and Central Cyclades sister to western populations. The topology within the East mtDNA clade is simple, with Lesvos Island forming a separate unit and populations from Turkey, East Aegean islands and Thrace forming a monophyletic group (Figure 3). The West mtDNA clade presents three monophyletic groups in polytomy: (a) west Peleponnesian populations, (b) east Peloponnesian populations and Crete and (c) all other mainland and island populations (Figure 3).

According to the estimated divergence times, the radiation within *trilineata+pamphylica* ocuured approximately 3.1 Mya (95% HPD intervals: 2.4-4.0 Mya) in the Late Pliocene (Figure 5). Similarly, the radiations of all other green-lizard groups (*L. agilis* complex, *L. media* complex, *L.viridis/bilineata* complex), but also the radiation of the eastern and western units of *trilineata+pamphylica*, took place roughly at the same time (mean values from 3.0 to 3.5 Mya; combined intervals: 2.2-4.4 Mya).

4. DISCUSSION

4.1. Phylogenetic patterns and mitonuclear discordance

The discordant phylogenetic patterns derived from mtDNA and genomic markers fall into three main categories. First, several relationships between major mtDNA lineages of the *trilineata+pamphylica* clade are unresolved or even misleading. Collecting thousands of independent loci and applying coalescence approaches that incorporate incomplete lineage sorting assisted with their resolution. Second, using genome wide SNPs allowed us to discover fine-scale genetic structure that was not apparent from mtDNA alone. Finally, the use of independent nuclear markers facilitated the detection of genetic admixture in several samples, which is linked to biogeographic mechanisms, such as dispersals and secondary contacts.

STRUCTURE analyses clearly cluster Aegean *Lacerta* into Eastern and Western groups, in full agreement with both coalescence and concatenated trees. In the East, *L. pamphylica* and East Aegean *L. trilineata* are monophyletic. MtDNA phylogenies have shown that *L. pamphylica* is nested within *L.trilineata*, but the clear *L. trilineata* paraphyly was never supported by nodal support on mtDNA gene-trees or alterative topology tests (Godinho, Crespo, Ferrand & Harris, 2005; Ahmadzadeh et al., 2013; Sagonas et al., 2014; current study). The peculiar relationship of the easternmost *L. pamphylica* to the Central Cyclades populations (Sagonas et al., 2014; current study), is most probably artifactual, the result of attraction between these long mtDNA branches. Genomic markers demonstrate that *L. trilineata* is paraphyletic in relation to *L. pamphylica*, providing a clearer picture of the group's history but also reinforcing the need for a general taxonomic re-evaluation. A thorough species-delimitation and systematic review of this group is underway.

East *L. trilineata* populations form two groups, one corresponding to the island of Lesvos and the other to all remaining populations (Figure 2b). One individual from the neighbouring part of Turkey is admixed between the two. The unique genetic identity of the samples from Lesvos is also demonstrated in the mtDNA tree (Sagonas et al., 2014; current study). However, all phylogenomic trees show that these populations, although genetically distinct, are nested within the east Aegean clade and do not represent an ancient split.

Within the West Aegean, the Central Cyclades populations split very early in the group's diversification history. The remaining western populations differentiate into a southern and a northern assemblage, a pattern that is not clear in the mtDNA tree. The situation in the south is very interesting: SVDquartets using individuals, SNAPP and

concatenated analyses return unresolved relationships among three southern lineages, i.e. west Peloponnesos, east Peloponnesos and Crete, while the SVDquartets species-tree and the mtDNA gene-tree favour the relationship between Crete and east Peloponnesos, rendering the Peloponnesian populations paraphyletic (Figure 2 and 3). STRUCTURE and DAPC do not differentiate the Peloponnesian populations into two clusters but identify the Cretan ones as separate. Finally, an individual from the central parts of Peloponnesos is nested in west Peloponnesos in the ddRAD concatenated tree but in east Peloponnesos in the mtDNA tree (Figure 3). The biogeographic background behind these complex patterns is not clear. Based on these results, Cretan populations are closely related to the Peloponnesian ones and probably those of the southeastern parts.

A similar interesting pattern is found within the north group. STRUCTURE and DAPC cluster populations into three subgroups: West Cyclades, West mainland and islands and East mainland and islands. While SVDquartets analysis of individuals and concatenated ML show the West Cyclades phylogenetically close to the East mainland, SNAPP and SVDquartets species-trees show unresolved relationships among the three lineages. The distinction of mainland Balkan populations into West and East (of Pindos Mt. range; Figure 1) is not clear in the mtDNA gene-tree but it is evident in the genomic markers. However, it is also clear that the southwestern populations show proportions of admixture with the east, implying genetic admixture between the two lineages (Figure 1 and 2).

4.2. Synchronous *Lacerta* radiations in the Pliocene

Our dating analysis included all green-lizard species and their internal mtDNA diversity throughout their range, with the exception of *L. strigata* and *L. schreiberi*, since we used all available *cytb* sequences, the most popular marker of choice for lacertid phylogenies. According to our results, all *Lacerta* radiations seem to have occurred during the same time (late Pliocene; Figure 5), in the same area (west Eurasia), and may have been triggered by the same factor, most probably the Pliocene climate transition (Salzmann et al., 2011). The Pliocene climatic fluctuations, coupled with the geomorphology and geological history of Anatolia, were responsible for the radiation of oriental green lizards into multiple lineages (Ahmadzadeh et al., 2013). *Lacerta media* further diversified in the east during the same time interval, while the *trilineata+pamphylica* group radiated in west Anatolia and the circum-Aegean region.

Our dating results agree almost perfectly with those from Ahmadzadeh et al. (2013), but largely deviate from Sagonas et al. (2014), even though all three studies

included the same dataset (taxon-wise), type of marker (mtDNA) and calibration strategy and implementation. However, when the latter study's dataset was re-analysed here, in exactly the same manner, the published result could not be replicated (Appendix S3). On the contrary, in all runs, the result always matched Ahmadzadeh et al. (2013) and the present study. Even so, the more recent nodes were still slightly older (Appendix S3). These slightly older dates were inferred when all sequences were included in the re-analysis, but after representing each putative genetic cluster with only one individual, the estimated times decreased and all three studies came to perfect agreement with regard to the divergence-times (Appendix S3). It is clear that the violation of the Yule speciation model, by including a great number of conspecific sequences, led to an overestimation of divergence times for the MRCAs of the respective entities. Additionally, the mean substitution rate for our *cytb* dataset was estimated to be 0.0153 (95% HPD: 0.0123-0.0185), very close to the mean value estimated for lacertids (0.0164; Carranza & Arnold, 2012), reinforcing the validity of our dating results.

4.3. Testing the biogeographic hypotheses

The ancestral-area reconstruction shows that the ancestor of Aegean green lizards was distributed in Anatolia while that of the populations west of the AB was distributed in the Central Cyclades. Hence, the favoured biogeographic scenario includes a dispersal event from Anatolia to the central Aegean islands and a second dispersal from the islands to the southern part of the Balkan Peninsula (Figure 4). Consequently, the biogeographic analysis supports the ISD hypothesis, according to which green lizards crossed the Aegean and used the central Aegean islands as stepping-stones (Figure 4 and Appendix S1). Very different outcomes would have been expected if the other two hypotheses were true. If the V scenario was favoured, then the ancestor of all studied populations would be distributed in the entire Aegean area, on both sides of the AB (not just Anatolia in the east), and the ancestor of the populations west of the AB would be distributed in all or most of the regions defined west of the AB (not just the central Aegean islands). Finally, if the OD hypothesis was favoured, then the ancestor of all Aegean green lizards would be distributed in Anatolia but the ancestor of the populations west of the AB would be distributed in at least the continental parts of the Balkan Peninsula and possibly other regions too (not just the central Aegean islands without the continental parts). The ISD hypothesis explains the outcome of all analyses from both mtDNA and genome-wide markers and fits all present and published results, with no need for assumptions of additional extinction events and secondary dispersals.

The two rejected hypotheses are in conflict with the phylogenomic and genetic-structure results and/or demand multiple assumptions. The V hypothesis (Appendix S1), favoured in Sagonas et al. (2014), follows the "Greek limbless skink paradigm" (*Ophiomorus punctatissimus*), according to which vicariance is the mode of diversification, as the formation of the AB fragments an old and wide distribution (Poulakakis, Pakaki, Mylonas & Lymberakis, 2008). This hypothesis should be rejected for the Aegean green lizards, not just because it is not supported by the biogeographical analysis on genomic data, but also because the AB was completely formed several millions of years before the estimated radiation time.

The OD hypothesis from Anatolia to the west after the formation of the AB (Appendix S1), which is favoured by Ahmadzadeh et al. (2013), is more compatible with the estimated divergence-times. This biogeographical pattern follows the "Snake-eyed skink paradigm" (*Ablepharus kitaibelii*), a demonstrated case of dispersal from Anatolia to the Balkan Peninsula around the AB (Skourtanioti et al., 2016). However, the phylogenetic results for *Lacerta* largely deviate from those of the snake-eyed skink. The latter phylogeny presents a branching pattern that follows the dispersal route from the northeast to the south of the peninsula (Skourtanioti et al., 2016). Yet, the opposite pattern is seen here for the Aegean *Lacerta*: high genetic diversity and structure is found in the south and the branching in the tree follows a south to north route. Ahmadzadeh et al. (2013) attribute this to an extinction event after the dispersal to the south of the Balkans and a secondary dispersal from south to north. However, the major lineage of the Central Cyclades was not represented in that study, hiding an important part of the underlying history. Although the scenario of OD around the Aegean with a subsequent extinction and a secondary northward dispersal cannot be definitively rejected, it is not supported by the biogeographical analysis and it is much less parsimonious (Figure 4).

4.4. Green lizards crossed the Aegean Barrier

The emergence of the AB had a profound impact on the eastern Mediterranean biodiversity. It is believed to be an impermeable barrier between the East and the West and animals that have been found to have crossed it are considered as exceptions (Lymberakis & Poulakakis, 2010). Most reptiles and amphibians that can be found on both sides, having crossed over the AB instead of by-passing it in the north via overland dispersal, have the bulk of their distribution on one side and appear on a single or few locations on the other (e.g. one or a few islands). Usually these single crossings are very recent, so that, besides rafting, human induced dispersals cannot be excluded (Lymberakis & Poulakakis, 2010). There are only two cases of old dispersals of

herpetofauna species over the AB, but they too should be characterised as single crossings. The dice snake *Natrix tessellata* mtDNA phylogeny presents an interesting relationship between west Anatolia and Crete, dating back to the Plio-Pleistocene boundary, so that a cross-barrier transmarine colonisation should be assumed (Kyriazi et al., 2013). Second, the populations of the green toad *Bufo viridis* on one Cycladic island (possibly more) belong to a distinct lineage that seems to be related to east Aegean populations, a relationship that dates back to the early Pleistocene (Dufresnes et al., 2018). Finally, other species are found on both sides of the AB as evident results of anthropogenic dispersals (e.g. Kornilios et al., 2010).

In recent years, an emerging literature has demonstrated similar ancient cross-barrier relationships for invertebrates (e.g Kornilios, Poulakakis, Mylonas & Vardinoyannis, 2009; Allegrucci, Trucchi & Sbordoni, 2011) that should also be considered single crossings of the AB. Until now there has only been one clear case, that of the trapdoor spider *Cyrtocarenum*, which has a large distribution on both sides of the AB, a result of two synchronous eastward dispersals, via two different routes (Kornilios et al., 2016). Interestingly the timing of those colonisations coincides with the one found here for *Lacerta* and one of the routes through the Central Cyclades is the same but on the opposite direction.

How did these recent and ancient single crossings and rare major expansions over the AB happen? How did green lizards cross the Aegean? As already mentioned, landbridges connecting the east and west parts of the Aegean have been depicted in some, albeit few, palaeogeographic reconstructions (Creutzburg, 1963). If actual landbridges existed during specific periods, these must have probably been very limited, and they could have facilitated this dispersal. However, transmarine dispersal via rafting combined with island-hopping is not an uncommon mode of long-distance dispersal in reptiles, often favoured by winds, sea-currents and the influence of major rivers (Hawlitschek, Ramírez Garrido & Glaw, 2017). While no information is available for the palaeowinds and palaeocurrents of the region, several major rivers flow from west Anatolia into the Aegean Sea that could promote westward dispersals. Moreover, overseas dispersal would have been more possible during periods of glacial maxima as early as in the late Pliocene, when the sea-level dropped dramatically expanding land surfaces and reducing the distances between islands and lands (Van Andel & Shackleton, 1982).

4.5. Conclusions and historical biogeography of Aegean green lizards

Our phylogenomic, biogeographic and mtDNA gene-tree analyses, together with the estimated divergence-times, strongly support the Island Stepping-stone Dispersal scenario. Green lizards offer a rare paradigm of diversification in the Aegean, where populations largely expanded their geographical distribution and crossed the Aegean Barrier by using the central Aegean islands as stepping stones. As already mentioned, green lizards dispersed from Anatolia to the central Aegean islands by either using actual landbridges or via rafting combined with island-hopping, if there were no land connections at that time.

Our estimated divergence times show that the radiation of Aegean green lizards (including the aforementioned dispersal over the AB) occurred sometime between 4.0 and 2.4 Ma (mean value: 3.1 Ma) (Figure 5). This time-period corresponds very well to the Late Pliocene (Piacenzian: 3.6-2.6 Ma). During the early Pliocene, the central Aegean islands formed a united landmass connected to the mainland; this connection was disrupted during the late Pliocene and was never re-established (Anastasakis et al., 2006). Hence, when green lizards dispersed to the central Cycladic landmass, the latter was or was not connected to the mainland, depending on the exact time this phylogenetic event happened (sometime between 4.0 and 2.4 Ma, according to the molecular clock results) and the exact time this land connection was disrupted (sometime during the period from early to late Pliocene). In this sense, and according to our biogeographical analysis, green lizards further dispersed to the southern Balkan Peninsula from the central Aegean island-mass, following either an overland or an overseas dispersal mode. The opposite pattern of reptiles dispersing to the central Aegean islands from the west (Balkan Peninsula) and then being isolated there is not at all uncommon. In fact, in several cases, the phylogenetic split between west mainland and central Aegean clades is estimated to have happened in the same time as for green lizards, which is attributed to the aforementioned land connection and subsequent disconnection (Ursenbacher et al., 2008; Kyriazi et al., 2013; Kornilios et al., 2014). The opposite pattern of colonisation found here for *Lacerta* is unique among vertebrates, but is dated during the same time.

Ahmadzadeh et al. (2013) conclude that also during the late Pliocene, climatic factors coupled with the geology of Anatolia differentiated oriental green lizards, including those of the studied group into two ancestral units. Consequently, during a relatively short period, sometime during the late Pliocene, an ancestor spread westwards from Anatolia to the central Aegean islands and from there to the southern Balkan Peninsula, while the Anatolian populations further diversified because of climatic factors. At the end of the Pliocene, this rapid radiation led to four ancestral lineages in

607 the area, now identified as the four major clades of the mtDNA and phylogenomic trees:
608 *L. pamphylica*, East *L. trilineata*, Central Cyclades *L. trilineata* and West *L. trilineata*.
609 Later on, during the late Pleistocene, Crete was colonised from the southeastern
610 parts of the Peloponnese (Figure 4 and 5). At that time Crete had no connections with
611 the mainland (Dermitzakis, 1990), which means this colonisation probably also
612 happened overseas during times of lower sea-level, with Kythera and Antikythera
613 islands acting as stepping stones (Figure 1). In fact, green-lizards from Crete and
614 Kythera form a single mtDNA clade (Sagonas et al., 2014). The situation within
615 Peloponnesos implies an old occurrence and isolation of *Lacerta*, with the existence of
616 two lineages that might even be paraphyletic. Peloponnesos, the southernmost tip of the
617 Balkan Peninsula, is an important biodiversity hotspot, as a result of its complicated
618 geological history and its role as a refugial area (Thanou, Giokas & Kornilios, 2014 and
619 references therein). Additionally, the West Cyclades do not host the Central Cyclades
620 lineage: they were colonised overseas from the east parts of mainland Greece in the late
621 Pleistocene, with a mechanism similar to that of Crete. Finally, the distinction of Greek
622 mainland populations, west and east of the Pindos mountain ridge (Figure 1b),
623 highlights the biogeographical role of this formation, as shown in other studies of
624 reptiles (e.g. Thanou et al., 2014).
625

627 **ACKNOWLEDMENTS**
628 This project has received funding from the European Union's Horizon 2020 research
629 and innovation programme under the Marie Sklodowska-Curie grant agreement No
630 656006 (Project Acronym: CoPhyMed). This work used the Vincent J. Coates
631 Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 OD018174
632 Instrumentation Grant. We are grateful to Kevin Epperly for his help with lab-work and
633 J. Howlett for the linguistic support of the text. We declare that we have no conflict of
634 interest.
635
636 **Data Accessibility**
637 Input files for the analyses based on ddRAD data are stored in Dryad
638 (doi:10.5061/dryad.15d8dq8). Newly generated mtDNA sequences have been deposited
639 in GenBank (Accession Numbers: MK330103 – MK330131).
640
641 **Supporting information**
642 Additional Supporting Information may be found in the online version of this article:
643
644 Appendix S1 The three biogeographic hypotheses discussed in the manuscript.
645 Appendix S2 Summary of the iPyRAD pipeline.
646 Appendix S3 Re-visiting the molecular-dating analysis of Sagonas et al. (2014).

REFERENCES

Ahmadzadeh, F., Carretero, M. A., Harris, D. J., Perera, A., & Böhme, W. (2012). A molecular phylogeny of eastern group of ocellated lizard genus *Timon* (Sauria: Lacertidae) based on mitochondrial and nuclear DNA sequences. *Amphibia-Reptilia*, 33, 1–10.

Ahmadzadeh, F., Flecks, M., Rödder, D., Böhme, W., Ilgaz, Ç., Harris, J.D., … Carretero, M.A. (2013). Multiple dispersal out of Anatolia: biogeography and evolution of oriental green lizards. *Biological Journal of the Linnean Society*, 110, 398–408.

Allegrucci, G., Trucchi, E., & Sbordoni, V. (2011). Time and mode of species diversification in *Dolichopoda* cave crickets (Orthoptera, Rhaphidophoridae). *Molecular Phylogenetics and Evolution*, 60, 108–121.

Anastasakis, G., Piper, D.J.W., Dermitzakis, M.D., & Karakitsios, V. (2006). Upper Cenozoic stratigraphy and paleogeographic evolution of Myrtoon and adjacent basins, Aegean Sea, Greece. *Marine and Petroleum Geology*, 23, 353–369.

Böhme, M.U., Schneeweiss, N., Fritz, U., Schlegel, M., & Berendonk, T.U. (2006). Small edge populations at risk: genetic diversity of the green lizard (*Lacerta viridis viridis*) in Germany and implications for conservation management. *Conservation Genetics*, 8*,* 555-563.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., … Drummond, A.J. (2014). Beast 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, 10, e1003537.

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A., & RoyChoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, 29, 1917–1932.

Burbrink, F.T., Lawson, R. &, Slowinski, J.B. (2000). Mitochondrial DNA phylogeography of the polytypic North American rat snake (*Elaphe obsoleta*): A critique of the subspecies concept. *Evolution*, 54, 2107–2118.

Busack, S. D., & Lawson, R. (2008). Morphological, mitochondrial DNA and allozyme evolution in representative amphibians and reptiles inhabiting each side of the Strait of Gibraltar. *Biological Journal of the Linnean Society*, 94, 445–461.

680 Camargo, A., Sinervo, B., & Sites, J.W.Jr. (2010). Lizards as model organisms for
681 linking phylogeographic and speciation studies. *Molecular Ecology*, 19, 3250–70.
682 Carranza, S., & Arnold, E.N. (2012). A review of the geckos of the genus
683 *Hemidactylus* (Squamata: Gekkonidae) from Oman based on morphology,
684 mitochondrial and nuclear data, with descriptions of eight new species. *Zootaxa*,
685 3378, 1–95.
686 Carranza, S., Arnold, E. N., & Amat, F. (2004). DNA phylogeny of *Lacerta*
687 (*Iberolacerta*) and other lacertine lizards (Reptilia: Lacertidae): Did competition
688 cause long-term mountain restriction? *Systematics and Biodiversity*, 2, 57-77.
689 Chernomor, O., von Haeseler, A., & Minh, B.Q. (2016). Terrace aware data structure
690 for phylogenomic inference from supermatrices. *Systematics Biology*, 65, 997–
691 1008.
692 Chifman, J., & Kubatko, L., (2014). Quartet inference from SNP data under the
693 coalescent model. *Bioinformatics*, 30, 3317–3324.
694 Creutzburg, N. (1963). Paleogeographic evolution of Crete from Miocene till our days.
695 *Cretan Annales*, 15/16, 336–342. (in Greek).
696 Dermitzakis, M.D. (1990). Paleogeography, geodynamic processes and event
697 stratigraphy during the late Cenozoic of the Aegean area *Accademia Nazionale*
698 *dei Lincei*, 85, 263–288.
699 Drummond, A.J., Suchard, M.A., Xie, D., & Rambaut, A. (2012). Bayesian
700 phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and*
701 *Evolution*, 29, 1969–1973.
702 Dufresnes, C., Lymberakis, P., Kornilios, P., Savary, R., Perrin, N. & Stöck, M. (2018).
703 Phylogeography of Aegean green toads (*Bufo viridis* subgroup): continental
704 hybrid swarm *vs*. insular diversification with discovery of a new island endemic.
705 BMC Evolutionary Biology, 18, 67.
706 Eaton, D.A.R. (2014). PyRAD: assembly of de novo RADseq loci for phylogenetic
707 analyses. *Bioinformatics*, 30, 1844–1849.
708 Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and
709 high throughput. *Nucleic Acids Research*, 32, 1792–1797.

710 Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of
711 individuals using the software STRUCTURE: a simulation study. *Molecular*
712 *Ecology*, 14, 2611–2620.
713 Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure:
714 Extensions to linked loci and correlated allele frequencies. *Genetics*, 164, 1567–
715 1587.
716 Godinho, R., Crespo, E.G., Ferrand, N., & Harris, D.J. (2005). Phylogeny and evolution
717 of the green lizards, *Lacerta* spp. (Squamata: Lacertidae) based on mitochondrial
718 and nuclear DNA sequences. *Amphibia-Reptilia*, 26, 271–285.
719 Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O.
720 (2010). New algorithms and methods to estimate maximum-likelihood
721 phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, 59,
722 307–321.
723 Hawlitschek, O., Ramírez Garrido, S., & Glaw, F. (2017). How marine currents
724 influenced the widespread natural overseas dispersal of reptiles in the Western
725 Indian Ocean region. *Journal of Biogeography*, 44, 1435-1440.
726 Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic
727 markers. *Bioinformatics*, 24, 1403–1405.
728 Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal
729 components: A new method for the analysis of genetically structured populations.
730 *BMC Genetics*, 11, 94.
731 Kapli, P., Lutteropp, S., Zhang, J., Kobert, K., Pavlidis, P., Stamatakis, A., & Flouri, T.
732 (2017). Multi-rate Poisson tree processes for single-locus species delimitation
733 under maximum likelihood and Markov chain Monte Carlo. *Bioinformatics*, 33,
734 1630–1638.
735 Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., & Mayrose, I. (2015).
736 Clumpak: a program for identifying clustering modes and packaging population
737 structure inferences across K. *Molecular Ecology Resources*, 15, 1179–1191.
738 Kornilios, P., Kyriazi, P., Poulakakis, N., Kumlutaş, Y., Ilgaz, Ç., Mylonas, M., &
739 Lymberakis, P. (2010). Phylogeography of the ocellated skink *Chalcides ocellatus*
740 (Squamata, Scincidae), with the use of mtDNA sequences: A hitch-hiker's guide to
741 the Mediterranean. *Molecular Phylogenetics and Evolution*, 54, 445-456.

Kornilios, P., Poulakakis, N., Mylonas, M., & Vardinoyannis, K. (2009). The phylogeny and biogeography of the genus *Zonites* Montfort, 1810 (Gastropoda: Pulmonata): preliminary evidence from mitochondrial data. *Journal of Molluscan Studies*, 75, 109–117.

Kornilios, P., Thanou, E., Kapli, P., Parmakelis, A., & Chatzaki, M. (2016). Peeking through the trapdoor: historical biogeography of the Aegean endemic spider *Cyrtocarenum* Ausserer, 1871 with an estimation of mtDNA substitution rates for Mygalomorphae. *Molecular Phylogenetics and Evolution*, 98, 300–313.

Kornilios, P., E. Thanou, P. Lymberakis, R. Sindaco, C. Liuzzi & S. Giokas. (2013). Mitochondrial phylogeography, intraspecific diversity and phenotypic convergence in the four-lined snake (Reptilia, Squamata). *Zoologica Scripta*, 43, 149-160.

Kyriazi, P., Kornilios, P., Nagy, Z.T., Poulakakis, N., Kumlutaş, Y., Ilgaz, Ç., ... Lymberakis, P. (2013). Comparative phylogeography reveals distinct colonization patterns of Cretan snakes. *Journal of Biogeograsphy*, 40, 1143–1155.

Le Pinchon, X., & Angelier, J. (1981). The Hellenic arc and trench system: a key to the neotectonic evolution of the eastern Mediterranean area. *Philosophical Transactions of the Royal Society of London*, 300, 357–372.

Lymberakis, P., & Poulakakis, N. (2010). Three continents claiming an archipelago: the evolution of Aegean's Herpetofaunal diversity. *Diversity*, 2, 233–255.

Marzahn, E., Mayer, W., Joger, U., Ilgaz, Ç., Jablonski, D., Kindler, C., Kumlutaş, Y., Nistri, A., Schneeweiss, N., Vamberger, M., Žagar, A., & Fritz, U. (2016). Phylogeography of the *L. viridis* complex: mitochondrial and nuclear markers provide taxonomic insights. *Journal of Zoological Systematics and Evolutionary Research*, 54, 85–105.

Miller, M.A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Proceedings of the Gateway Computing Environments Workshop (GCE)*, New Orleans, LA, 1–8.

Minh, B.Q., Nguyen, M.A.T., & von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*, 30, 1188–1195.

772 Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., & Minh, B.Q. (2015). IQTREE: A fast
773 and effective stochastic algorithm for estimating maximum likelihood phylogenies.
774 *Molecular Biology and Evolution*, 32, 268–274.
775 Paulo, O. S., Pinheiro, J., Miraldo, A., Bruford, M. W., Jordan, W. C., & Nichols, R. A.
776 (2008). The role of vicariance vs. dispersal in shaping genetic patterns in ocellated
777 lizard species in the western Mediterranean. *Molecular Ecology*, 17, 1535–1551.
778 Pavlicev, M., Mayer, W. (2009). Fast radiation of the subfamily Lacertinae (Reptilia:
779 Lacertidae): History or methodical artefact? *Molecular Phylogenetics and*
780 *Evolution*, 52, 727-734.
781 Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., & Hoekstra, H.E. (2012). Double
782 digest RADseq: an inexpensive method for de novo SNP discovery and
783 genotyping in model and non-model species. *PLoS ONE*, 7, e37135.
784 Poulakakis, N., Lymberakis, P., Antoniou, A., Chalkia, D., Zouros, E., Mylonas, M., &
785 Valakos, E. (2003). Molecular phylogeny and biogeography of the wall-lizard
786 *Podarcis erhardii* (Squamata: Lacertidae). *Molecular Phylogenetics and*
787 *Evolution*, 28, 38–46.
788 Poulakakis, N., Pakaki, V., Mylonas, M., & Lymberakis, P. (2008). Molecular
789 phylogeny of the Greek legless skink *Ophiomorus punctatissimus* (Squamata:
790 Scincidae): The impact of the Mid-Aegean trench in its phylogeography.
791 *Molecular Phylogenetics and Evolution*, 47, 396–402.
792 Pritchard, J.K., Stephens, M., & Donnelly, P (2000). Inference of population structure
793 using multilocus genotype data. *Genetics*, 155, 945–959.
794 Rambaut, A., & Drummond, A.J. (2007). Tracer v1.5. Available from
795 http://tree.bio.ed.ac.uk/ software/tracer/.
796 Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). A versatile open
797 source tool for metagenomics. *Peerj*, 4, e2584.
798 Sagonas, K., Poulakakis, N., Lymberakis, P., Parmakelis, A., Pafilis, P., & Valakos,
799 E.D. (2014). Molecular systematics and historical biogeography of the green
800 lizards (*Lacerta*) in Greece: Insights from mitochondrial and nuclear DNA.
801 *Molecular Phylogenetics and Evolution*, 76, 144–154.

802 Salzmann, U., Williams, M., Haywood, A.M., Johnson, A.L.A., Kender, S., &
803 Zalasiewicz, J. (2011). Climate and environment of a Pliocene warm world.
804 *Palaeogeography, Palaeoclimatology, Palaeoecology*, 309, 1–8.
805 Skourtanioti, E., Kapli, P., Ilgaz, Ç., Kumlutaş, Y., Avci, A., Ahmadzadeh, F., ...
806 Poulakakis, N. (2016). A reinvestigation of phylogeny and divergence times of the
807 *Ablepharus kitaibelii* species complex (Sauria, Scincidae) based on MtDNA and
808 NuDNA genes. *Molecular Phylogenetics and Evolution*, 103, 199–214.
809 Swofford, D.L. (2003). *PAUP∗: Phylogenetic Analysis Using Parsimony, Version 4.0*
810 *b10*. Sinauer Associates, Sunderland.
811 Thanou, E., Giokas, S., & Kornilios, P. (2014). Phylogeography and genetic structure of
812 the slow worms *Anguis cephallonica* and *Anguis graeca* (Squamata: Anguidae)
813 from the southern Balkan Peninsula. *Amphibia – Reptilia*, 35, 263–269.
814 Ursenbacher, S., Schweiger, S., Tomović, L., Crnobrnja-Isailović, J., Fumagalli, L. &
815 Mayer, W. (2008). Molecular phylogeography of the nose-horned viper (*Vipera*
816 *ammodytes*): evidence for high genetic diversity and multiple refugia in the Balkan
817 peninsula. *Molecular Phylogenetics and Evolution*, 46, 1116–1128.
818 Van Andel, T., & Shackleton, J. (1982). Late Paleolithic and Mesolithic Coastlines of
819 Greece and the Aegean. *Journal of Field Archaeology,* 9(4), 445-454.
820 Yu, Y., Harris, A.J., Blair, C.. & He, X.J. (2015). RASP (Reconstruct Ancestral State in
821 Phylogenies): a tool for historical biogeography. *Molecular Phylogenetics and*
822 *Evolution*, 87, 46–49.

824 **Biosketch**:
825 **Panagiotis Kornilios** is a postdoctoral fellow working on animal phylogenetics,
826 phylogeography, molecular ecology and evolution. This work is part of a Marie Curie-
827 Skłodowska Global Fellowship awarded to PK and hosted by the Leaché Lab
828 (University of Washington, Seattle, USA) and the Carranza Lab (IBE, Barcelona,
829 Spain). PK's current work focuses on the application of genome-targeted benchwork
830 and analytical methods on comparative phylogeography. The authors have a long-
831 lasting collaboration in phylogeography and phylogenetics of animal taxa, especially
832 reptiles, from the eastern Mediterranean.
833
834 **Author contributions**: P.K. conceived the idea; P.K., E.T. and A.L. designed the work;
835 P.K., E.T., P.L. Ç.I., and Y.K. collected and/or contributed specimens; P.K. and E.T.
836 carried out laboratory work and analyses; P.K. led the writing and all authors were
837 involved in the writing process.

838 **Table 1.**

839 Information regarding the samples analysed in the current study. The fist column (left)
840 shows the working codes of the samples and, in some cases, the GenBank Accession
841 Numbers of the respective sequences that we used as working codes. The first 34
842 samples (bold) are the ones included in the ddRAD analyses. Other information
843 includes the species names and sampling localities (with regions and country) of the
844 analysed specimens. For the ones retrieved from GenBank, the source (literature) is
845 reported, instead of the sampling locality. NHMC = Natural History Museum of Crete.
846 Accession numbers of sequence data that were generated here or retrieved from
847 GenBank are shown in the last column of the table. Input files for the analyses based on
848 ddRAD (double-digest restriction-site-associated DNA) data are stored in Dryad
849 (doi:10.5061/dryad.15d8dq8).

850

851

Working Code	Species	Locality, Region / Source (Literature)	GenBank Accession Numbers Cytochrome b
LT391	*L. trilineata*	Skyros isl., Greece	MK330107
LT393	*L. trilineata*	Skyros isl., Greece	MK330108
LT394	*L. trilineata*	Milos isl., Greece	MK330112
LT402	*L. trilineata*	Kimolos isl., Greece	MK330113
LT441	*L. trilineata*	Crete isl., Greece	MK330117
LT446	*L. trilineata*	Crete isl., Greece	MK330118
LT371	*L. trilineata*	Naxos isl., Greece	MK330119
NHMC80.3.61.76	*L. trilineata*	Serifos isl., Greece	MK330111
LT416	*L. trilineata*	Tinos isl., Greece	MK330121
LT375	*L. trilineata*	Naxos isl., Greece	MK330120
LT353	*L. trilineata*	Kaiafas, Peloponnesos, Greece	MK330114
LT360	*L. trilineata*	Selinountas, Peloponnesos, Greece	MK330103
LT377	*L. trilineata*	Argolis, Peloponnesos, Greece	MK330115
LT378	*L. trilineata*	Argolis, Peloponnesos, Greece	MK330104
LT345	*L. trilineata*	Feneos, Peloponnesos, Greece	MK330116
LT382	*L. trilineata*	Evvoia, Greece	MK330109
LT469	*L. trilineata*	Aitoloakarnania, Greece	MK330106
NHMC80.3.60.276	*L. trilineata*	Rhodos isl., Greece	MK330127
NHMC80.3.60.277	*L. trilineata*	Rhodos isl., Greece	MK330128
NHMC80.3.60.316	*L. trilineata*	Xanthi, Greece	-
NHMC80.3.60.217	*L. trilineata*	Dodoni, Greece	MK330105
NHMC80.3.60.418	*L. trilineata*	Kerkyra isl., Greece	MK330110
NHMC80.3.60.419	*L. trilineata*	Kerkyra isl., Greece	-
NHMC80.3.60.185	*L. trilineata*	Lesvos isl., Greece	MK330122
NHMC80.3.60.186	*L. trilineata*	Lesvos isl., Greece	-
LT408	*L. trilineata*	Samos isl., Greece	MK330123
LT2	*L. trilineata*	Kertil, Balıkesir, Turkey	MK330129
LT3	*L. trilineata*	Evciler, Bayramiç, Çanakkale, Turkey	MK330131
LT6	*L. trilineata*	Çığrı village, Başmakçı, Afyon, Turkey	MK330130
LT8	*L. trilineata*	Koçarlı, Aydın, Turkey	MK330124
LP2	*L. pamphylica*	Akseki, Antalya, Turkey	MK330125
LP3	*L. pamphylica*	Karakışla, Akseki, Antalya, Turkey	MK330126
LV480	*L. viridis*	Pertouli, Greece	-
NHMC80.3.60.413	*L. viridis*	Evros, Greece	-
KC897016	*L. trilineata*	Ahmadzadeh et al. (2013)	KC897016
KC897015	*L. trilineata*	Ahmadzadeh et al. (2013)	KC897015

KC897014	*L. trilineata*	Ahmadzadeh et al. (2013)	KC897014
KC897017	*L. trilineata*	Ahmadzadeh et al. (2013)	KC897017
KC897018	*L. trilineata*	Ahmadzadeh et al. (2013)	KC897018
KC897021	*L. trilineata*	Ahmadzadeh et al. (2013)	KC897021
KC897020	*L. trilineata*	Ahmadzadeh et al. (2013)	KC897020
KC897019	*L. trilineata*	Ahmadzadeh et al. (2013)	KC897019
LN835029	*L. trilineata*	Marzahn et al. (2016)	LN835029
LN835028	*L. trilineata*	Marzahn et al. (2016)	LN835028
KC897013	*L. pamphylica*	Ahmadzadeh et al. (2013)	KC897013
LN835022	*L. pamphylica*	Marzahn et al. (2016)	LN835022
KC896975	*L. media*	Ahmadzadeh et al. (2013)	KC896975
KC897005	*L. media*	Ahmadzadeh et al. (2013)	KC897005
KC896982	*L. media*	Ahmadzadeh et al. (2013)	KC896982
KC896988	*L. media*	Ahmadzadeh et al. (2013)	KC896988
LN835021	*L. media*	Marzahn et al. (2016)	LN835021
GQ142118	*L. agilis*	Pavlicev and Mayer (2009)	GQ142118
LN834626	*L. bilineata*	Marzahn et al. (2016)	LN834626
LN834643	*L. bilineata*	Marzahn et al. (2016)	LN834643
LN834669	*L. bilineata*	Marzahn et al. (2016)	LN834669
LN834710	*L. bilineata*	Marzahn et al. (2016)	LN834710
LN834705	*L. bilineata*	Marzahn et al. (2016)	LN834705
LN834743	*L. viridis*	Marzahn et al. (2016)	LN834743
LN834723	*L. viridis*	Marzahn et al. (2016)	LN834723
LN834769	*L. viridis*	Marzahn et al. (2016)	LN834769
LN834818	*L. viridis*	Marzahn et al. (2016)	LN834818
LN835024	*L. strigata*	Marzahn et al. (2016)	LN835024
LN835023	*L. schreiberi*	Marzahn et al. (2016)	LN835023
AY616290	*L. agilis*	Kalyabina-Hauf and Ananjeva (2004). Direct submission	AY616290
AY616335	*L. agilis*	Kalyabina-Hauf and Ananjeva (2004). Direct submission	AY616335
KC665499	*L. agilis*	Andres,C. (2013) Direct submission	KC665499
LN835020	*L. agilis*	Marzahn et al. (2016)	LN835020
AY616263	*L. agilis*	Kalyabina-Hauf and Ananjeva (2004). Direct submission	AY616263
KC665498	*L. agilis*	Andres,C. (2013) Direct submission	KC665498
AY616305	*L. agilis*	Kalyabina-Hauf and Ananjeva (2004). Direct submission	AY616305
AY616241	*L. agilis*	Kalyabina-Hauf and Ananjeva (2004). Direct submission	AY616241
KC665477	*L. agilis*	Andres,C. (2013) Direct submission	KC665477
KC665471	*L. agilis*	Andres,C. (2013) Direct submission	KC665471
DQ902143	*Timon tangitanus*	Busack and Lawson (2008)	DQ902143
AF378967	*Timon pater*	Paulo et al. (2008)	AF378967
GQ142119	*Timon nevadiensis*	Pavlicev and Mayer (2009)	GQ142119
DQ902142	*Timon lepidus*	Busack and Lawson (2008)	DQ902142
JQ425836	*Timon kurdistanicus*	Ahmadzadeh, F., Carretero, M. A., Harris, D. J., Perera, A., & Böhme, W. (2012	JQ425836
AY151838	*Gallotia stehlini*	Carranza, S., Arnold, E. N., & Amat, F. (2004)	AY151838
AY151836	*Gallotia atlantica*	Carranza, S., Arnold, E. N., & Amat, F. (2004)	AY151836
AY151844	*Gallotia intermedia*	Carranza, S., Arnold, E. N., & Amat, F. (2004)	AY151844
AM489592	*Gallotia galloti*	Klassert,T.E., Suarez,N.M., Almeida,T., Lopez,M., Pestano,J.J. and Hernandez,M. (2007) Direct submission	AM489592
AY154903	*Gallotia caesaris*	Carranza, S., Arnold, E. N., & Amat, F. (2004)	AY154903
AY154902	*Gallotia caesaris*	Carranza, S., Arnold, E. N., & Amat, F. (2004)	AY154902

Table 2.

Summary of the ddRAD (double-digest restriction-site-associated DNA) data matrices, as resulted from the iPyRAD pipeline. This includes information on the number of prefiltered and filtered loci, the total number of base pairs, and the number of Single Nucleotide Poymorphisms (SNPs), unlinked uSNPs and biallelic uSNPs. This information is presented for three different datasets built with different percentages of missing data, expressed as % of individuals for a given locus: D0 (0% missing data, i.e. all loci present for all samples), D10 (10% missing data, all loci present for at least 90% of the samples) and D25 (25% missing data, all loci present for at least 90% of the samples). Dataset D25 also included the outgroup *Lacerta viridis*. These datasets were used in different downstream analyses: DAPC = Discriminant Analysis of Principal Components implemented in the R package ADEGENET; STRUCTURE = Bayesian clustering analysis; SNAPP = SNP and AFLP Package for Phylogenetic analysis implemented in BEAST2 (Bayesian Evolutionary Analysis Sampling Trees); SVDquartets implemented in PAUP* (Phylogenetic Analysis Using Parsimony); Concatenation = maximum likelihood analysis in IQ-TREE. Next to each analysis, we report which type of dataset was used: uSNPs, biallelic uSNPs, ddRAD (the entire sequence of the ddRAD loci).

Dataset name		**D0**	**D10**	**D25-outgroup**
iPyRAD parameters	Minimum % of individuals for a given locus	100	90	75
Results Statistics	Number of prefiltered loci	67,999		73,575
	Number of filtered loci	599	5,944	11,620
	Total number of base pairs	24,036	238,331	466,138
	Number of SNPs	1,141	10,993	25,715
	Number of uSNPs	470	4,767	9,873
	Number of biallelic uSNPs	-	2,199	-
Final Analyses	DAPC	uSNPS		
	STRUCTURE	uSNPS		
	SNAPP		uSNPS biallelic	
	SVDquartets			uSNPS
	Concatenation			ddRAD

FIGURE 1.

The Aegean region, situated in the east Mediterranean (see inset at the lower right of the figure) and the geographical distributions of the study taxa *L. trilineata* and *L. pamphylica*. The sampling localities are shown as black dots, with the sample codes next to them. These sampling codes are also given in Table 1, together with other information on the analysed material. The ddRAD (double-digest restriction-site-associated DNA) assemblies included a total of 34 samples, with two *L. viridis* individuals used as outgroups (Table 1). The mitochondrial DNA maximum likelihood phylogeny was built on 47 complete cytochrome *b* sequences (41 ingroup), while 31 additional sequences were used in the final molecular clock analysis (Table 1). (b) The sampling localities of the present study: small circles/squares = mitochondrial DNA (mtDNA) data, large circles/squares = mtDNA and ddRAD data (double-digest restriction-site-associated DNA). Colour of the circles corresponds to the identified clusters from the genetic structure analyses, i.e. DAPC (Discriminant Analysis of Principal Components implemented in the R package ADEGENET) and the hierarchical Bayesian population clustering with STRUCTURE), and the identified genetic lineages from SVDquartets, implemented in PAUP* (Phylogenetic Analysis Using Parsimony) and SNAPP (SNP and AFLP Package for Phylogenetic analysis), implemented in BEAST2 (Bayesian Evolutionary Analysis Sampling Trees), as shown in Figure 2. In three cases, admixed populations, as derived from STRUCTURE, are shown as pies, with different colours representing the admixture proportion of each cluster. In the same Figure 1b, the seven geographical areas that were defined for the ancestral-area reconstruction analysis are also shown, with different colours corresponding to those of Figure 4. Names of islands and other geographic elements, mentioned in the manuscript, are also presented.

FIGURE 2.

Species-tree and genetic clustering results, using the unlinked SNPs dataset (unlinked Single Nucleotide Polymorphisms). (a) Tree derived from the Bayesian coalescence analysis of SNAPP (SNP and AFLP Package for Phylogenetic analysis) implemented in BEAST2 (Bayesian Evolutionary Analysis Sampling Trees). Values on nodes represent statistical support (SNAPP posterior probability / Bootstrap values from SVDquartets species-tree analysis) with filled circles representing full support (1.0 and 100). (b) The results from the Discriminant Analysis of Principal Components (DAPC) (optimal K=8 and number of PCA=3) implemented in the R package ADEGENET. (c) The results from the hierarchical Bayesian population clustering with STRUCTURE, shown next to the corresponding clades of the species-tree. Horizontal bars represent individuals with

the colour of each one representing the proportion of that individual's membership in each cluster. From right to left: all populations (optimal *K*=2); populations from east and west of the Aegean Barrier (optimal *K*=3, in both cases); populations from the north and south (optimal *K*=3 and *K*=2, respectively). The colours of each STRUCTURE cluster correspond to those of the clusters in DAPC (Figure 2b) and the clades of the species-tree (Figure 2a). They also correspond to the colours of the circles/pies that represent sampling localities in Figure 1.

FIGURE 3.

Left: Results from the phylogenomic analyses using individuals as terminal branches: Maximum likelihood (ML) tree on the concatenated ddRAD (double-digest restriction-site-associated DNA) loci from IQ-tree and SVDquartets, implemented in PAUP* (Phylogenetic Analysis Using Parsimony) on the unlinked SNPs (Single Nucleotide Polymorphisms). The tree presented is the one from the concatenated ddRAD loci with the statistical support on the nodes from both analyses (SH-aLRT, i.e. Shimodaira–Hasegawa approximate Likelihood Ratio Test, from IQ-tree / ultrafast bootstraps from IQ-tree / bootstrap values from SVDquartets). Filled circles indicate full support. Right: the ML tree from IQ-tree based on the mitochondrial DNA. Filled circles on nodes indicate full support (1.0 and 100) while open circles represent nodes with weaker support (SH-aLRT / ultrafast bootstraps / standard bootstraps). Dashed lines highlight specimens with interesting phylogenetic placements discussed in the text. Sample codes for terminals are provided in Table 1 and in Figure 1. In both trees, outgroups are not shown.

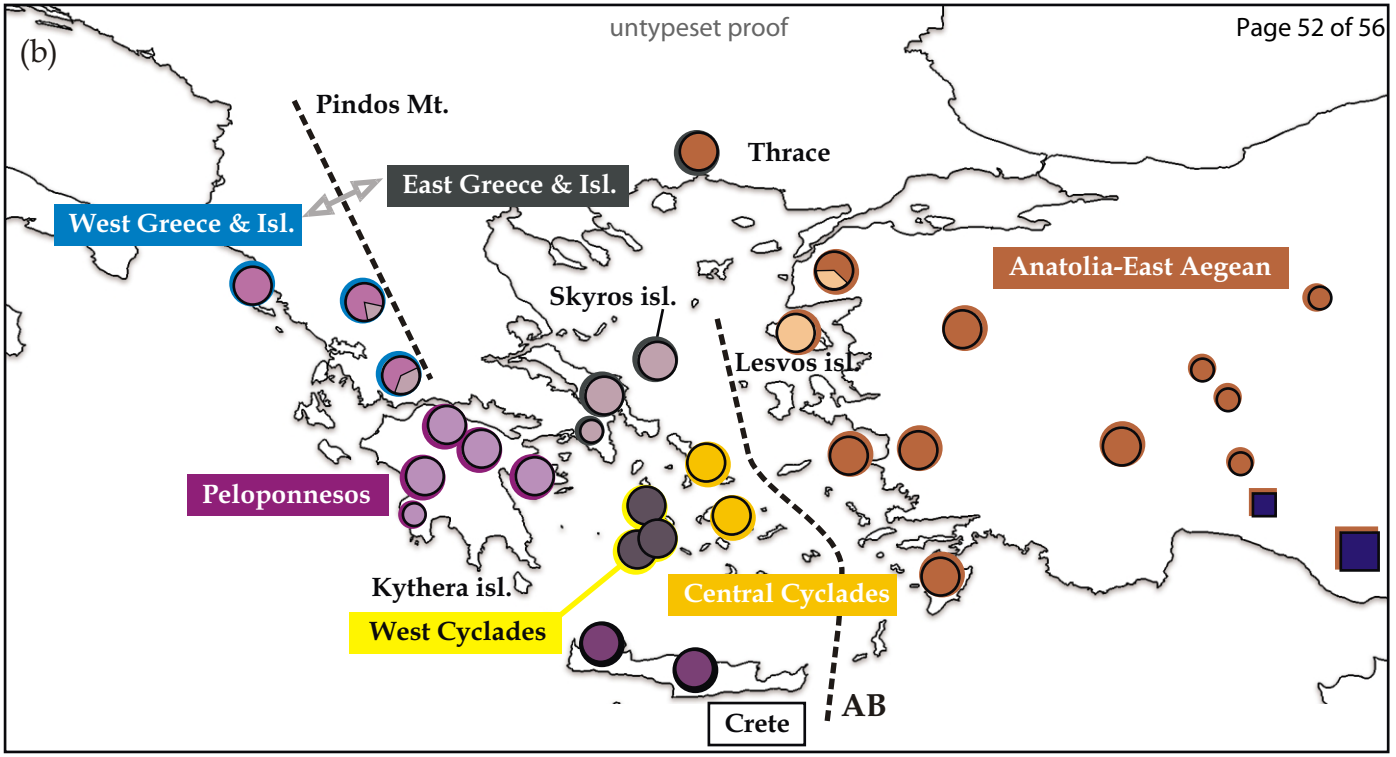
FIGURE 4.

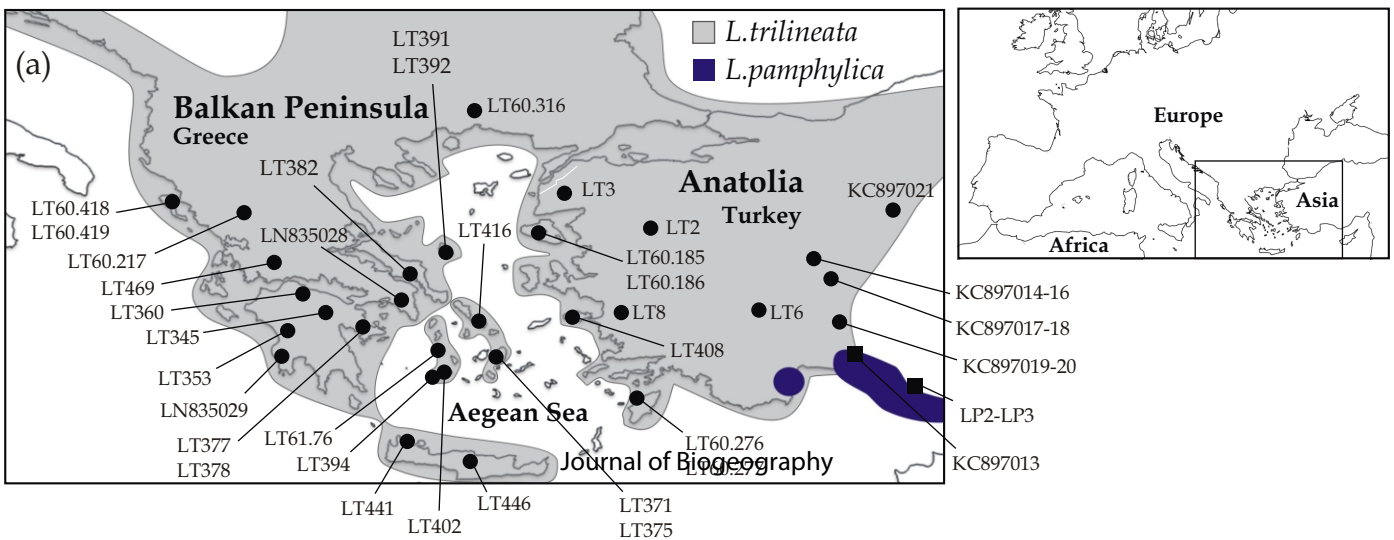
The ancestral distribution areas obtained with the Bayesian binary Markov chain Monte Carlo (BBM) in RASP (Reconstruct Ancestral State in Phylogenies). Terminal nodes were assigned to seven geographical areas presented in the inset with different colours (see also map in Figure 1). Probabilities for reconstructions are given as percentages next to the proposed ancestral areas and are symbolised as pies. Terminal branches refer to the names of the geographic location/area of occurrence for the respective lineages.

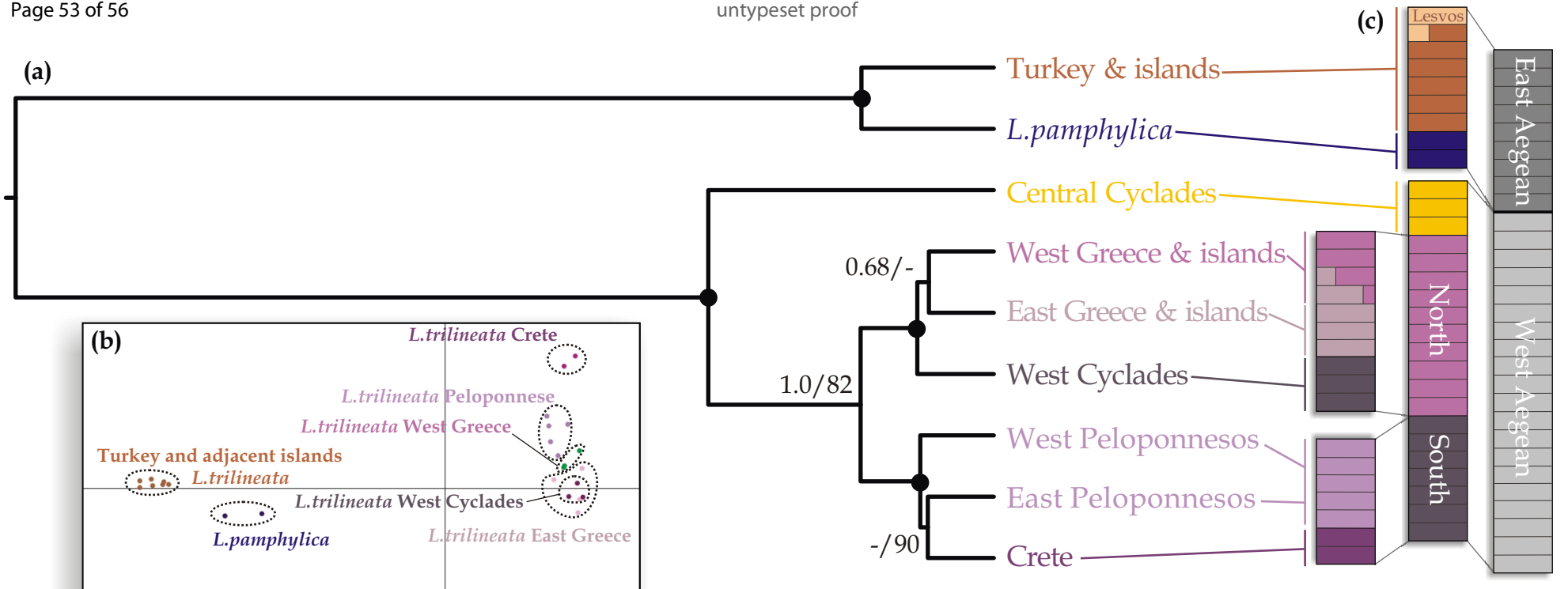
FIGURE 5.

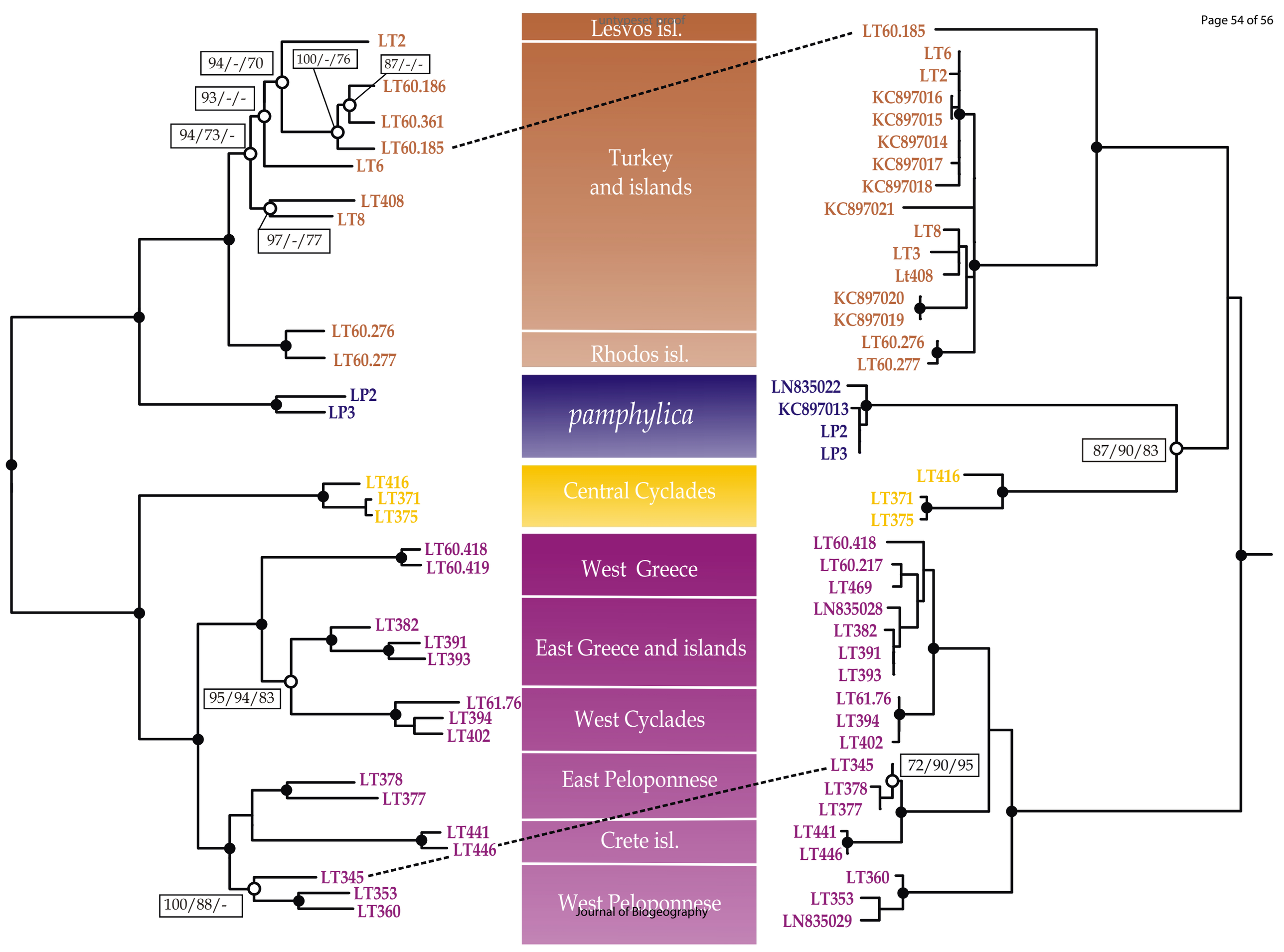
Dated mitochondrial gene-tree, based on complete cytochrome *b* sequences, inferred with BEAST (Bayesian Evolutionary Analysis Sampling Trees). Numbers next to nodes are mean ages and 95% intervals, both in Ma. Axis on the bottom refers to My. Nodes that represent the ancestors of species other than the target-species of the present study,

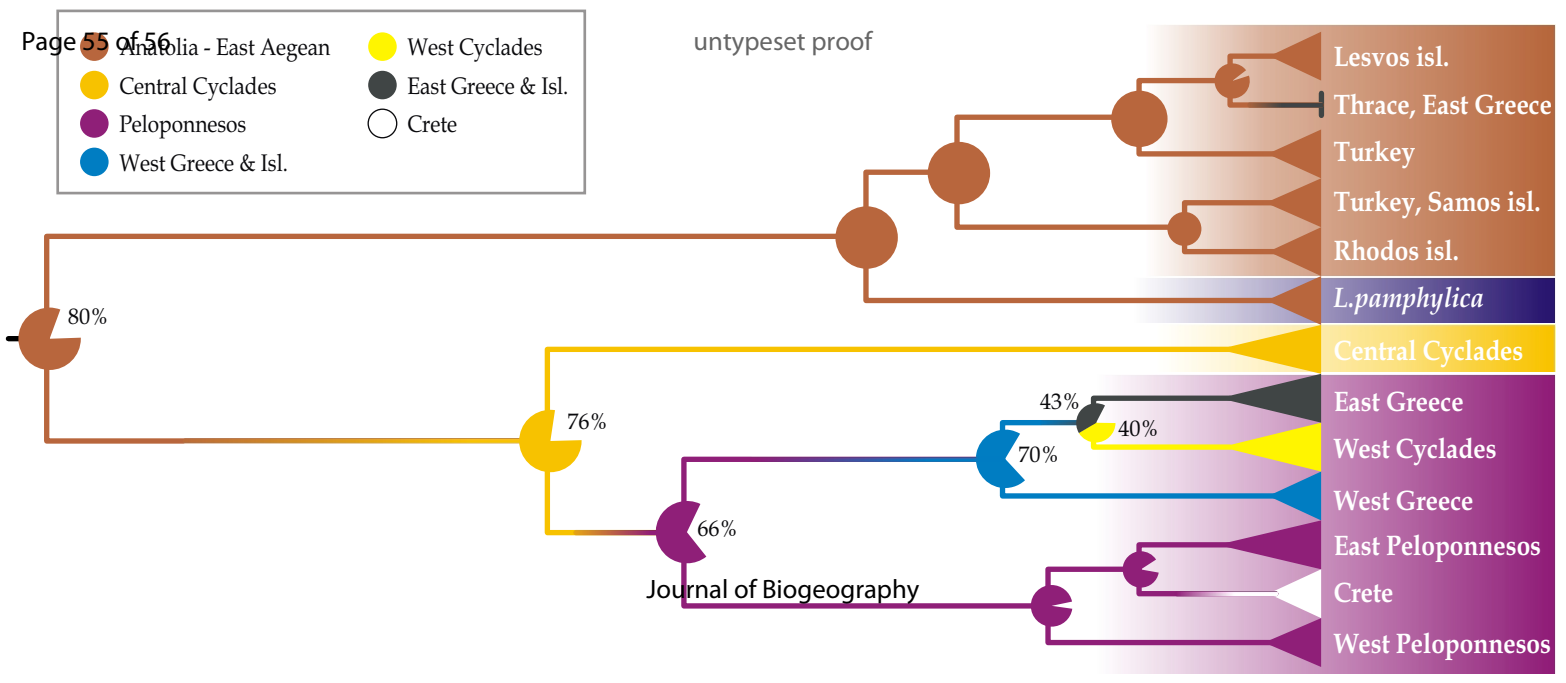
948 have been collapsed. Within the studied-group, each clade is a representative of each
949 mitochondrial cluster derived from the mPTP analysis (Multi-rate Poisson tree
950 processes). Names next to terminal branches refer to the geographic location/area the
951 respective lineages occur.
952

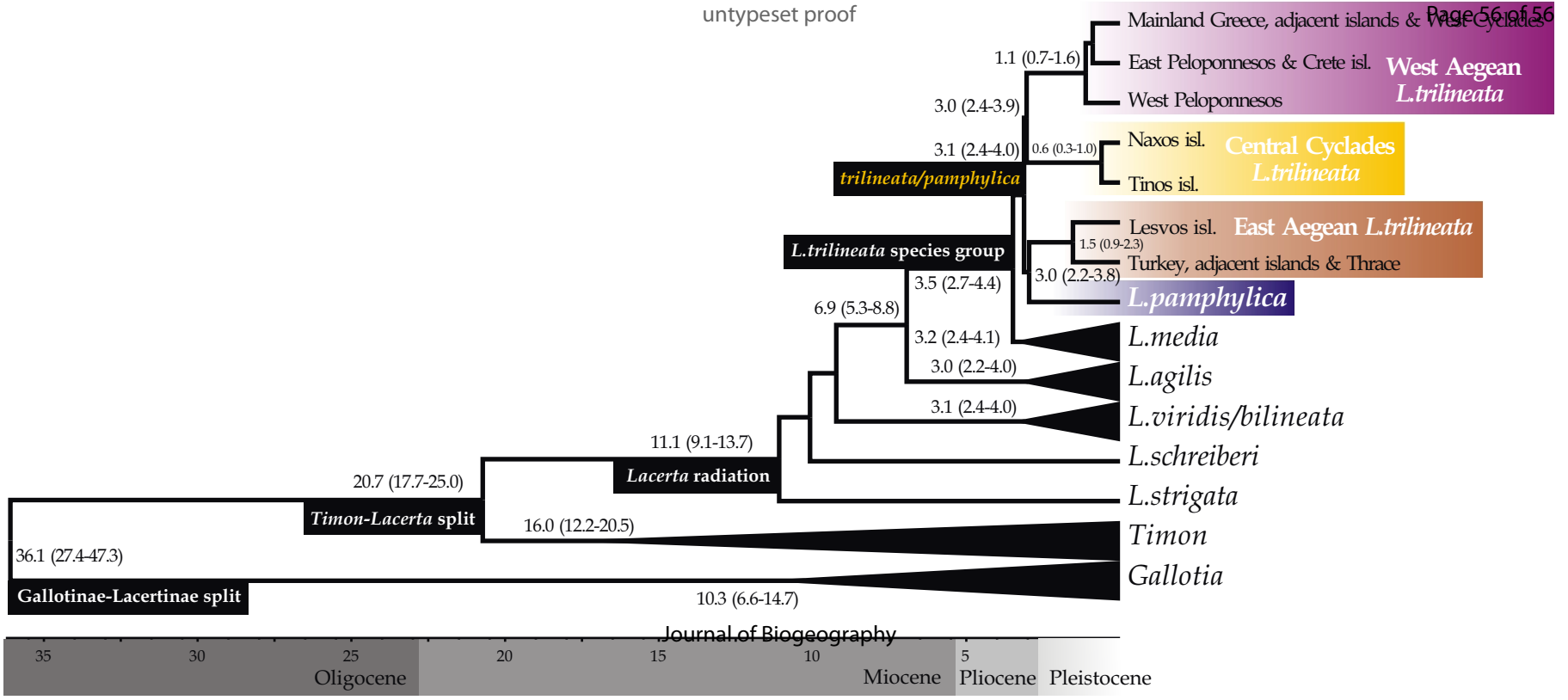

(b)
Pindos Mt.
Thrace
East Greece & Isl.
West Greece & Isl.
Anatolia-East Aegean
Skyros isl.
Lesvos isl.
Peloponnesos
Kythera isl.
Central Cyclades
West Cyclades
Crete
AB


(a)
L.trilineata
L.pamphylica
Balkan Peninsula
Greece
Anatolia
Turkey
Aegean Sea
LT391
LT392
LT60.316
LT382
LT3
KC897021
LT60.418
LT60.419
LN835028
LT416
LT2
LT60.185
LT60.186
LT60.217
LT469
LT360
LT345
LT8
LT6
LT408
LT353
LN835029
LT377
LT378
LT61.76
LT394
LT60.276
LT441
LT402
LT446
LT371
LT375
KC897014-16
KC897017-18
KC897019-20
LP2-LP3
KC897013
Europe
Asia
Africa


(a)
Turkey & islands
L.pamphylica
Central Cyclades
West Greece & islands
East Greece & islands
West Cyclades
West Peloponnesos
East Peloponnesos
Crete
0.68/-
1.0/82
-/90
(b)
L.trilineata Crete
L.trilineata Peloponnese
L.trilineata West Greece
Turkey and adjacent islands
L.trilineata
L.trilineata West Cyclades
L.pamphylica
L.trilineata East Greece
L.trilineata
Central Cyclades
(c)
Lesvos
East Aegean
North
South
West Aegean






Mainland Greece, adjacent islands & West Cyclades
East Peloponnesos & Crete isl.
West Peloponnesos
West Aegean
L.trilineata
1.1 (0.7-1.6)
3.0 (2.4-3.9)
3.1 (2.4-4.0)
0.6 (0.3-1.0)
Naxos isl.
Tinos isl.
Central Cyclades
L.trilineata
trilineata/pamphylica
Lesvos isl.
East Aegean L.trilineata
1.5 (0.9-2.3)
Turkey, adjacent islands & Thrace
L.trilineata species group
3.0 (2.2-3.8)
L.pamphylica
3.5 (2.7-4.4)
6.9 (5.3-8.8)
3.2 (2.4-4.1)
L.media
3.0 (2.2-4.0)
L.agilis
3.1 (2.4-4.0)
L.viridis/bilineata
11.1 (9.1-13.7)
Lacerta radiation
L.schreiberi
20.7 (17.7-25.0)
L.strigata
Timon-Lacerta split
16.0 (12.2-20.5)
Timon
36.1 (27.4-47.3)
Gallotia
Gallotinae-Lacertinae split
10.3 (6.6-14.7)
35
30
25
20
15
10
5
Oligocene
Miocene
Pliocene
Pleistocene