Title: Cryptic diversity and phylogeographic patterns of *Mediodactylus* species in the Eastern Mediterranean region

Panayiota Kotsakiozi[a,b], Aglaia Antoniou[c], Nikolaos Psonis[a,b1], Kostas Sagonas[a,b2], Emmanouela Karameta[a,b], Çetin Ilgaz[d], Yusuf Kumlutaş[d], Aziz Avcı[e], Daniel Jablonski[f], Diego Darriba[g], Alexandros Stamatakis[h,i,j], Petros Lymberakis[a], Nikos Poulakakis[a,b]

a Natural History Museum of Crete, School of Sciences and Engineering, University of Crete, Knosos Avenue, Heraklion 71409, Greece

b Department of Biology, School of Sciences and Engineering, University of Crete, Vassilika Vouton, Heraklion 70013, Greece

c Hellenic Centre for Marine Research, Institute of Marine Biology, Biotechnology and Aquaculture, Heraklion 71003, P.O. Box 2214, Crete, Greece

d Department of Biology, Faculty of Science, Dokuz Eylül University, Buca/İzmir 35160, Türkiye

e Department of Biology, Faculty of Science, Aydın Adnan Menderes University, Aydın 09010, Türkiye

f Department of Zoology, Comenius University in Bratislava, Mlynská dolina, Ilkovičova 6, 84215 Bratislava, Slovakia

g Computer Architecture Group, Centro de investigación CITIC, University of A Coruña, A Coruña, Spain

h Institute of Computer Science, Foundation for Research and Technology-Hellas, Greece

i Computational Molecular Evolution group, Heidelberg Institute for Theoretical Studies, 69118 Heidelberg, Germany

j Department of Informatics, Institute of Theoretical Informatics, Karlsruhe Institute of Technology, Karlsruhe 76128, Germany

Corresponding author: Panayiota Kotsakiozi, email address: pkotsakiozi@gmail.com

[1] Current address: Foundation for Research and Technology - Hellas (FORTH), Institute of Molecular Biology and Biotechnology (IMBB), Ancient DNA Lab, N. Plastira 100, Vassilika Vouton, Irakleio GR70013, Greece

[2] Permanent address: Department of Zoology, School of Biology, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

Abstract

Cryptic diversity poses a great obstacle in our attempts to assess the current biodiversity crisis and may hamper conservation efforts. The gekkonid genus *Mediodactylus*, a well-known case of hidden species and genetic diversity, has been taxonomically reclassified several times during the last decade. Focusing on the Mediterranean populations, a recent study within the *M. kotschyi* species complex using classic mtDNA/nuDNA markers suggested the existence of five distinct species, some being endemic and some possibly threatened, yet their relationships have not been fully resolved. Here, we generated genome-wide SNPs (using ddRADseq) and applied molecular species delimitation approaches and population genomic analyses to further disentangle these relationships. The, so far, most extensive nuclear dataset encompassing 2,360 loci and ~699,000 bp from across the genome of *Mediodactylus* gecko, enabled us to resolve previously obscure phylogenetic relationships among the five, recently described, *Mediodactylus* species and to support the hypothesis that the taxon includes several new, undescribed species. Population genomic analyses within each of the proposed species showed strong genetic structure and high levels of genetic differentiation among populations.

Keywords: Cryptic species, ddRADseq, Gekkonidae, Phylogenomics, Population genomics, SNPs

1. Introduction

Cryptic species constitute a major challenge in studies assessing biodiversity and a great obstacle in the global efforts to preserve species diversity. The term "cryptic species" is used to describe two or more distinct lineages that have been classified as a single nominal species due to their superficially indistinguishable morphology (Bickford *et al.* 2007). Although the concept has been known since the 18th century (Winker 2005), advances in DNA sequencing, including high-throughput sequencing, now allow for elucidating complex evolutionary histories, shedding light on complex speciation processes in non-model organisms and revealing a plethora of cryptic species in mammals (Herrera *et al.* 2022), reptiles (Engelbrecht *et al.* 2019), fishes (Guimarães *et al.* 2022), annelid worms (Bolotov *et al.* 2022), insects (Schär *et al.* 2022), mollusks (Sun *et al.* 2016), plants (Nitta & Chambers 2022), fungi (Wyrębek *et al.* 2021), and bacteria (Williamson *et al.* 2022).

With current species extinction rates being up to 1,000 times higher (Pimm *et al.* 2014) than the background rate (that is, the pre-human extinction rate or the extinction rate that is not related with anthropogenic factors), the discovery of such unrecognized species is now more important than ever (Dirzo & Raven 2003) in order to reevaluate conservation actions and optimize conservation strategies to protect what remains. This is particularly important for the focal region of our study, the Mediterranean basin, one of the world's biodiversity hotspots (Myers *et al.* 2000). Being at the crossroad of three continents (Africa, Europe, Asia) and exhibiting a complex geological history that left an imprint on the biogeography of many taxa (Lymberakis & Poulakakis 2010; Poulakakis *et al.* 2015), the Mediterranean basin is also a "scientific research hotspot" due to its species richness, its high levels of endemism (Lymberakis & Poulakakis 2010; Tierno de Figueroa *et al.* 2013; Lymberakis *et al.* 2018) and its susceptibility to climate change (Vogiatzakis *et al.* 2016). The herpetofauna of the region counts 398 reptile species with 54% of them being endemic and distributed throughout the basin (Kerim & Oğzukan 2017) and 13% being threatened (i.e., categorized by the IUCN as vulnerable-VU, endangered-EN, or critically endangered-CR, (IUCN, 2008)). A fraction of these species has been discovered during the last 20-25 years (e.g. 79 reptile species have been added to the herpetofauna of the European region between 2000 and 2020 (Uetz *et al.* 2022)).

One of the most characteristic examples are the wall lizards of the genus *Podarcis* in southern Europe, here the initial number of species [17 in Harris and Arnold (1999)] has increased by over 50% (Poulakakis *et al.* 2005; Pinho *et al.* 2007; Carretero 2008; Lymberakis *et al.* 2008; Larbes *et al.* 2009; Salvi *et al.* 2017; Psonis *et al.* 2018; Senczuk *et al.* 2019; Kiourtsoglou *et al.* 2021; Psonis *et al.* 2021) reaching the 26 species that are recognized today (Uetz *et al.*

2022). Similarly, recent studies have identified new species within various taxa that are distributed in the Mediterranean basin and were long considered as being single species or species complexes, including the blind snake *Xerotyphlops* (Kornilios *et al.* 2020a), the Roughtail Rock Agama *Laudakia* (Karameta et al. 2022), the skink lizard *Ablepharus* (Skourtanioti et al. 2016), and the green lizard *Lacerta* (Kornilios *et al.* 2020b). Many of these "newly described" species are endemics and/or have extremely narrow distribution ranges. For example, the wall lizard *P. levendis* is a steno endemic species only found on two islets (south of the Peloponnese) and is classified by the IUCN as VU, whereas *L. cypriaca* is endemic to Cyprus and has not been classified by the IUCN yet due to its recent elevation to the species level, but it may be considered as being threatened under criterion B (IUCN criteria for the Red List categories; Criterion B refers to the geographic range of a species).

Evidently, cryptic diversity, the "biodiversity wildcard" (Bickford *et al.* 2007), constitutes a major challenge in our efforts to devise conservation actions since the discovery of new species, especially endemic ones, substantially affects the conservation strategies as it changes the species richness indicators and the levels of endemism in a given region. Biodiversity parameters such as species richness and endemism are taken into account in the design/identification of protected areas and Key Biodiversity Areas (IUCN 2016).

The Mediterranean thin-toed gecko (*Mediodactylus kotschyi* complex) was, until recently, one case of a "species complex" (Böhme *et al.* 2009). Its taxonomy was reevaluated (Kotsakiozi *et al.* 2018) based on nuDNA and mtDNA data, recognizing five distinct species within the complex (Fig. 1), some of them being endemic to geographically restricted areas; *M. kotschyi* (Steindachner, 1870) distributed in the mainland Balkans, most of the Aegean Islands and Italy, *M. orientalis* (Štěpánek, 1937) in Levant, Cyprus, southern Anatolia and the south-eastern Aegean Islands, *M. danilewskii* (Strauch, 1887) in the Black Sea region and in south-west Anatolia, *M. bartoni* (Štěpánek, 1934) in Crete, and *M. oertzeni* (Boettger, 1888) occurring only in the southern Dodecanese Islands. This taxonomy was recently adopted by the 2020 update of the Species list of the European herpetofauna (Speybroeck *et al.* 2020). Nevertheless, the inter- and intra-phylogenetic relationships of these species remain mostly uncertain.

While DNA-based species delimitation methods have proved to be useful, the identification of speciation events under incomplete lineage sorting (ILS) is challenging (Bamberger *et al.* 2021). Modern sequencing approaches [such as RADseq (Davey & Blaxter 2010), ddRADseq (Peterson *et al.* 2012), ezRAD (Toonen *et al.* 2013)] can generate sufficient data to address this challenge. Recent investigations in the lacertid genus *Podarcis* using genomic data revealed hidden patterns of genetic diversity and provided an improved resolution of their phylogenetic

relationships (Garcia-Porta *et al.* 2019; Psonis *et al.* 2021; Yang *et al.* 2021), also suggesting the need for taxonomic revisions. Likewise, genome-wide SNPs have revealed a clearer picture of the phylogenetic relationships and provided a more stable taxonomy for eastern Mediterranean taxa including a) the Aegean green lizards of the genus *Lacerta*, leading to the recognition of *Lacerta citrovittata* and *L. diplochondrodes* (Kornilios *et al.* 2019, 2020b), b) the *Bufotes* toads in the eastern Mediterranean (Dufresnes *et al.* 2019), and c) the land snail *Albinaria cretensis* in the western part of the island of Crete (Bamberger *et al.* 2021).

In this study, we employed a ddRAD sequencing approach and analyzed genome-wide SNP data to elucidate the phylogenetic relationships among the eastern Mediterranean lineages of the genus *Mediodactylus* as defined in Kotsakiozi *et al.* (2018). Our objective was to re-evaluate the current taxonomy as well as assess the genomic diversity and the geographic structure of the populations using species-level genomic data.

2. Materials and Methods

2.1. Samples, DNA extraction, ddRAD-seq libraries preparation

In total, we used 94 specimens (Table S1) from 60 sampling localities (Fig. 1), covering the largest part of the distribution range of the five species (*M. danilewskii*, *M. kotschyi, M. oertzeni*, *M. bartoni*, *M. orientalis;* also see Table S1 for the number of individuals sampled per species) in the eastern Mediterranean and representing all major clades and subclades revealed in previous phylogenetic studies (Kasapidis *et al.* 2005; Kotsakiozi *et al.* 2018). Total genomic DNA was isolated from tail or tongue tissue of specimens that were preserved frozen (−80 °C) or in ethanol. DNA was isolated using either the DNeasy Blood & Tissue Extraction kit (Qiagen®, Hilden, Germany) according to the manufacturer's instructions, or an Ammonium Acetate based DNA extraction procedure (Bruford *et al.* 1998). The quality of the extracted DNA was evaluated using agarose gel electrophoresis (TAE, 1.5% gel) and quantification of the DNA extracts was performed using the Qubit® 2.0 Fluorometer (Invitrogen®, Carlsbad, California, USA).

The double-digest restriction site-associated DNA (ddRAD) libraries were prepared following the protocol of Peterson *et al.* (2012). Briefly, for the ddRAD library preparation, ~750 ng of high-quality DNA was simultaneously double-digested using SbfI and MspI (New England

147 BioLabs®, Ipswich, MA, USA) restriction enzymes following the manufacturer's instructions.
148 The individual barcoding was followed by the selection of fragments using the Blue Pippin
149 electrophoresis platform (Sage Science, Beverly, MA, USA) under the range selection of 415-
150 515 bp. Targeted fragments were amplified through 11 cycles of Polymerase Chain Reaction
151 (PCR) using the Phusion® Polymerase kit (New England BioLabs®, Ipswich, MA, USA).
152 Libraries were pooled and sequenced (paired-end sequencing, 150-bp reads long) on an
153 Illumina Hi-Seq 2000 lane at the Yale Center for Genome Analysis (Yale University, New Haven,
154 USA).

155 *2.2. Sequence Data processing*

156 Raw Illumina reads were processed using ipyRAD v.0.9.77 (Eaton & Overcast 2020). Samples
157 were demultiplexed using their unique sequence barcodes and Illumina indexes allowing no
158 mismatches between the barcodes of the two reads (Illumina paired-end sequencing). Base
159 calls with Phred quality scores below 20 (default setting; precision of the base call is 99%)
160 were converted into undetermined characters (N) and reads including more than five (default
161 setting) Ns were discarded. The minimum genotype depth was set to 6 (according to the ipyrad
162 manual this is approximately the minimum depth at which a heterozygous base call can be
163 distinguished from a sequencing error). The clustering threshold for the *de novo* assembly was
164 set to 0.90 based on a preliminary analysis (not shown) of our data while following a similar
165 reasoning used by Razkin *et al.* (2016) and Viricel *et al.* (2014), we also tested the clustering
166 thresholds of 0.85 and 0.95. The remaining parameters were left at their default settings,
167 including the minimum number of individuals that have a given locus (set to 4). As a result,
168 we got a sparse matrix, including loci for which at least four samples contain data. Thus, a
169 high proportion of missing data was present in the assembled dataset. To assess the impact
170 of missing data in getting a resolved phylogeny, for the final data assembly, we applied an
171 extra filtering criterion (i.e., the min_taxa; as in https://github.com/ddarriba/ddrad-seq; see
172 below), aiming to determine the minimum amount of data retaining sufficient phylogenetic
173 information for a resolved phylogeny. This is described in detail in recent studies dealing with
174 the effect of missing data on phylogenomic inference of lizard species (Psonis *et al.* 2018;
175 Psonis *et al.* 2021). Thus, instead of discarding all loci with missing data above a particular
176 threshold (as one would do by adjusting the min_samples_locus parameter in ipyrad), we
177 retained loci that are phylogenetically informative for parts of the phylogeny with the aim to
178 increase the potential to retain additional phylogenetic information for distinguishing among
179 more divergent taxa at deeper splits in the tree (e.g. see Eaton *et al.* (2017)).

We generated four different datasets with distinct fractions of phylogenetically informative loci by varying the min_taxa threshold. In the first dataset, we set min_taxa:=4 (dataset: Med100) considering that this dataset contains 100% of the loci. Subsequently, we gradually decreased the amount of missing data by requiring more phylogenetically informative loci to be present [i.e., min_taxa:= 8 (dataset: Med50), min_taxa:= 13 (dataset: Med25), min_taxa:= 17 (dataset: Med12) that correspond to about 50%, 25% and 12.5% of the loci of the initial Med100 dataset, respectively]. For each one of these datasets, we estimated the missing data per individual and per locus using the propTyped function of the adegenet package in R.

To evaluate these datasets with respect to the impact of missing data and justify our choice of the most stable dataset for comprehensive and final analyses, prior to the phylogenomic analyses, we used Pythia (Haag *et al.* 2022). Pythia is an open source software tool (https://github.com/tschuelia/PyPythia), that predicts *a priori* the expected behavior or difficulty of phylogenetic tree searches. We predicted this difficulty for each of the four datasets. Given that a Maximum Likelihood analysis, especially on a large genomic dataset, is time and resource intensive, it is helpful to predict *a priori* the "potential" of a given dataset to either converge to topologically similar tree topologies or to result in multiple statistically indistinguishable yet topologically highly distinct trees. In other words, Pythia predicts and quantifies, on a scale ranging between 0.0 (easy dataset) and 1.0 (extremely difficult), the difficulty of analyzing a given dataset. As such, it increases user awareness and allows to devise an effective as well as appropriate analysis strategy (e.g., increase the number of independent tree searches to construct a reliable tree on a "difficult" dataset). Although Pythia predicted the dataset with the least missing data (Med12; score 0.07; see Results) as being least difficult, the scores provided for the other three datasets were low (easy-to-analyze datasets) as well (0.09-0.16; see Results). Therefore, we also performed i) preliminary DAPC (see Section "Population Genomics Analyses" below) and ii) Maximum Likelihood analyses (for settings see Section "Phylogenomic Analyses"), on all four datasets. Then, we used the --rfdist option to compute the topological Robinson-Foulds (RF) distance (Robinson & Foulds 1981) among 50 ML trees, in a preliminary investigation on how the amount of missing data (See Results Section) affects the results.

2.3. Phylogenomic Analyses

For the dataset that Pythia suggested (Dataset Med12 including the 94 samples and the full sequences with a length of 698,737bp; see below) as having the best potential for a resolved phylogeny, we used ModelTest-NG (https://github.com/ddarriba/modeltest; (Darriba *et al.* 2019), to predict the best model of evolution for the phylogenetic analyses. We performed a

Maximum Likelihood (ML) tree inference using RAxML-NG (v.1.0.3; (Kozlov *et al.* 2019)) under the GTR + gamma model, with 50 random starting trees using 25 random and 25 parsimony-based starting trees (the default value for this step is 20 tree searches, but we increased this number to 50 to explore the tree space more thoroughly). To check the bootstrap convergence of the best scoring tree in each analysis we used the --bsconvergence option and the bootstrap support (BS) was also calculated and mapped onto the best-scoring ML tree of the selected dataset. We also performed an ML analysis using only the unlinked SNPs (one SNP per locus, the dataset was assembled using the R scripts available at https://github.com/ddarriba/ddrad-seq) of the selected dataset (Med12) using the Lewis (Lewis 2001) ascertainment bias correction. The command lines used for the ML analysis using RAxML-NG are provided in the Supplementary Material (Code for analyses). A Bayesian Inference (BI) analysis was performed for the selected (Dataset Med12 see below) dataset using MrBayes v.3.2.7 (Ronquist *et al.* 2012) and under the GTR + gamma model. The MCMC analysis ran for 1,000,000 generations using two independent runs with four chains each. The result was saved every 1,000 generations and for the "burn in" we discarded the first 25% of samples. Apparent convergence of the BI analysis was evaluated using the Estimated Sample Size (ESS>200) and the Potential Scale Reduction Factor (PSRF=1.0). The produced trees were visualized using FigTree v.1.4.4.

To test if the uneven representation of species and relevant missing data (see Results Section; Tables 2 and S2) affect our phylogenomic analyses, we performed an additional ML analysis on a pruned version of the Med12 dataset. The distributional pattern of missing data in our dataset is due to the overrepresentation of *M. kotschyi* (Table S1; ~60% of the samples) with respect to the remaining species (see also 3.4. Species Delimitation Section). Thus, we pruned the dataset down to 22 samples used in order to keep between 4 to 6 samples per species (except *M. bartoni* for which only 2 samples are available). The samples were selected such as to have similar proportions of missing data (Table S2).

In order to account for incomplete lineage sorting (ILS) that can induce gene trees / species trees incongruences that in turn might heavily impact phylogenetic reconstructions we also performed a coalescent based phylogenetic analysis with SVDquartets (Chifman & Kubatko 2014) as implemented in *PAUP (Swofford *2003*)using the multi locus data of the 94 sample dataset (Med12). SVDquartets infers the species tree directly from the site patterns and therefore bypasses the impact of gene tree estimation error. The analysis was executed i) considering the two best supported 8- and 12-species delimitation schemes and ii) based on the current taxonomy considering the five species. Runs were performed using exhaustive Quartet sampling with 200,000 random quartets and 1,000 bootstrap replicates.

The trees inferred by all phylogenetic inference methods were unrooted. Initially, we attempted to root the tree, using the Mediterranean house gecko (*Hemidactylus turcicus*) as outgroup. However, due to the high amount of missing data, the *Hemidactylus turcicus* sequences were excluded from the final dataset. To determine the most probable root of the tree, we used the RootDigger tool (Bettisworth & Stamatakis 2021) using as input the ML tree. Rootdiger can indicate the most likely root location on a given unrooted tree and infers a confidence value for the possible root placement. We kept the parameters as default and the exhaustive mode which evaluates the likelihood of placing the root into every branch of the tree, and as such it allows us to quantify root placement uncertainty.

2.4. Population Genomic Analyses

The population structure within each species was evaluated using the Bayesian clustering method implemented in STRUCTURE v.2.3.4 (Pritchard *et al.* 2000) via the STRUCTURE_THREADER pipeline v.1.3.10 (Pina-Martins *et al.* 2017). We used STRUCTURE to identify genetic clusters and assign individuals to these clusters without prior information about the sampling location. The analysis was performed on the Med12 dataset (based on the Pythia score and the RF distances) and was conducted on a per species basis (as defined in Kotsakiozi *et al.* (2018) and currently adopted by the 2020 update of the Species list of the European herpetofauna), though only for those species where more than six samples were available (*M. Kotchyi, M. orientalis, M. oertzeni*). To comply with the assumption of independence across loci, we subsampled our dataset by selecting one SNP per locus using respective R scripts (https://github.com/ddarriba/ddrad-seq). This filtered dataset was also used in all population genomic analyses (see below) and from now on, we will refer to it as Med12_1snp dataset. For each analysis the most likely allocation of samples to clusters (K), was determined by conducting 10 independent runs for each K ranging from 1 to 10. Each run assumed an admixture model and independent allele frequencies and used a burn-in period of 100,000 and 500,000 generations. The best K was selected based on the deltaK method of Evanno *et al.* (2005) using STRUCTURE_THREADER (Earl & vonHoldt 2012). Results were summarized and plotted with CLUMPAK that accounts for label switching and multimodality (Kopelman *et al.* 2015).

To complement the Bayesian analysis, we also performed a Principal Component Analysis (PCA) with the R package LEA (Frichot & François 2015), and a Discriminant Analysis of Principal Components (DAPC) of ADEGENET R package, using the Med12_1snp dataset. We used the *find.clusters* option of the ADEGENET R package (Jombart *et al.* 2010) in order for individuals to be assigned to DAPC-defined clusters, without *a priori* defining samples to

populations/groups. The number of DAPC-clusters is chosen based on the lowest BIC value. DAPC transforms the raw data using a PCA and then a DA is applied on the retained principal components to provide an efficient description of the genetic clusters using a few synthetic variables (discriminant functions) that are linear combinations of the original variables (raw data) (Jombart *et al.* 2010). Thus, the among-group variance is maximized while the within-group variance is minimized.

Same Med12_1snp dataset was then used to estimate the FST distans and perform AMOVA analyses. Pairwise genetic differentiation (FST) between groups of populations and their statistical support (p-value: 0.05) were calculated in Arlequin v3.5.2.2 (Excoffier & Lischer 2010), using 16,000 permutations (according to the manual that guarantees to have less than 1% difference with the exact probability in 99% of the cases).The partitioning of the genomic variation among and within populations was evaluated through a hierarchical Analysis of MOlecular VAriance (AMOVA) (Excoffier *et al.* 1992), as implemented in Arlequin, using 16,000 permutations. Details on the grouping for the AMOVA analyses are provided in Table S3.

2.5. Species delimitation Analysis

Acknowledging that species delimitation can be challenging and that different approaches may yield conflicting results, we conducted species delimitation using two methods; a) the BFD* method (Leache *et al* 2014) and b) the multi-rate PTP (mPTP) (Kapli et al 2017).

Species delimitation with the BFD* method was performed on a subset of the Med12_1snp dataset (selected dataset; see Results Section) using SNAPP (Bryant *et al.* 2012) package in BEAST2 v.2.7.5 (Bouckaert *et al.* 2019). This was deemed necessary since BFD* is computationally demanding, and thus we reduced our dataset based on the tree topology inferred via ML and BI to contain fewer representatives from every major clade or subclade (named Med12_snapp dataset; see Table S1 for the samples used in this reduced dataset). However, given the substantially more *M. kotschyi* samples used compared to the other species (see Fig. 1 and Table S1), this resulted in an overrepresentation of the *M. kotschyi* haplotypes. This overrepresentation might be an issue that should be taken into account in a species delimitation analysis [for details see (Magoga *et al.* 2021)] since the higher the number of sampled haplotypes, the higher the probability to find intermediate haplotypes among closely related species becomes. The case of *M. orientalis* is analogous, though less evident. Thus, to test if this unbalanced Med12_snapp dataset affects our results, we performed one more BFD* analysis (on the dataset named Med12_snapp2) by randomly subsampling the Med12_snapp dataset, in order for each of the species to be equally represented by 4-7 samples (except *M.*

bartoni that had only two samples). SNP data was converted to binary format with phrynomics R package (https://github.com/bbanbury/phrynomics/). BFD*uses a Yule prior with a parameter lambda (λ) representing the speciation rate. We estimated the λ value using the pyule script (https://github.com/joaks1/pyule). The script required the tree height (estimated based on the tree produced by the analysis of the concatenated sequences of the most stable dataset: Med12; see Results; Phylogenomic Analyses) and the number of tips/species as input. The number of tips/species varied from four to twelve depending on the species model scheme (see Table 1 for details in the scenarios), thus resulting in different λ values (from 40.1 considering four species to 77.9 considering twelve species). Mutation rates u and v were set to one 1 and were not sampled, while intraspecific variance was set to 0.1 (10%, α = 1, β = 10, Rateprior = gamma) and coalescence rate was sampled with a starting value of 10, following the settings used in similar studies for lizard species (e.g. Psonis *et al.* (2018)). The BFD* analysis was run with a chain length of 100,000 generations, alpha = 0.3, 50% burn-in percentage and 48 steps. The analyseswere executed in BEAST using a chain length of 1,000,000 generations and samples were stored every 10,000 generations. Apparent convergence for each delimitation scheme analysis as well as species tree estimation was assessed using Tracer and ESS values (ESS>200).

Specimens were assigned to the following alternative species delimitations (i) Model 1 (RunA), the four groups revealed by preliminary PCA on the entire 94 sample dataset, (ii) Model 2 (RunB) the five groups revealed by DAPC analyses on the entire 94 sample dataset, (iii) Model 3 (RunC), the five currently recognized species model, (iv) Model 4-8 (RunD-H), the groups revealed by the phylogenetic, DAPC and STRUCTURE analyses, in which the species number ranged from four to twelve (Table 1). More specifically, Model 4 (RunD), six species model with two species within *M. kotschyi*, Model 5 (RunE), six species model with two species within *M. orientalis*, Model 6 (RunF), seven species model with three species within *M. kotschyi*, Model 7 (RunG) eight species model with three species within *M. kotschyi*, and two species within *M. orientalis*, and Model 8 (RunH) twelve species model with three species within *M. kotschyi*, five species within *M. orientalis* and two species within *M. oertzeni*. Following Leaché *et al.* (2014), Bayes factor Delimitation (BFD*) was used to select among alternative delimitations and estimated as follows: BF = 2 × (MLE1-MLE0) where MLE0 was the marginal likelihood estimate value of the best model (Table 1) and MLE1 was the marginal likelihood estimate value for each alternative model evaluated against model 0. The strength of support from BF comparisons of competing models can be evaluated using the framework of Kass and Raftery

(1995). The BF scale is as follows: 0 < BF < 2 is not worth more than a bare mention, 2 < BF < 6 is positive evidence, 6 < BF < 10 is strong support, and BF > 10 is decisive.

The second species delimitation approach we employed, mPTP (Kapli et al., 2017), is an improved version PTP and does not require the user to define any analysis parameters (i.e. similarity thresholds, cutoffs, etc). The method uses a Markov chain Monte Carlo (MCMC) sampling approach, and computes support values for each delimitation of the input tree. Those values can be used to assess the confidence of the inferred ML delimitation scheme. For the mPTP analysis we used the concatenated sequence data of the Med12 dataset and the respective ML tree which we uploaded to the mPTP web server (https://mcmc-mptp.h-its.org/mcmc/).

3. Results

3.1 ddRADseq data metrics

After quality filtering, the sequencing of the ddRAD libraries resulted in 1,046,505 reads on average per sample (Table S2). The number of loci per sample after applying the clustering threshold, the average depth of clusters per individual, and the percentage of complete genotypes per individual are presented in Table S2. The filtered ipyrad assembly included 32,964 loci, each being present in at least four samples (MinCov = 4, paralogs removed) with an average of 3,554 loci per sample (Table S2). The SNPs matrix produced by ipyrad included 419,529 variable sites (84.6% missing) with a total of 30,517 unlinked SNPs.

The application of the extra filter to the ipyrad dataset resulted in four datasets (Med100, Med50, Med25, Med12), where the number of loci, the length of sequences in the assembly, and metrics regarding the percentage of missing data are presented in Table 2. For the selected by Pythia dataset (Med12 see below), the fraction of missing genotypes per sample ranged from 37% to 91% (Table S2). The proportion of missing data per locus (see also Table 2) ranged from 0% (i.e. some loci and specifically 86 out of the 2360 loci, were present in all 94 samples) to 83% (i.e. 3 out of the 2360 loci have missing data in 83% of the individuals).

3.2. Phylogenomic Analyses

Pythia suggested the Med12 dataset while the scores for all four datasets (Med100, Med50, Med25, and Med12) were also low i.e. 0.16, 0.09, 0.15, and 0.07 respectively. This dataset was also suggested by estimating the RF distances among all pairs of 50 inferred ML trees of each dataset. Therefore, this dataset was used for subsequent ML, BI and the SVDquartets analyses. ML analysis converged after 400 trees (cut-off threshold 0.01) and resulted in the

robustly supported tree (average BS on the tree equals to 92.2) presented in Fig. 2. BI analysis resulted in a tree with high BS Posterior Probabilities (PP; 0.96-1.00) and with identical topology (PP values are also presented in Fig. 2) to the one from ML. The phylogenomic inference confirmed the presence of five major clades within the eastern Mediterranean *Mediodactylus* taxa, each with high statistical support [PP=1.00, BS=100], which correspond to the five currently recognized species. The SVDquartets analysis (Fig. 3) resulted in a tree with the same topology as the ML/BI trees presented in Fig.2. Importantly, the species tree inferred with SVDquartets showed twelve highly supported clades that are geographically separated i.e. species occupy non overlapping regions, as shown inFig.3. The ML analysis on the SNPs matrix (not shown) also robustly supported (BS values 94-100) the presence (and the grouping of samples within each one) of the twelve clades (see Fig. 3). Finally, the tree topology remained unaltered for the ML analysis on the pruned dataset with 22 samples.

The rooted tree produced by RootDigger analysis placed, with high probability (lwr=0.99), *M. danilewskii* (Fig. 2), a species that ranges from Crimea to the coastline of Türkiye, and to the East Aegean islands (Fig. 1; blue), as being the most likely root of the tree. The most densely sampled clade, the one of *M. kotschyi,* can be robustly subdivided into three subclades; one hosts samples from continental Greece and the north/central Aegean Islands (called A1), another one includes the Kythira/Antikythira Islands samples (A2), and the third one comprises the Cyclades and the island of Kos that geographically belongs to the east Aegean Islands (A3). *M. oertzeni* which is distributed in the southeast Aegean Islands (Fig. 1; violet) seems to be a sister clade of *M. kotschyi* and closely related to the Crete's clade, *M. bartoni.* Last, *M. orientalis* (Fig. 2) which is further subdivided into two subclades; one including samples from western Türkiye (i.e. Aydin) and the east Aegean Islands (called B2) and one including samples from southern Türkiye (i.e. Adana, Gaziantep), Cyprus, and Israel (called B1).

3.3. Population Genomics Analyses

Genetic structure: The Evanno method (2005) on the population STRUCTUREanalysis for *M. kotschyi* (Fig. 2, S3) supported the presence of two clusters (K=2; Q values>0.95), which correspond to the A1/A2 and A3 clades ofthe phylogenetic tree that contains a split within this clade forming two monophyletic lineages; A1/A2 and A3 (Fig. 3). Hierarchical STRUCTURE analysis then showed the separation of A1 from A2 (STRUCTURE on the A1/A2 cluster; K=2) and then clear geographic differentiation within each subclade (Figs. 2, S3) -that were also supported by BFD* and mPTP as possibly different species (see Results section-Species delimitation) . Specifically, the three population clusters supported by DeltaK for A1, coincide with the split observed within this subclade (Fig. 2) separating the islands from continental

Greece and the north Aegean Islands as well as from the Peloponnese (Fig. 2-). The DeltaK method resulted in similar conclusions for subclades A2 and A3 as in both cases K=2 is returned as the most likely choice. In both cases the clustering (A2=Kythira and Antikythira Islands and satellite islets; A3=north and south part of the Cyclades) coincide with the splits observed in the tree of Fig. 2. For *M. oertzeni* (Figs. Fig2,S4), Delta K method supported two clusters of populations, while for *M. orientalis* (Figs. 2, S4) K=6 was the best supported value according to the deltaK method, albeit five major clusters were plotted by CLUMPAK (Fig. S4; 1-Cyprus, 2-Adana/Gaziantep, 3-Israel, 4-Aydin/East Aegean Islands, 5-Muğla) since the 6th cluster (Q value of the 6th cluster in major cluster equals 0.0004) appears only in the minor clustering (in 3 out of the 10 CLUMPAK runs) scheme. In both cases the results of STRUCTUREanalyses supported the geographic differentiation and are in agreement with the tree topology of Fig. 2.

Regardless of the filtering used (retaining 12.5% to 100% of the loci), PCA (Fig. S1.A) suggested the differentiation of *M. kotschyi* from all the remaining ones, while DAPC suggested the differentiation of *M. kotschyi* and *M. oertzeni* (Fig. S1.B) from the remaining ones. Based on this finding we proceeded to the next two DAPC analyses using the Med12 dataset (as indicated by Pythia) and filtered as to keep one SNP per locus. For *M. kotschyi*, the DAPC analysis (Fig. 4A) supported the presence of eight DAPC-groups that are in agreement to both ML/BI and coalescent trees 2, 3). In particular, we found a clear distinction according to the first axis, between the samples that originated from the north/central Aegean Islands and continental Greece (A1) from the remaining samples. Based on the second axis of DAPC, the samples from the Kythira/Antikythira Islands (A2) are differentiated from those from the southern Cyclades Islands (A3). The DAPC-groups defined within clade A1 (Fig. 4A; groups 2 to 5; Peloponnese, Kythnos Isl., continental Greece, and central Aegean Islands, respectively) largely coincide with the distinct clusters defined by the hierarchical STRUCTURE analysis within clade A1 (Fig. 2). The DAPC on the other group of species, indicated a clear distinction between species as *M. bartoni*, *M. oertzeni*, and *M. danilewskii* which form distinct groups. Interestingly, *M. orientalis* showed substantial differentiation (Fig. 4B) that is also similar to the *Structure* clustering for this species (Figs. 2, S4) forming five groups; i) Adana-Israel, ii) Cyprus, iii) Muğla (Türkiye), and iv) Ikaria-Fournoi Islands (east Aegean), and v) one sample from the Muğla clusters within the danilewskii-group.

The results of the AMOVA analysis are presented in Table S3. The vast majority of the genetic variation (66%-91,7%) was observed among groups. More specifically, when we considered six to twelve groups (grouping as in Schemes D, F, G, H of the BFD* analysis; See Table 1)the

variation among groups exceeded 90% compared to a variation of 68% among groups that the current taxonomy scheme (presented in RunC of Table 1) attained.

Genetic differentiation: Pairwise *FST* estimates between the major clades of the tree (Fig.2; A, B, C, D, and E; current taxonomy) receive high values (*FST*>0.68). Also high values (*FST*>0.75) were received among the three clades of *M. kotschyi* that coincide with delimited species (see below). Regarding clade A1, high *FST* values (0.46-0.57) were estimated between groups of populations (i.e. north Cyclades / north Aegean-continental Greece / Peloponnese). A lower level of differentiation (*FST*=0.36) was recorded between the two clusters of south Cyclades Islands (Clade A3) and a higher level of differentiation (*FST*=0.63) was observed between the Kythira/Antikythira Islands (Clade A2). The two subclades of Clade B (Figs.2, 3; B1/B2) showed a moderate compared to the rest level of differentiation (*Fst*=0.34), whereas high differentiation (*FST*=0.81) was observed between the two subclades of Clade E (Fig.2; E1/E2).

3.4 Species delimitation

The Marginal Likelihood Estimates (MLE) that were obtained from the first BFD* analysis, which was based on the PCA, DAPC, STRUCTURE and ML/BI tree topology (dataset; Med12_snapp), are presented in Table 1 and the results of the second BFD* analysis (dataset; Med12_snapp2) aiming to avoid overrepresentation of the sample-rich clades (i.e., including 4-7 samples per clade) are presented in Table S4. Both analyses supported the twelve lineages scheme (Table 1; RunH) as the delimitation of choice (BF values > 10; decisive) coinciding with well supported lineages in ML/BI and coalescent trees (Figs.2, 3). The mPTP alsosupported the presence of twelve delimited species. More specifically, both analyses supported the two recently recognized species (*M. danilewski, M. bartoni*) and supported additional delimited species within *M. kotschyi* (subclades A1, A2, A3 as being distinct species)*, M. orientalis* and *M. oertzeni* clades lso . Within the *M. orientalis* clade, BFD* and mPTP supported the delimitation of five species (1-Cyprus, 2-Israel, 3-Adana, 4-Muğla, 5-Ikaria-Fournoi-Aydin) and within*M. oertzeni* clade supported the delimitation of two species (Karpathos-Rhodes and Symi-Tilos).

4. Discussion

During the last decade, the feasibility to use thousands of genome-wide DNA markers in non-model organisms opened a new era in phylogenomics, revolutionized the field and revealed complex evolutionary processes and biogeographic patterns. In this study, using an extensive nuclear dataset including thousands of loci across the genome of the five recently recognized *Mediodactylus* species of the eastern Mediterranean not only we were able to confirm their monophyly but also reveal additional hidden species diversity in the study area. Our analyses

produced a clearer picture of the evolutionary relationships and intraspecific population structure and revealed that three species (*M. kotschyi, M. orientalis* and *M. oertzeni*) comprise species complexes. For *M. kotschyi*, the presence of three species is robustly supported by our results. For *M. orientalis* and *M. oertzeni* our data support the presence of five and two species within each complex respectively, however, these findings should be interpreted with caution given the high proportion of missing data for these two species. Last, the twelve delimited species seem to have non overlapped distributional ranges and that the paleogeography of the region played an important role on shaping their distributions.

Species delimitation and phylogenetic relationships

The five recognized *Mediodactylus* species (*M. kotschyi, M. orientalis, M. danilewskii*, *M. bartoni*, and *M. oertzeni*) of the Eastern Mediterranean region form well-supported, monophyletic clades, confirming the morphological grouping of Beutler (1981) and the recent raising of those groups to species level (Kotsakiozi *et al.* 2018). More specifically, the *kotschyi*, *bartoni* and *oertzeni* morphological groups proposed by Beutler (1981) represented exactly these taxa, while the *danilewskii* group was split into two species; *M. danilewskii* and *M. orientalis*. The most differentiated species is *M. kotschyi* with a relatively broad geographic range and high levels of genetic differentiation among populations. The species tree produced by SVDquartets revealed that the three highly supported lineages within *M. kotschyi* -which were also supported by both species delimitation methods as being different species- comprise sister taxa with A3 from south-eastern Cyclades being the most differentiated one. *Mediodactylus danilewskii*, and *M. orientalis* seem to be more closely related to each other compared with the other species as they cluster close to each other in DAPC (Fig. 4). However, it is important to note that the conclusions about these species should be interpreted with caution, as these species, particularly the relatively widespread *M. danilewskii*, are undersampled.

The species delimitation analyses supported the scheme of twelve putative species. Specifically, mPTP and BFD*analyses indicated *M. danilewskii* and *M. bartoni* clades as distinct species while they supported the presence of three species within *M. kotschyi*, the five-species scheme within *M. orientalis* and the presence of two species within *M. oertzeni*. Note that the species delimitation supported by BFD* and mPTP for *M. orientalis* and *M. oertzeni* completely coincide with the DAPC and STRUCTURE results for these two species. AMOVA analysis further supported the pattern indicated by BFD* as the percentages of variation among groups were maximized (>90%) when we considered the species delimitation schemes that were best supported by BFD*. This enhanced the validity of this specific grouping of lineages.

We emphasize that the high percentage of missing data for a number of samples, which is anticipated since our dataset includes several distinct species, and the fact that a couple of species are undersampled considering their distribution range, did not allow us to draw strong conclusions regarding a possible taxonomic revision. We tested the effect of missing data on the analyses by producing four datasets (Med100, Med50, Med25, Med12) containing different percentages of missing data (from 61% to 86%; Table 2). Among the four datasets, and as expected, the most stable dataset was the one (Med12) with the lowest proportion of missing data. However, Pythia predicted low scores for all four datasets and identical (or almost identical) results were obtained for the four datasets during preliminary analyses (tree topology and population clustering). These observations supported the idea that the percentage of missing data, although it was -on average- relatively high, does not affect the main results. This can be attributed to the fact that the filters applied here, aim to retain phylogenetic informativeness and preserve the phylogenetic signal in the data. The inclusion of more missing data among more divergent taxa increased the probability of encompassing more phylogenetic information for deeper cladogenetic events in a tree (Eaton *et al.* 2017). Similar findings regarding the effect of missing data on phylogenomics have been observed in other studies (Takahashi *et al.* 2014; Wang *et al.* 2017; Psonis *et al.* 2018; Psonis *et al.* 2021). Nonetheless, we do have strong evidence that more species complexes exist within the taxon. For example, one of the species that appeared to be a species complex with possibly five species is *M. orientalis*. However, this species is undersampled and exhibited a high percentage of missing data. This indicates that a denser sampling strategy is needed which will result in a more complete genomic dataset for this species before strong conclusions about taxonomic revisions can be drawn. The other species complexes revealed were *M. oertzeni*, a species with a restricted geographic distribution and *M. kotschyi*. Although these species cannot be considered as being undersampled, complementary studies (e.g. ecology, traditional taxonomy etc) are needed prior to issuing taxonomic recommendations. However, it is important to note that the higher number of loci used here, allowed us to unravel hidden diversity that remained undetected before. In particular, two out of the five species were confirmed (*M. danilewskii*, *M. bartoni*) and the presence of three other species complexes is suggested; *M. kotschyi* with three, *M. orientalis* with five and *M. oertzeni* with two robustly supported distinct groups of lineages (putative species), respectively.

Phylogeography

For *M. kotschyi*, the three distinct groups of lineages (subclades) supported by SVDquartets results are also geographically distinct (see Fig. 3) with the first (A1) being present in mainland

549 Greece and the north/central Aegean islands, the second (A2) in the Kythira island group, and
550 the third (A3) in the central and southern Cyclades, with high *FST* values among them. Based
551 on the subspecies taxonomy, the first subclade includes different morphological subspecies
552 [e.g. *M. kotschyi skopjensis* (Karaman, 1965) and *M. k. kotschyi* (Steindachner, 1870] from the
553 A3 lineage [e.g. A3 samples had been assigned to *M. k. concolor* (Bedriaga, 1881, 1882)]. The
554 absence of a clear intraspecific structure and fully resolved differentiation within *M. kotschyi* in
555 the study of Kotsakiozi *et al.* (2018) can potentially be attributed to the use of few loci in
556 contrast to the genome wide information of a high number of loci of the present study. Here,
557 the higher resolution that the SNPs data offer, allowed for a more fine-grained species
558 delimitation and detection of three distinct groups of lineages within *M. kotschyi*, two of which
559 are located solely on islands. This further supports the theory that islands (as here the
560 Mediterranean islands) harbor hidden diversity (Pérez-Delgado *et al.* 2022).

561 Focusing on the geographic distribution of *M. kotschyi* subclades, there is a north-to-south and
562 an east-to-west differentiation. Within *M. kotschyi*, the split of the Cyclades into north-western
563 (in subclade A1) and south-eastern islands (subclade A3), has also been observed in other
564 animal species, such as the *Euscorpius* scorpions (Parmakelis *et al.* 2006), vipers (*Vipera*
565 *ammodytes* complex; Thanou *et al.* (2023)), and partially in *Podarcis* lizards (Poulakakis *et al.*
566 2005; Yang *et al.* 2021) and *Mesobuthus* scorpions (Parmakelis *et al.* 2006), reflecting the
567 known geological separation of the Cyclades (~3.5 Mya) (Popov *et al.* 2004). In subclade A1,
568 there are three distinct subgroups (Figs. 2, 4) exhibiting clear geographical differentiation: the
569 Peloponnese, the northwestern Cyclades Islands/Skyros Island, and mainland Greece. The
570 inclusion of the north Aegean Islands (Lemnos Island and Thassopoula islet) within the
571 mainland populations is likely due to their recent geological separation from the nearby
572 mainland (Popov *et al.* 2004). Subclades A2 and A3 exclusively consist of island populations.
573 Subclade A2 consists of individuals originated from a biogeographical distinct island group at
574 the southwest edge of the Aegean archipelago, which includes Kythira Island (to the north),
575 Antikythira Island (to the south), and the Pori and Lagouvardos islets (in between).
576 Phylogenetically, subclade A2 is closely related to subclade A1, which includes the Peloponnese
577 and shares geographical proximity with the Kythira island group. This biogeographical pattern
578 is observed in several animal taxa, such as *Podarcis* (Spilani *et al.* 2019) and *Ablepharus*
579 (Skourtanioti *et al.* 2016).

580 From a paleogeographic perspective, this island system is quite interesting. While the larger
581 island of Kythira to the north submerged during the Pliocene (Meulenkamp 1985), there is no
582 evidence of similar tectonic movements affecting Pori, Lagouvardos, and Antikythira (a larger

islet south of Pori and Lagouvardos) during that period. The presence of *M. kotschyi* on all the islands within this group suggests that either the group was colonized from the Peloponnese following the re-emergence of Kythira or that the populations of Pori, Lagouvardos, and Antikythira islets remained unaffected by the Pliocene tectonic rearrangements, giving rise to the population of Kythira following its re-emergence. These findings align with the divergence times inferred in Kotsakiozi *et al.* (2018), supporting the divergence of the Kythira island group from the mainland at the end of the early Pliocene (3.9 Mya), while the differentiation within this group occurred during the Pleistocene.

The subclade A3 consists of individuals that originated from the central and south Cyclades Islands. Of particular note, the island of Kos that biogeographically belongs to the east Aegean Islands clustered within A3. The Aegean Sea constitutes a major contemporary barrier to biotic exchange between mainland Greece and Türkiye. The palaeogeographic evolution of the Aegean region has been described in detail in several studies (Parmakelis *et al.* 2006; Papadopoulou *et al.* 2010; Poulakakis *et al.* 2015; Kornilios *et al.* 2019), starting from the united landmass (Ägäis) of the middle Miocene to the formation of the Aegean Barrier (AB) in the late Miocene (10-9 Mya) [for more details see Kornilios *et al.* (2019)], causing the separation of the west Aegean (Cyclades Islands) from the east Aegean Islands. This pattern is imprinted in the biogeography and phylogeny of animal species (especially in those with limited overseas dispersal abilities). However, there are several cases of animal species, characterized as 'naughty' in Poulakakis *et al.* (2015), that have passed the Aegean Barrier (e.g., *Ablepharus kitaibelii, Podarcis erhardii, Pelophylax bedriagae, Trachelipus aegaeus, Albinaria brevicollis, Dichomma dardanum, Zonites rhodius*). So, the presence of Kos (east Aegean) in the subclade of Cyclades can be either one more case of 'naughty' animal or a human-aided dispersal, which is not uncommon for *Mediodactylus* (Koynova *et al.* 2017; Mares & Novarini 2020; Urošević *et al.* 2021).

On the other hand, *M. orientalis,* with a broad distribution area (Fig. 1), is divided into two robustly supported clades (see Fig. 2. B1; Türkiye, Cyprus, and Israel and B2; western Türkiye and East Aegean Islands). The first subclade showed a clear geographic differentiation as it is divided into four lineages (Fig. 3) exhibiting an east-west differentiation. The first lineage is from Israel, the second is from southeast Türkiye, the third from southwestern Türkiye and the fourth from Cyprus, which can be explained by the geomorphology of the area (e.g., the Taurus Mountains, Anatolian Diagonal, Nur Mountains) and the isolation of Cyprus. The Anatolian Diagonal is a line of mountain ranges that run from the south of Gümüşhane – Bayburt in the north, southwest across Türkiye to the Taurus Mountains (Mutun 2010). It is thought to consist

a significant geographic barrier shaping the distribution of various species across Türkiye and dividing lineage distribution into east and west (Ciplak *et al.* 1993; Rokas *et al.* 2003; Sengor *et al.* 2003; Mutun 2010; Bilgin 2011). The uplift of the Nur Mountains (during late Pliocene) seems to explain the isolation of the populations distributed at the southern-east edge of the taxon's distribution. Of particular interest is the region of southwestern Türkiye (Muğla region in our case), in which two different species of *Mediodatylus* are present (*M. orientalis* and *M. danilewskii*). This area is extremely rich in biodiversity, with the presence of distinct phylogenetic lineages, even at species level, in particular for several reptile species, such as *Ophiomorus kardesi* (Kornilios *et al.* 2018), *Laudakia stellio* (Karameta *et al.* 2022), *Xerotyphlops vermicularis* (Kornilios 2017), and *Blanus strauchi* (Sindaco *et al.* 2014).

Considering the lineage from Cyprus, it seems that Cyprus is more closely related to south Türkiye. Cyprus has been isolated for at least 5.3 Ma from the surrounding continental regions, with which it has probably never been connected, making it one of the very few and by far the largest, oceanic islands of the Mediterranean Sea (Dimitriou *et al.* 2022). Kotsakiozi *et al* (2018) estimated that the isolation of the Cyprus lineage occurred in the late Miocene (~6 Mya). This time corresponds to the Messinian Salinity Crisis (~6-5.3 Mya; (Krijgsman *et al.* 1999)), when the island was connected with, or being closer to Anatolia either through a land bridge or via a series of intermediate islets used as stepping-stones.

The second subclade of *M. orientalis* corresponds to the area of east Aegean islands and western Türkiye, which can be attributed to the paleogeographic history of this region where the east Aegean islands were connected to Türkiye even during the late Pleistocene (Perissoratis & Conispoliatis 2003; Lykousis 2009; Sakellariou & Galanidou 2017). It is worth noting that the five groups of populations within *M. orientalis* clade indicated by STRUCTURE analysis and coinciding with the subclades and lineages of the phylogenetic tree were also supported as delimited species by mPTP and BFD* analyses. Interestingly, these five lineages (see Figs. 2 and 5) also correspond to distinct subspecies; the lineage from Israel to *M. orientalis orientalis* (Štěpánek, 1937), the lineage of Cyprus to *M. orientalis fitzingeri* (Štěpánek, 1937) while the lineage from Adana-Gaziantep has been suggested to belong to *M. orientalis bolkarensis* (Rösler, 1994). However, a finer-scale sampling strategy along the Middle East coastline might help to disentangle the phylogenetic relationships within the taxon and provide insights into its phylogeographic history. In any case, we stress the need for additional studies focusing on the ecology, the morphology, and the biology of the taxa under study to fully conclude on the suggested species status.

Two other clades that appear in the tree (*M. bartoni* and *M. oertzeni*), are island species with restricted geographic distribution (southeast Aegean for *M. oertzeni* and Crete and surrounding islets for *M. bartoni*). The presence of only two individuals for *M. bartoni* did not permit us to assess the intraspecific diversity of this species in more detail. For *M. oertzeni* on the other hand, there was a clear differentiation between the islands of Rhodes and Karpathos from Tilos and Symi islands. Populations of these two groups had previously been described to belong to two distinct subspecies [*oertzeni:* Rhodes-Karpathos and *beutleri:* Symi; See Valakos *et al.* (2008)] that completely coincide with the two delimited species supported by our analysis. The close phylogenetic affinity of *Mediodactylus* geckos from Rhodes and Karpathos islands is a common pattern in animal species [e.g. water frogs of the genus *Pelophylax* (Lymberakis *et al.* 2007) and ground beetles of the genus *Dendarus* (Trichas *et al.* 2020)]. Karpathos, which was an island during the Miocene, was joined with Rhodes and Anatolia in the Early Pliocene (Daams & Van de Weerd 1980) and it was permanently isolated during the Late Pliocene (Böger & Dermitzakis 1987). Taking into account the estimated time of divergence of Karpathos and Rhodes by Kotsakiozi *et al.* (2018) in the Middle Pleistocene (~1 Mya), the distribution of *M. oertzeni* on Karpathos Isl. is the result of the dispersal of an ancestral form of *M. oertzeni* from Rhodes Isl. to Karpathos Isl., when Karpathos already was an island.

Mediodactylus danilewskii was estimated as the most probable root of the tree (Fig. 2). The species is distributed in a broad geographic area, expanding from Crimea to south Türkiye. Unfortunately, our small sample size did not allow us to investigate the genetic structure of its populations. However, given the concordance between our findings and those of Kotsakiozi *et al.* (2018) and by taking into account the presence of this species in Bulgaria and its subsequent introduction to different areas of Bulgaria (Koynova *et al.* 2020) and along the Turkish coasts of the Black Sea (Bülbül *et al.* 2023), we can hypothesize that this species covers a much broader area than the one sampled here, as samples from north Greece, north Türkiye, and Bulgaria [included in Kotsakiozi *et al.* (2018), but not in the present study] cluster within it. Thus, given the substantial morphological variation that this species exhibits within its range (Ajtić 2014; Pulev *et al.* 2014), a finer sampling strategy and a subsequent population genetic/omic analysis within *M. danilewskii* will shed more light on its evolutionary history.

Conclusion

Genomic data and current species delimitation methods are powerful tools for the study of cryptic diversity (Bickford *et al.* 2007; Chattopadhyay *et al.* 2016; Tang *et al.* 2022). These tools enabled us to reveal the relationships among *Mediodactylus* species at almost the entire distribution range of *Mediodactylus* populations in the eastern Mediterranean region and

revealhidden diversity. More specifically, genomic data confirmed the validity of the recent raising of *M. kotschyi, M. orientalis, M. danilewskii*, *M. bartoni*, and *M. oertzeni* lineages to species level and revealed three species complexes that need further investigation. Our results suggest that in the Eastern Mediterranean region there are possibly twelve *Mediodactylus* species with non overlapping distributional ranges, since *M. kotschyi*, *M. orientalis* and *M. oertzeni* seem to consist species complexes with three, five and two species within each complex respectively. However, we stress the need of additional studies before a possible systematic revision, given the high proportion of missing data and the low number of localities sampled for two of these species. Some of the newly suggested species are island endemics (e.g., *M. bartoni* endemic to Crete and satellite islets, *M. oertzeni E1 or E2 lineage* endemic to southeast Aegean Islands etc) and some of them may be classified as threatened in upcoming IUCN evaluations. Given the rate of species discovery since the adoption of the phylogenetic species concept, the distribution and the number of hotspots around the globe (Peterson & Navarro-Siguenza 1999) might still change substantially. Unraveling cryptic diversity contributes to addressing several of the shortfalls that Hortal *et al.* (2015) identified as biodiversity knowledge gaps. These shortfalls (e.g., Linnean Shortfall; knowledge gaps in taxonomy, Wallacean Shortfall; in species distribution, Prestonian Shortfall; in abundance and population dynamics etc.) severely affect our efforts to preserve biodiversity, which is critical for the ecosystems and human societies (Díaz *et al.* 2018). The use of genomic data and current species delimitation methods serve as a first step to unravel cryptic diversity, even for taxa that display complex evolutionary relationships.

Author contribution

Panayiota Kotsakiozi: Investigation, Formal analysis, Visualization, Writing - original draft; Aglaia Antoniou: Data curation, Resources, Investigation, Writing - review & editing; Nikolaos Psonis: Resources, Writing - review & editing; Kostas Sagonas: Resources, Writing - review & editing; Emmanouela Karameta: Resources, Writing - review & editing; Çetin Ilgaz: Resources, Writing - review & editing; Yusuf Kumlutaş: Resources, Writing - review & editing; Aziz Avcı: Resources, Writing - review & editing; Daniel Jablonski: Resources, Writing - review & editing; Diego Darriba: Methodology, Data curation, Writing - review & editing; Alexandros Stamatakis: Methodology, Data curation, Writing - review & editing; Petros Lymberakis: Resources, Writing - review & editing; Nikos Poulakakis: Conceptualization, Funding acquisition, Project administration, Supervision, Resources, Writing - review & editing.

Data availability

The Radseq data used in the analyses are available on the NCBI SRA in demultiplexed form, under BioProject number PRJNA1051519 and BioSample accessions SAMN38777126-SAMN38777219.

Funding

This study was part of a broader research project entitled "Adding more Ecology and Genomics in understanding of the biological Diversity: the role of islands as natural laboratories" and was funded by NSFR 2007–2013 Program for the development, European Social Fund, operational program, education and life learning Ministry of Education and Religious Affairs, managing authority, Co-financed by Greece and The European Union. A.S. was financially supported by the European Union (EU) under Grant Agreement No 101087081 (Comp-Biodiv-GR). DD was supported by the Ministry of Science and Innovation of Spain (PID2019-104184RB-I00/AEI/10.13039/501100011033) and by the Xunta de Galicia and FEDER funds of the EU under the Centro de Investigación de Galicia accreditation 2019-2022 (ED431G 2019/01).

Acknowledgements

We would like to thank Dr. Shai Meiri, Dr. Spyros Sfenthourakis, Dr. Steven Roussos and Dr. Ilias Strachinis for providing us *Mediodactylus* samples. We would also like to thank the two anonymous reviewers for their comments that greatly benefit our manuscript.

References

Ajtić, R. (2014). Morphological, biogeographical and ecological characteristics of Kotschy's gecko (*Cyrtodactylus kotschyi* Steindachner, 1870 Gekkonidae) from the mainland portion of its distribution range. Fauna Balkana 3, 1-70.

Bamberger, S., Xu, J., Hausdorf, B. (2021). Evaluating Species Delimitation Methods in Radiations: The Land Snail *Albinaria cretensis* Complex on Crete. Systematic Biology 71, 439-460.

Bettisworth, B., Stamatakis, A. (2021). Root Digger: a root placement program for phylogenetic trees. BMC Bioinformatics 22, 225.
Beutler, A. (1981). *Cyrtodactylus kolschyi* (STEINDACHNER 1870) - Ägäischer Bogenfingergecko. Cyrtodactylus kotschyi 1, 53-74.
Bickford, D., Lohman, D.J., Sodhi, N.S.*, et al.* (2007). Cryptic species as a window on diversity and conservation. Trends in Ecology & Evolution 22, 148-155.
Bilgin, R. (2011). Back to the Suture: The Distribution of Intraspecific Genetic Diversity in and Around Anatolia. International journal of molecular sciences 12, 4080-4103.
Böger, H., Dermitzakis, D. (1987). Neogene paleogeography in the Central Aegean region. Annals of the Hungarian Geological Institute 70, 217-220.
Böhme, W., Lymberakis, P., Ajtic, R.*, et al.* (ed ^ (eds)) (2009). *Mediodactylus kotschyi*. The IUCN Red List of Threatened Species 2009: e.T157281A5069008. Accessed at
Bolotov, I.N., Kondakov, A.V., Eliseeva, T.A.*, et al.* (2022). Cryptic taxonomic diversity and high-latitude melanism in the glossiphoniid leech assemblage from the Eurasian Arctic. Scientific reports 12.
Bouckaert, R., Vaughan, T.G., Barido-Sottani, J.*, et al.* (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol 15, e1006650.
Bruford, M.W., Hanotte, O., Brookfield, J.F., Burke, T. (1998). Multilocus and single-locus DNA fingerprinting. Molecular genetic analysis of populations: a practical approach 2, 287-336.
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.
Bülbül, U., Zaman, E., Özkan, H., Koç Gür, H. (2023). New Records of the Bulgarian Bent-toed Gecko Mediodactylus danilewskii (Strauch, 1887) (Reptilia: Gekkonidae) from Turkey. Acta Zoologica Bulgarica 75, 61-65.
Carretero, M.A. (2008). An integrated Assessment of a group with complex systematics: the Iberomaghrebian lizard genus *Podarcis* (Squamata, Lacertidae). Integrative Zoology 3, 247-266.
Chattopadhyay, B., Garg, K.M., Kumar, A.K.*, et al.* (2016). Genome-wide data reveal cryptic diversity and genetic introgression in an Oriental cynopterine fruit bat radiation. BMC Evol Biol 16, 41.
Ciplak, B., Demirsoy, A., Bozcuk, A.N. (1993). Distribution of Orthoptera in relation to the Anatolian Diagonal in Turkey. Articulata 8, 1-20.
Daams, R., Van de Weerd, A. (1980). Early Pliocene small mammals from the Aegean island of Karpathos (Greece) and their paleogeographic significance. Geol. Mijnbouw 59, 327-331.
Darriba, D., Posada, D., Kozlov, A.M.*, et al.* (2019). ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. Molecular Biology and Evolution 37, 291-294.
Davey, J.W., Blaxter, M.L. (2010). RADSeq: next-generation population genetics. Brief Funct Genomics 9, 416-423.
Díaz, S., Pascual, U., Stenseke, M.*, et al.* (2018). Assessing nature's contributions to people. Science 359, 270-272.
Dimitriou, A.C., Antoniou, A., Alexiou, I.*, et al.* (2022). Diversification within an oceanic Mediterranean island: Insights from a terrestrial isopod. Mol Phylogenet Evol 175, 107585.
Dirzo, R., Raven, P.H. (2003). Global State of Biodiversity and Loss. Annual Review of Environment and Resources 28, 137-167.
Dufresnes, C., Mazepa, G., Jablonski, D.*, et al.* (2019). Fifteen shades of green: The evolution of Bufotes toads revisited. Mol Phylogenet Evol 141, 106615.
Earl, D.A., vonHoldt, B.M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4, 359-361.

795 Eaton, D.A., Spriggs, E.L., Park, B., Donoghue, M.J. (2017). Misconceptions on missing data in RAD-seq
796 phylogenetics with a deep-scale example from flowering plants. Systematic Biology 66, 399-
797 412.
798 Eaton, D.A.R., Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets.
799 Bioinformatics 36, 2592-2594.
800 Engelbrecht, H.M., Branch, W.R., Greenbaum, E.*, et al.* (2019). Diversifying into the branches: Species
801 boundaries in African green and bush snakes, Philothamnus (Serpentes: Colubridae).
802 Molecular Phylogenetics and Evolution 130, 357-365.
803 Evanno, G., Regnaut, S., Goudet, J. (2005). Detecting the number of clusters of individuals using the
804 software structure: a simulation study. Molecular Ecology 14, 2611-2620.
805 Excoffier, L., Lischer, H.E. (2010). Arlequin suite ver 3.5: a new series of programs to perform
806 population genetics analyses under Linux and Windows. Mol Ecol Resour 10, 564-567.
807 Excoffier, L., Smouse, P.E., Quattro, J.M. (1992). Analysis of molecular variance inferred from metric
808 distances among DNA haplotypes: application to human mitochondrial DNA restriction data.
809 Genetics 131, 479-491.
810 Frichot, E., François, O. (2015). LEA: An R package for landscape and ecological association studies.
811 Methods in Ecology and Evolution 6, 925-929.
812 Garcia-Porta, J., Irisarri, I., Kirchner, M.*, et al.* (2019). Environmental temperatures shape thermal
813 physiology as well as diversification and genome-wide substitution rates in lizards. Nature
814 Communications 10, 4077.
815 Guimarães, K.L.A., Lima, M.P., Santana, D.J.*, et al.* (2022). DNA barcoding and phylogeography of the
816 Hoplias malabaricus species complex. Scientific reports 12.
817 Haag, J., Höhler, D., Bettisworth, B., Stamatakis, A. (2022). From Easy to Hopeless—Predicting the
818 Difficulty of Phylogenetic Analyses. Molecular Biology and Evolution 39.
819 Harris, D.J., Arnold, E.N. (1999). Relationships of Wall Lizards, Podarcis (Reptilia: Lacertidae) Based on
820 Mitochondrial DNA Sequences. Copeia 1999, 749-754.
821 Herrera, N.D., Bell, K.C., Callahan, C.M.*, et al.* (2022). Genomic resolution of cryptic species diversity
822 in chipmunks. Evolution 76, 2004-2019.
823 Hortal, J., Bello, F.d., Diniz-Filho, J.A.F.*, et al.* (2015). Seven Shortfalls that Beset Large-Scale
824 Knowledge of Biodiversity. Annual Review of Ecology, Evolution, and Systematics 46, 523-549.
825 IUCN. (2016). A Global Standard for the Identification of Key Biodiversity Areas, Version 1.0. IUCN,
826 Gland, Switzerland.
827 Jombart, T., Devillard, S., Balloux, F. (2010). Discriminant analysis of principal components: a new
828 method for the analysis of genetically structured populations. BMC Genetics 11, 94.
829 Kapli, P., Lutteropp, S., Zhang, J.*, et al.* (2017). Multi-rate Poisson tree processes for single-locus
830 species delimitation under maximum likelihood and Markov chain Monte Carlo.
831 Bioinformatics 33, 1630-1638.
832 Karameta, E., Lymberakis, P., Grillitsch, H.*, et al.* (2022). The story of a rock-star: multilocus phylogeny
833 and species delimitation in the starred or roughtail rock agama, *Laudakia stellio* (Reptilia:
834 Agamidae). Zoological Journal of the Linnean Society 195, 195-219.
835 Kasapidis, P., Magoulas, A., Mylonas, M., Zouros, E. (2005). The phylogeography of the gecko
836 *Cyrtopodion kotschyi* (Reptilia: Gekkonidae) in the Aegean archipelago. Mol Phylogenet Evol
837 35, 612-623.
838 Kerim, Ç., Oğzukan, C. (2017). Amphibians and Reptiles of the Mediterranean Basin. In:
839 *Mediterranean Identities* (ed. Borna F-B), p. Ch. 9. IntechOpen, Rijeka.
840 Kiourtsoglou, A., Kaliontzopoulou, A., Poursanidis, D.*, et al.* (2021). Evidence of cryptic diversity in
841 Podarcis peloponnesiacus and re-evaluation of its current taxonomy; insights from genetic,
842 morphological, and ecological data. Journal of Zoological Systematics and Evolutionary
843 Research 59, 2350-2370.

Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resour 15, 1179-1191.
Kornilios, P. (2017). Polytomies, signal and noise: revisiting the mitochondrial phylogeny and phylogeography of the Eurasian blindsnake species complex (Typhlopidae, Squamata). Zoologica Scripta 46, 665-674.
Kornilios, P., Jablonski, D., Sadek, R.A*., et al.* (2020a). Multilocus species-delimitation in the *Xerotyphlops vermicularis* (Reptilia: Typhlopidae) species complex. Molecular Phylogenetics and Evolution 152.
Kornilios, P., Kumlutaş, Y., Lymberakis, P., Ilgaz, Ç. (2018). Cryptic diversity and molecular systematics of the Aegean *Ophiomorus* skinks (Reptilia: Squamata), with the description of a new species. Journal of Zoological Systematics and Evolutionary Research 56, 364-381.
Kornilios, P., Thanou, E., Lymberakis, P*., et al.* (2019). Genome-wide markers untangle the green-lizard radiation in the Aegean Sea and support a rare biogeographical pattern. Journal of Biogeography 46, 552-567.
Kornilios, P., Thanou, E., Lymberakis, P*., et al.* (2020b). A phylogenomic resolution for the taxonomy of Aegean green lizards. Zoologica Scripta 49, 14-27.
Kotsakiozi, P., Jablonski, D., Ilgaz, Ç*., et al.* (2018). Multilocus phylogeny and coalescent species delimitation in Kotschy's gecko, *Mediodactylus kotschyi*: Hidden diversity and cryptic species. Molecular Phylogenetics and Evolution 125, 177-187.
Koynova, T., Doichev, D., Natchev, N. (2020). New data on the distribution of the Bulgarian Bent-toed Gecko (*Mediodactylus danilewskii* Strauch, 1887) in Shumen town (NE Bulgaria). Biharean Biologist.
Koynova, T., Tzankov, N., Popgeorgiev, G., Naumov, B., Natchev, N. (2017). A new distribution record of the Kotschy's Gecko (*Mediodactylus kotschyi*) from inland north-eastern Bulgaria. Herpetology Notes 10, 1-2.
Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35, 4453-4455.
Krijgsman, W., Hilgen, F.J., Raffi, I., Sierro, F.J., Wilson, D.S. (1999). Chronology, causes and progression of the Messinian salinity crisis. Nature 400, 652-655.
Larbes, S., Harris, D.J., Pinho, C*., et al.* (2009). Relationships of *Podarcis* wall lizards from Algeria based on mtDNA data. Amphibia-Reptilia 30, 483-492.
Leaché, A.D., Fujita, M.K., Minin, V.N., Bouckaert, R.R. (2014). Species Delimitation using Genome-Wide SNP Data. Systematic Biology 63, 534-542.
Lykousis, V. (2009). Sea-level changes and shelf break prograding sequences during the last 400ka in the Aegean margins: Subsidence rates and palaeogeographic implications. Continental Shelf Research 29, 2037-2044.
Lymberakis, P., Pafilis, P., Poulakakis, N., Konstantinos, S., Valakos, E. (2018). The Amphibians and Reptiles of the Aegean Sea. In: *Biogeography and Biodiversity of the Aegean. In honour of Prof. Moysis Mylonas*. Broken Hill Publishers Ltd, Nicosia, Cyprus.
Lymberakis, P., Poulakakis, N. (2010). Three Continents Claiming an Archipelago: The Evolution of Aegean's Herpetofaunal Diversity. Diversity 2, 233-255.
Lymberakis, P., Poulakakis, N., Kaliontzopoulou, A., Valakos, E., Mylonas, M. (2008). Two new species of *Podarcis* (Squamata; Lacertidae) from Greece. Systematics and Biodiversity 6, 307-318.
Lymberakis, P., Poulakakis, N., Manthalou, G*., et al.* (2007). Mitochondrial phylogeography of Rana (*Pelophylax*) populations in the Eastern Mediterranean region. Molecular Phylogenetics and Evolution 44, 115-125.
Magoga, G., Fontaneto, D., Montagna, M. (2021). Factors affecting the efficiency of molecular species delimitation in a species-rich insect family. Molecular Ecology Resources 21, 1475-1489.

Mares, G., Novarini, N. (2020). A likely population of the alien gecko Mediodactylus kotschyi (Steindachner, 1870) in the province of Belluno (Northeastern Italian Alps). Bollettino del Museo di Storia Naturale di Venezia, 71: 83-88. 71, 83-88.

Meulenkamp, J.E. (1985). Aspects of the Late Cenozoic Evolution of the Aegean Region. In: *Geological Evolution of the Mediterranean Basin: Raimondo Selli Commemorative Volume* (eds. Stanley DJ, Wezel F-C), pp. 307-321. Springer New York, New York, NY.

Mutun, S. (2010). Intraspecific genetic variation and phylogeography of the oak gallwasp Andricus caputmedusa (Hymenoptera: *Cynipidae*) Effects of the Anatolian Diagonal. Acta Zoologica Academia Scientiarum Hungaricae 56, 153-172.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Kent, J. (2000). Biodiversity hotspots for conservation priorities. Nature 403, 853-858.

Nitta, J.H., Chambers, S.M. (2022). Identifying cryptic fern gametophytes using DNA barcoding: A review. Applications in Plant Sciences 10.

Papadopoulou, A., Anastasiou, I., Vogler, A.P. (2010). Revisiting the Insect Mitochondrial Molecular Clock: The Mid-Aegean Trench Calibration. Molecular Biology and Evolution 27, 1659-1672.

Parmakelis, A., Stathi, I., Chatzaki, M.*, et al.* (2006). Evolution of *Mesobuthus gibbosus* (Brullé, 1832) (Scorpiones: Buthidae) in the northeastern Mediterranean region. Molecular Ecology 15, 2883-2894.

Pérez-Delgado, A.J., Arribas, P., Hernando, C.*, et al.* (2022). Hidden island endemic species and their implications for cryptic speciation within soil arthropods. Journal of Biogeography 49, 1367-1380.

Perissoratis, C., Conispoliatis, N. (2003). The impacts of sea-level changes during latest Pleistocene and Holocene times on the morphology of the Ionian and Aegean seas (SE Alpine Europe). Marine Geology 196, 145-156.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E. (2012). Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. PLoS One 7, e37135.

Peterson, T., Navarro-Siguenza, A. (1999). Alternate species concepts as bases for determining priority conservation areas. Peterson and Navarro-Siguenza 1999.

Pimm, S.L., Jenkins, C.N., Abell, R.*, et al.* (2014). The biodiversity of species and their rates of extinction, distribution, and protection. Science 344, 1246752.

Pina-Martins, F., Silva, D.N., Fino, J., Paulo, O.S. (2017). Structure_threader: An improved method for automation and parallelization of programs structure, fastStructure and MavericK on multicore CPU systems. Molecular Ecology Resources 17, e268-e274.

Pinho, C., Harris, D.J., Ferrand, N. (2007). Comparing patterns of nuclear and mitochondrial divergence in a cryptic species complex: the case of Iberian and North African wall lizards (Podarcis, Lacertidae). Biological Journal of the Linnean Society 91, 121-133.

Popov, S.V., Rögl, F., Rozanov, A.Y.*, et al.* (2004). Lithological-paleogeographic maps of Paratethys : 10 maps late Eocene to Pliocene.

Poulakakis, N., Kapli, P., Lymberakis, P.*, et al.* (2015). A review of phylogeographic analyses of animal taxa from the Aegean and surrounding regions. Journal of Zoological Systematics and Evolutionary Research 53, 18-32.

Poulakakis, N., Lymberakis, P., Valakos, E., Zouros, E., Mylonas, M. (2005). Phylogenetic relationships and biogeography of *Podarcis* species from the Balkan Peninsula, by bayesian and maximum likelihood analyses of mitochondrial DNA sequences. Molecular Phylogenetics and Evolution 37, 845-857.

Pritchard, J.K., Stephens, M., Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics 155, 945-959.

Psonis, N., Antoniou, A., Karameta, E.*, et al.* (2021). The wall lizards of the Balkan peninsula: Tackling questions at the interface of phylogenomics and population genomics. Molecular Phylogenetics and Evolution 159, 107121.

945 Psonis, N., Antoniou, A., Karameta, E*., et al.* (2018). Resolving complex phylogeographic patterns in
946 the Balkan Peninsula using closely related wall-lizard species as a model system. Molecular
947 Phylogenetics and Evolution 125, 100-115.
948 Pulev, A., Domozetski, L., Sakelarieva, L. (2014). Distribution of Kotschy`s Gecko *Mediodactylus*
949 *kotschyi* (Steindachner, 1870) (Reptilia: Gekkonidae) in South-West Bulgaria. Ecologica
950 Balcanica.
951 Razkin, O., Sonet, G., Breugelmans, K*., et al.* (2016). Species limits, interspecific hybridization and
952 phylogeny in the cryptic land snail complex *Pyramidula*: The power of RADseq data.
953 Molecular Phylogenetics and Evolution 101, 267-278.
954 Robinson, D.F., Foulds, L.R. (1981). Comparison of phylogenetic trees. Mathematical biosciences 53,
955 131-147.
956 Rokas, A., Atkinson, R., Webster, L., Csóka, G., Stone, G. (2003). Out of Anatolia: Longitudinal
957 gradients in genetic diversity support an eastern origin for a circum-Mediterranean oak
958 gallwasp Andricus quercustozae. Molecular Ecology 12, 2153-2174.
959 Ronquist, F., Teslenko, M., van der Mark, P*., et al.* (2012). MrBayes 3.2: efficient Bayesian
960 phylogenetic inference and model choice across a large model space. Syst Biol 61, 539-542.
961 Sakellariou, D., Galanidou, N. (2017). Aegean Pleistocene Landscapes Above and Below Sea-Level:
962 Palaeogeographic Reconstruction and Hominin Dispersals. In: *Under the Sea: Archaeology and*
963 *Palaeolandscapes of the Continental Shelf* (eds. Bailey GN, Harff J, Sakellariou D), pp. 335-359.
964 Springer International Publishing, Cham.
965 Salvi, D., Pinho, C., Harris, D.J. (2017). Digging up the roots of an insular hotspot of genetic diversity:
966 decoupled mito-nuclear histories in the evolution of the Corsican-Sardinian endemic lizard
967 *Podarcis tiliguerta*. BMC Evol Biol 17, 63.
968 Schär, S., Talavera, G., Rana, J.D*., et al.* (2022). Integrative taxonomy reveals cryptic diversity in North
969 American Lasius ants, and an overlooked introduced species. Scientific reports 12.
970 Senczuk, G., Castiglia, R., Böhme, W., Corti, C. (2019). *Podarcis siculus latastei* (Bedriaga, 1879) of the
971 Western Pontine Islands (Italy) raised to the species rank, and a brief taxonomic overview of
972 *Podarcis* lizards. Acta Herpetologica 14.
973 Sengor, A.M.C., Ozeren, S., Zor, E., Genc, T. (2003). Tast Anatolian high plateau as a mantle-
974 supported, N-S shortened domal structure. Geophys. Res. Lett. 30.
975 Sindaco, R., Kornilios, P., Sacchi, R., Lymberakis, P. (2014). Taxonomic reassessment of *Blanus strauchi*
976 (Bedriaga, 1884) (Squamata: Amphisbaenia: Blanidae), with the description of a new species
977 from southeast Anatolia (Turkey). Zootaxa 3795, 311-326.
978 Skourtanioti, E., Kapli, P., Ilgaz, Ç*., et al.* (2016). A reinvestigation of phylogeny and divergence times
979 of the *Ablepharus kitaibelii* species complex (Sauria, Scincidae) based on mtDNA and nuDNA
980 genes. Molecular Phylogenetics and Evolution 103, 199-214.
981 Speybroeck, J., Beukema, W., Dufresnes, C*., et al.* (2020). Species list of the European herpetofauna –
982 2020 update by the Taxonomic Committee of the Societas Europaea Herpetologica.
983 Amphibia-Reptilia 41, 139-189.
984 Spilani, L., Bougiouri, K., Antoniou, A*., et al.* (2019). Multigene phylogeny, phylogeography and
985 population structure of *Podarcis cretensis* species group in south Balkans. Molecular
986 Phylogenetics and Evolution 138, 193-204.
987 Sun, S., Li, Q., Kong, L*., et al.* (2016). DNA barcoding reveal patterns of species diversity among
988 northwestern Pacific molluscs. Scientific reports 6.
989 Takahashi, T., Nagata, N., Sota, T. (2014). Application of RAD-based phylogenetics to complex
990 relationships among variously related taxa in a species flock. Molecular Phylogenetics and
991 Evolution 80, 137-144.
992 Tang, Q., Burri, R., Liu, Y*., et al.* (2022). Seasonal migration patterns and the maintenance of
993 evolutionary diversity in a cryptic bird radiation. Molecular Ecology 31, 632-645.

Thanou, E., Jablonski, D., Kornilios, P. (2023). Genome-wide single nucleotide polymorphisms reveal recurrent waves of speciation in niche-pockets, in Europe's most venomous snake. Mol Ecol 32, 3624-3640.
Tierno de Figueroa, J.M., López-Rodríguez, M.J., Fenoglio, S., Sánchez-Castillo, P., Fochetti, R. (2013). Freshwater biodiversity in the rivers of the Mediterranean Basin. Hydrobiologia 719, 137-186.
Toonen, R.J., Puritz, J.B., Forsman, Z.H.*, et al.* (2013). ezRAD: a simplified method for genomic genotyping in non-model organisms. PeerJ 1, e203.
Trichas, A., Smirli, M., Papadopoulou, A.*, et al.* (2020). Dispersal versus vicariance in the Aegean: combining molecular and morphological phylogenies of eastern Mediterranean Dendarus (Coleoptera: Tenebrionidae) sheds new light on the phylogeography of the Aegean area. Zoological Journal of the Linnean Society 190, 824-843.
Uetz, P., Freed, P., Aguilar, R., Hošek, J. (ed ^ (eds)) (2022). The Reptile Database. http://www.reptile-database.org Accessed at
Urošević, A., Maričić, M., Vučić, T.*, et al.* (2021). New findings of Kotschy's gecko, Mediodactylus kotschyi (Steindachner, 1870) in Serbia, with a particular focus on recently recorded populations in Niš and Sremska Mitrovica. 12, 151-257.
Valakos, E.D., Pafilis, P., Sotiropoulos, K.*, et al.* (2008). Amphibians and Reptiles of Greece.
Viricel, A., Pante, E., Dabin, W., Simon-Bouhet, B. (2014). Applicability of RAD-tag genotyping for interfamilial comparisons: empirical data from two cetaceans. Mol Ecol Resour 14, 597-605.
Vogiatzakis, I.N., Mannion, A.M., Sarris, D. (2016). Mediterranean island biodiversity and climate change: the last 10,000 years and the future. Biodiversity and Conservation 25, 2597-2627.
Wang, X., Ye, X., Zhao, L.*, et al.* (2017). Genome-wide RAD sequencing data provide unprecedented resolution of the phylogeny of temperate bamboos (Poaceae: Bambusoideae). Scientific reports 7, 1-11.
Williamson, C.H.D., Stone, N.E., Nunnally, A.E.*, et al.* (2022). Identification of novel, cryptic Clostridioides species isolates from environmental samples collected from diverse geographical locations. Microbial Genomics 8.
Winker, K. (2005). Sibling species were first recognized by William Derham (1718). The Auk 122, 706-707.
Wyrębek, J., Molcan, T., Myszczyński, K.*, et al.* (2021). Uncovering Diagnostic Value of Mitogenome for Identification of Cryptic Species Fusarium graminearum Sensu Stricto. Frontiers in Microbiology 12.
Yang, W., Feiner, N., Salvi, D.*, et al.* (2021). Population Genomics of Wall Lizards Reflects the Dynamic History of the Mediterranean Basin. Molecular Biology and Evolution 39.

Tables

Table 1. BFD* analysis results for *Mediodactylus* species delimitation models. Clades coding refers to Fig. 3. Bayes Factor (BF) delimitation was used for model selection and was estimated based on the marginal likelihood estimate (MLE) value for each model. Positive BF values indicate support for the alternative model, and negative BF values indicate support for the null model (the model with the highest MLE). The most highly supported scheme is shown in bold.

Model [partition of clades/subclades]	**Species**	**MLE**	**Rank**	**BF**
RunH [A1 / A2 / A3 / B1-Muğla/B1-Cyprus/B1-Israel/B1-Adana B2 / C / D / E1/ E2] - *tree topology-mPTP delimitation*	**12**	**-115.97**	**1**	**-**
RunG [A1 / A2 / A3 / B1 / B2 / C / D / E] - *tree topology*	8	-368.53	2	-158.88
RunF [A1 / A2 / A3 / B / C / D / E] - *tree topology*	7	-457.96	3	-683.98
RunE [A / B1 / B2 / C / D / E] - *tree topology*	6	-504.88	4	-777.82
RunD [A1A2 / A3 / B / C / D / E] - *tree topology*	6	-549.96	5	-867.98
RunC [A / B / C / D / E] – *current taxonomy*	5	-900.16	6	-1568.38
RunB [A1 / A2 / A3 / BCD / E] - *DAPC groups*	5	-1866.68	7	-3501.42
RunA [A1 / A2 / A3 / BCDE] - *PCA groups*	4	-4142.14	8	-7547.22

Table 2. Number of loci and length of sequences (bp) in the assembly for the four filtered datasets (min_taxa = 4, 8, 13, 17).Estimates for the missing data per dataset is also provided and specifically the percentage of missing data for the total matrix (% of missing data), the average percentage of missing data per individual (% average per indv) and the range of missing genotypes (from the 94-sample dataset) per locus and the number of loci that are missing in more than 50% and 75% of the samples (>50% and 75%) in each dataset.

dataset	**min_taxa 4**	**min_taxa 8**	**min_taxa 13**	**min_taxa 17**
% of loci	100%	50%	25%	12.5%
Number of loci	18,300	8,664	4,182	2,360
Sequence length (bp)	5,342,266	2,547,210	1,234,877	698,737
% of missing data[1]	86.8%	78.7%	69.5%	61.7%
% of missing data[2]	84%	76,3%	68%	61.5%
% average per indv[1]	86.6%	78.6%	69.7%	61.7%
Range miss per locus[1]	0%-98%	0%-95%	0%-90%	0%-84%
> 50% and 75%[1]	12,884 and 11,463	5,792 and 4,393	2,603 and 1,394	1,761 and 439

[1] for the SNPdataset where one SNP per locus was retained and used in the population genomics analyses

[2] for the complete sequence dataset used in the phylogenomics analyses

Figure Legends

Figure 1. The sampling locations of the studied specimens. The green shaded area indicates the distribution range of what was previously considered as the *Mediodactylus kotschyi* species complex according to the IUCN database. Numbers correspond to the sampling location codes provided in Table S1. The differently colored sampling locations indicate the most recently proposed species-level taxonomy as described in Kotsakiozi et al. (2018): *M. kotschyi* (red; 1-35)*, M. orientalis* (yellow; 36-49)*, M. danilewskii* (blue; 50-53)*, M. bartoni* (light blue; 54-55)*, and M. oertzeni* (purple; 56-60)*.*

Figure 2. Maximum Likelihood (ML) tree reconstructed using ddRAD data. Bayesian Inference (BI) analysis resulted in an identical topology, bootstrap (BS) support values and Posterior Probabilities (PP) from the ML and the BI analyses respectively, are shown on the branch nodes of the tree. Individual codes follow those in Table S1 with the first two digits representing the map codes of Figure 1. Letters and numbers on the nodes are used to label the respective clade/subclade of the tree and are consistent using the coding used for the BFD* grouping schemes in Table 1. The vertical STRUCTURE bar plots on the right, indicate the groups of populations identified by the respective analysis on each clade/subclade of the tree. The spots on the emended maps indicate the geographic distribution of the STRUCTURE defined groups.

Figure 3. Bootstrap 50% majority-rule consensus tree from SVDQuartets analysis for twelve lineages/species as they were delimited by BFD* and mPTP analyses and their respective distribution on the maps.

Figure 4. Discriminant Analysis of Principal Components (DAPC) for *Mediodactylus* populations that belong **(A)** to *M. kotschyi* and **(B)** to other *Mediodactylus* lineages of the eastern Mediterranean. Individuals are represented as dots with the different colors representing the DAPC-groups defined. A bar plot of eigenvalues for the discriminant analysis (DA eigenvalues) is displayed in each inset. The plots are made using the first two DAs in both cases.